# Using computer vision to aid navigation for people with visual impairements

Kai Ferdelman

March 2021

# Chapter 0

# Abstract

There are many people suffering from visual impairments around the world who still rely on canes and guiding dogs to help them navigate outside. Current navigation methods are however flawed and don't take advantage of strides in technology that could allow for better navigation. In this project a team of three students of the university of Twente will attempt to develop a navigational aid for people with visual impairments, using computer vision and haptic feedback. Unlike many attempts made before the development and design process will be performed in close collaboration with workers from that sector and potential end users themselves.

# Contents

# Chapter 1

# Introduction

A 2008 estimate found that in the Netherlands alone an estimated 311 000 people are suffering of some type of visual impairment. Of these about 77 000 are legally blind. Since then these numbers are estimated to have increased by 18 percent, putting them at 367 000 and 91 000 people respectively. This disability affects nearly every aspect of these peoples lives, however especially their navigation ability outdoors is negatively effected. Without their sight it becomes incredibly difficult to sense obstacles and navigate through them, especially if the surrounding is new and unknown to them. This situation can be even worse when there are other pedestrians around them, who might not pay attention themselves. This all together put anyone with a visual impairment at serious risk.

While this problem is not new, there have not been many 'new age' solutions to it. Still the best walking aid for the visually impaired is either a cane or a guide dog, but both of these options come with their own problems. While dogs can be trained very well to assist a visually impaired person and can help them in much more than just navigation, not everyone can keep a dog. This can be due to allergies, animals not being allowed in their living space or simply the fact that a dogs upkeep might be to expensive. Using a guide cane instead precludes a user from all these problems and is with reason the most popular navigation aid used by the visually impaired. While it is a simple and cheap solution it is far from perfect. A cane can only detect obstacles that are right in front of its user and only if it is on level with the ground. This means that a person who is using a cane is more likely to miss an obstacle and walk into it or be hit from the side. Additionally this person would not be able to sense farther ahead, leaving them open to sudden surprises.

Currently the best there is to offer, are improved guiding canes and new special techniques such as echolocation. Echolocation can be very effective at overcoming the previously mentioned flaws of the other aids, but again not everybody can use it due to bad hearing and for those that could learn to use it, it takes years to master, if they are so lucky to find a teacher. Improved guiding canes often feature a small scanning device at the bottom end of the cane that can detect obstacles at a greater distance and in multiple directions and give its user feedback on any obstacles either trough an earpiece or through haptic feedback incorporated into the canes grip.

## 1.1 Reserach Questions

Currently existing devices show a trend into the right direction, using small sensors and smart devices for further assistance, however I believe this can be taken further. Using RGB and depth cameras and more sophisticated wearable haptic feedback devices a person who is visually impaired could be made much more aware of their surroundings, improving their navigation skills and keeping them safer all without using their sense of hearing. To guide the project the following question must be asked:

*How can a haptic wearable be developed to enhance the navigation capabilities of people with visual impairment?*

To answer this, first another question must be answered though:

*What are the shortcomings people with visual impairments face in naviga-*

*tion?*

To achieve an implementation of this idea, the project will be split up into four distinct parts. First, there is the sensing part that focuses on the sensors used to detect the surroundings and process them accurately. Second, there is the haptic feedback wearable that using small actuators informs its wearer about their surroundings. Next there is the interface that translates the information from the sensing unit to the haptic wearable in a way that the user can understand the information. Finally various scenarios will be created for virtual reality that can be used to test the haptic wearable in a safe manor. Each of the parts will be worked on by a separate student, with some coordination between them to make the parts work together.

This project and thesis will focus on the sensing part. To help with the design of the sensor unit, the following question can be used to guide the choices:

*How would a wearable computer vision system need to be designed to detect important features to aid somebody with a visual impairment?*

To answer this question, the following question need to first be answered:

*What sensory inputs best contribute to a computer vision based navigation system for people who are visually impaired?*

*How can a wearable computer vision based navigation system for people who are visually impaired be designed to encompass comfortable and irritation free use?*

This thesis consists of multiple chapters, first delving deeper into currently existing devices and solutions that might be helpful in developing a helpful aid. Next a basic concept for the developed device will be proposed, followed by a further Ideation chapter, supported by expert interviews and a focus group. Once a concept has been accepted the thesis will detail the development and evaluation process. Finally the thesis will end with a conclusion and some recommendations for further development and research.

# Chapter 2

# State of the Art

The following Chapter consists of a literature review on five different topics. The first is about currently existing 'advanced' navigational aids for people with visual impairments, what they focus on and why. This section is followed by a section on modern sensing technologies, followed by object detection and classification, followed by Mapping and tracking techniques such as RGB-D Slam. Finally the last section focuses on wearable sensor units and how they are best used. This chapter is looking to give some insights and answer some of the proposed research questions. While the main research question about the development of a navigational aid for people with a visual impairment, will be treated throughout the project and the first sub question will be answered through background research and interviews, sub questions SQ-2, SQ-3 and SQ-4 should be to some degree answerable through some research on the state of the art of current hardware, software and methodologies. The questions respectively concern "How [...] a wearable computer vision system [would] need to be designed to detect important features to aid somebody with a visual impairment?", "What sensory inputs best contribute to a computer vision based navigation system for people who are visually impaired?" and finally "How can a wearable computer vision based navigation system for people who are visually impaired best be designed to encompass comfortable and irritation free use?".

Especially when focusing on SQ-3 and SQ-4 some useful information should be found, that can help answer SQ-2 and later can flow into the design of the device.

## 2.1 Existing navigational aids for people with visual impairments

There are many good reasons that people with visual impairments are looking towards advanced guidance aids. As mentioned by Ruxandra Tapu et al. [1] the currently most used obstacle detection tool is the famous white guiding cane. The cane in combination with memorizing a surrounding, according to Ruxandra Tapu et al. [1] is the only way of navigating successfully, while in an unfamiliar surrounding they are completely lost and at the mercy of others around them to reach their desired destination. While both Ruxandra Tapu et al. and Darius Plikynas et al [2] agree that GPS can provide invaluable information on the position of the user, they also agree on the shortcomings GPS faces in lacking accuracy, especially when inside, and its lack of obstacle classification ability. To counter this they agree that some type of additional input is needed to assist the navigation.

Shang Wenqin et al. [3] expands on the shortcomings of both established and modern navigational aids by classifying 3 problem groups. The first of this group is mentioned as having a restricted detecting scope. This means that a device does not have the needed types of sensors or the processing power to detect and possibly classify an obstacle in a short enough time span to permit effective mobility. The second group is defined as being unable to fully sense the spatial environment. This again could be due to missing sensing abilities, or could be due to poor placement of the sensors. Whatever the reason is this group of devices will commonly miss obstacles due to their spatial positioning, such as a hanging object. The final category is lacking a robust orientation approach. A device in this category will be missing the capability to realize its location on a larger scale.

Beginning on the simpler side of devices, there are many attempts to improve the capabilities of the basic guidance cane. Solutions such as presented by Sung Jae Kang et al. [4] use additional sensors built into the base of the stick to trace the movement of the stick and its user and adds additional detection capabilities for uneven ground using ultrasonic sensors. While this is proven to improve mobility its issues fall into the second category of devices shortcomings as mentioned by Ahang Wenqin et al. [3], as the advanced cane still misses the capacity to detect any obstacles that are raised from the ground. Additionally the device falls into the third category, as it only provides local information and also fits into the first category of shortcomings, as it is incapable of helping to understand abstract situation, such as recognizing stairs or an empty seat.

Slightly more advanced are devices such as the one presented by Ruxandra Tapu et al. [1]. These devices utilize a basic camera to detect and classify both static and dynamic obstacles. By using elegant algorithms this device is able to process an incoming stream of images without much delay, thereby passing Wenqins second problem. Despite this the device still does not give any non local information and thereby falls into the third issue group. Additionally the device does not offer any service but obstacle avoidance, again missing out on abstract situations, placing it also in the first issue category. While this approach seems to have a lot of drawbacks it presents some advantages as well that arise from its simplicity. Due to the system only needing a video stream, the software can be run on a smartphone, which makes the device extremely portable and cheap. On this Darius Plikynas et al. [2] expands briefly adding to the advantages of smartphone based systems, mentioning that they are already in use as an accessory by most people, especially the youth, and that acceptance of these devices is especially high compared to other system. Additionally with the expansion of 5G networks a phone can be used as sensor only and wirelessly relay the data to another device for improved processing.

The most advanced navigational devices are those combining multiple sensors and providing the system with enough processing power to give real time feedback. Devices, such as the one presented by Young Hoon Lee et al. [5], often use RGB-D cameras in combination with dedicated computers and different types of haptic feedback devices. This type of aid of course collects the most information and gives a much more complete picture of the surroundings. Due to the sensing capabilities these devices are able to mostly sidestep the issues mentioned by Wenquin et al. [3]. With its camera it is able to track and avoid both stationary and dynamic obstacles, it can be aware of where the user exactly is and it can be designed to understand abstract situations. Young Hoon Lee et al. [5] mentions however that these systems can come with some drawbacks. Due to the amount of data being processed these devices need a large processing capacity, which makes the device bulkier and more expensive.

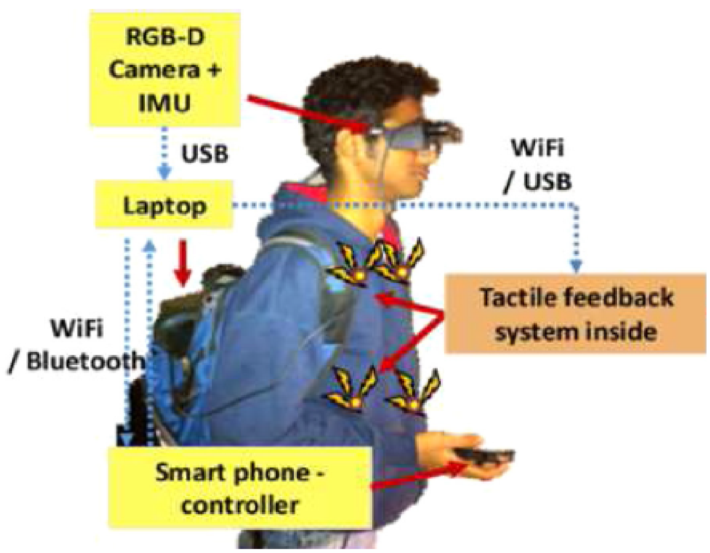



Figure 2.1: RGBD Haptics device by Young Hoon Lee et al. [5]

## 2.2 Sensor Devices

As mentioned before there are many different types of sensors that can be used to gather the needed data. According to Plikynas et al. [2] the usable solutions can be classified into two groups: sensor based and video camera based. Both the sensor classification and the camera classification have a lot of differences within them, however the sensor group is broader, giving more options to descide between.

In the domain of sensors there are many options. On the lower end there are ultra sonic sensors such as those used in devices such as the one by Kang et al. [4]. These simple sensors use ultra sonic sound waves to calculate the distance to the first obstacle that they are pointed at. While these sensors can be very small and cheap they in many ways are lacking for applications like these, as they do not monitor a large area making it likely for them to put a device they are use on into the second issue category brought up by Wenqin et al. [3].

A step up are the more advanced range finding sensors such as radar and lidar. Both of these sensors can use radio waves to scan a larger area. While radar has the potential to detect obstacles and calculate the range to them, its accuracy is lacking, due to which it would not be able to recreate an accurate surrounding. Lidar however can achieve much higher precision, which can be used to create a point cloud, effectively a recreation of the sensors surroundings. This type of sensor has in recent years seen much development and more use largely in robotics but also in navigational aids, such as the one developed by Michael Miles et al. [6]. While Lidar sensors can have a great range and accuracy, they are still very expensive and typically also very large. This of course makes them much less useful for a wearable navigation solution.

On the other side of Plikynas et al. [2] classification are the camera based solutions. Cameras can be found in many different forms, but many can be differentiated as RGB cameras or depth cameras. Regular RGB cameras come with much less functionality as they only provide a 2D image, however they are much cheaper and simpler than their counterpart. Additionally, RGB cameras are already built into nearly every smart phone, which makes them attractive to use as no new hardware is needed. While RGB cameras struggle with depth perception they can still be used effectively though, when combined with effective image recognition software, as has been demonstrated by Ruxandra Tapu et al. [1] with their smartphone based system.

Within the group of depth cameras on more separation can be made between the three processes of 3D calculation.

The first approach is using Time of flight calculations. By lighting up the surrounding of the sensor with light invisible to the human eye, the camera can determine for each of its pixels how long it took the light to bounce of an object and return to the camera sensor. With this information it can calculate the distance at each pixel. According to Jos´e Gomes da Silva Neto et al. [7], these cameras are especially effective outside, as sunlight has little to no effect on the performance. Despite this, ToF is rarely used in combination with depth cameras and is typically only found in combination with lidar sensors.

The second category of depth cameras is the so called structured light sensor. It uses a small beamer to place an intricate pattern of invisible light on the surrounding of the sensor. With the sensor being aware of the pattern it can detect distortions in the pattern, caused by changes in the surrounding. From this information it can calculate a point cloud to represent its surroundings. This method is a favourite for depth cameras as it is precise but computationally still quite simple. Neto et al [7] warn however that this type of sensing can quickly have problems when detecting complex structures on which the pattern might be obstructed to much.

