university of groningen / Johann Bernoulli institute

Intelligent Systems Research Group

Master's thesis

Using the HoloLens' Spatial Sound System to aid the Visually Impaired when Navigating Indoors

by

Arjen Teijo Zijlstra

Supervisor:

dr. M.H.F. Wilkinson

External supervisor:

drs. A. de Jong

Monitoring and Control Services

Knowledge, Expertise & Innovation

October 31, 2017

Abstract

In this study, the use of the HoloLens to aid visually impaired people when navigating indoors was explored. The HoloLens is an augmented reality system in the form of a cordless head mounted computer. The device has a spatial sound system that could be used in order to guide a visually impaired person through an unfamiliar building.

To find out how accurately people can localise sounds produced by the HoloLens' spatial sound system we developed an augmented reality application that can generate a sound at different angles and distances from the participant. In the first experiment the user had to point in the direction of the sound. The angle of deviation was used as a measure of the accuracy. In the second experiment a sound was used to guide a person over a path through a room without any visual cues or obstacles. The combination of both experiments was implemented in a proof of concept that guides a person through a building. Different types of sounds were tried and both visually impaired and sighted people participated in the experiments.

Results showed that people can consistently localise spatial sounds with a deviation less than 10° and often less than 5°. The results of the second experiment showed that all participants are capable of walking the optimal path at a convenient pace. This result was repeated in the proof of concept.

Taken together our results demonstrate the great potential of the HoloLens to help visually impaired people navigate indoors with high accuracy.

Acknowledgements

First of all, I would like to thank drs. Arnoud de Jong, for giving me the chance to do an internship at TNO and get to know all the inspiring people working there. It was a great opportunity to do a project that involved many new technologies I had never worked with before.

Furthermore, I would like to thank dr. Michael Wilkinson, for being my supervisor and for his feedback at an academic level.

Moreover, I would like to thank Royal Dutch Visio, for providing me with valuable input and the possibility to get to know much more about the world of the visually impaired.

And finally, I would like to thank all my family, friends and fellow students for always being there for me and for the feedback you guys gave me.

Arjen Teijo Zijlstra

Groningen,

October 2017.

Table of Contents

Chapter 1

Introduction

Worldwide an estimated number of 39 million people are blind and 246 million have low vision [Org14]. This adds up to a total of 285 million people are visually impaired. Being *blind* is defined as having a vision of less than 5%, whereas *low vision* means having a vision of less than 20%. In the Netherlands around 77 thousand people are blind and 234 thousand have a low vision [LBH+05; LK09].

A visual impairment brings along many challenges. Simple tasks like frying an egg, knowing what is in an aluminium can or checking an expiration date, when having a visual impairment, are not simple at all. Also, a visit to the doctor or finding the way in a public building is not easy for blind people. When outdoors, visually impaired people often use a white cane to follow tactile paving and some blind people rely on a guide dog. Also, most pedestrian crossings have sounds that signal them to the other side of the road when its safe. Unfortunately, in most public buildings support for the visually impaired is still nowhere to be seen. Most directions are only given visually and sometimes even for someone without a visual impairment these are hard to find. Therefore, other tools, such as the HoloLens, are needed. The Microsoft HoloLens is an augmented reality system renowned for its impressive holograms. For a large part, the holograms are so convincing due to the well-performing spatial sound system. In combination with the HoloLens' ability to create a spatial mapping from the sensory data it collects, the spatial sound system could be used in order to guide a visually impaired person through an unfamiliar building by creating sounds that the person could use to find its way.

1.1 Background

The HoloLens is a mixed reality system developed by Microsoft [Mic16] that consist of a see-through head-mounted display, built around an adjustable headband to make it wearable. The device is self-contained, which means that it does not need extra hardware or cables to operate. Besides the graphical processing unit, the HoloLens contains multiple sensors to obtain an understanding of its surroundings. The combination of its portability and its capability of spatial understanding makes the HoloLens particularly interesting as an instrument to aid the visually impaired.

Mixed Reality

Nowadays *virtual* reality and *augmented reality* are concepts known to most people, but Microsoft describes the HoloLens as a *mixed reality* system. Since these terms are used throughout this work, it is important to sharpen our definitions. We will follow the reasoning as presented in [MK94]. The term *mixed reality* is described in [MK94] by means of the *reality-virtuality continuum* as shown in Figure 1.1. To be able to do this, a third term is introduced, namely *augmented virtuality*.

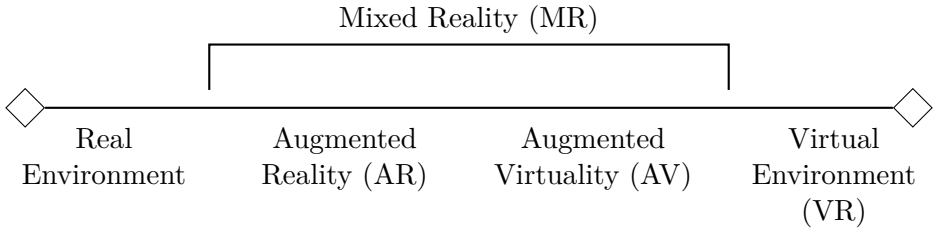


Figure 1.1: Reality-virtuality Continuum [MK94]. *Real Environment* is the "real world" as we know it, consisting of real objects that have an actual existence. *Virtual Environment* is an environment in which the observer is totally immersed, consisting of virtual objects that only exist in effect. *Augmented Reality* is a real environment to which virtual objects are added. *Augmented Virtuality* is a virtual world into which objects from the real world are mapped. *Mixed Reality* describes the set of environments in which both real and virtual objects are present i.e. augmented reality and augmented virtuality environments.

Besides defining the reality-virtuality continuum, [MK94] also defines the Extent of World Knowledge (EWK) dimension. This shows the level of understanding of the surroundings by the augmented reality device. In the EWK dimension there are two

ways to gain more knowledge about the real world. One way is to determine *where* the objects are, relative to the device. The other way is to know *what* the objects are that surround the device. The extremes of the EWK dimension are either to know nothing about the world, or to have the world completely modelled.

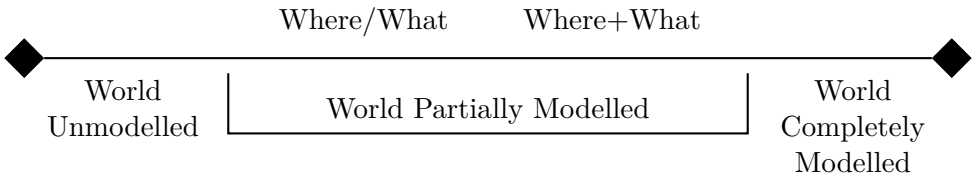


Figure 1.2: Extent of World Knowledge dimension [MK94]. The HoloLens by default only knows *where* objects are, relative to the device itself, by using Spatial Mapping.

The remaining two dimensions defined in [MK94] focus on the level of reality of mixed reality systems. The first one is called the *Reproduction Fidelity* dimension, which is mainly focussed on how photo-realistically virtual objects are rendered. The second one is called the *Extent of Presence Metaphor* dimension, which describes to what level the observer feels present in the virtual environment. Unfortunately, in neither of these last two dimensions audio is taken into account. It would be interesting to see research focussing on defining the dimensions reproduction fidelity and the extend of presence metaphor for (virtual) sound.

Hardware

This section gives an overview of the hardware of the HoloLens and what each part is used for. The device runs a complete version of Windows 10 Holographic including Windows app Store. This makes it possible to run almost any application compatible with Windows 10.

Input Modules

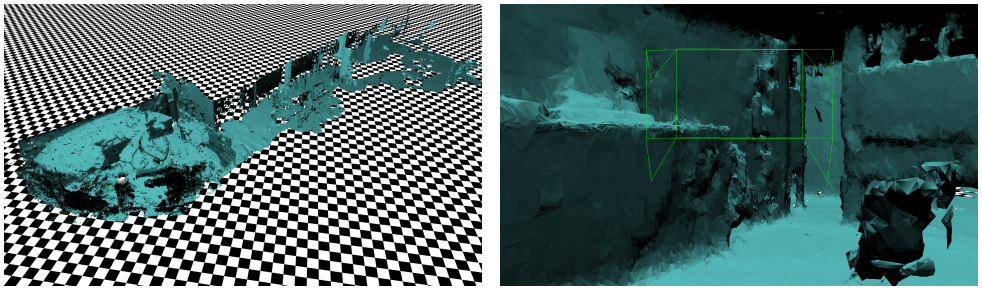
To acquire a high level of spatial understanding, the HoloLens is equipped with many sensors. The most important sensors are shown in Figure 1.3. The HoloLens uses these sensors, among others, to create a spatial mapping, to determine its orientation, to respond to voice commands and to provide the possibility to identify objects.

Figure 1.3: Sensors present in the HoloLens

However, the number of available sensors in the HoloLens are limited and restricted to its current set, since the device does not have a modular design. The current version of the device does not have a built-in GPS and is therefore dependent on other ways of localisation, based on spatial mapping. It is unknown whether or not GPS will be added in a future version of the HoloLens.

Spatial Mapping The depth (grey-scale) cameras present in the HoloLens are used to create a spatial mapping of the surroundings. An example of such a mapping is shown in Figure 1.4. The HoloLens uses this mapping for several things. Firstly, it can be used to render holograms in such a way that it looks like the holograms are actually located at certain places in the room. Besides that, the HoloLens also uses the mapping as a positioning system. It measures the distance towards objects using the infra-red sensors and it recognises turns and movements of the head using the gyroscope, accelerometer and magnetometer. By doing this, the HoloLens can determine where in the spatial mapping it is present. This way of tracking its location is called *inside-out* tracking. Furthermore, the HoloLens can match the spatial mapping with earlier created mappings. This information can be used to determine whether the HoloLens has been at a certain location before and consequently whether holograms were placed in that location the last time it was there.

Although the Spatial Mapping is very convincing and usually accurate in its current form, the HoloLens still needs a significant amount of time to create a complete spatial mapping of a room. This happens continuously and in the background, so the user usually does not even notice. In places where many transparent, reflective or dark objects are present the HoloLens has a hard time to identify these object as obstacles. Furthermore, when walking back and forth a hallway wearing the HoloLens, the device tries to combine all of its mappings into one world, which



(a) Third-person (b) First-person

Figure 1.4: First- and third-person view of a Spatial Mapping created by the HoloLens. In the third-person view you can clearly identify a hallway with at the end an office room. In the first-person view some objects can be recognised. On the bottom right you see a desk chair, on the left there is a cabinet against the wall and in the middle you can identify a doorway.

could result in the mapping to move relative to the real world, sometimes for over a meter. This results in all objects anchored relative to the spatial mapping to move for a meter (or more) as well. Figure 1.5 shows a mapping that has moved into the hallway for almost a meter. These errors makes it unreliable to use in buildings the HoloLens has not seen before, but by increasing the time scanning the surroundings this problem can be prevented.

Figure 1.5: Spatial Mapping of the HoloLens moved by almost a meter. The mapping is visualised by a grid, which can be triggered in the HoloLens main screen when clicking on mapped floor or walls.

User input The HoloLens can be controlled using certain actions that are captured using the aforementioned sensors. Three main manipulations exist, these are: *gestures*, *voice* commands or the HoloLens *clicker* [Mic17b].