The final category is the active stereo camera sensor. Instead of relying on its own light source to illuminate the surrounding, it uses two cameras that are spatially distant from each other. By comparing the two images produced by the two cameras the system can calculate the depths of the surroundings in a similar way to how animals use binocular vision. This process is sometimes improved though by again using invisible light to highlight key points. Neto et al. [7] mention that this process can collect the most accurate data even at distance, but it also needs the most processing power.

## 2.3 Image Processing

The collection of depth data is of incredible value for the mapping of the terrain detection of stationary obstacles. However for the detection of objects, whether moving obstacle or specifics, such as doors or a light switch, an RGB image has to be processed using an object recognition algorithm. There are some amazing object detection services by providers such as Google cloud and Microsoft azure. These services are run on dedicated servers and use machine learning and are train on massive data sets, leading to high success rates. These services have two major drawbacks however: The first is, that these services are not free and with their per use charge are not suitable for a cheap personal navigational aid. The second reason against these services is, specifically that they are run on remote dedicated servers. While this centralization of data and processing improves the efficacy of the service, it also requires any device making use of it to be permanently connected to the internet. For an indoor application this might be possible, any outdoor application would suffer under this restriction.

Less powerful object detection methods can however be run successfully on a local device. In the method presented by Chongyi Li et al. [8] a combination of RGB images and depth images is successfully used to detect objects within the frame and rank them on their perceived importance, filtering out any background noise. The proposed ASIF-NET algorithm proves to accurately detect the most significant objects in frame versus the ground truth. This of course is only on step in the process though. To accurately tell what the system sees, the detected object has to also be classified. As proposed proven in the research of Imania Ayu Anjani et al. a well trained convolutional neural network (CNN) is well suited to process limited data input to classify the content into a preset list of options. By first selecting all objects using Li's et al. ASIF-NET [8] algorithm and then feeding the output of that to a CNN, objects can be effectively be detected and subsequently classified with accuracies of up to 96 percent according to Anjani et al. [9].

## 2.4 RGB-D Slam

Collecting data from the environment is of course not the only thing a navigational aid must be able to do. In addition to it, the system must be able to analyze and process the gathered data. To work effectively the system will not only need to be able to avoid obstacles, but also localize itself even when there is no GPS signal for it as is common inside. To help with this process SLAM algorithms can be used. SLAM, as presented by Sylvie Naudet-Collette et al. [10], is the simultaneous localization and mapping of the systems environment. By creating 3D point clouds the algorithm recreates the sensors surroundings and when presented with new data either localizes itself within the already known map or adds to it. The SLAM algorithms might differ between implementations, but are all designed around the same central concept.

Algorithms like DP-SLAM, according to Aiwu Sun [11], only work on grid maps, excluding them from work on 3D surroundings, but it is able to correct its generated map over time and keeps errors from accumulating. Other algorithms such as OpenCV RGB-Odometry are specifically build for C++ with OpenCV, making them very efficient, but in RGB-Odometrys case keeps it from creating point clouds. RGB-D SLAM brings an additional feature, by allowing to combine a 3D point cloud with a colored image allowing for a colored point cloud and using all the available data from an RGB-D camera.

In the research by Sylvie Naudet-Collette et al. [10], a further advanced version of RGB-D SLAM, Constrained RGB-D SLAM is discussed. This method couples available 3 dimensional data with the standart SLAM algorithm to reinforce the localization process. Using this improved algorithm can, according to Naudet-Collette et al. [10], reduce drift from nine percent to only three percent. While doing this the algorithm is still able to, on a standard CPU, achieve a frame processing time of only 25ms.

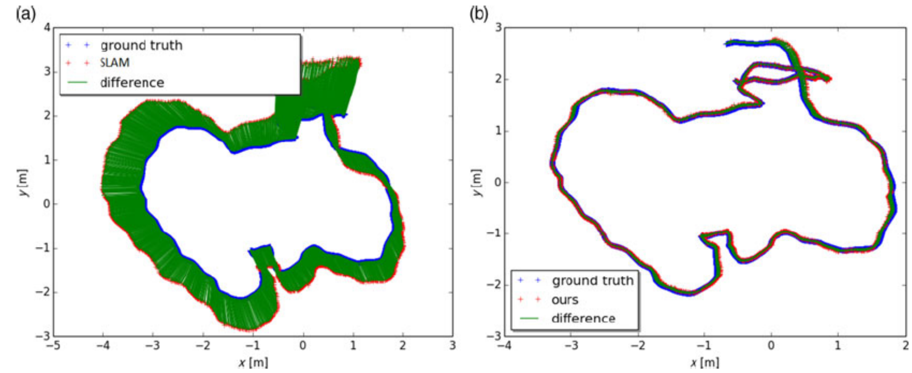



Figure 2.2: Accuracy improvement of Constrained Slam [10]

## 2.5 Wearable sensor arrays

A final consideration has to be made to the placement of the sensor unit. While the cameras can be made quite small, it still is placed on a human and therefore has to be designed with certain aspects in mind. At the same time technical aspects have to be considered though to not waste the potential of the sensing device.

When placing the RGB-D Camera it has to, from a technical standpoint, be placed to minimize the dead zone to create a well defined point cloud. According to Garen Haddeler et al. [**?**], in most current applications the sensors are place intuitively and based on the designers choice. This can however lead to unintended dead zones. Indeed the best location to place the sensor device is as high as possible. This might seem intuitive, but also goes against the common placement on the chest. Specifically when placing the camera on a human the forehead would present a good placement option as it reduces the chance of the users hands or arms getting in the way and also at its greater angle to the ground improves the mapping of it and avoids the risk of having bumps in the ground obstruct the beamers light from returning as warned by Neto et al. [7].

This purely technological standpoint is not enough though when considering a wearable device. The most important design requirements, given by Leire

Francés-Morcillo et al. [12], are found to be in order, comfort ease of use and simplicity. None of these necessarily exclude the forehead, but must all be seriously taken into account. A human limit though is how much a human user can and would be willing to carry on their head. Medically speaking, according to Moen et al. [13] a healthy human can carry up to 20 percent of their body weight on their head without extra exertion or medical issues. This of course exceeds the weight of a small camera by far, however as mentioned by Frances-Morcillo et al. [12] the wearable must also be comfortable. While the is no clear consensus on how much weight is still comfortable, an average hat weighs in at about 150 to 200 grams. Given this any design should not exceed this value by much. Importantly Francés-Morcillo et al. [12] mentions, that there exist no clear evaluation tool for wearability, which means that any design has to be tested thoroughly to be acceptable for the end user.

## 2.6 Conclusion

While there are clearly many options to design a navigational aid for people with visual impairments and there have been a lot of attempts at creating a successful aid, there are currently, according to Young Hoon Lee et al. [5], no standardized or complete systems on the market that are effective. This could be due to many reasons, but is likely due to poor design choices, especially in making it user friendly.

Addressing sub question three "What sensory inputs best contribute to a computer vision based navigation system for people who are visually impaired?", it has become clear that the best type of sensory input is a combination of RBG images and depth data collected respectively by an RGB and depth camera. This data can be combined to effectively detect individual obstacles and if needed the type of the obstacle and on a larger scale can recreate the users surroundings creating a map for point to point navigation.

Sub question four "How can a wearable computer vision based navigation system for people who are visually impaired best be designed to encompass comfortable and irritation free use?", has disappointingly lead to less information, giving a limitation of about 150 grams for a comfortable head mounted wearable, but not giving any indication on how to specifically design for comfort. This will have to be overcome with a prolonged human centered design phase and evaluation, supported by rapid prototyping.

# Chapter 3

# Proposed Development

In this project we are proposing to develop a head band or alternatively a type of smart glasses, each with built in sensors. The sensors would include an RGB camera and a stereoscopic depth camera as found in that combination in the Intel Realsense. To achieve this the Realsense D435i has been selected for its effectiveness and small size. To minimize the devices weight and size on the users head the device will be further disassembled and powered and supported by a small computer that will be either back mounted or carried in an additional bag. The computer will run a Python program implementing Constrained RGBD-SLAM, the ASIF-NET algorithm and a convectional neural network to process all collected data. The collected data streaming through these three parts of the program will need to reveal usable information for both obstacles and objects of interest on their direction, their distance from the user and their type.

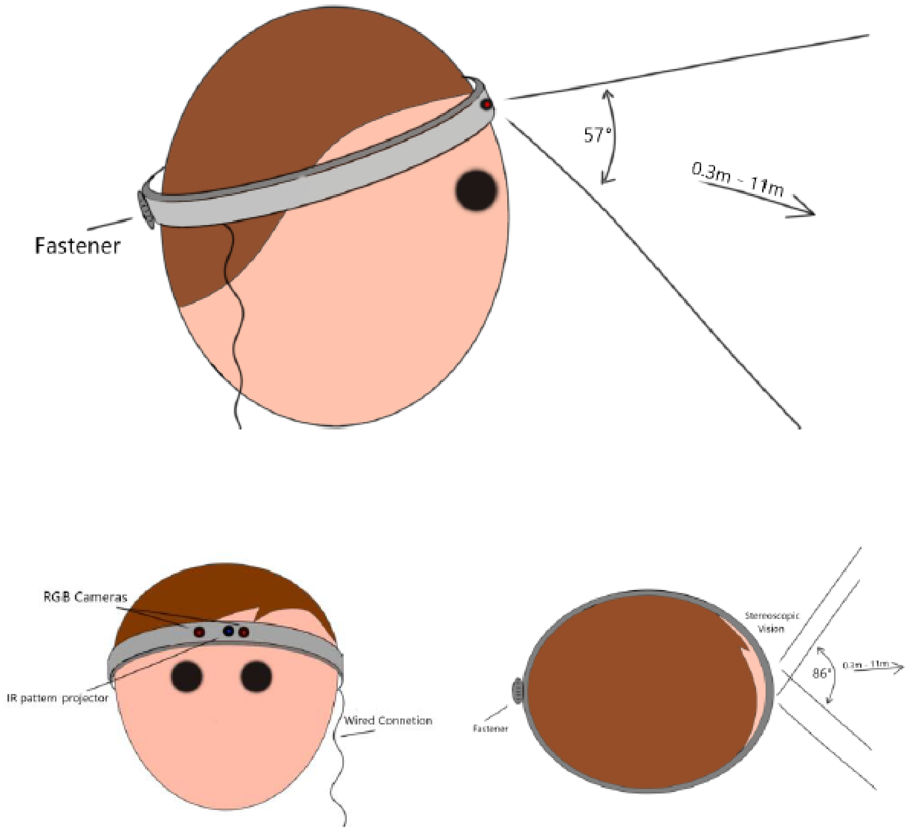



Figure 3.1: Proposed Head band wearable

The hardware and software components are, based on the research on the state of the art, decided upon and will not be changed, however the specific use cases still stand to be picked. This will be done together with a focus group formed of visually impaired participants. Additionally the design of the wearable will be decided upon during the ideation phase and rely on some feedback from potential end users. Final changes to the design may also be made during the evaluation phase, while testing for comfort and general acceptability with test subjects.

# Chapter 4

# Ideation

During the ideation phase, different options for the development of the final device, as a whole, were proposed and explored. For this input from interviews with experts in their field and end users themselves were used.

In this chapter the different proposals and methods will be discussed.

## 4.1 Usecases

To define what circumstances the device should be used in and therefore be developed for, the team conducted interviews with experts in the field of navigating while blind from the Visio organisation NL and further interviews with a range of people suffering from visual impairments themselves.

Firstly, from the interviews, we were able to find that on many occasions before this, groups and companies have attempted to solve similar problems, using modern technology. These groups so far have usually failed at delivering a desirable product. From looking at these products and finding out what held them back we were able to find and therefore avoid pitfalls these devices encountered.

An apparently common mistake is to develop the device as an extension of the users cane. While in theory this is a good idea, the added weight to the cane makes it itself harder to use, increasing the likely hood of missing an obstacle. This leads to the device solving a problem that it creates itself in the first place. This criticism came up often enough during the interviews to completely eliminate such development as a possibility.

Another common issue that came up during the interviews was the so called information overload. That would be, if a device gives the user so much information about their surrounding, that the user would not be able to distinguish between different signals and consequently would miss out on most of the information given to them. This very quickly can occur when sound is used to convey information thereby competing for attention with ambient sounds, but also from haptic feedback if the signals are to complex.

The final issue brought up commonly was the issue of expensive equipment breaking and or being stolen. Especially the breaking, again mostly affecting guiding canes as unobservant cyclists or other participants in traffic would hit the cane and break it, becomes problematic. But also theft seems to be a problem as the clearly valuable equipment and defenseless user can be a tempting target.

With the major pitfalls mostly not directly affecting the environment or the cause for using a navigational aid, focus was shifted to what the most difficult situations are, that a person with visual impairments would encounter. From this we found 3 main problems in which some kind of navigational aid could immensely help its user:

The first case would be a person, who is visually impaired, being 'lost' in their surrounding, due to the fact that there are no clear markers around them, that they would be able to recognize, leading to them being in a sense lost in an environment that a sighted person would be able to navigate. This use case can also be extended to include the inability to sense objects of interest at a larger distance than their cane provides, which can often be a problem when navigating a lesser known environment. By detecting both obstacles and objects that the user might be looking for and passing this information on to the user the person can be given a better sense of their surroundings, improving their confidence and effectiveness in navigating.