Three main gestures are implemented by default, these are: the *bloom*, *gaze* and the *air-tap* gesture. The bloom gesture returns you to the desktop equivalent of the HoloLens by exiting or suspending all applications. To perform a bloom gesture, you must hold your hand in front of you with your fingertips together and then open your hand. When blooming from the desktop, the start menu is opened from which installed applications can be run. Gaze can be compared with moving the cursor by moving the mouse on a desktop computer. Gaze is equal to the rotation which the HoloLens has at every moment in time. This way, the HoloLens can determine in which direction the user points its face. This can be used to move a cursor around, but also numerous other applications are possible. The air-tap gesture roughly translates to a mouse click on a traditional computer. Air-tap also gives the opportunity to implement *double-tap* and *tap and hold*, which gives a wide range of possibilities for controlling applications and objects.

Voice commands can also be used to control your device. Many buttons respond when the text on it is said out loud. This makes some takes a lot easier, since gazing and tapping at a small button can be hard. Furthermore, the personal assistant *Cortana* is also available for all your questions. Its feature set is currently still somewhat limited, but this will most probably expand in the future. Finally, voice commands can be used to dictate messages. This is especially useful, since typing on the HoloLens requires gazing at every key you would like to press on the virtual keyboard.

The HoloLens clicker is available as a replacement for the air-tap gesture. The clicker comes in handy when your arm gets tired from air-tapping a lot and if the HoloLens does not register your air-tap when it is not performed slightly incorrect or just outside of the gesture frame.

Figure 1.6: HoloLens clicker [Mic17c]

It is possible to define your own hand gestures and voice commands when programming an application. However, Microsoft recommends not to implement any complex gestures and to keep the number of voice commands limited, such that the HoloLens will not have any problems to distinguish different commands.

Output Modules

The two most important output modules of the HoloLens are the speakers and the heads-up display (see Figure 1.7). Since this research is focussed on aiding the visually impaired, which includes blind people, we are mainly interested in the speakers. However, in Chapter 4 we also discuss some possibilities for using the display to assist visually impaired people.

(a) Left-speaker (b) Projection lenses

Figure 1.7: Output modules of the HoloLens [Mic16]

Holograms The projection lenses are used to display holograms in front of the user his eyes as if the are present in the room. The lenses are configured in such a way that the eyes of the user need to focus at a two meter distance to see the holograms optimally. Therefore, it is most comfortable to look at holograms that are actually at two meters distance. This configuration will also be taken into account in this research, when configuring the distance at which a sound is played.

One of the biggest issues with the HoloLens is the limited Field of View where holograms are visible. The part of the view where holograms are visible is only about 30% of that of the view of a normal seeing person. This could result in holograms being shown as cut off in mid-air, especially when standing close to the hologram. It is expected the field of view will be increased in future versions of the HoloLens, but for now it limits the experience. Figure 1.8 shows in what part of the field of view holograms are visible.

Another disadvantage of the projection lenses is that holograms are hardly visible in bright (sun)light. Indoors most holograms can easily be seen and recognised, especially when set to the highest brightness, but outdoors it can be hard to see the holograms that are shown on the display.

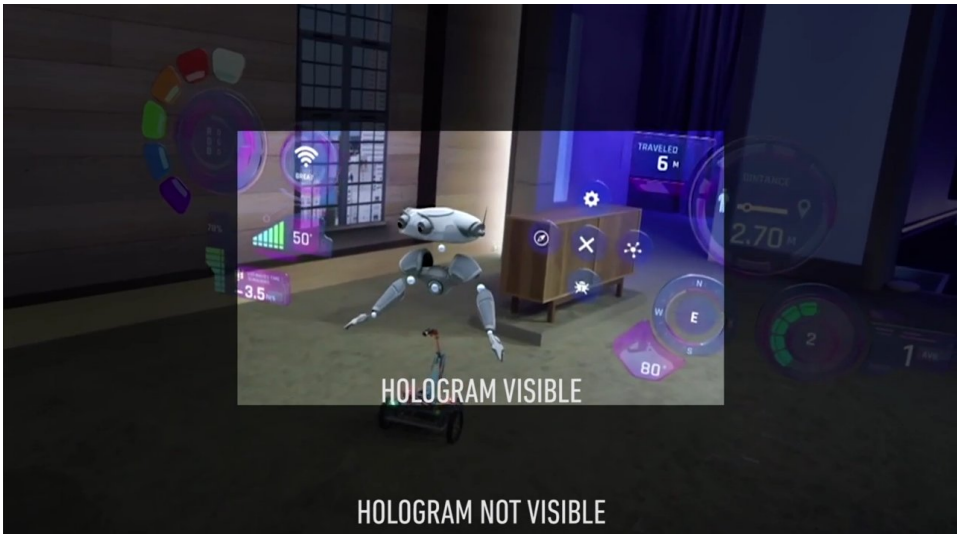


Figure 1.8: HoloLens' Field of View visualised [Rob15]. The part in which holograms are visible is about 30%.

Spatial Sound Besides using the Spatial Mapping to show holograms at the correct location in a space, the HoloLens also uses the mapping to create sounds from the location which it is expected to come from. To understand the underlying mechanism, sounds in general are briefly introduced.

Several ways of sound reproduction exist; after mono sound nowadays also stereo, 5.1 and 7.1 surround sound have made their way to the market. Binaural recordings are used to give the user the idea that, for example, the taxi he just heard really drives by. The HoloLens makes use of 3D sound played on the small speakers just above the ears to make holograms feel more present. The HoloLens determines the sounds by using Head Related Transfer Functions (HRTF). Since the HoloLens can determine when the wearer is moving its head and it knows the distance towards the virtual sound source, it can recompute the HRTF and therefore make the audio played on the headphones change according to the orientation of the HoloLens. Usually, HRTFs are recorded for a particular person, but the HoloLens uses a set of HRTFs that are suitable for most people. The audio calibration is fine-tuned based on the interpupillary distance, which is determined during the HoloLens display calibration. During the display (or visuals) calibration the user is asked to align its finger with holographic targets shown on the HoloLens. Since the calibration highly depends on visual feedback, it is currently impossible for a blind person to calibrate the HoloLens. Using non-individualized HRTFs is known to reduce the accuracy of

sound localisation, but it should still be accurate enough for this research [WAK+93].

Sound localisation is a passive skill that is continuously remodified in humans during their daily life [HVV+98]. A lot is known about how well people can localise sounds in the real world [MG91], but there are some known issues with spatial hearing through auditory displays [MM90]. One of them is front-back confusion [COC06], which means a subject thinks the sound is in front of him, while it is actually behind him, and vice-versa. Another problem with auditory displays is that it often interferes with other audio signals, for example with speech, which in turn has an effect on reaction time [Sim90]. Research conducted on the localisation of virtual sounds often includes virtual objects to choose from, not just a direction [STG+06]. This makes it a lot less flexible than the research performed with sounds in the real world, where speakers that can silently be moved around the subject are used and experiments are performed in darkened anechoic chambers.

It is well known that sound with a broader spectrum is easier to locate than sound with a narrower spectrum [HW74]. White noise has an equal intensity at different frequencies and therefore a broad spectrum. This makes it useful and much used for sound localisation experiments. Pink noise has an inverse proportional intensity to the frequency, which preserves the broad spectrum but makes it easier to listen to by human listeners. Furthermore, research has shown that sounds from a lower point are easier to localise than sounds from a higher point [Lan02]. It is however unknown whether the distance of a virtual sound has an impact on the ability of a subject to localise the sound. One could argue that sounds closer to the person are louder, assuming they are played at the same volume, and therefore should be more easy to localise. We have to note that the HoloLens' speakers are located just above the users' ears, which might influence this fact. In future versions of the HoloLens this could possibly be improved by using bone conducting headphones. These do not limit hearing other sounds like normal headphones, but since this technology is also relatively new, it asks for more research before being applied in practice.

A large part of what the ideas presented in this work are derived from brainstorm sessions with Royal Dutch Visio. Visio is a Dutch centre of expertise for everything related to being blind or visually impaired[1]. Visio collaborates with research institutes and supports them by providing knowledge about blindness. Some of their recent projects involve the use of the Google Glass and the Apple Watch to aid the visually impaired. In addition projects like the use of Bluetooth beacons that

[1]See: `https://www.visio.org`

provide the smartphone of a blind person at a specific location with information was developed [Hav10]. Visio was directly involved in the idea to apply the HoloLens spatial sound system for indoor navigation, since it closely relates to their own research, but many problems were not solvable without the features offered by the HoloLens. During the rest of this research the expertise of Visio was often included in certain design decisions and the experiments presented in this thesis have been performed at two Visio special schools in the Netherlands.

In the next chapter the research methods and experiments are described, followed by the obtained results. In Chapter 4 the results are evaluated and finally some options for future work using the HoloLens to aid visually impaired people are discussed.

1.2 Related Work

Nowadays, a lot of research is dedicated to facilitating daily life activities for the visually impaired. This has resulted in more public places to be equipped with tactile paving, some public signs offer Braille and some (Dutch) television programs have audio description. Installing these facilities can be time consuming and expensive, even though they do not always prove their use. For example, not all blind people can read Braille because the visual impairment developed at a later age. Therefore, research is now directed towards mobile solutions, often using the newest technology. Many smartphone applications exist to help blind people in their daily life.

In this section we focus on solutions that help blind people with navigation. We identified three different problems that need to be solved to make indoor navigation possible: *location determination*, *knowledge of surroundings* and *information transmission*. We specifically look at how the relating solutions solve these three problems.

The vOICe

An interesting system that makes use of augmented reality to help blind people navigate is The vOICe [Mei92]. The vOICe[2] is a vision substitution system, built on custom hardware, that converts camera footage into soundscapes.

[2]OIC needs to be said out loud: *Oh I See!*. See: `https://www.seeingwithsound.com/winvfaq.htm#faq_oic`

The vOICe does not perform an autonomous way of location determination, but feeds the subject with information such that the subject can do this positioning himself. To create knowledge of the surroundings, the device uses a camera located between the subject's eyes, to continuously create low-resolution footage of 64×64 pixels of what is in front of the subject at one frame per second. Since the camera is located at the front of the subject and facing forward the system only knows what is in front of the subject. Furthermore, since the footage is not stored in any way, it cannot be used to build an understanding of the environment nor to determine its location relative to earlier locations.

To translate the knowledge about the surroundings into something understandable for the subject, the pictures taken by the camera are mapped into sound using a one-to-one function. A pixel higher in the image corresponds to a higher frequency and a higher brightness to a higher amplitude (i.e. loudness). The sound is composed for each column of pixels from left to right. Transmission of this information is achieved by playing the composed sound over headphones worn by the subject.

This system is supported by the argument that it is possible to find an inverse of the mapping from image to sound. A disadvantage is that the sound produced by the system interferes strongly with other sounds. The system takes some time to get used to, but according to field tests it can be very useful for some people. The vOICe is still actively developed and researched Auvray, Hanneton, and O'Regan [AHO07], Ward and Meijer [WM10], and Haigh et al. [HBM+13].

Project Tango

An augmented reality system which will soon come to Android phones is Google Tango [LD+15; Goo14]. It makes use of depth sensors and the smartphone's camera to create a spatial representation of the phones surroundings. Geodan recently presented a system designed to help visually impaired with indoor navigation, that makes use of the Google Tango system. The phone running the software is used to provide feedback using sounds and haptic feedback to the visually impaired person [Geo17].

The system lets the Google Tango system determine the location of the subject. The Google Tango system determines the location relative to its representation of the space. Besides that, the system combines this relative position with GPS and WiFi to improve the location determination.

Geodan built a system that adds routing information to the representation. It is unclear how this routing information is obtained, but it could be from an earlier run or it could be received over the internet. This routing information is used to compute the shortest path to where the subject wants to go.

According to Geodan, the system provides the user with feedback in three ways; *vibrations*, *sound* and *image*. It is unclear how these cues are exactly used for information transmission, but the user should be able to follow the path using only this information. The sound is played over headphones, which limits user in hearing other sounds in his surroundings.