The second use case would assume that the user is lost in a surrounding that they don't know or at least don't recognize. This again could be a result of the shortcomings of the guidance cane, but also from the person following some other sense, such as smell, and getting lost. By implementing a checkpoint system, the user would be able to while moving around freely place checkpoints at places they would recognize again if taken there. When they then get lost the device would lead them back to one of these checkpoints, in the process avoiding collisions. From there the user could again move freely to where ever they would want to go.

The final use case, built on the information gathered from the interviews, would help a visually impaired person in their own home or outside. A common issue described by the interviewees was that when dropping an item on the floor or even negligently placing something on a table without noting where exactly, they might not easily find that item again. When this happens, they described how they would either need to ask for help finding it or drop to the floor themselves moving around in increasing circles until they have found the item. To help with this problem the device could be told what to search for and once the item is detected, it could guide the user to it, while again avoiding other obstacles.

After discussing the three use cases with the team, the focus was placed on the first case, concerning the users awareness of their surroundings. This use case was mainly chosen as it seemed to be the best addition to the guiding cane and seemed to fill the most concerning hole in blind navigation that could be found.

In the next step of defining the use cases, five different specific scenarios, were designed, based on more specific feedback from the interviewees. The five scenarios were all designed to fit within the first given use case but all address different smaller issues that could be regularly encountered.

The first scenario concerns a street crossing. No matter if it is only a cross walk or a crossing with a traffic light, these situations can be difficult and possibly even dangerous for a person without sight. While the person might be walking down the street being navigated by google maps they would be able to follow the street, but no navigational service lists where exactly a street crossing is. Due to this, sometimes it becomes difficult for a visually impaired person to cross the a street, when they don't know the area well. By detecting a traffic light and guiding the user toward it, the system can improve the safety of the user and remove the time spent searching for it. The desired behaviour can be seen in figure 4.1.

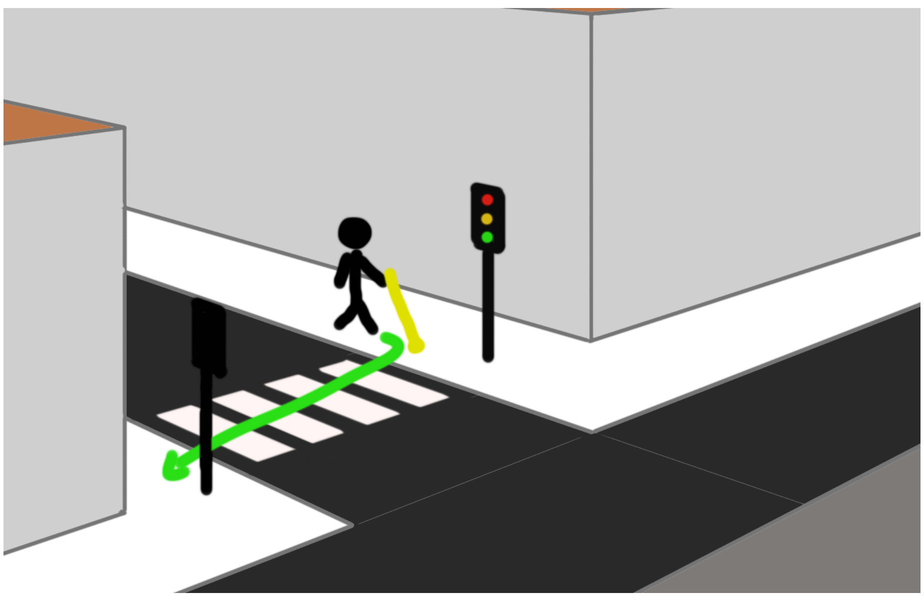


Figure 4.1: Scenario 1: Street crossing

The second scenario is focused on finding a staircase in a public space, when desired. While looking for an object of interest this scenario also has a large focus on accurately detecting walk able areas and keeping the user to these, to keep them out of harms way. To focus on this we decided to settle on a train station platform, as it is a common public place typically including stair cases and also potentially life threatening situations. For a successful navigation of this scenario the user would have to move along the platform without being guided onto the rail tracks and eventually be lead down a set of stairs to remove them from the station platform as can be seen in figure 4.2.

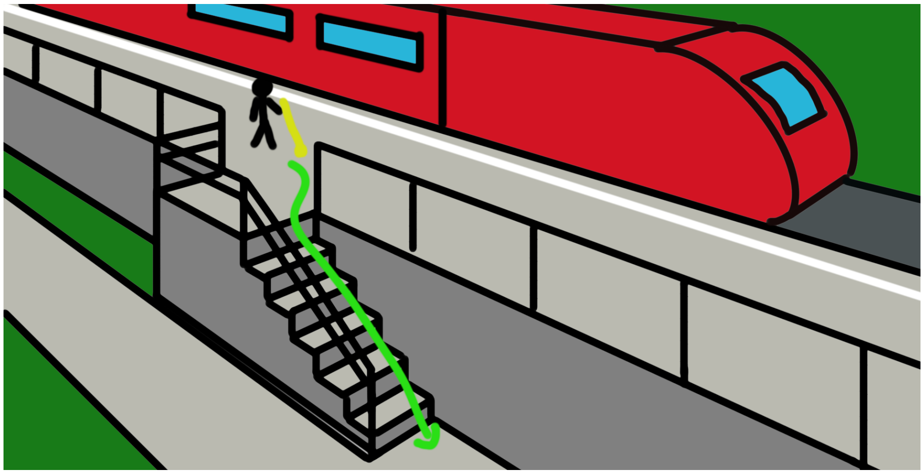


Figure 4.2: Scenario 1: Train station platform

The next scenario is set around the simpler task of keeping the visually impaired user on a side walk moving forward. As can be seen in figure 4.3 the side walk would have a small drop off towards the street while having no easily detectable marker to the other side where it would have a gradual slope down into a ditch. The path in this scenario would be straight and without any further obstacles.

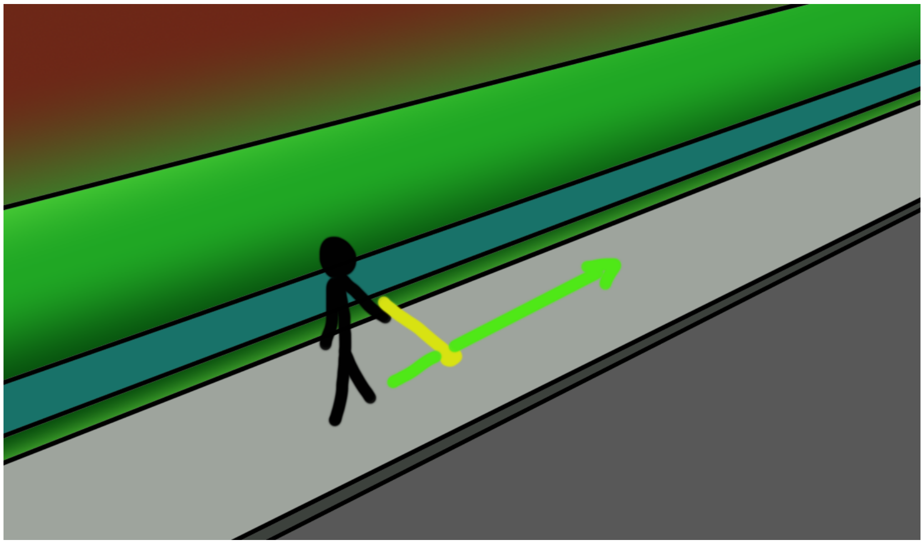


Figure 4.3: Scenario 1: Side walk

The fourth scenario is set in a mall. The idea would be that a visually impaired person needs to move through the wider than usual space of a malls main walkway. The user would navigate by themselves towards where they think the desired store might be, but when close the device would help out in locating and guiding the user towards the stores entrance. As would be expected in a mall there are many obstacles, such as benches, plants and kiosks in the middle of the path. Due to this, the scenario was focus mainly on avoiding obstacles while moving through the mall space. The desired behaviour can be seen in figure 4.4.

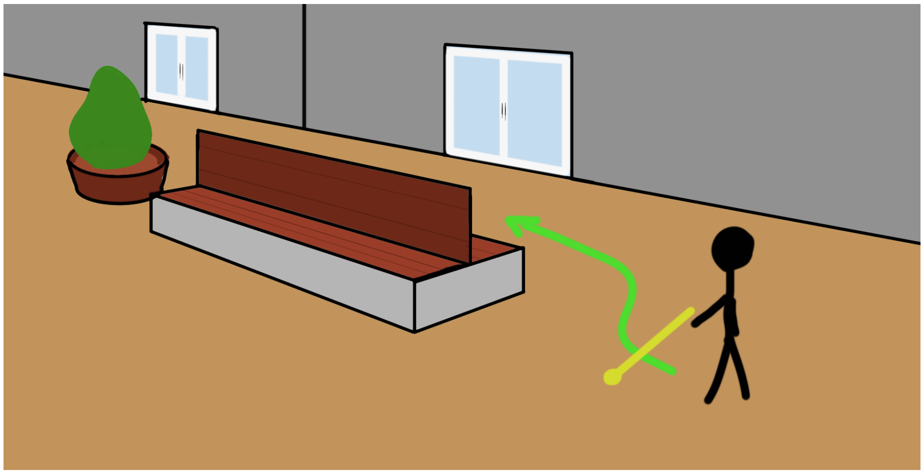


Figure 4.4: Scenario 1: Mall

For the final scenario the visually impaired person would be placed in a park. This park would be made up from grass fields and curved paths through it, as can be seen in figure 4.5. These paths being dirt or sand paths would have fading edges to the grass, which in combination with them being curved would make following them difficult and slow. By informing the user about the slope of the path they can be kept on track without getting lost on a grass field.

Figure 4.5: Scenario 1: Park paths

After creating these five scenarios, the scenarios were inspected to see which would address the worst of the issues, that had been found during the interviews and were presented to another interviewee. By doing so the scenarios were cut down to three, to specify the scope more.

The three scenarios determined to be the most relevant, were the first scenario, the street crossing, the second, the train station and the fourth scenario, the mall.

With this the use cases that the final device would have to be able to successfully guide a user through and achieve the specified goals, were defined. Based on these scenarios we decided to in effect focus on the last five meters, this becoming the common theme throughout this work.

## 4.2 Software

With changes in the overall design and usage of the entire system the software running the the sensing unit had to be adjusted to match as well.

The original idea, as described earlier in chapter 3, saw the software build a virtual map from what the sensor detects over time. This would be done using a SLAM algorithm. The second part of the software, the object recognition, was to built around the ASIF-Net algorithm. The detected object would then be classified using a simple concurrent neural network. This way the system could build a map with both obstacles and objects of interest saved in it.

With the specification of the use case however it became clear that a full SLAM built map would not be necessary as the device is not intended to surpass what a human would be able to detect. Instead a simple short term memory saving a set of previous frames would be much more helpful for correctly assessing objects on the edge of the field of view and to help with filtering the incoming information. By removing the full SLAM process the software would also be able to run at higher speeds or on a smaller device. To further assist in classifying objects the choice was also made to incorporate the full suite of tensorflow, a large machine learning library.

In a further step a switch was made in how the information about objects and obstacles was treated. Rather than calculating their positions and passing a vector on for each of these to the feedback device, a grid map would be created consisting of columns, going outwards from the user, and rows, rings of certain radius centered around the user. By using this approach each grid cell would contain information about whether it is free, occupied or contains an object of interest.

Due to the desire to in some cases detect an object of interest that might be partially obscured by an obstacle, a final change was made to the object recognition. Rather than using the ASIF-Net algorithm which would only be able to detect a single most important object, a switch was made to use the Yolo V4 algorithm to detect and classify up to 50 objects. This would come at a small sacrifice to the systems speed, however it being the fastest object detection algorithm for multiple objects, as can be seen in figure 4.6, it should still be able to satisfy.

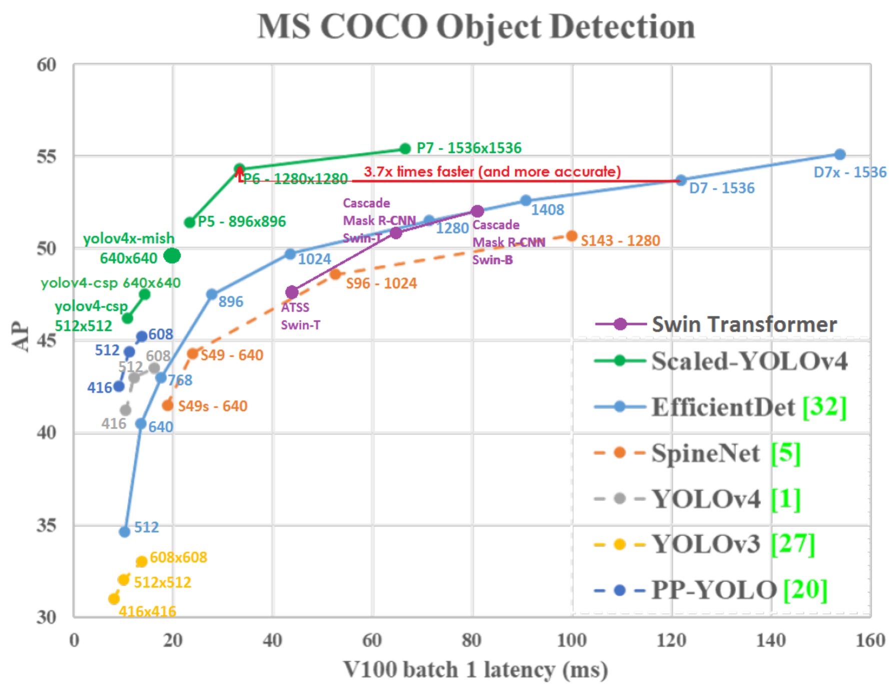



Figure 4.6: Speeds of various object detection algorithms

## 4.3 Hardware

As with the software the hardware, being the headset containing the sensor unit went through multiple iterations of design.

The first version of the headset was inspired by the head strap of the Microsoft hololens seen in figure 4.7. This design provides the flexibility in size to fit all, fits comfortably due to its inner padding and is ideal for mounting any sensor to its front and holding it tight.

Figure 4.7: Microsoft hololens head strap

As for the mount of the sensor there were two possible options as presented in figure 4.8: The first option would mount the sensor at a slightly downwards facing angle onto the headband. The second option would place the sensor directly and straight on the headband with a reflector in front of it, redirecting its view downward. This second method would allow the sensor to be better attached, placed further into the headband and therefore be less exposed, protecting it from the weather, but also collisions with other objects. While the first option would not offer these benefits it would simplify the design and thereby keep the end product at a lower cost. Due to this the first option was chosen to continue with.

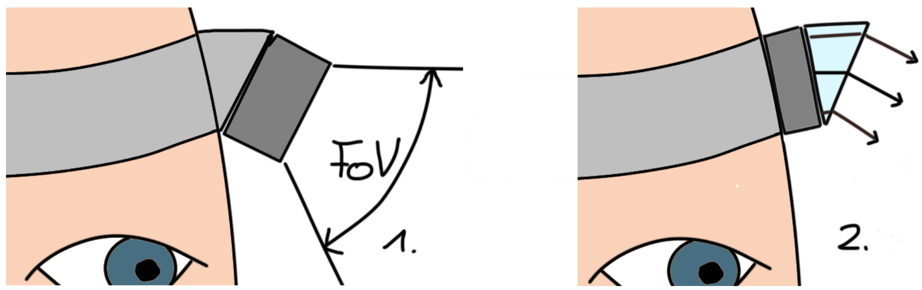



Figure 4.8: Sensor mounting positions

Another functionality that was considered, was the addition of a button to the headset. When pressed the button would deliver information on the object right in front of the user, while suppressing any other object classification information while not pressed. The idea behind this would be to prevent the user from being overloaded with information, as was brought up during the interviews. While this option was briefly experimented with, it was in the end removed again as it did not provide enough benefit.

As the headband for Microsoft's Hololens needs to support a much bigger weight than the here developed headset, it is way to over engineered for this purpose. While in concept it checks all boxes to make a good headband it is for this application simply to big to be justified. Additionally the tightening mechanism featured on the Hololens headband is more complex than it would need to be. Both these issues, that were found could be solved with a combined solution however. By keeping a more robust but padded sensor holder and a counter part for the back of the head and connecting the two pieces with Velcro strips as shown in figure 4.9. By doing so a large part of the headsets weight is eliminated while it can still be easily and effectively adjusted. By removing the tightening mechanism, another weakpoint is eliminated, removing another possibility for breaking.

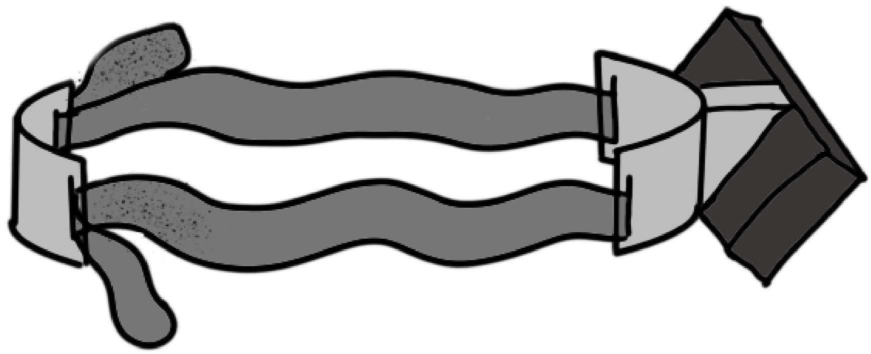


Figure 4.9: Velcro supported headset

## 4.4 Ideation results

As becomes evident from this chapter a well functioning comfortable and most importantly effective sensing headset should be possible to achieve in the realization process. The System, with its changes to the design, should be able to keep up in all the designed scenarios.

# Chapter 5

# Specification

The specification chapter names for both the hardware and software concrete goals that must be achieved to guarantee a successful headset. If the specified values are achieved the minimal viable product would be completed. Any performance better than specified would improve the product and thereby exceed the MVP.

## 5.1 Software

A consideration for the entire software is how far ahead the device should analyse the depth data. The maximal sensing range of the D435i sensor is 11 meters, however based on the use case and its scenarios we can limit the sensing range to five meters, to also fit with the theme of 'the last five meters'. The software should also be able to analyse the full image it receives each frame, equating to a field of view of 89 degrees horizontally and 58 degrees vertically, only reducing the size of the input itself while filtering. If a smaller field of view is desired it should be possible to adjust to this, but the software should be able to handle a full field of view.

The further software can once again be spilt up into Obstacle detection and object avoidance. These two parts, while being part of the same program, must run parallel to each other so that they can each run at different speeds to each other.

The faster of the two parts is the obstacle avoidance. To protect the user from walking into an obstacle it has to run at a higher frame rate than the object detection. An average human walks at about a speed of 1.5 meters per second. Assuming every 10 centimeters a new measurement should be taken, this would require a minimal frame rate of 15 FPS. This does not take into account slight variations in the time each individual frame takes. To ensure that at no point the moved distance between measurements is longer than 10 centimeters, the frame rate can be adjusted to a minimum of 20 FPS. The accuracy of the depth data between the minimal sensing range and the maximum of five meters should also be at its lowest 95 percent. This means, that in a single frame not more than five percent of pixels should report significantly incorrect or no data.

As for the object detection, the algorithm should not take more than 100 milliseconds to analyze a single frame. This time limit would translate to a minimal frame rate of 10 FPS, half that of the obstacle avoidance. This slower part is allowed as we cannot expect the much more complex object detection to run at the same speed as the obstacle avoidance. Based on figure 4.6, versions of the Yolo algorithm can reach latencies of less than 80 milliseconds. As before though, to avoid occasional slower frames a 100 millisecond latency should be allowed.

For interfacing the feedback device the software should be able to format all obstacles and objects of interest into a grid with a width of three columns, each representing a third of the view and five rows each with a depth of one meter. The number of rows and columns should be easily adjustable in order to match variations in the layout of the haptic vest. When a specific column is requested by the haptic device, by sending an uppercase letter i.e. 'A', 'B', 'C', 'D' ... , the software should return a message in shape of a single string. The format of the string is a lowercase letter to identify the row followed by a single digit implying if the cell is empty, full or unknown. Finally, if there is an object, this is followed by a 3 digit number that represents the object, repeating for every object detected in the cell. Once a single cell has been described in full the next cell in the column is treated in the same way and appended to the string. A detailed overview of this code can be seen in figure 5.1.

| **Row ID** | | **Fill status** | | **Object id** | |
|---|---|---|---|---|---|
| **char** | **distance (m)** | **int** | **meaning** | **int** | **object** |
| a | 0.2 - 1 | 0 | empty | 001 | traffic light |
| b | 1 - 2 | 1 | occupied | 002 | road |
| c | 2 - 3 | 2 | unknown | 003 | stairs up |
| d | 3 - 4 | | | 004 | stairs down |
| e | 4 - 5 | | | 005 | doors |
| | | | | 006 | normal obstacle |
| | | | | 007 | low hanging object |
| | | | | | |
| **Example:** | a0b0c1003d2e2 | | | | |
| **Meaning:** | 0.2 - 2 meters free; stairs up at 2 - 3 meters; 3 - 5 meters unknown | | | | |

Figure 5.1: Interface code

## 5.2 Hardware

The hardware, being the headset, has fewer hard limits. The main measurable specific is that the headset must not weigh more than 150 grams. This is necessary to keep it light enough to be worn for potentially hours at end without becoming a nuisance. As discussed before 150 grams is about the weight of a hat, therefore being a good number to aim for.

To keep the device from becoming to large, thereby standing out to much or becoming harder to handle, the headband should not exceed a thickness of one centimeter and should not be wider, from top to bottom, than two centimeters. To ensure that the device is not to much of a hassle when putting on, it needs to be able to be put on or removed in a maximum of 30 seconds.

Additionally, the the headset must of course be comfortable to be worn for longer times, but as discussed earlier this cannot be measured in specific values and instead has to be evaluated through open feedback by the users. As an evaluation metric the testers should fill in a likert scale to give feedback on different variables. For the headset to be successful no single variable should score lower than a six out of 10 and combined the score should not be lower than a seven out of 10.

# Chapter 6

# Realization

The following chapter focuses on the realization of the sensing part of the device. As before, the development is split up into the software and the hardware part. Parts of this development phase outline work that was done in parallel to the ideation phase.

## 6.1 Software

The entirety of the program is written in the programming language python, the language being chosen as it provides good general functionality and is common enough to have wrappers written for any external software that might be needed. The program running in the background of the device is built up of the two parts running in parallel. to achieve this the two parts of the program are run on separate threads.

The first of the two parts, the obstacle avoidance, is entirely self build only using some basic math libraries and the realsense library to extract data from the D435i sensor. The data the sensor provides, can be retrieved as two arrays describing first a grey scale image with values between 0 and 11 and second an RGB image. The RGB image is simply the direct result of a basic RGB camera mounted in the sensor. The grey scale image, on the other hand, is a depth image with each pixels value being the distance from the camera to that point.

The first step in processing the data, is cutting out the floor. If the software went on believing the floor it detects was an obstacle it would at all times warn of an obstacle immediately in front of the user. This of course would make the device useless, which is why the floor needs to be filtered out. While there are multiple ways of figuring out whether an area is floor or not, the one chosen in this project is a purely mathematical approach. As can be seen in figure 6.1, by setting the users height, the impact angle can be calculated from it and the depth found by the sensor.

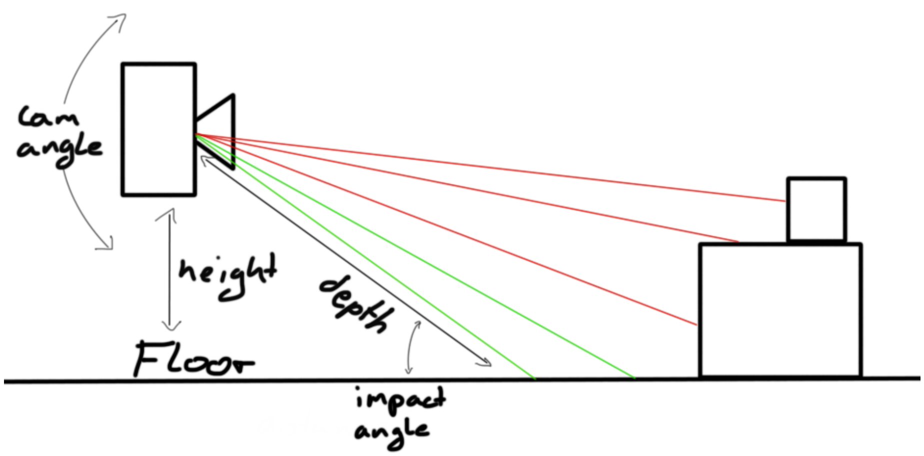



Figure 6.1: Floor detection

Using these values together, it can be determined if a point detected by the sensor is at the correct height to be part of the floor. Using the built in inertial measurement unit of the D435i this method also works when the sensor is rotated. This is done by calculating the downwards axis from gravity detected by the accelerometer and then based on this, removing or adding the camera angle to the previous calculations. To avoid false calculations caused by other motion detected by the accelerometer, the data from it can be filtered to smoothen its output a bit. The combination of these steps will then result in a view as shown in the following figure 6.2.

Figure 6.2: Depth view with floor detection

The same process can't quite be used in the same way for the ceiling, as different rooms have different ceiling heights. This thankfully does not matter to much for us, as first of all the use case for the device is outdoors and second we can assume that anything 30 centimeters taller than the user can be ignored. Therefore, the same process can be used in principle, but inverting the calculations to match an upwards calculation and replacing the users height with a constant of 30 centimeters. This, if found necessary, is of course still adjustable.

The next step in developing the obstacle avoidance, is detecting the closest obstacle in each direction. Since the sensor has a horizontal resolution of 640 pixels, that means an equivalent 640 directions have to be calculated. This is quite simply done by finding the closest point out of the 360 points in a vertical column of the image. During this, of course, the previous points classified as floor will be ignored. By repeating this process for each column the 2D image is effectively compressed to a 1D array. A big problem using this process however, is that any noise that is falsely detected close to the sensor will be treated as the closest obstacle. To avoid any such cases, the input data for each column needs to first be filtered. In this case, each column is treated for outliers, removing any if they are found. After these extremes have been removed from the columns array, the column is run through a Savitzky-Golay filter which smoothens the array. This can, under some circumstances, omit some detail but also massively reduces the risk of very wrong data points. The implementation of this filter is taken from Scipy library. Once this process is performed the resulting data can be visualized in a similar way as common radar installations, as can be seen in the resulting figure 6.3.