The system is precise up to a centimetre, but a disadvantage is that the device is hand-held and therefore limits movements. Toyota announced Project Blaid with the same goals (helping the visually impaired with navigation), but with a custom piece of hardware that is worn on the blind persons shoulders, so it does not have the limitation in movements [Pau16; Toy16]. Recently, Apple released ARkit, which works similar to Google Tango, available for all iOS devices with a processor newer than A9[3].

Microsoft Research

Recently, studies that use the technology of the HoloLens to help blind people were initiated. Microsoft Research presented a system in which auditory cues that encode information about distances to obstacles, walls and the orientation of the room are provided using Spatial Sound [BCS+15]. The system does not perform any way of location determination, but wall and floor detection is an important part of this work, which could be extended to be used for relative positioning.

As said before, wall and floor detection is the way the system obtains knowledge of its surroundings. The method first detects the planes it sees from the footage of the depth-camera mounted on the forehead of the subject. These planes are updated regularly using new footage. From these planes, the system detects the floor and the walls using the normals it detects.

For information transmission, the system uses spatial sound beacons to give auditory cues to the subject. It gives provides these cues for different reasons. The system provides cues for side walls, the computed vanishing point of the floor,

[3]See: `https://developer.apple.com/arkit/`

openings and obstacles. These cues can then be identified and used by the subject to navigate.

Related to this, Microsoft awarded a grant for a research involving the use of augmented reality with the HoloLens to aid the visually impaired [Kpm15]. Furthermore, an experimental app appeared in the new Windows Holographic app store[4], which plays a sound from the nearest surface detected by the HoloLens in front of the person, with a pitch depending on the shape of the object [Dav17]. These systems are still very conceptual, but it shows that a lot of research effort is invested in the development of new technologies that help to improve the lives of visually impaired people.

1.3 Scope

At the moment of writing the HoloLens is a rapidly developing product that has not turned mainstream yet. As we have already discussed, many limitations of the device can be identified. The lack of GPS, its slightly worse performance in extreme brightness and its limited field of view make it very likely that the device is mainly focussed on in-door use. These limitations are mainly focussed on the visual aspects of the device. In this thesis, we focus on the location fidelity of the 3D sound produced by the spatial sound system. The main research question is

- How can Spatial Sound be used to aid the visually impaired when navigating indoors?

To answer this question, the following sub-questions are formulated:

- With what accuracy can people localise spatial sound produced by the HoloLens?
- How efficiently can spatial sound guide a person over a path?
- How well can a person follow a route through a building using only spatial sound?

It is known that people can accurately localise real sounds and even spatial sounds [Lan02]. In this work we show how well people do on spatial sounds produced

[4] *White Cane* is available for HoloLens. See: `https://www.microsoft.com/en-us/store/p/white-cane/9p96g5q8sqgc`

by the HoloLens. This result is used to design a system in which people will follow a sound while moving over a path. In addition to this, a proof of concept is built that lets people walk a route through a building only by following a virtual sound.

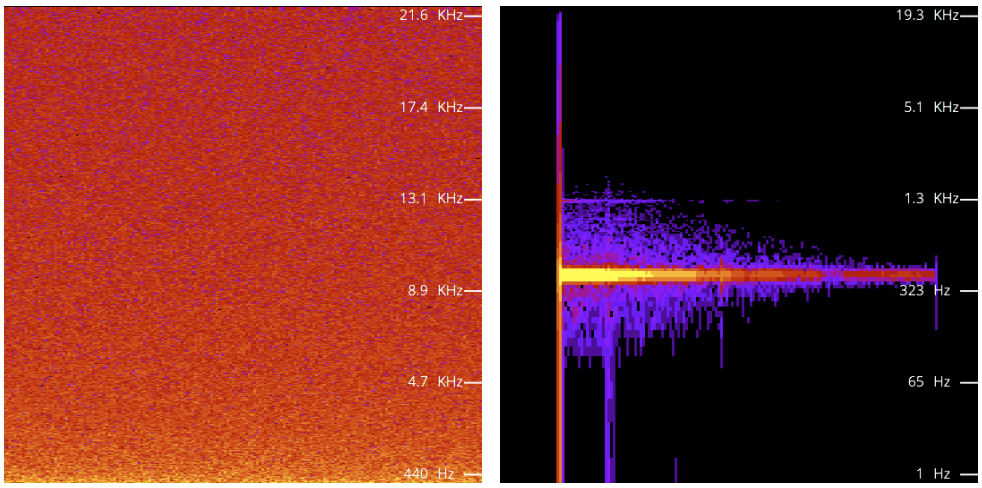
Chapter 2

Methods

The objective of the experiments is to measure the ability of a person to localise and follow a certain sound, played by the HoloLens, in a space. Different experiments are performed to test different abilities. The first experiment is only focused on the direction of a sound played by the HoloLens. In the second experiment we test how well a person can follow a sound over a pre-set path. The third experiment can be seen as a proof of concept, in this experiment we test how well our concept can guide someone through a building.

To determine the effect of the sound used, different sounds are used for the experiments. Pink noise is used since its broad spectrum is usually good for localisation, but slightly easier to listen to than white noise. In order to find a more comfortable sound to listen to, the other sound that is used sounds like a *sonar* ping. Pink noise can either be played constantly, at an interval (where the interval is dependent on the distance to the beacon) or just once for half a second. The sonar sound can only be played at an interval or just once. In Figure 2.1 the spectra of both sounds is displayed.

Furthermore, to determine whether or not a sound will be useful in a future final product, we also ask the visual impaired participants a few questions regarding the sounds they have listened to. The possible answers are on a scale from 1 to 5 and are used to determine the usability of the particular sound/playing mode combination. Question 1 denotes the perceived accuracy using a particular sound setting. Question 2 denotes how comfortable the sounds were. Question 3 denotes the level of interference with the surroundings. The following questions are asked:



(a) Pink Noise (b) Sonar sound (logarithmic scale)

Figure 2.1: Sound spectrum of the two sounds used during the experiments. As can be seen, the sonar sound has a very narrow spectrum compared to the pink noise.

1. How well do you think you did?
2. How pleasant was the sound?
3. How well were you able to hear your surroundings?

During the experiments the orientation (position and rotation) of the HoloLens is recorded at the start and end of each step and continuously at half a second intervals during each step. This position is relative to the spatial mapping and position when the application was started. It is possible to store the rotation because of the gaze functionality of the HoloLens. Furthermore, the angle of deviation between the direction the subject is facing and the actual direction where the sound comes from and also the distance between the location of the subject and the actual location of the sound beacon is continuously computed and stored. Finally, the time a user needs to complete the tasks is also included in the results. When the subject finishes one set of tests, the data is submitted to the database.

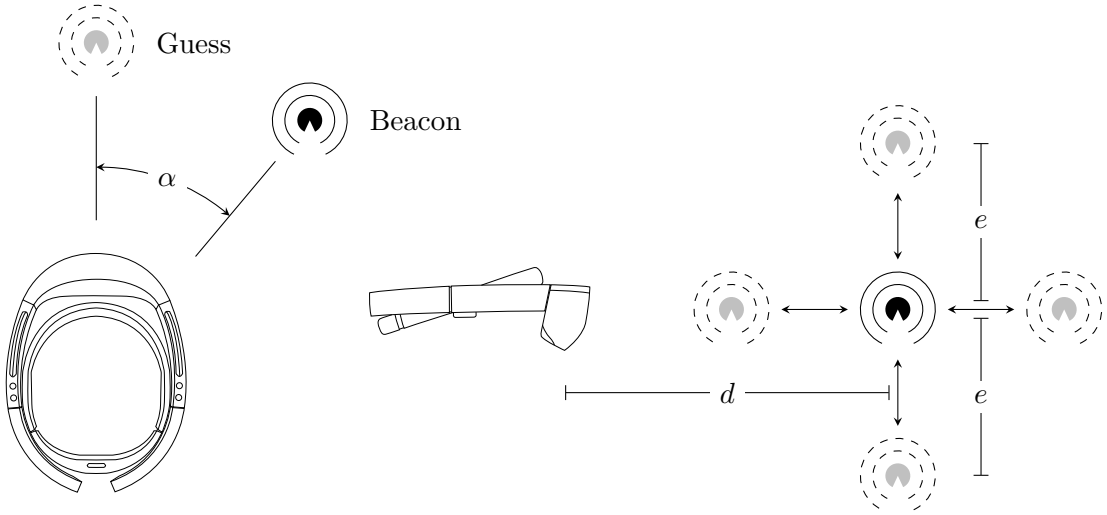
Note that, since some of the participants are blind, the HoloLens is not calibrated for them, but one independent person that has full vision is used as a model. The same model is also used for the tests with participants that have a fully or partially functional vision. Therefore, the scores are to be considered to be achieved with non-individualised HRTFs.

2.1 Experiment I: Direction

Experiment I is performed to determine the ability of a (blind) person to find the direction from which a sound is played. The results will also be used to find the best practices and sounds to guide a blind person in a certain direction.

The subject stands in the middle of a room wearing a HoloLens. The experiment is started from the computer of the examiner. At the start of the experiment a sound is played from a random direction at a given distance. The subject is asked to point its face in the direction where he thinks the sound is coming from. When the subject is certain about the direction of the sound the subject presses the clicker once to record the result. The angle of deviation from the actual direction is computed and stored. The next sound will automatically start after a short delay. This process is repeated 10 times per set of experiments.

In this experiment three parameters are variable. First of all there is the sound parameter. As mentioned earlier, there are two types of sound, *pink noise* and *sonar-like*, and three ways of playing the sound, *constant*, at an *interval* (playing for half a second, then silent for 1 second) and *once* for half a second.



(a) Angle of deviation. (b) Positioning of the beacon.

Figure 2.2: Graphical display of experimental details. In (a) α is the angle between the direction the subject thinks the sound comes from and the direction the sound actually comes from. In (b) setting d is the *distance* between the subject aan the beacon, and setting e is the *elevation* of the beacon relative to the subject.

Moreover, we have the distance and elevation at which a sound is played. To determine which distance and elevation works best, for both distance and elevation, three different settings are used; distance is either at 1, 2 or 3 meters from the listener, elevation is 2 meters above, same elevation, 2 meters below the listeners head elevation. Settings are not cross-varied, so when the elevation or distance is varied, other variables are set to default. The default sound is constant pink noise.

Figure 2.2a shows that the angle of deviation is measured in the horizontal plane (the so-called azimuth plane), since for our concept it eventually only matters that the subject walks in the right direction on a flat floor. If the system proves to be working well, future research could enhance it by adding measurements for sounds coming from above or below. Figure 2.2b shows the settings for the experiment regarding positioning. The beacons can spawn lower or higher and closer or father from the subject.

The settings of all 13 experiments are listed in Table 2.2. Experiment 1 and 8 have the exact same settings, to find out whether people score better after more experiments. Experiments 1 to 8 are performed using sighted participants. These participants are employees from TNO age older than 25. Experiments 9 to 13 are performed with students of Visio special schools with a vision between 0% and 20%.

ID	**Sound**	**Play type**	e **(m)**	d **(m)**	**Subject**
1	Pink Noise	Constant	0	2	Full Vision
2	Pink Noise	Constant	-2	2	Full Vision
3	Pink Noise	Constant	2	2	Full Vision
4	Pink Noise	Once	0	2	Full Vision
5	Pink Noise	Constant	0	1	Full Vision
6	Pink Noise	Constant	0	3	Full Vision
7	Pink Noise	Interval	0	2	Full Vision
8	Pink Noise	Constant	0	2	Full Vision
9	Pink Noise	Constant	0	2	Visually Impared
10	Pink Noise	Interval	0	2	Visually Impared
11	Pink Noise	Once	0	2	Visually Impared
12	Sonar-like	Interval	0	2	Visually Impared
13	Sonar-like	Once	0	2	Visually Impared

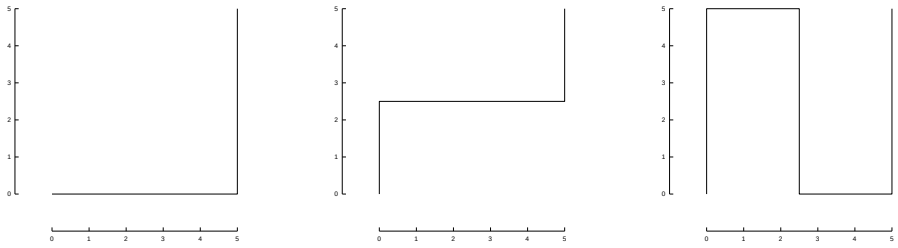
Table 2.2: Settings of experiments I. e denotes the elevation difference from the listener's head in meters, d denotes the distance from the listener's head in meters

2.2 Experiment II: Routes

Experiment II is performed to determine the ability of a blind person to walk a route guided by a sound played in front of him. The results will also be used to find the best practices and sounds to guide a blind person in a certain direction.

The subject stands at a start position in a room with at least 5 meters of free space in front and at the right. The experiment is started from the computer of the examiner. When the experiment is started a sound starts playing from the direction in which the subject should walk. The task of the subject is to follow the sound by continuously walking towards it. The sound moves along a pre-set path maintaining a fixed distance from the subject, until the end of the path is reached. When the subject reaches the end of the route, the subject hears a confirmation sound and the results are submitted to the database.

In this experiment only two parameters are variable. Equal to experiment I, there is the sound parameter, that can play pink noise or a sonar-like sound and can play either constant, at an interval or just once for half a second. The other parameter is the pre-set path the subject has to walk. Three pre-defined paths are used, with increasing complexity. The different paths are shown in Figure 2.3. The subjects do not know anything about the shape of the path before starting the experiment.



(a) Path 1 (b) Path 2 (c) Path 3

Figure 2.3: Pre-configured paths. Coordinate $(0, 0)$ is the starting point and $(5, 5)$ the end point. Distances are in meters.

Altogether the settings of Experiment II are shown in Table 2.4. During the experiments the position of the subject is continuously logged (at an interval of 0.5s). This way, the subject can be accurately followed and its behaviour can be closely analysed. The performance of a subject will be scored on two indicators;

ID	Route	Sound	Play type	Subject
14	1	Pink Noise	Constant	Visually Impared
15	2	Pink Noise	Constant	Visually Impared
16	3	Pink Noise	Constant	Visually Impared
17	1	Sonar-like	Interval	Visually Impared
18	2	Sonar-like	Interval	Visually Impared
19	3	Sonar-like	Interval	Visually Impared

Table 2.4: Settings of experiments II

distance travelled and *time elapsed*. Since the subject has to pass all "checkpoints" the distance travelled could be seen as the deviation from the path. The experiments from all routes and settings are performed with students of Visio special schools with a vision between 0% and 20%.

2.3 Proof of Concept

The results of Experiment I and II are used to develop a proof of concept, which shows the potential of the method developed in this research. This proof of concept consists of two stages.

Before the start of this experiment, a route through a building is set out by walking the route while wearing the HoloLens and placing way-points at the corners of the turns. This route is stored and will be used as the pre-set path for the experiments. The way-points are stored as World Anchors relative to the spatial mapping.

After the route is built, the HoloLens is passed on to a test-subject. The subject stands at the starting point of the route. The experiment is started from the computer of the examiner. When the experiment is started a sound starts playing from the direction in which the subject should walk. The task of the subject is to follow the sound by continuously walking towards it. When the subject reaches the end of the route, the subject hears a confirmation sound and the results are submitted.SS

The same route can be walked unlimited times, until a new route is set. In this experiment only one set of parameters is tested. The sound used is constantly playing pink noise, since this resulted in the best score in Experiment II.

This experiment could be performed with visually impaired people or blindfolded, but since the system is not yet capable to consistently place the world anchors at the exact same spot relative to the real world we cannot guarantee the safety of the test-subject and therefore this proof of concept is only performed as an experiment with people without a visual impairment. However, it should be noted that the subjects have no prior knowledge about the route and at several points they have multiple physically possible routes to choose from.

The performance of a subject will be scored on two indicators; *distance travelled* and *time elapsed*, similar as in Experiment II.

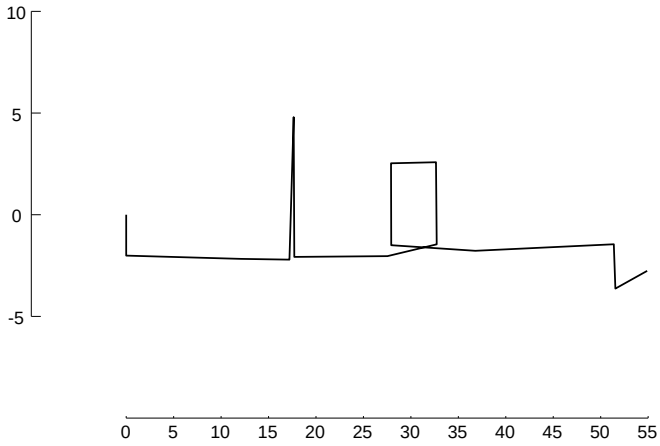


Figure 2.4: Path set using the proof of concept. Coordinate $(0, 0)$ is the starting point and ($\approx (55, 3)$ the end point. The axes show the distance in meters.

2.4 Implementation

Since the HoloLens runs the Holographic version of Windows 10, we can use the Windows SDK to develop our methods. The game engine Unity3D[1] makes it possible to build the application as an augmented reality application. In Unity all entities are represented as GameObjects. GameObjects can either be virtual objects (e.g. visual or audio) or so-called managers that control certain functionality. A useful open source toolkit providing some specific functionality for the HoloLens is the *MixedRealityToolkit*[2] (MRTK). It provides functionality ranging from a complete cursor (including animations, etc.) to a library with options to share data between separate HoloLens devices. In Unity, a number of functions are executed in a specific order for every GameObject, such as `Start()`, `OnEnable()` and `OnDestroy()`[3].

In an augmented reality application a scene represents the real world and the camera is the user wearing the HoloLens. Virtual objects added to the scene will be visible to the user in the real world. Objects can be given functionality in the form of a script as one of its components. GameObjects always have at least one component: the *Transform* component, which defines the location and orientation of the object relative to the parent object. GameObjects can be build up out of multiple components, e.g. it can have a Mesh component defining its 3D shape, a Physics component giving it gravity or a Material component defining how an object is displayed. Therefore a GameObject can also multiple script components. This way functionality can easily be split into multiple scripts and reused when needed.

The only noticeable virtual object that exists in our platform is a *Beacon*. In theory we could have multiple of those, but in our experiments we only use one at any given time. This beacon is implemented as an *AudioSource* that can play in certain modes (constant, interval, once). The Beacon listens to messages `OnPlay` and `OnStop`. In Listing 2.1 the implementation details of playing at an different playing types are shown. In the direction experiment the Beacon is given a component that makes sure it always stays at a set position relative to the camera (i.e. the user) by updating its position every frame. This way the user cannot walk towards or away from the beacon, the elevation relative to the user does not change as the user moves, but the user can rotate in the direction beacon. The AudioSource component

[1] `https://www.unity3d.com`
[2] Prev. *HoloToolkit*. `https://github.com/Microsoft/MixedRealityToolkit-Unity`
[3] See `docs.unity3d.com` for detailed documentation on the order of events.

Listing 2.1: `Beacon`

```
void OnPlay()
{
   switch (sound.Type)
   {
      case SoundInfo.PlayType.Interval:
         Invoke("PlayAtInterval", 0f);
         break;
      case SoundInfo.PlayType.Once:
         gameObject.GetComponent<AudioSource>().PlayOneShot(sound.Clip);
         break;
      default: // SoundInfo.PlayType.Constant
         gameObject.GetComponent<AudioSource>().Play();
         break;
   }
}

void ComputeInterval()
{
   // interval in seconds = distance in meters, clamped to 0.2 and 1.5 seconds
   interval = JSONHelper.Distance2D(Camera.main.transform, transform);
   interval = Mathf.Clamp(interval, .2f, 1.5f);
}

void PlayAtInterval()
{
   ComputeInterval();
   gameObject.GetComponent<AudioSource>().PlayOneShot(sound.Clip);
   Invoke("PlayAtInterval", interval);
}

void OnStop()
{
   switch (sound.Type)
   {
      case SoundInfo.PlayType.Interval:
         CancelInvoke("PlayAtInterval");
         goto default; // goto is fallthru in C#
      default:
         source.Stop();
         break;
   }
}
```

uses the Microsoft HRTF Spatializer plug-in to create 3D sounds. A second virtual object is a *Path* that consists of objects at certain waypoints (either relative to the start position or to the virtual world) to represent the path the user should be able to walk. These waypoints can be used by a third object: the *Guide*. The Guide always shows the user the direction in which the next waypoint is. If an object has both the *Beacon* and the *Guide* component, this means that the user can hear this object and because of the spatial sound know where it is.

To be able to control all of this, some managers are needed. The most important one is the *ClickManager* to capture the air-tap gesture. The MRTK provides a component that can be added to an object such that it receives an event when it is tapped upon, however, since we are interested in any air-tap and we do not have a visual target the user can tap on, we cannot use this. Therefore, we implemented the

Listing 2.2: `DoubleTap`

```
const float DELAY = .5f; // 500ms is the standard double click threshold

void Start()
{
   recognizer = new GestureRecognizer();
   recognizer.StartCapturingGestures();

   recognizer.SetRecognizableGestures(GestureSettings.Tap |
      GestureSettings.DoubleTap);
   recognizer.TappedEvent += (source, tapCount, ray) =>
   {
      if (tapCount == 1)
         Invoke("SingleTap", DELAY);
      else if (tapCount == 2)
      {
         CancelInvoke("SingleTap");
         DoubleTap();
      }
   };
}

void SingleTap()
{
   // Single Tap
}

void DoubleTap()
{
   // Double Tap
}
```

ClickManager to capture every air-tap event and use the orientation of the camera to compute how well a person did. The user is provided with audio feedback after every tap. The ClickManager also makes it possible to use the HoloLens clicker, which makes it easier for a subject to make sure that the tap is actually registered. Besides a single air-tap, also a double tap is implemented, which had to be done using timers, since the HoloLens does not support any built-in way of doing this. Listing 2.2 shows the C# code for registering a double tap. Doing it this way implies that single taps are always delayed by half a second.

Another important manager is the *AnchorMan*, which manages all the WorldAnchors in the scene across application start and stop. The *WorldAnchor* component is provided by the SDK, which gives an object a static position relative to the spatial mapping created by the HoloLens. After adding a WorldAnchor component to an object it is impossible to move it. To be able to move it, the WorldAnchor needs to be removed first and added after moving the object. It is recommended to give objects further than 6 meters apart separate anchors, therefore all waypoints in the Path object have their own WorldAnchor. The waypoints only get WorldAnchors when they are placed relative to the spatial mapping (as in the proof of concept). When they are placed relative to the starting point (as in Experiment II), they do

Listing 2.3: `Heartbeat`

```
void Start()
{
   client = new MqttClient(SERVER, PORT);
   client.Connect("1");

   InvokeRepeating("HeartBeat", 60f, 60f); // every 60 seconds
}

void HeartBeat()
{
   JSONObject obj = Message("HeartBeat");
   obj.AddField("sound", "\"bahdum\"");

   client.publish(TOPIC, Encoding.ASCII.GetBytes(obj.ToString()));
}
```

not need this.