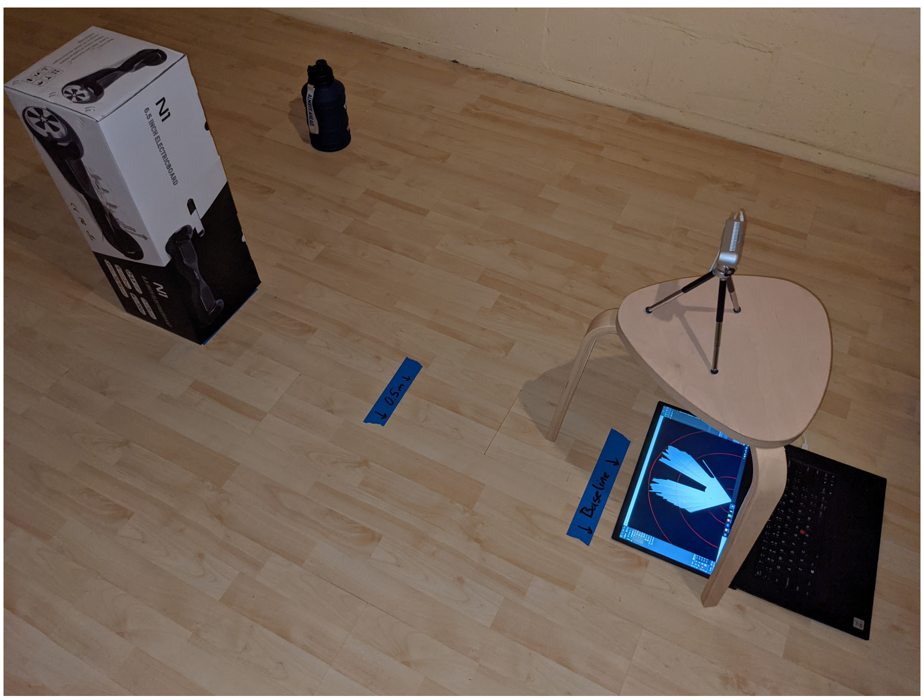


Figure 6.3: Radar like obstacle detection

The final step to the obstacle avoidance is placing the found data into the format desired for the haptic feedback device. The agreed upon standard format seeks to place all information into a two dimensional array representing a grid in front of the user. This grid should be, as mentioned before, three cells wide and five cells deep. The representation of this grid should imitate a smaller scale version of figure 6.4.

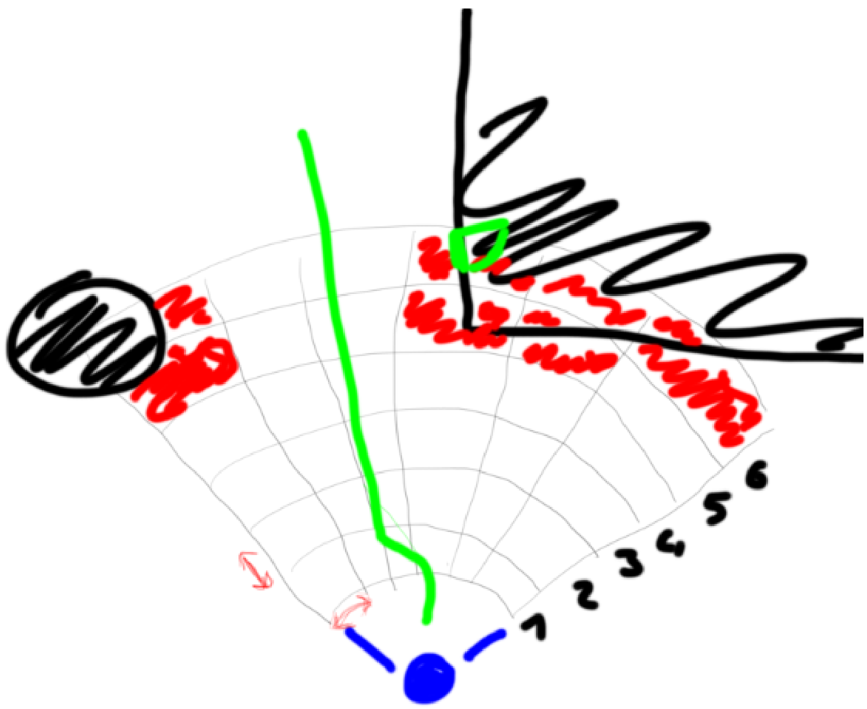



Figure 6.4: Representation of the grid view

To achieve this the 640 values of the 1D depth array have to be split into thirds. For each third, as before, the closest value is determined, but not before each third is subjected to the Savitzky-Golay filter again.

With the obstacle avoidance complete, the focus can be shifted to the object recognition.

Based on the state of the art, the decision was made to use a combination of the ASIF-Net algorithm developed by Chongyi et al. [8], to detect an object and following that use a convolutional neural network (CNN) to analyze what object it is. The initial implementation of the ASIF-Net algorithm was done based on the documentation on its dedicated GIT hub page. A big problem with its implementation however presented itself, due to the algorithm being build on older infrastructure, requiring outdated packages that in part did not work with other newer ones. The CNN was built using the tensorflow library with its prebuilt structures. The setup that worked well with the selected scenarios was a structure of two layers of two dimensional convolutional neural networks, each followed by pooling the array. Following the two layers, an additional 5 layers of dense networks are added, but their number of neurons per layer can be kept at a low number of 64, due to the previous two layers. This results in a solid network with high speed and accuracy.

Due to changes in the use case, the software needed to be changed to detect more than one object at a time. The ASIF-Net algorithm is not capable of this so a change to a completely different object detection method was performed. The algorithm chosen to replace ASIF was the 'you only look once' (YOLO) algorithm version four [14]. As a full object detection algorithm, it is not necessary to first find the object and then classify it, as it does all that on its own. When presented an RGB image it can detect up to 50 objects classifying them at high accuracy and noting their bounding boxes, within the object supposedly is. A visualization of this can be seen in figure 6.5.

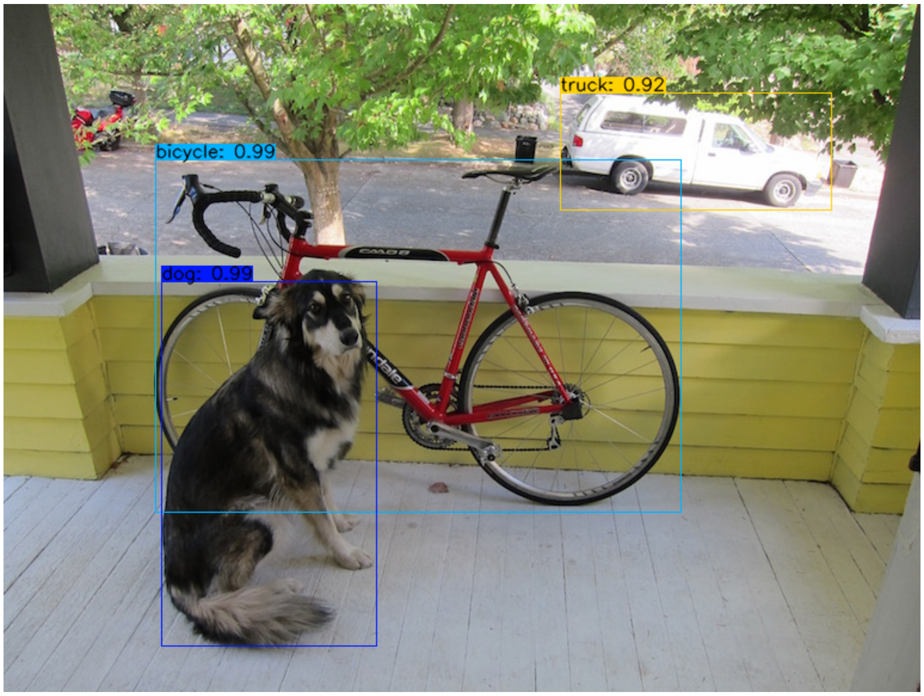



Figure 6.5: Object detection using YOLO V4

The disadvantage of the YOLO algorithm over ASIF is that YOLO does not consider the depth data and thereby take more computing power to detect objects with the same accuracy as ASIF. Despite this, it is powerful enough to return results at high speed and accuracy.

While Yolo4 can be implemented on its own, as demonstrated by Bochkovskiy et al. [14], there are also options of implementing it while using tensorflow as a support. This method is demonstrated by [15], with the code basis for it given on the associated Git hub page. Starting out with code base by TheAiGuy, the code can be adjusted to further fit the needs of this project. When detecting an object of interest the software will determine the center of the object and collect the depth of points within the bounding box. As the bounding box often includes little bits that are not part of the object (see figure **??**), the points of which the depth is used are weighted with points closer to the center being weighted higher. By again removing any outliers and finally averaging the depth measurement points the distance at which an object is located can be determined. This process is repeated for every object. With all objects being assigned X, Y and Z positions they can be added to the previously created two dimensional grid cell array.

For accurately detecting objects any object detection algorithm first needs to be trained on a set of example data. For the purposes of this project, a dataset of 9000 images has been created from the Googles 'Open Images Dataset' which provides millions of images with predefined bounding boxes. The selected dataset is a combination of images highlighting the previously selected objects of interest, such as doors and stairs. Using the dataset of images a model for the algorithm could be trained.

The final step to developing the software, is implementing the interface to the haptics device. As descided with the team the sensor side of the project would only send information upon request. Once requested, the program sends back a string with the depth information for a single column encoded. The exact protocol for this can be found in the ideation chapter and in figure 5.1. The communication runs using basic serial communication, on the side of the sensor implementing the Pyserial library. When any data is received, the software converts the character to an integer with the corresponding Ascii value. Rather than using checks for every case, the program can directly access the corresponding column, by using the input character as the array index, thereby making it more flexible. When the correct column is selected any needed information is copied from the two dimensional grid cell array into a return message string. Iteratively, each row is checked for information and appended to the string. Once complete, the program returns the message informing the haptic device about the requested column.

## 6.2 Hardware

The largest concern in the design of the hardware was to keep the size small and the weight low. As mentioned in the specification the band could not be wider than two centimeters, not thicker than one centimeter and not heavier than 150 grams. To achieve this the greatest tool at hand is 3D printing. Using PLA plastic, complex shapes can be designed and quickly prototyped, it has great strength, a bit of flexibility and is very light. For this project, all parts were designed by myself in solidworks, based of measurements take of a varied group of heads.

Based on the concept developed during the Ideation phase, a frontal camera holder and a back of the head counter piece was needed. The back piece could be easily designed as a slight curved square with two attachment points on either side to fasten the Velcro strips. The front part, the sensor holder, was a bit more complicated. The sensor comes with a standard 6mm threaded attachment point, which was determined to be the best point to connect to. However, with the first design there was not enough space to screw the sensor onto the holder without them colliding. To solve this, an adapter piece was designed to first screw into the sensor and then be clipped into the main holder piece, as can be seen in figure 6.6.

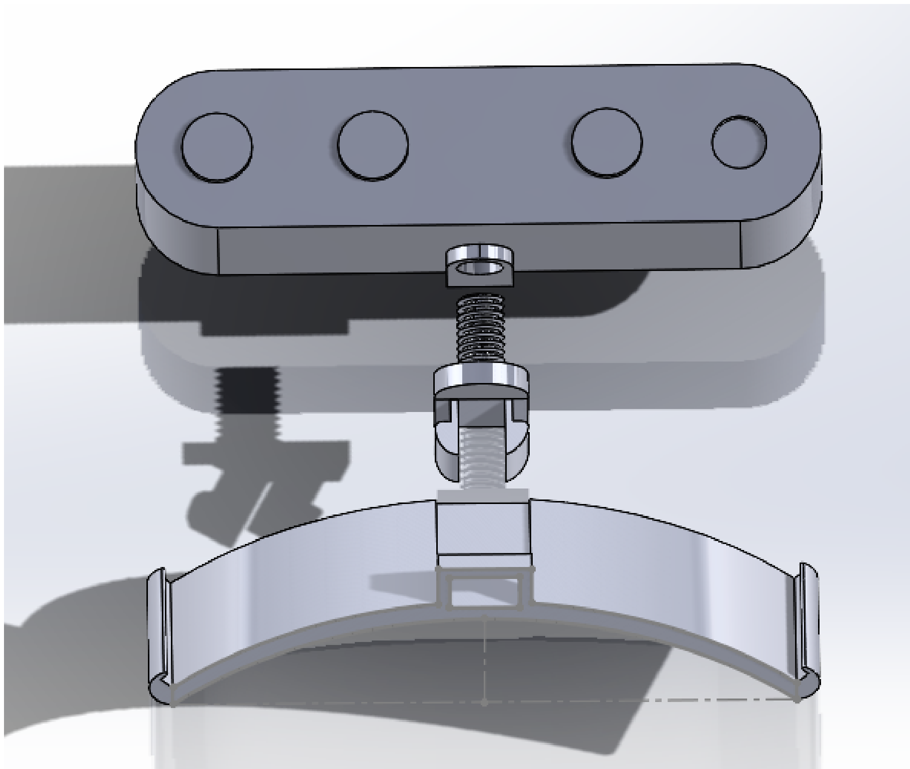


Figure 6.6: Sensor mount version 1

The headset design in concept was good, not to big and easily wearable as demonstrated in figure 6.7.