Since the Windows SDK does not provide an easy way to control applications from outside of the device, an MQTT connection is added to the application to make it controllable from another computer. MQTT is a publish/subscribe messaging protocol designed for lightweight applications and is particularly popular in applications in the Internet of Things. The *MqttManager* is responsible for this connection and makes use of the `M2Mqtt` library to do this. It subscribes to relevant topics and passes messages it receives from the application back over the MQTT connection. To verify that the application is still active a so-called heartbeat is sent on the connection every minute. All messages are sent as JSON. An example of how such a message is constructed is shown in Listing 2.3 in (pseudo) C# code. An important manager dealing with these messages is the *CommandsManager*, which parses the *start* and *stop* commands, and initiates experiments from the data it gets.

Another manager important for debugging purposes is the *SettingsManager*. The CommandsManager can receive a command that enables debug mode, built into the application. Debug mode is set and the SettingsManager is responsible for all objects to know. If debug mode is enabled this means that all beacons, guides and waypoints are visible. This is achieved by giving objects a *DebugMarker* component.

DebugMarker listens to the SettingsManager regarding whether debug mode is enabled or not. The Mesh component, which defines the shape of the object, is disabled at start-up and therefore the object is not visible to the user. As soon as debug mode is enabled it sets the mesh of the object to enabled and thereby making the object visible to the user. This way it is easier to find out what is wrong if the application does not behave as it should. The SettingsManager class

Listing 2.4: `SettingsManager`

```
public delegate void DebugDelegate();
private bool debugEnabled = false;
public DebugDelegate DebugCallback;ěd

public bool DebugEnabled
{
   get { return debugEnabled; }
   set { debugEnabled = value; DebugCallback(); }
}

void Start() // Singleton uses Awake()
{
   DebugCallback += Dummy;
}

void Dummy()
{
   // I do nothing but preventing null reference exceptions
}
```

Listing 2.5: `DebugMarker`

```
void Start()
{
   SettingsManager.Instance.DebugCallback += UpdateSphere;
   UpdateSphere();
}

void OnPlay()
{
   gameObject.GetComponent<Renderer>().material.color = Color.red; // active
}

void OnStop()
{
   gameObject.GetComponent<Renderer>().material.color = Color.white;
}

void UpdateSphere()
{
   gameObject.GetComponent<Renderer>().enabled =
      SettingsManager.Instance.DebugEnabled;
}
```

is shown in Listing 2.4, it is implemented as a Singleton as provided by the MRTK. DebugMarker is shown in Listing 2.5

Some of the functionality used for this application does not work in the HoloLens emulator and only works on the device itself. Therefore, it is quite hard to debug the platform and log errors during the experiments. To circumvent this a *LogsManager* is implemented that hooks the application's log messages and sends them on a specific MQTT "log" topic to make it easier to store and read the logging. This can be done by composing the built-in delegate `Application.logMessageReceived` as shown in Listing 2.6. Similarly, the results of the experiments are collected at runtime and sent over a reserved MQTT topic. On the other end of the MQTT

Listing 2.6: `Logging`

```
1  void OnEnable()
2  {
3     Application.logMessageReceived += LogMessage;
4  }
5
6  void OnDisable()
7  {
8     Application.logMessageReceived -= LogMessage;
9  }
10
11 public void LogMessage(string message, string stackTrace, LogType type)
12 {
13    MqttManager.Instance.Publish(message, LOGS_TOPIC);
14 }
```

connection a Node.js application is listening on the different topics and handling them accordingly. The results are stored per session as JSON and post-processed using Python. All in all, the MQTT connection makes it a lot easier to control, debug and use the HoloLens for these experiments.

The logical view of the most important GameObjects of the HoloLens are shown in Figure 2.5. It shows the interaction between the different objects and includes the relation between the available components. Some functions might be called different in the actual implementation or require more details than shown in the logical view.

To avoid being dependent on the MQTT connection during demonstrations, a *KeywordManager*[4], provided by the MRTK, was used. Underwater it uses the *KeywordRecognizer* provided by the Windows SDK, but the MRTK makes it much easier to set up. The commands triggered by the users voice can be used to start and stop a demo, or pick some a set of installed settings. Note that without the MQTT connection it is impossible to store results. Another way of giving feedback to the user is to make use of the *TextToSpeechManager* provided by the MRTK. The TextToSpeechManager is used to verify the keywords recognized by the KeywordManager by repeating it to the user. This way another dependency on the MQTT connection is removed for demonstration purposes. Lastly, a third-person view is implemented that can be used in the HoloLens emulator, to check whether all objects are in the right location relative to the first-person camera. This third-person view is achieved by adding a second camera that follows the first camera and reserving all of the viewport for the second camera. Furthermore, when enabling third-person view, the first camera gets a non-symmetrical shape added as a child, such that it is visible and it is possible to see in which direction it is rotated.

[4]The KeywordManager component was deprecated on June 11, 2017.

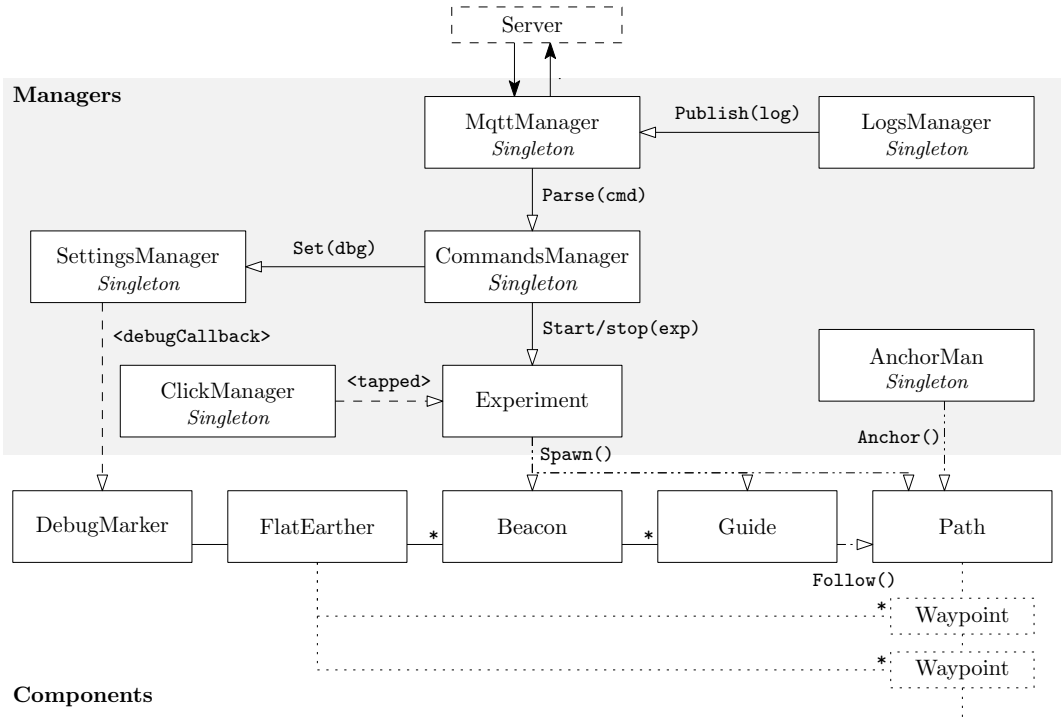


Figure 2.5: Logical view of the implementation of the HoloLens application. The *MqttManager* is the only connection with the server and both in- and outward communication is possible. Here —▷ indicates a direct call to another object, - -▷ indicates an event or callback, accompanied by the name of the event between <>'s, ··▷ indicates a call to a component of the object (e.g. its location), —— indicates that a GameObject can have both components. The $\star$ denotes the optional side in the relationship, i.e. every Guide is a Beacon, but not every Beacon is also a guide. The waypoints are shown as dashed (- - -) because they are not represented as separate components since they have no unique functionality.

From the other side of the MQTT connection, experiments can be started, stopped and settings can be changed by sending it as a command over the topic. At the moment, simple experiments can be started by packing settings in JSON format before sending. This could be extended by building a control panel with buttons for corresponding settings and listening to topics for logs, results and heartbeats showing instant results in graphs and figures.

Chapter 3

Results

After collecting the results from the MQTT connection the raw data needs to be processed. All locations and rotations are recorded relative to the HoloLens spatial mapping and its position when the application was started. After restarting the application the HoloLens location is set to $(0, 0)$ and therefore all data points are translated back to the position relative to their starting point by applying rotation and translation matrices. This chapter lists the result obtained after post-processing the raw data.

3.1 Experiment I: Direction

This section will display the results of Experiment I. The result will be shown per sub-experiment as dots on a circle, where each dot represents the accuracy of one participant (adapted from [WAK+93], FIG. 1. explaining confusion errors). The individual accuracy is measured by the absolute mean of all tries for one setting. The mean accuracy is displayed using a triangle with an angle equal to twice the accuracy (both negative and positive). The triangle is green if the mean accuracy is smaller than 10 degrees, yellow if the accuracy is between 10 and 15 degrees and red if it is bigger than 15 degrees.

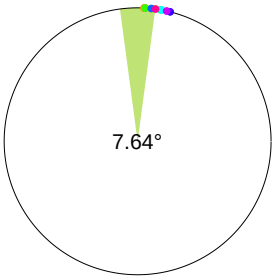


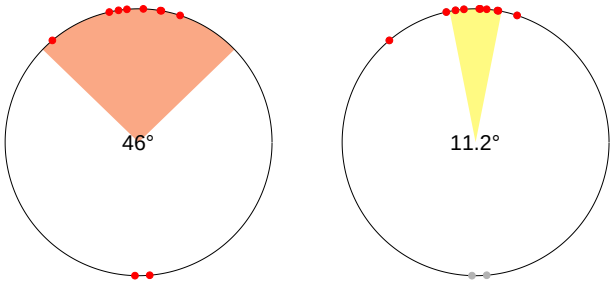
Figure 3.1: Example display of results. Each dot is the mean of one participant's results.

Figure 3.1 shows an example of what the results will look like. Note that in this example all subjects have a different colour. However, in the rest of this section all subjects will be coloured red for anonymity. In this case the triangle showing the mean accuracy is green since it is smaller than 10 degrees (i.e. 7.64°).

Front-back confusion

It is well known that front-back confusion is common within the world of virtual 3D sound. In Figure 3.2a an example of this can be seen. In general the subject performs decently, but because of the front-back confusion its accuracy looks pretty off. What could be done is ignore these confusions as outliers, but instead we consider these as fixable confusions that could be corrected in various ways in a practical application[1]. Therefore, instead of measuring the angle of deviation all the way from 0°, the angle of deviation is measured as the smallest angle from the y-axis and this way the front-back confusion is ignored when computing the accuracy.

Figure 3.2 shows the results of one test-subject with this front-back confusion. The accuracy of this test-subject as used in this chapter is shown in Figure 3.2b. As can be seen, the obtained accuracy now better describes the performance of the subject.



(a) Before (b) After

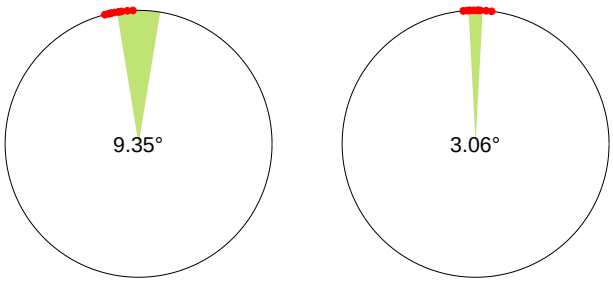
Figure 3.2: Front-back confusion. a) shows the raw results of one test-subject with two front-back confusions. b) shows the results as being used for the mean. The grey dots are the original position of the two outliers.