Figure 6.7: Headset version 1 being worn

This design had an additional big advantage, as the sensor could easily be removed to place it on another mount, making development and testing easier. However, the repeated movement of the clip produces to much stress, finally breaking the clip apart.

To prevent such damage in any newer versions, the switch was made to remove part of the mount in a way that the sensor could be screwed directly onto it, with the final design shown in figure 6.8.

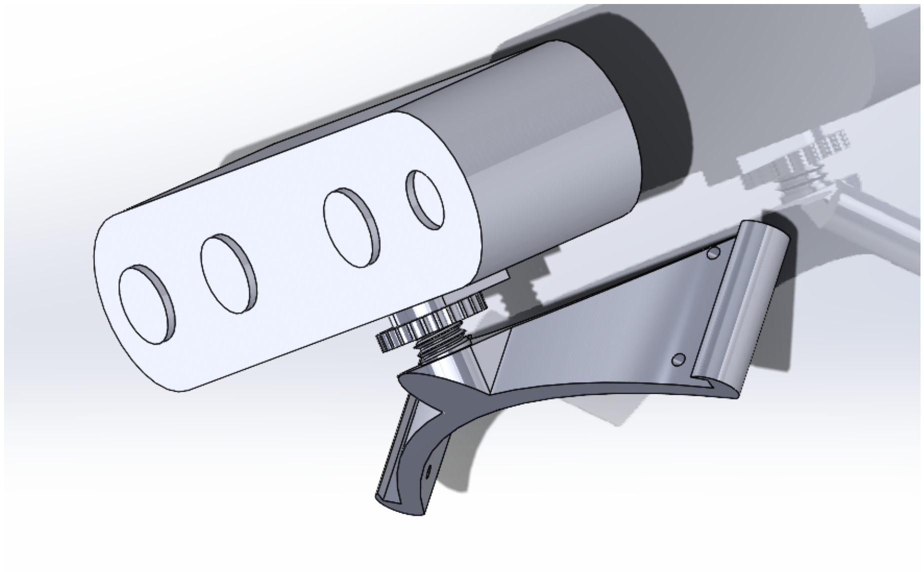


Figure 6.8: Sensor mount version 2

This could be done as the sensor with its weight is producing a rotational force on the headset. The force produced by the sensor is most noticeable at the bottom edge of the mount, while the force needed to hold the rotation back needs to be applied at the top edge of the mount. This can be done, since the cut out part is only in the center of the top edge with the side edges and bottom completely untouched. That way the Velcro straps pulling backwards, still have a good grip to hold back the mount.

Additional changes made to the sensor mount are, that the Velcro strips are tied down to the front piece, rather than relying on tension to hold them in place. For this, the sensor mount in its latest version comes with small holes on the sides for string to be threaded through.

The final change to the design came with an additional small fastener wheel that is placed underneath the sensor to lock it in place when it is screwed in far enough, reducing the chance of the sensor over or under fitting the printed screw.

To make the headband wearable for longer times both the front and back pieces are padded with half a centimeter of foam taped directly to the plastic. A second wearability issue concerning the cable coming from the camera getting in the way, could be solved by printing a small cable holder which threads onto the Velcro strip and allows the cable to be clipped into it, to lead it behind the users head without bothering them. These additions to the design can be seen in figure 6.9.

Figure 6.9: Final headset design

At this point, the main weight of the headset was coming from the sensor itself. By removing the glass lens, the Mask and the aluminum front and back seen in figure 6.10, the sensors weight can be massively reduced, in the process also reducing its size.

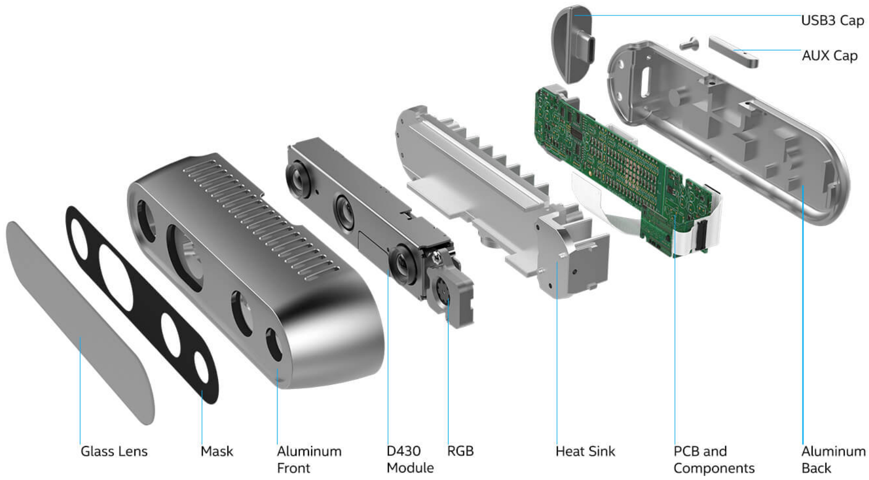



Figure 6.10: Stripdown of the D435i sensor array

The final version of the headset with the cutback sensor mount, cable holder and stripped sensor can be seen worn in figure 6.11.

Figure 6.11: Final headset design being worn

# Chapter 7

# Evaluation

The evaluation phase was split up into three phases. The first phase being the system testing phase, during which the device was placed in a clean environment to collect quantitative data from its ideal use case. The second phase focuses on partial system user tests. During these tests the research focus was mostly placed on the wearbility of the device. The third and final stage of the evaluation phase was a full system user test in which the sensing part of the device was connected with the hapic feedback device. With the two connected, test participants were then subjected to simulated use case scenarios.

## 7.1 System testing

During the first phase of testing, the sensing device was placed in a clean environment. During this test the device was tested for four different variables: the speed of the obstacle avoidance, the speed of the object detection, the accuracy of the object detection and finally, the robustness to movement.

Testing for the speed of the two parts of the software was straight forward enough. The time it took to run through one cycle could simply be output by saving the previous cycles system time and subtracting it from the current. With the delay being measured in milliseconds, 1000 could simply be divided by the result to find the corresponding frame rate.

When testing the obstacle avoidance part for a couple of minutes with different numbers of obstacles at various distances, the software does not seem to slow down. While it varies slightly with every frame, most delays are within a range of 45 to 50 milliseconds no matter the amount of objects, as can be seen in figure 7.1. This makes sense as the number of operations performed by the computer do not change with the number of obstacles. As can be seen, the frame rate on average does not drop below 20 frames per second, thereby passing the speed requirement set out in the specifications.

| **Obstacle detection** | | |
|---|---|---|
| **# of obstacles** | **avg. delay (ms)** | **avg. FPS** |
| 1 | 47 | 21,28 |
| 2 | 49 | 20,49 |
| 3 | 46 | 21,74 |

Figure 7.1: Speed data of obstacle avoidance

The speed of the object recognition can be found in the same way as that of the obstacle avoidance. However, looking at the results in figure 7.2, it can be seen that the speed of the algorithm slightly drops of with a larger amount of objects in frame. This makes sense, as the algorithm for every additional object needs to detect its bounds and classify it. Additionally, the more objects are in frame the less 'clean' the environment becomes, making it harder for the algorithm to differentiate one object from another, further slowing it down. Despite this, the object recognition algorithm at its worst drops its speed only to 15 frames per second. This speed is still much higher than the desired 10 frames per second, as laid out in the specification section. Importantly to note, for the evaluation of the algorithm some of the objects were replaced by smaller more manageable objects for an inside space, as any outside space would not be considered clean anymore and it could be difficult finding the required number of objects next to each other.

| Object recognition | | |
|---|---|---|
| # of obstacles | avg. delay (ms) | avg. FPS |
| 1 | 55 | 18,18 |
| 3 | 59 | 16,94 |
| 5 | 58 | 17,24 |
| 10 | 66 | 15,15 |

Figure 7.2: Speed data of object detection

As expected, the accuracy with which the objects are detected, drops off the further away the object is from the sensor. Up close the accuracy reaches on average 94%, although depending on the type of object and the situation the accuracy can jump up to 99% or even 100%. As shown in figure 7.3, the accuracy drops of to 80% at three meters and even 62% at five meters. While the 62% could be better, the accuracy at this distance is still good enough to satisfy, especially considering the vastly improved accuracy at shorter distances.

| Yolo4 accuracy | |
|---|---|
| distance (m) | avg. accuracy (%) |
| 1 | 94 |
| 2 | 89 |
| 3 | 81 |
| 5 | 62 |

Figure 7.3: Accuracy of YOLO V4

Finally, to measure the devices resistance to movement, the device was moved back and forth along a 10 centimeter line for 5 minutes, each at varying frequencies. At the lowest speed of 30 movements per minute, only 4 errors were detected throughout the span of 5 minutes. Compared to that, at the fastest speed of 120 movement per minute, 118 errors were detected in 5 minutes. This, as can be seen in figure 7.4, leads to a frequency of about 20 errors in 1000 measurements, roughly 30 times more, showing the importance of a stable camera mount.

| Resistance to movement | | |
|---|---|---|
| # of moves/min | Errors | Errors/1000 |
| 30 | 4 | 0,66666667 |
| 60 | 31 | 5,16666667 |
| 120 | 118 | 19,6666667 |

Figure 7.4: Resistance to movement

## 7.2 Usability and wearability testing

The wearability tests were performed in two stages. The first stage consisted of testing small changes to the design in parallel to the development phase. This had wide reaching influence on the design of the headset. The second phase was performed in parallel with the full system user tests. This was done by having the participants fill in a questionnaire with various choice and open questions.

The first phase of testing mainly brought to light different aspects of the designs ergonomics. This largely concerned the curvature of the front and back pieces of the headset. While the PLA plastic used for the parts is not completely stiff and can flex a little bit, it should still resemble the curvature of an average head. From the testing it was found that over the eight centimeter wide holder the ideal curvature was 1.2 centimeters deep. With this curvature it provided enough contact area for smaller heads and could still flex enough for larger heads.

One concrete parameter that could be tested for and was defined during the specification phase, was the time it took to put on the headset. Positively, out of all attempts the time to put it on was never more than 15 seconds, with the average being even lower at 7.3 seconds. The speed with which the headset can be put on was great as it completely crushed the maximum of 30 seconds. This speed can be mostly attributed to the simplified fastening mechanism of using Velcro strips, which most people are very familiar with.

The second parameter that could be measured was the weight of the headset. In the specification the maximum weight was set at 150 grams. The final version of the headset however only weighed in at about 60 to 70 grams. With this, the headsets weight is less than half the permitted weight, making it possible to add more parts for further functionality in the future, without risking it to become to heavy. This light weight, as expected before, is due to the use of the PLA plasic as the weight without the sensor drops to about 15 grams.

In the second phase, connected to the full system test, the feedback was very positive. The questionnaire tested for four variables and was open for further input and clarification. The variables tested for were wearablility, weight, sweat resistance and exerted pressure.

The wearability of the headset was rated at an averaged 9.4 out of a maximum of 10. On the provided likert scale from one to seven, out of the five responses four gave the highest rating with only one participant giving a five. None of the participants felt any need to further comment on the headsets wearability.

The second variable tested for, was the weight. By the users it was scored at an average of 9.1 out of 10. Again, a likert scale from one to seven was provided with three users scoring a six and two of them scoring a seven. For further notes, the users only commented on the headsets low weight, which they liked, as it did not draw much attention from them.

The third variable being tested for, was the amount of sweating induced by the headset. For the test it was clarified that the highest number would mean no sweating induced and the lowest score would be heavy sweating. This variable scored the best out of all the variables at a 9.7 out of 10. On the seven point likert scale the the users scored it at a seven four times and at a six once. This resistance is likely due to the effective use of the foam padding. This prevents the plastic sitting directly on the skin allowing it to continue breathing.

The pressure being exerted by the headset was the final variable tested for. This resulted in the lowest score at an 8.3 out of 10, which is still a very good result though. This lower score is mostly caused by one user who scored the pressure at 3 on the seven point scale, while the others evenly split their votes between six and seven.

From these usability test it clearly shows that the headset is well designed to keep its user comfortable and not cause any nuisances. While there is still some space for improvements, the design satisfies and even exceeds all expectations.

## 7.3 Full system test

The final system test were performed with users wearing the full system including both the headset and the haptic device. The tests were performed in a controlled environment to protect the users from injuries. The test space was a large room of a size of about seven meters length and four meters width. The participant would be placed on one side of the room, blindfolded and wearing the navigational aid. Without the user seeing, paper box obstacles were placed throughout the room. The user would then have to navigate from one side of the room to the other without hitting any obstacles, as can be seen in figure 7.5. After completing the first task, they were presented with different objects of interest and were told to find a specific one. Throughout these tests the software would also record its own performance.