[1] e.g. by never using sounds behind a subject and letting it turn 180° by using a sound at 90° and move it accordingly

In some experiments it might be the case that a subject does not have a high accuracy, but has a high precision. This could be due to many factors; a person might have a personal bias, the software/hardware could be have some deviations but it could also be due to the fact that the device is worn slightly skewed. It should be kept in mind that the interpupillary distance in the HoloLens cannot be calibrated for blind people, since vision is required for this process. Therefore, the device is calibrated once by an experienced user and used by all participants with the same configuration.

Precision

The precision is computed by the mean absolute deviation per participant. This is computed for all tests at once because there are no breaks nor reconfigurations in between of the tests for a single participant. Figure 3.3 shows the accuracy and precision of one participant. In this case the precision was computed using only the values from this test, but in the remainder of this chapter the results from all tests will be taken into account.



(a) Accuracy (b) Precision

Figure 3.3: Example of a relatively low accuracy and high precision. a) shows the accuracy of one test-subject. b) shows the precision computed using the mean absolute deviation. The dots are the same guesses as in a) but shifted around the mean, to illustrate the precision.

Throughout the rest of this chapter, the precision will be reported together with to the accuracy as $accuracy \pm precision$.

Practical optimum

To find out how well someone could score optimally in this experiment, a run of the experiment is performed with the sound sources being visible. The participant was asked to point exactly at the centre of the source before submitting the answer. The absolute mean of the results gives a practical optimal angle of deviation of 0.17°, as shown in Figure 3.4. Combined with the minimum audible angle of 1° found in [Mil58] this gives an idea of the best possible score in this experiment.

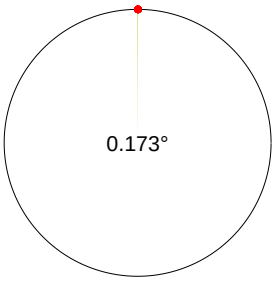
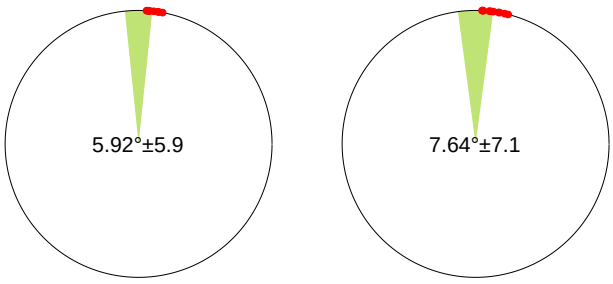


Figure 3.4: Practical minimum score

Base

The accuracy and precision of the first (base) experiment is shown in Figure 3.5a next to the last experiment in Figure 3.5b. The results show that the accuracy does not increase between the first and last experiment.

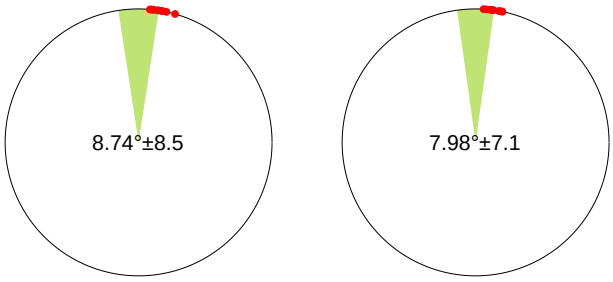


(a) ID: 1, first (b) ID: 8, last

Figure 3.5: Results of the first and last experiment of the participants with full vision.

Virtual Elevation

The accuracy and precision of the experiment set up with sounds coming from a lower and higher elevation is shown in Figure 3.6. The results show that the accuracy is not better for sounds from a lower elevation than for sounds from a higher elevation.

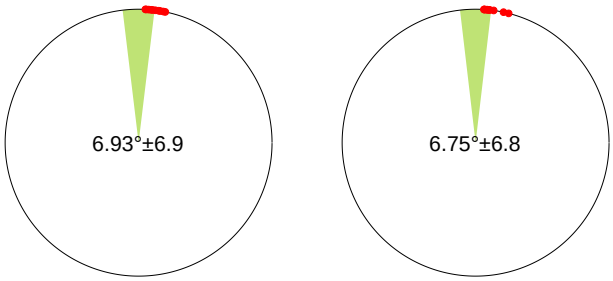


(a) ID: 2, lower (b) ID: 3, higher

Figure 3.6: Results of the experiments with lower and higher virtual location of the sound.

Virtual Distance

The accuracy and precision of the experiment set up with sounds coming from a closer and further point in space is shown in Figure 3.7. The results show that the accuracy does not differ for sounds from a closer point in space and for sounds from a further point in space.

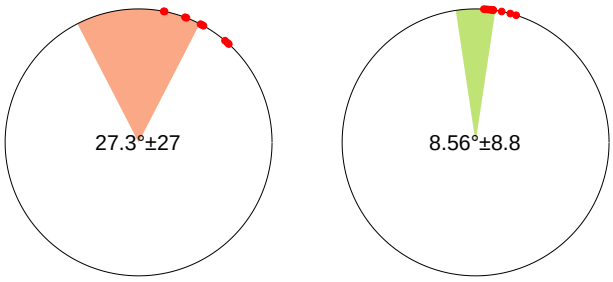


(a) ID: 5, closer (b) ID: 6, further

Figure 3.7: Results of the experiments with lower and higher virtual distance of the sound.

Playing Type

The accuracy and precision of the experiment set up with sounds playing just once and in an interval is shown in Figure 3.8. The results show that the accuracy does decrease a lot when the sound is only played once, but not when played at an interval.

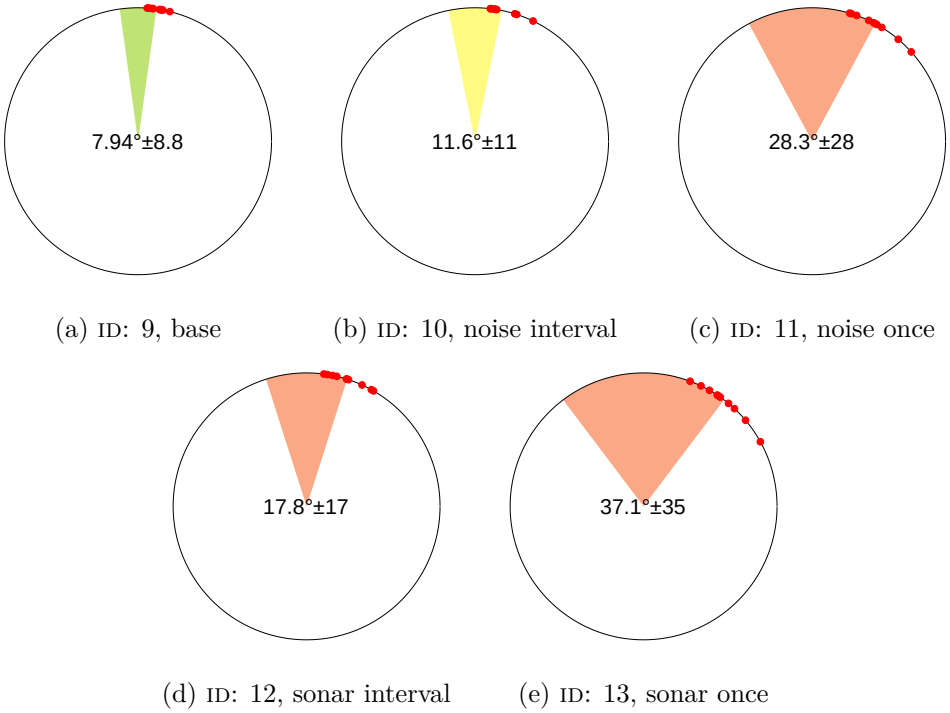


(a) ID: 4, once (b) ID: 7, interval

Figure 3.8: Results of the experiments with different play duration and interval setting of the sound.

Playing Type and Sound

The accuracy and precision of the experiment set up with noise and sonar sounds playing just once and in an interval with visually impaired participants is shown in Figure 3.9. The results show that the accuracy does decrease a lot when the sound is only played once, at an interval and even more when the sonar-like sound is used.

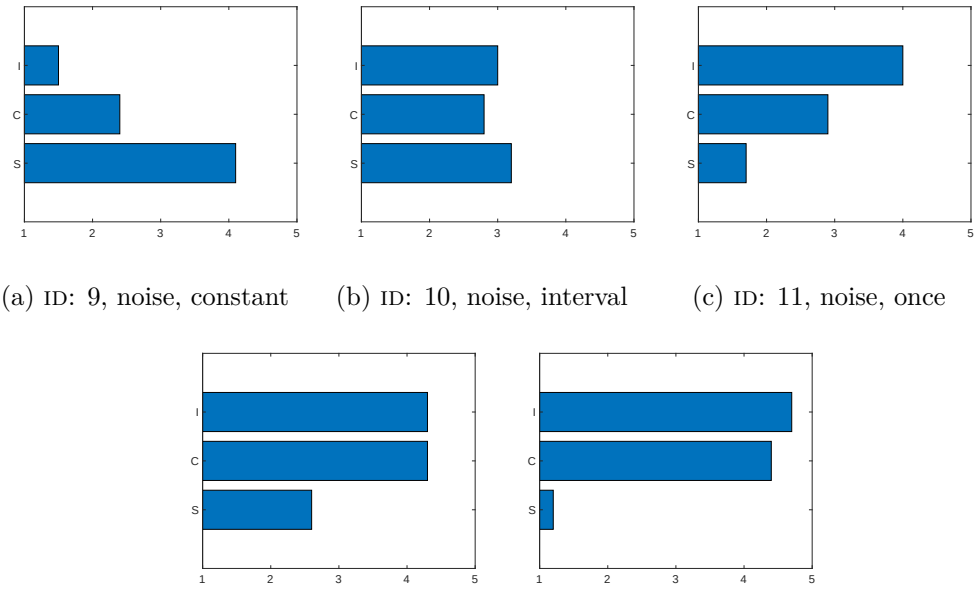


(a) ID: 9, base (b) ID: 10, noise interval (c) ID: 11, noise once

(d) ID: 12, sonar interval (e) ID: 13, sonar once

Figure 3.9: Results of the experiments with different play duration and interval setting of the sonar sound of the visually impaired participants.

The results of the questionnaire are shown in Figure 3.10. As can be seen, the noise sound scores higher on perceived accuracy than the sonar-like sound. However, the sonar-like sound is rated as more comfortable by the participant and it also interferes less with the surroundings.



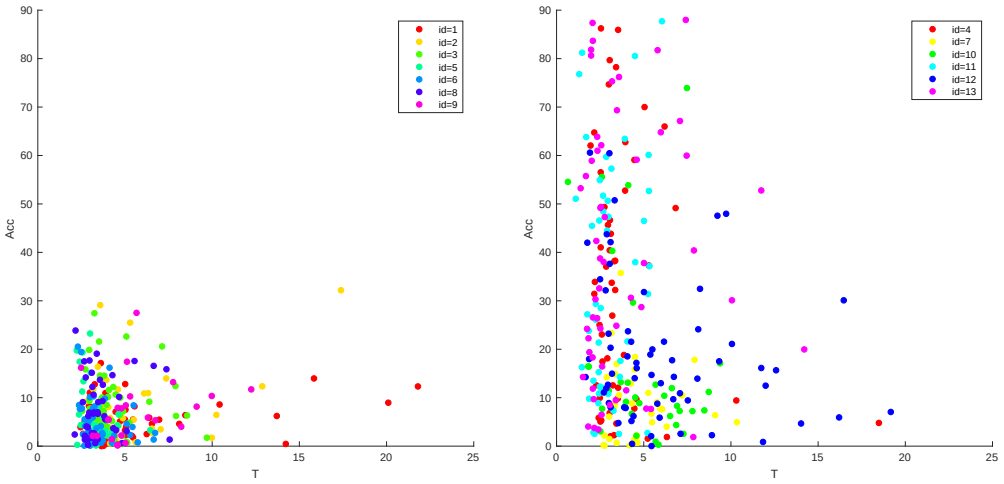
(a) ID: 9, noise, constant (b) ID: 10, noise, interval (c) ID: 11, noise, once

(d) ID: 12, sonar, interval (e) ID: 13, sonar, once

Figure 3.10: Results of the questionnaire regarding the sound settings of the experiments for the visually impaired participants. In the I denotes how much the sound interfered with the surrounding sound, C denotes how comfortable the sounds were and S denotes how well the participants believed they scored. The horizontal axis denotes the averaged rating.

Reaction time

The time taken by all participants the experiments is shown in Figure 3.11. The results in Figure 3.11a show that the participants need more time in the first few experiments to make their guess. Figure 3.11b shows that for sounds played at an interval more time is needed than for sounds that are played just once. However, sounds played at an interval also yields a higher accuracy.

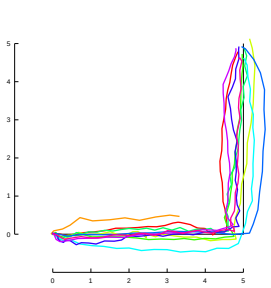


(a) Constant (b) Interval and Once

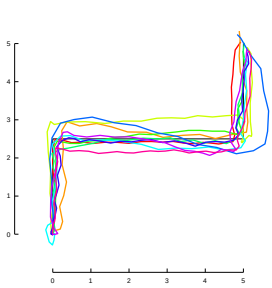
Figure 3.11: Time taken for all experiments. a) shows the experiments with constantly playing sounds b) shows the experiments with sounds playing at an interval and just once. The y-axis represents the accuracy with which the participant was able to guess. Every dot is one guess of one participant. Experiments are grouped by colour. Experiments with ID <= 8 are from sighted participants Experiments with ID > 8 are from visually impaired participants. Experiments with ID 4, 11 and 13 are configured with sounds that are played just once. Experiments with ID 7, 10 and 12 are configured with sounds that are played at an interval.

3.2 Experiment II: Routes

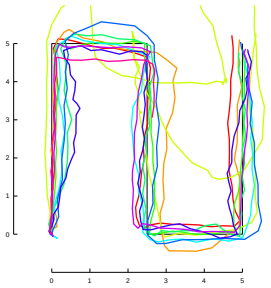
The paths walked by the participants are plotted over the route. Figure 3.12 displays the results per route. Here different participants are coloured differently to be able to distinguish the routes from each other. The results show that more participants get lost when the sonar-like sound is used, while most are able to follow the route when noise is played.



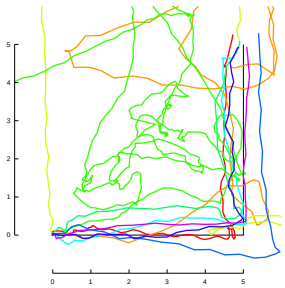
(a) ID: 14, route 1, noise



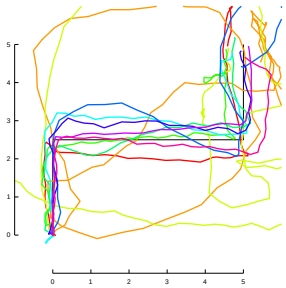
(b) ID: 15, route 2, noise



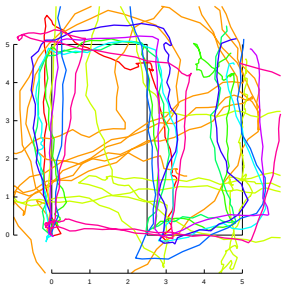
(c) ID: 16, route 3, noise



(d) ID: 17, route 1, sonar, interval



(e) ID: 18, route 2, sonar, interval



(f) ID: 19, route 3, sonar, interval

Figure 3.12: Tracks walked by the participants for each route. Each participant is distinguished by a different colour. The axes denote the grid on which the participants walk where $(0, 0)$ is the starting point and $(5, 5)$ is the end point.

Table 3.2 shows the mean distance travelled and time needed per route. The results show that when using the sonar-like sound, on average, the participants travel twice as much distance and also twice the time.

ID	**Route**	$\mu_{\text{dist.}}(m)$	$\mu_{\text{time}}(s)$	$\min_{\text{dist.}}(m)$
14	1	9.3	20.2	10
15	2	10.4	19.3	10
16	3	21.9	34.4	20
17	1	22.2	58.0	10
18	2	19.6	42.5	10
19	3	37.2	80.3	20

Table 3.2: Results of experiment II. $\min_{\text{dist.}}$ is the shortest distance between the corners of the path. Note that the average distance walked ($\mu_{\text{dist.}}$) can be smaller than $\min_{\text{dist.}}$ because the participant could optimally cut the corners a little. Experiments with ID 14, 15 and 16 are configured with constant noise as a guide and experiments with ID 17, 18 and 19 are configured with the sonar sound that is played at an interval.

3.3 Proof of Concept

The paths walked by the participants are plotted over the route. Figure 3.13 displays the results for the proof of concept. Here different participants are coloured differently to be able to distinguish the routes from each other. The results show that more participants were all able to find the way from start to finish.

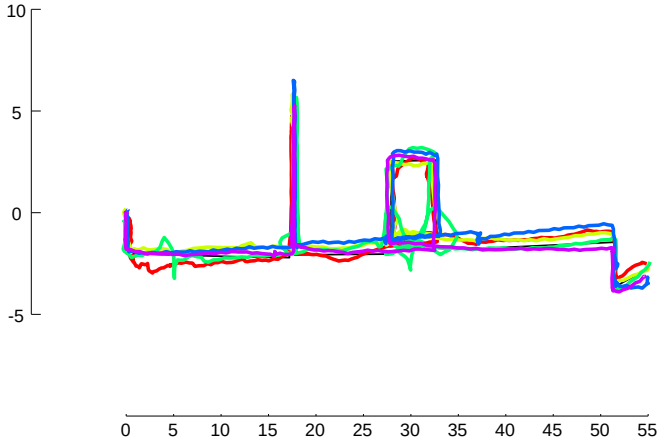


Figure 3.13: Tracks walked by the participant for each route using the proof of concept. Each participant is distinguished by a different colour. The axes denote the grid on which the participants walk where $(0, 0)$ is the starting point and ($\approx (55, 3)$ is the end point.

Table 3.4 shows the distances travelled and time needed to complete the route per participant. The results show that some participants take longer than others, yet still all are able to make it to the end without any problems.

P_{ID}	$\mu_{dist.}(m)$	$\mu_{time}(s)$	$min_{dist.}(m)$
1	100.7	256.0	*90.1*
2	102.2	124.0	*90.1*
3	125.7	160.7	*90.1*
4	100.5	134.9	*90.1*
5	93.7	117.6	*90.1*
mean	104.6	158.6	90.1

Table 3.4: Results of proof of concept. $min_{dist.}$ is the shortest distance between the anchors of the route, which is approxamitely 90.1 meters, but due to world anchors shifting it might differ between participants.

Chapter 4

Findings

In this research, we studied how accurately people can localise spatial sound produced by the HoloLens. Furthermore, we studied how efficiently people can be guided over a path by following spatial sound and we built a proof of concept that can be used in a building to set up a route virtually guided by sound.

It was expected that people could accurately localise spatial sound produced by the HoloLens because it is known that sounds in the real world can be localised naturally and virtual sounds closely match this accuracy. Regarding the efficiency of being able to follow a moving sound no literature was available. It is a novel idea but the expectations were high.

Our results showed that people can consistently localise spatial sounds produced by the HoloLens with a deviation less than 10° and often less than 5°. Because of the possibility to continuously update the guess in a system where the user wants to reach the location of the sound, this accuracy is expected to be good enough to be able to guide a person in that particular direction. Experiment II was used to verify this hypothesis.

Variations in height and distance did not have a positive effect on the accuracy. Regarding the height this can be explained by the fact that people tended to experience that the sound was coming from a high point in space (even when it was actually projected to come from a lower point), which might be due to the fact that the speakers are located above the ear instead of inside it. Therefore, the fact that sound from a lower point in space are easier to localise might not apply in the case that the HoloLens is used. Regarding the distance, the fact that there is no positive

effect could imply that a sound coming from a closer or further point in space only has an impact on the volume and that this has no significant impact on how accurate people can localise the sound. Overall, the results showed that participants do not get significantly better over time, but they do react faster after a few experiments. This is probably due to the participants getting accustomed with the experiments.

The sound used and the way of playing them does have a big impact on the accuracy and reaction time. The best results were obtained by using constantly playing noise. To lower interference with surrounding sounds, a test was done with playing noise at an interval. This way, the noise was not constantly interfering with the surroundings and participants would feel more comfortable while doing the experiment. They turned out to do slightly worse, but they still achieved an acceptable accuracy. However, it also caused a slightly slower reaction time. If the sound was only played once for half a second, the participants would often have no time to move their head, in order to localise the sound better, during the half a second and therefore a much lower accuracy was achieved. The delay in direction time was to be expected, since the participants know they get to hear the sound again if they wait a little bit longer. For the experiments where the sounds only played once this was not the case, since the participants only had one chance to hear the sound anyway.

In order to find a more comfortable sound to listen to, a sonar-like sound was used. The results from the questionnaire showed that the sonar-like sound was a lot more comfortable to listen to and that it interfered less with the surroundings. However, it was rated with a lower perceived accuracy by the participants. This mirrored the actual results, which were a lot worse than when using noise. For future applications, it is important to find a sound with a broad spectrum, which is easy to localise, that is also comfortable to the ears. A “clap” sound was often proposed, since it also works well during lessons for visually impaired children, but after some pilot sessions for this research it was not selected because the sonar sound was rated as more comfortable. If a sound with high accuracy and comfort is found, an interval or *on request* mode could be implemented to direct visual impaired people in the right direction, with less interference than a continuously playing sound. The on-request mode just plays the sound for half a second (or longer, if desired), every time the user requests it (e.g. by clicking a button). This way, the sound is not continuously heard, but the user can request a cue when needed.

The results also show that, without the help of any other devices, a visually

impaired person can be guided by a sound over a path. There were no obstacles in the room during Experiment II, so the participants could walk in any direction, and they were able to stay on track. When using noise, nearly every person was able to complete the route walking the optimal distance at a convenient pace.

The sonar-like sound also performed a lot worse when walking routes. Many participants got lost and had to walk extra meters. Some were not even able to get back on track and were not able to finish the path. This has show that the sonar-like sound is not suitable for this application.

To illustrate the potential of this approach, a proof of concept was developed that could guide a person over a route through a building. It was the task of the participant to follow the sound without any other cues or help. Since the system is not yet reliable enough to use for a visually impaired person, without the help of any other devices, the participants were people with a fully functional vision. The results showed that every participant was able to walk the correct route and make it to the end. Participants took different approaches; a person might walk fast but make a lot of mistakes and therefore walk a lot more meters. Another person might walk slower but make fewer mistakes. One person ended up in the wrong room (on the other side of the wall), which was probably due to a shift of the spatial mapping of the HoloLens. An important thing to note was that participants were unable to hear any instructions during the experiment, which shows the high level of interference.