Figure 7.5: Tester moving through an obstacle course

The obstacle avoidance was unsurprisingly not effected in its speed, as it already in the previous test, did not get effected by more or less obstacles. The object detection however, was affected much more. Throughout the tests it performed at an average speed of about 16 frames per second. This is slightly slower than in the clean environment, but still fast enough to satisfy the requirements. As with the speed, the average accuracy also decreased from its ideal environment. The average accuracy throughout the tests dropped to about 60%, while it at times could drop to as low as 35%, which was the cut off limit. This is unfortunate, as it could lead to mistakes, showing that there is still some work that could be done to improve the object detection. Despite this, from evaluating the user, the test still seemed successful.

Throughout all the test, the users typically walked about nine meters from start to end. They did this in an average of about one minute and 15 seconds. This leads us to a speed of 0.12 meters per second, or about a tenth of the average persons walking speed. While this seems slow the system still shows its potential through the fact that only about half of all testers hit an obstacle and when they did, only barely strafed it. This issue of hitting boxes can mainly be attributed to the systems bad vision in the half meter in front of it, something that a future continuation could work on.

The object recognition test was much more successful with participants being able to correctly distinguish the searched for object about 80% of the time. Once correctly identified they could then also usually move towards the object successfully.

A notable phenomenon to mention is the learning effect, which could clearly be observed here with participants performing better at both parts of the test when they had used the device at least once before.

# Chapter 8

# Further development

The project has proven its concept and shown a lot of potential at that. Despite this, there are still many things that would need improvement. The following final chapter will focus on what can still be done to further build on the device presented so far.

## 8.1 Complete device

To improve the device as a whole, there are many things that could be done. First of all, the final user tests were, out of ethical and security reasons, not performed with people who actually suffer from visual impairdness themselves. Testing with the actual target group could bring to light issues that were not discovered during the test runs with the not visually impaired test users. On the other hand, because our testers did not know how to use a cane they relied entirely on the device to guide them. The system of course is not actually designed to fully replace the cane but to add to it, meaning that issues experienced by the testers might not have been experienced by a visually impaired person. All in all, testing and evaluating with the actual target group could have given the project more credibility.

In a similar sense the testing was also performed in a simulated scenario to protect the participants. Testing in the true scenario could bring issues to light that were unknown to us or put noticed issues into perspective.

Furthermore this project focused on a very limited set of use case scenarios. In the real world there are of course more environments, obstacles, objects of interests and combinations of all of these. Expanding the dataset of objects to process and generally expanding the usefulness of the device to more situations could truly further the development.

One main functionality that might make sense to add as well, would be a 'pointer'. This could be part of the haptic gauntlet, that could be pointed in a certain direction to get specific information for that direction only, taking the idea of the button that had been considered during ideation and further building on it.

A final improvement that could be made on the full device would be to miniaturize the computing module. Throughout this project the sensor and the haptic device were connected to a laptop that was running the sensor software and communicating with the haptic device through serial communication. If the code could be optimized to run with lower computing power a smaller computer could be built into the haptic vest as the main controller. Alternatively, it could also be possible to offload the heavier processing to a mobile device as almost everyone carries a phone with them anyways. The phone could be connected to the device through Bluetooth and would decrease the size and weight of the device and also massively drop the cost.

## 8.2 Sensing

For the sensing side of the work specifically, there are of course also still many possible improvements.

Firstly there are some additional functions that could be added to the headset itself. The mount for example, could be redesigned to be adjustable in its angle so that it could change for taller or shorter users. Additionally, the mount could be stabilized to decrease the amount of smaller movements coming from the user walking along.

A second simple but important improvement that could be done would be waterproofing the sensor. While removing the sensors protective case, removed weight and made it much smaller, it of course also made it more vulnerable. If the protective covering could be build directly into the headband it could perform the same function while being smaller and lighter. In addition to this, the sensor could also be disassembled further with its components being distributed over a larger are of the headband but making it smaller in the front. Alternatively, in a simpler way the sensor could be sprayed with a water resistant coating, leaving it unprotected to other damage but ensuring it to be weather proof.

Another improvement that could help immensely, would be increasing the field of view along the vertical axis. The issue here is that the sensor only has a vertical angle of 58 degrees. Assuming the sensor is worn by a person of a height of 1.7 meters and the sensor is angled such that its top most view is horizontal to the ground, the closest point on the ground it could detect would still be more than a meter away. By using a special lens in front of the sensor, as seen in figure 8.1, the field of view could be increased at the cost of detail in the image. The loss of quality on the vertical axis would not be such a big problem as it would, if at all, make only minor changes to the estimated distances that objects are at.

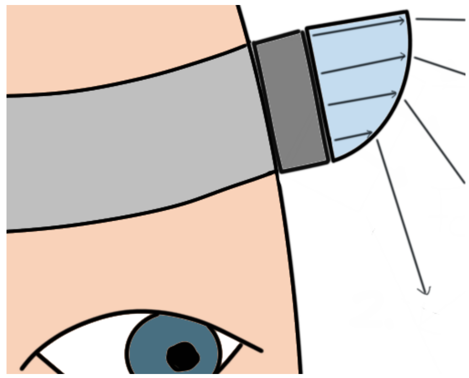


Figure 8.1: Sensor mount with special wide angle lens

Finally, the largest improvement, other than building a better object detector, would be to improve the filtering of the incoming data. While the filtering achieved in this project is already quite good, noise passing through the filter is still the most common issue for false readings. Could this be entirely stopped, it would improve the device greatly.

# Bibliography

[1] R. Tapu, B. Mocanu, and T. Zaharia, "A computer vision system that ensure the autonomous navigation of blind people," in *2013 E-Health and Bioengineering Conference (EHB)*, 2013, pp. 1–4.

[2] A. Zvironas, M. Gudauskis, and D. Plikynas, "Indoor electronic traveling aids for visually impaired: Systemic review," in *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2019, pp. 936–942.

[3] S. Wenqin, J. Wei, and C. Jian, "A machine vision based navigation system for the blind," vol. 3, 06 2011.

[4] Sung-Jae Kang, Young Ho, and In Hyuk Moon, "Development of an intelligent guide-stick for the blind," in *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)*, vol. 4, 2001, pp. 3208–3213 vol.4.

[5] Y. H. Lee and G. Medioni, "Rgb-d camera based wearable navigation system for the visually impaired," *Computer Vision and Image Understanding*, vol. 149, pp. 3–20, 2016.

[6] M. Miles and J. Jetter, "A lidar-based navigation system for the visually impaired," *The University of Arizona*, 11 2016.

[7] J. G. da Silva Neto, P. J. da Lima Silva, F. Figueredo, J. M. X. N. Teixeira, and V. Teichrieb, "Comparison of rgb-d sensors for 3d reconstruction," in

*2020 22nd Symposium on Virtual and Augmented Reality (SVR)*, 2020, pp. 252–261.

[8] C. Li, R. Cong, S. Kwong, J. Hou, H. Fu, G. Zhu, D. Zhang, and Q. Huang, "Asif-net: Attention steered interweave fusion network for rgb-d salient object detection," *IEEE Transactions on Cybernetics*, vol. 51, no. 1, pp. 88–100, 2021.

[9] I. A. Anjani, Y. R. Pratiwi, and S. Norfa Bagas Nurhuda, "Implementation of deep learning using convolutional neural network algorithm for classification rose flower," in *Journal of Physics: Conference Series*, vol. 1842, 2021. [Online]. Available: www.scopus.com

[10] S. Naudet-Collette, K. Melbouci, V. Gay-Bellile, O. Ait-Aider, and M. Dhome, "Constrained rgbd-slam," *Robotica*, vol. 39, no. 2, p. 277–290, 2021.

[11] A. Sun, "Building a virtual world in ros based on the robot's perception," August 2019. [Online]. Available: http://essay.utwente.nl/80005/

[12] L. Francés-Morcillo, P. Morer-Camo, M. I. Rodríguez-Ferradas, and A. Cazón-Martín, "Wearable design requirements identification and evaluation," *Sensors*, vol. 20, no. 9, p. 2599, May 2020. [Online]. Available: http://dx.doi.org/10.3390/s20092599

[13] B. Moen, "What causes musculoskeletal strain?" *mayoclinic.org*, 2020.

[14] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Scaled-YOLOv4: Scaling cross stage partial network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 13 029–13 038.

[15] TheAiGuy, "tensorflow-yolov4-tflite," July 2020. [Online]. Available: https://github.com/theAIGuysCode/tensorflow-yolov4-tflite

# Appendices

# Appendix A

# Interview / Usertest forms

Enschede, 11/03/2021

## Information brochure Department HMI

Dear reader,

In this letter, we would like to inform you about the research you have applied to participate in. The research will take place 18-03-2021, online. The proposed research, entitled “Using computer vision to aid navigation for people with a visual impairment”, focuses on developing a wearable device for people who are visually impaired, that uses computer vision to detect and map the users surrounding and guide them through it using haptic feedback. The research is done by four students: Adrian Hopfenspirger, Tim Yeung, Ruben Wierik and Kai Ferdelman. The research you will be taking part in is a basic interview in which we will ask prepared questions, but you are welcome to add any information that was not asked for.
The goal of the interview is to gain information about the visually impaired short comings and how our approach could help them in the best way.

The interview lasts for about an hour. You will not be paid for participating in the interview.

Yours sincerely, Kai Ferdelman

Coordinator: Kai Ferdelman

Department: Human Media Interaction

Faculty of EEMCS

University of Twente Tel: +31 (0)53 – 489 3740

Email: k.j.ferdelman@student.utwente.nl

Tel: +49 15756737328

Enschede, 11/03/2020

# Informed Consent for standard research

'I hereby declare that I have been informed in a manner which is clear to me about the nature and method of the research as described in the aforementioned information brochure for 'Using Computer vision to aid navigation for people with a visual impairment'. My questions have been answered to my satisfaction. I agree of my own free will to participate in this research. I reserve the right to withdraw this consent without the need to give any reason and I am aware that I may withdraw from the experiment at any time. If my research results are to be used in scientific publications or made public in any other manner, then they will be made completely anonymous. My personal data will not be disclosed to third parties without my expressed permission. If I request further information about the research, now or in the future, I may contact Kai Ferdelman. If you have any complaints about this research, please direct them to the secretary of the Ethics Committee of the Faculty of Electrical Engineering, Mathematics and Computer Science at the University of Twente, P.O. Box 217, 7500 AE Enschede (NL), email: ethics-comm-ewi@utwente.nl).

Signed in duplicate:

Timon van Hesselt .................................

Name subject Signature

I have provided explanatory notes about the research. I declare myself willing to answer to the best of my ability any questions which may still arise about the research.'

Kai Ferdelman .................................

Name researcher Signature

Enschede, 25/03/2021

# Information brochure Department HMI

Dear reader,

In this letter, we would like to inform you about the research you have applied to participate in. The research will take place ..-04-2021, online. The proposed research, entitled "Using computer vision to aid navigation for people with a visual impairment", focuses on developing a wearable device for people who are visually impaired, that uses computer vision to detect and map the users surrounding and guide them through it using haptic feedback. The research is done by four students: Adrian Hopfenspirger, Tim Yeung, Ruben Wierik and Kai Ferdelman. The research you will be taking part in is a simple focus group with three participants and a moderator.
The goal of the focus group is to understand the struggles of those with visual impairments when trying to navigate and to see how haptic feedback guided by computer vision could improve the situation.
The session will be recorded, but the recording will not be given to third parties and will only be used for this project.

The meeting lasts for about an hour.

You can at any point in time withdraw from this research without having to give a reason.

You will not be paid for participating in the focus group.

Yours sincerely, Kai Ferdelman

Coordinator: Kai Ferdelman

Department: Human Media Interaction

Faculty of EEMCS

University of Twente Tel: +31 (0)53 – 489 3740

Email: k.j.ferdelman@student.utwente.nl

Tel: +49 15756737328

Enschede, 11/03/2020

# Informed Consent for standard research

'I hereby declare that I have been informed in a manner which is clear to me about the nature and method of the research as described in the aforementioned information brochure for 'Using Computer vision to aid navigation for people with a visual impairment'. My questions have been answered to my satisfaction. I agree of my own free will to participate in this research. I reserve the right to withdraw this consent without the need to give any reason and I am aware that I may withdraw from the experiment at any time. If my research results are to be used in scientific publications or made public in any other manner, then they will be made completely anonymous. My personal data will not be disclosed to third parties without my expressed permission. If I request further information about the research, now or in the future, I may contact Kai Ferdelman. If you have any complaints about this research, please direct them to the secretary of the Ethics Committee of the Faculty of Electrical Engineering, Mathematics and Computer Science at the University of Twente, P.O. Box 217, 7500 AE Enschede (NL), email: ethics-comm-ewi@utwente.nl).

Signed in duplicate:

| | |
|---|---|
| | .................................. |
| Name subject | Signature |

I have provided explanatory notes about the research. I declare myself willing to answer to the best of my ability any questions which may still arise about the research.'

| | |
|---|---|
| Kai Ferdelman | .................................. |
| Name researcher | Signature |

University of Twente

25/06/2021

## User test information brochure

Navigation for the visually impaired - full system test

This user test is part of the development of a navigational aid for people suffering from visual impairments. The system uses computer vision and haptic feedback to safely guide the user through their surroundings. It also utilizes object recognition to guide the user to so called objects of interest.

This user test focuses on testing the final setup of the device.

The test consists of the two user tests, one utilizing virtual reality to simulate the input to the device and one using a sensor combination to accurately monitor the surroundings, both feeding their data to a haptic vest and an additional haptic glove. The user test will attempt to evaluate the effectiveness of each of the components and the wearability of the haptic devices and the sensing headset.