The HoloLens continuously updates the spatial mapping by scanning the environment using the available sensors. After the route to be walked was set out, the waypoints were anchored to the spatial mapping using so-called WorldAnchors. Because of the fact that this mapping is updated continuously, these anchors sometimes shift up to over a meter.

Concluding, the experiments have shown that a person can accurately localise and follow a virtual sound in a building. The convenience with which the participants completed the experiments shows that many opportunities lie in using similar ways of guiding people in future applications.

4.1 Perspectives

Guiding a person through a public building by using a sound as described in the proof of concept is just the first step. This system could be extended with obstacle detection and wall detection as they did in [BCS+15] to make users even less dependent on their white cane. This way the user could obtain knowledge about the surroundings, while being guided using the virtual sound. To do this in a sensible way, the spatial mapping accuracy of the HoloLens should be researched.

Nowadays, most people carry a smartphone around twenty-four seven, this makes smartphone applications a viable option for providing solutions for problems of visually impaired people without the need of additional hardware. Two examples of applications that can help blind people to identify objects are TapTapSee [Tap16] and Aipoly Vision [Aip17]. These applications classify objects found on the smartphone's camera footage using artificial neural networks and tell the user what is found by reading it out loud. TapTapSee makes use of the CloudSight.ai image recognition API, but technical details are not disclosed. Aipoly developed its own deep-learning engine that runs on a phone without the need for internet, which makes it particularly useful when not at home. If we could apply this kind of technology into a platform that uses the HoloLens' spatial sound system, it could not only tell you *what* it is, but also *where* it is, by just playing the sound from the location of the object. Vuforia is an augmented reality SDK, already available for the HoloLens, that makes it possible to track image targets in real time. This could be used to, for example, recognise the symbol for the toilet, or other symbols.

Microsoft Cognitive Services[1] is a platform consisting of a set of APIs, SDKs and other services in the field of machine learning. Some of the services that it offers include facial recognition, speaker recognition and even emotion recognition. This is technology that could greatly improve the amount of information that a visually impaired person gets from the surroundings. If this was built into an application, it could not only tell you *who* a person is that is walking through the room, *what* this person looks like, but even whether he is happy or not. Combining this with the technology of spatial sound

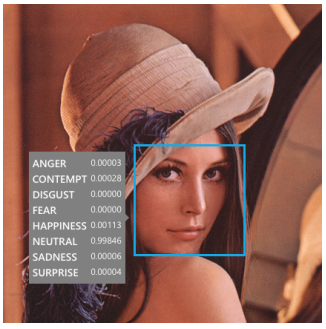


Figure 4.1: Example of expression recognition by Microsoft Cognitive Services

[1] https://azure.microsoft.com/en-us/services/cognitive-services/

it, the device could play the sound telling the user all this information from the location of this person (possibly even in a particular voice) and therefore the user would immediately know where he is. Altogether, the Microsoft cognitive Services offers a great variety of possibilities that could very well be used in as AR platform to aid the visually impaired.

An interesting idea to use the HoloLens' display to help the visually impaired is to show the user an enlarged version of text recognised by the camera. Or even an enlarged version of an object that is in front of the user. This way, it might be easier for the user to take in the information displayed in front of them. Another option would be to overlay important objects with a hologram in a colour with a higher contrast. This way, we could even help colour blind people, since it might be easier for these people to distinguish objects or colours.

Another example that can be used on a smartphone is the platform Be My Eyes [Be 17]. Be My Eyes sets up a video connection between a blind person and an available volunteer, such that the blind person can ask for help with small tasks as, to check an expiration date of a certain product or to recognise which aluminium can contains the right ingredient needed for cooking. Be My Eyes enables anyone to help a blind person from his home. This is another option that could be integrated into a platform aiding the visual impaired. The native Skype app for the HoloLens already offers some functionality that make it possible to see what the user sees and many companies are interested in applications in the field of remote assistance for the HoloLens and other AR systems. For example, the augmented reality helmet from DAQRI[2] focuses for a large part on calling in remote expertise. But one could also think of other applications, for example in the world of surgery, where experts are often uncommon.

A final application that could be developed using spatial sound to aid the visual impaired that came up in one of the brainstorm sessions with Visio is in the field of gamification. During one of the first times paying a visit to one of the Visio special schools it had snowed. Therefore they could not practice crossing the streets, because it was too dangerous for the students. An augmented reality game that simulates cars passing by could make it possible to practice indoors and prevent dangerous situations. This system would also make it possible to let students get in touch with situations that are not available near their schools, since these could be simulated by the system.

[2] `https://daqri.com/products/smart-helmet/`

All in all, this research has shown that using virtual spatial sounds as informative cues for the visually impaired has a high potential. The HoloLens' sound system is already accurate enough and its mapping system could be improved with existing technology to be too. A future version of a system using this technology could physically be much smaller and lighter than the current version of the HoloLens, but technologically have a much bigger impact than it has now. All efforts currently working on improving the independence of the visually impaired could be combined in one large platform, providing people with all the information they need, about directions, recognising people and signs on the fly, asking for help when needed, when wanted.

References

[AHO07] Malika Auvray, Sylvain Hanneton, and J Kevin O'Regan. Learning to perceive with a visuoauditory substitution system: localisation and object recognition with The Voice. *Perception* 36.3 (2007), pp. 416–430.

[Aip17] Aipoly. Vision Through Artificial Intelligence. 2017. URL: http://www.aipoly.com (visited on 03/03/2017).

[Bar14] Chris Baraniuk. Headset lets blind people navigate with sound. *New Scientist* 224.2995 (2014), p. 22.

[BCS+15] Simon Blessenohl et al. Improving Indoor Mobility of the Visually Impaired with Depth-Based Spatial Sound. *ICCV-ACVR workshop*. Dec. 2015. URL: https://www.microsoft.com/en-us/research/publication/improving-indoor-mobility-of-the-visually-impaired-with-depth-based-spatial-sound/.

[Be 17] Be My Eyes. Lend your eyes to the blind. 2017. URL: http://www.bemyeyes.org (visited on 03/03/2017).

[COC06] Sang Jin Cho, Alexander Ovcharenko, and Ui-pil Chong. Front-Back Confusion Resolution in 3D Sound Localization with HRTF Databases. *Strategic Technology, The 1st International Forum on*. IEEE. 2006, pp. 239–243.

[Dav17] Javier Davalos. White Cane. 2017. URL: https://www.microsoft.com/en-us/store/p/white-cane/9p96g5q8sqgc (visited on 03/28/2017).

[Geo17] Geodan. Geodan en Bartiméus ontwikkelen virtuele geleidelijn. 2017. URL: https://www.geodan.nl/virtuele-geleidelijn/ (visited on 03/22/2017).

[Goo14] Google. Project Tango. 2014. URL: https://google.com/atap/project-tango/ (visited on 04/05/2017).

[Hav10] E.M. Havik. The functionality of route information for visually impaired and blind people. 2010. URL: https://www.zonmw.nl/nl/onderzoek-resultaten/gehandicapten-en-chronisch-zieken/programmas/project-detail/inzicht/the-functionality-of-route-information-for-visually-impaired-and-blind-people/verslagen/ (visited on 02/14/2017).

[HBM+13] Alastair Haigh et al. How well do you see what you hear? The acuity of visual-to-auditory sensory substitution. *Frontiers in psychology* 4 (2013).

[HVV+98] Paul M Hofman, JG Van Riswick, A John Van Opstal, et al. Relearning sound localization with new ears. *Nat Neurosci* 1.5 (1998), pp. 417–421.

[HW74] Jack Hebrank and D Wright. Spectral cues used in the localization of sound sources on the median plane. *The Journal of the Acoustical Society of America* 56.6 (1974), pp. 1829–1834.

[Kpm15] A. Kpman. Meet the award recipients of the first Microsoft HoloLens academic research grants. 2015. URL: https://blogs.windows.com/devices/2015/11/11/meet-the-award-recipients-of-the-first-microsoft-hololens-academic-research-grants/ (visited on 02/28/2017).

[Lan02] Erno Hermanus Antonius Langendijk. Spectral cues of spatial hearing. PhD thesis. TU Delft, Delft University of Technology, 2002.

[LBH+05] H Limburg et al. Avoidable visual impairment in The Netherlands: the project "Vision 2020 Netherlands" of the World Health Organization. *Nederlands tijdschrift voor geneeskunde* 149.11 (2005), pp. 577–582.

[LD+15] JC Lee, R Dugan, et al. Google project tango. 2015.

[LK09] Hans Limburg and Jan EE Keunen. Blindness and low vision in The Netherlands from 2000 to 2020-modeling as a tool for focused intervention. *Ophthalmic epidemiology* 16.6 (2009), pp. 362–369.

[Mei92] P.B.L. Meijer. An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering* 39.2 (1992), pp. 112–121.

[MG91] John C Middlebrooks and David M Green. Sound localization by human listeners. *Annual review of psychology* 42.1 (1991), pp. 135–159.

[Mic16] Microsoft. HoloLens. 2016. URL: https://www.microsoft.com/en-us/hololens (visited on 04/15/2017).

[Mic17a] Microsoft Research. HoloLens for Research. 2017. URL: https://www.microsoft.com/en-us/research/academic-program/hololens-for-research/ (visited on 02/28/2017).

[Mic17b] Microsoft Support. Use gestures. 2017. URL: https://support.microsoft.com/en-us/help/12644 (visited on 06/20/2017).

[Mic17c] Microsoft Support. Use the HoloLens clicker. 2017. URL: https://support.microsoft.com/en-us/help/12646/hololens-use-the-hololens-clicker (visited on 10/17/2017).

[Mil58] Allen William Mills. On the minimum audible angle. *The Journal of the Acoustical Society of America* 30.4 (1958), pp. 237–246.

[MK94] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77.12 (1994), pp. 1321–1329.

[MM90] James C Makous and John C Middlebrooks. Two-dimensional sound localization by human listeners. *The journal of the Acoustical Society of America* 87.5 (1990), pp. 2188–2200.

[Org14] World Health Organization. Visual impairment and blindness. WHO Fact Sheet No. 282. Aug. 2014.

[Pau16] Jamie Pauls. Project BLAID: Toyota's Contribution to Indoor Navigation for the Blind (2016).

[Rob15] A. Robertson. Microsoft's HoloLens is new, improved, and still has big problems. 2015. URL: https://www.theverge.com/2015/5/1/8527645/microsoft-hololens-build-2015-augmented-reality-headset (visited on 03/10/2017).

[Sim90] J Richard Simon. The effects of an irrelevant directional cue on human information processing. *Advances in psychology* 65 (1990), pp. 31–86.

[STG+06] Jaka Sodnik et al. Spatial sound localization in an augmented reality environment. *Proceedings of the 18th Australia conference on computer-human interaction: design: activities, artefacts and environments*. ACM. 2006, pp. 111–118.

[Tap16] TapTapSee. About TapTapSee. 2016. URL: `http://taptapseeapp.com` (visited on 03/03/2017).

[Toy16] Toyota. Project BLAID. 2016. URL: `http://www.toyota.com/usa/story/effect/projectblaid.html` (visited on 04/05/2017).

[WAK+93] Elizabeth M Wenzel et al. Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America* 94.1 (1993), pp. 111–123.

[WM10] Jamie Ward and Peter Meijer. Visual experiences in the blind induced by an auditory sensory substitution device. *Consciousness and cognition* 19.1 (2010), pp. 492–500.