To conduct the test, the test subject, wearing the device, will be blindfolded and asked to move through a parkour set up from cardboard boxes. The participant will be closely monitored by the researchers and stopped in their place if they are about to walk into an obstacle, to avoid any injury.

The second part consists of the user again wearing the device and being blindfolded, navigating their way to given objects.

For both these parts the subject will be filmed and evaluated on their speed, their path and any additional complaints they might offer.

After completing the user test the participant is asked to complete a questionnaire consisting of Likert scale and open questions.

The participant will not be compensated for participating.

The participant must sign the accompanying consent form to participate in this research.

University of Twente

25/06/2021

# User test consent form

Navigation for the visually impaired - full system test

I consent to taking part in the user test as laid out by the information brochure handed out in combination with the consent form.

I am aware that if at any point during the experiment, for any reason I can opt out without giving a reason. Alternatively, I can also at any time for any reason pause the experiment and continue shortly after.

I am aware that at any point during the experiment I can ask for further clarification of the experiment.

I am aware that I will be blindfolded and be asked to move through a safe environment, while only receiving information from the system being tested.

I allow the researchers to film any performed experiments, for later analysis under the condition that I can at any time ask for the recordings to be deleted for any reason and the recordings in any case being deleted after the research has concluded.

I allow the researchers to use any verbal or written answers related to the test.

I am aware that all personal information will be kept private, will not be shared with 3rd parties and any related information will be kept anonymously.

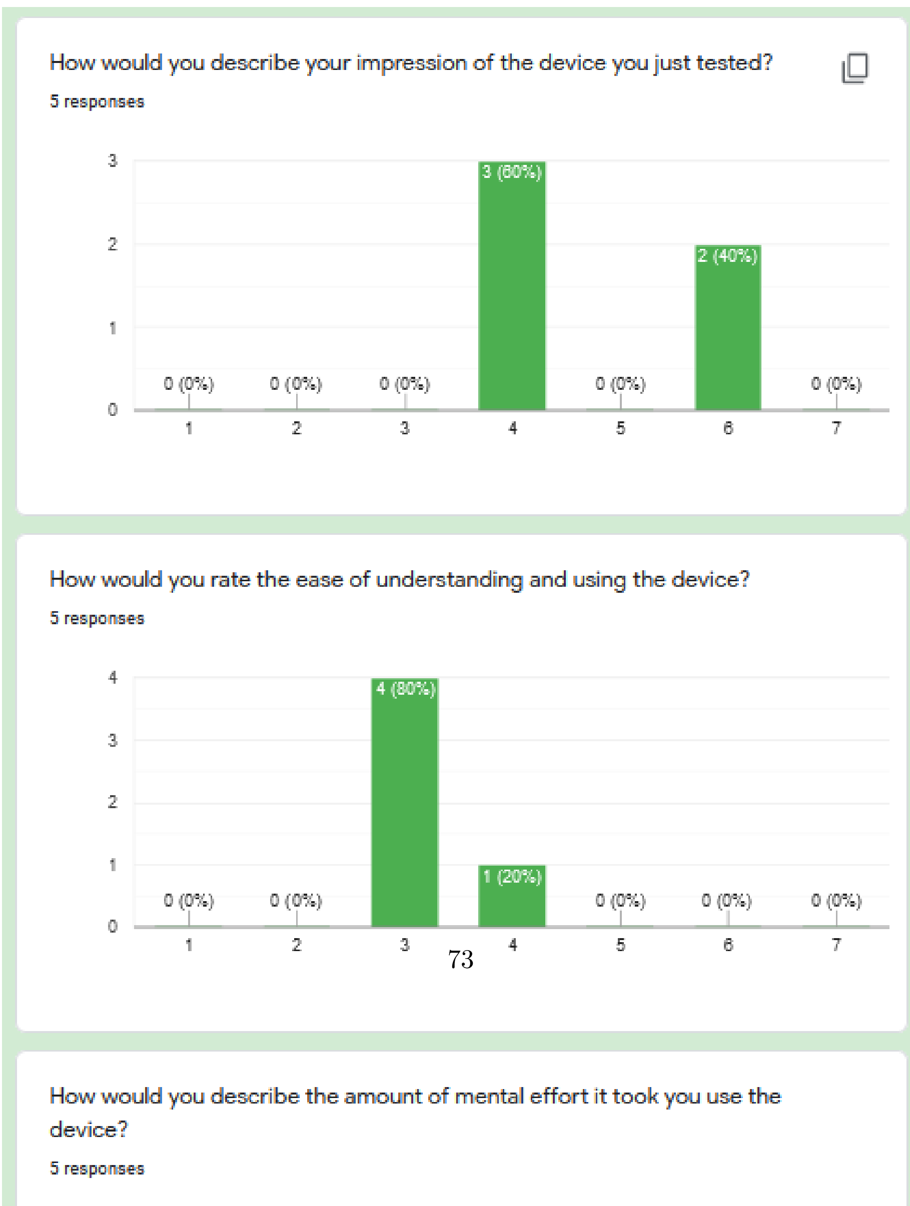
I am aware that I will not receive monetary compensation for taking part in the user test.

X ______________________

Research participant

# Appendix B

# Questionnaire



Which elements in particular would you say contributed to that view on the mental effort?

5 responses

I would say that the hand piece for object recognition was not reliable since I was not able to feel and hence recognize the vibration patterns properly even after multiple practice sessions. The back piece was easy to get used to but still not as reliable since the right vibration was not always working. Therefore I was not sure whether that one vibration meant that I could go or no.

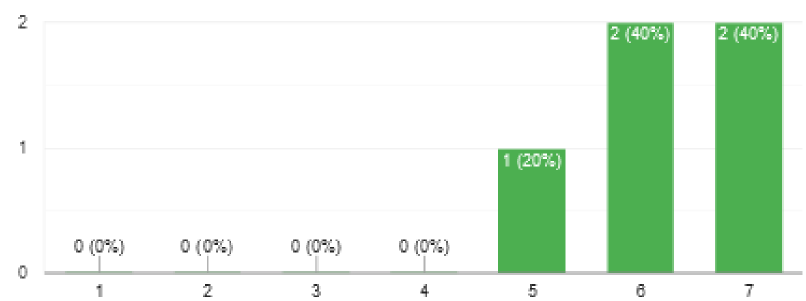
I had to pay a lot of attention to the different vibrations modes

sometimes unclear haptic feedback, dots being loosely attached to the skin mainly

the haptic code on the hand

How would you describe the amount of physical effort it took you to use the device?

5 responses




Which elements in particular would you say contributed to that view on the physical effort?

4 responses

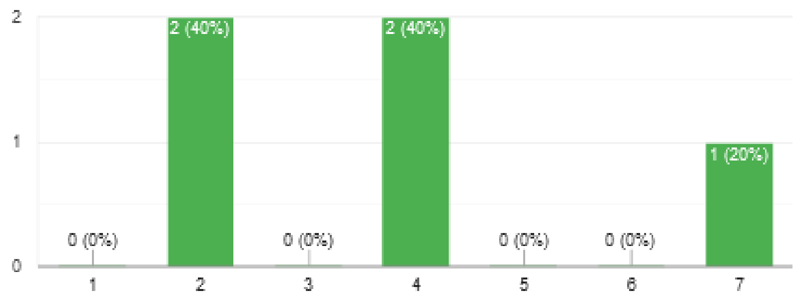
None

It did not really take physical effort, it fitted well and I was able to feel the vibrations

was all very easy to wear, no physical load

How would you rate the usability of the device?

5 responses




Do you have any notes about the usablity ?

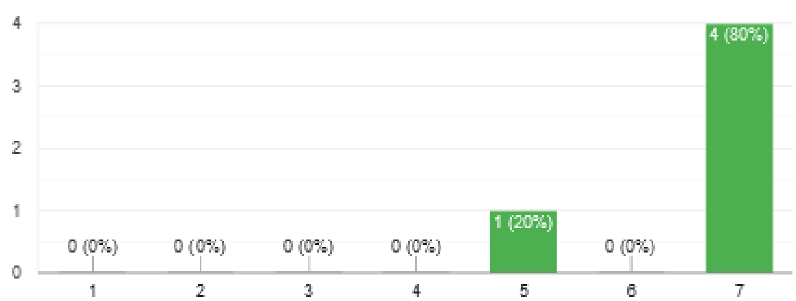
3 responses

I was able to get used to it and feel "safe" in a fixed environment but I would never feel comfortable fully trusting the device in real life. To elaborate, vibrations where not reliable and the idea that the vibration was softer or harder depending on the distance of the object was not something that me as a tester could feel or recognize.

it is still a prototype, but promising

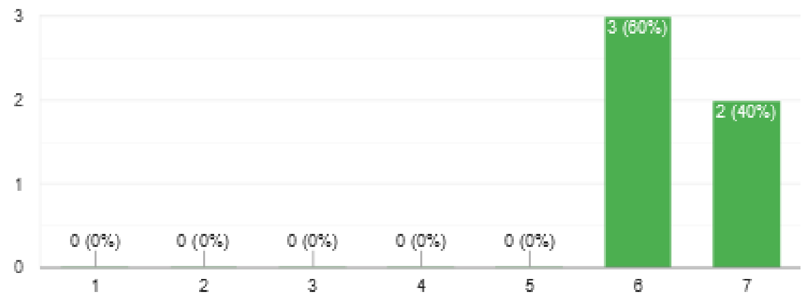
How would you rate the wearability of the Sensor?

5 responses



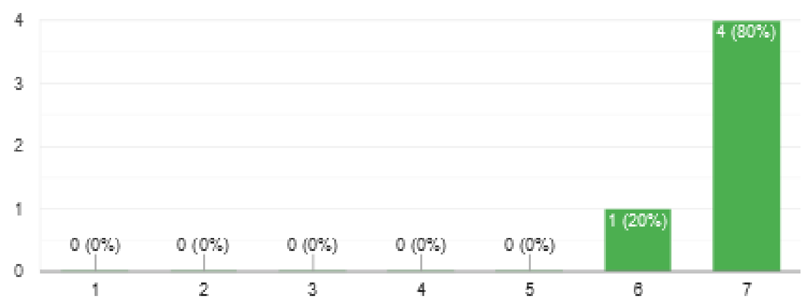
How would you rate the weight of the sensor ?

5 responses



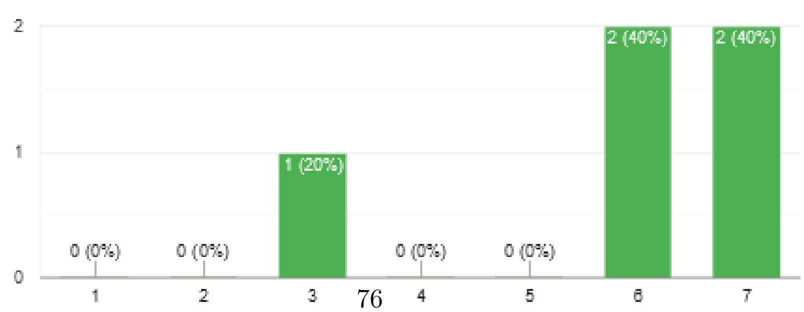

How would you rate the amount of sweating expericed while wearing the sensor?

5 responses



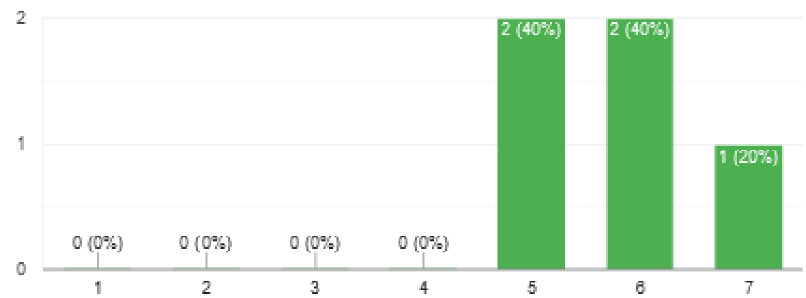

How would you rate the amount of pressure expericed while wearing the sensor?

5 responses



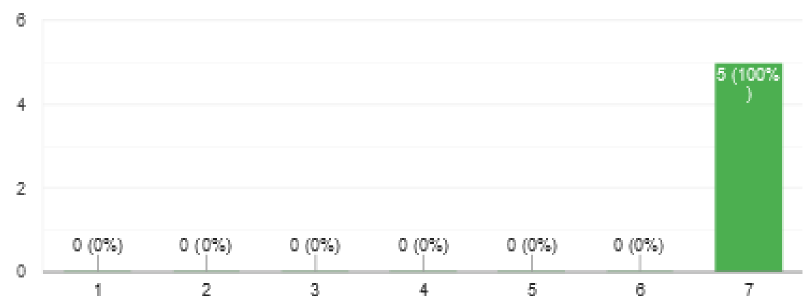
How would you rate the wearability of the Vest and Gauntlet?

5 responses




How would you rate the amount of sweating expericed while wearing the Vest/Gautlet

5 responses




Do you have any notes about the wearability of the Sensor ?

4 responses

None

no

no, as my focus went to understanding the vibration pattern, I even did not notice it any more.

What did you like or dislike about the Sensor? Do you have suggestions for improvements ?

4 responses

The fit was comfortable and flexible, it was small which was nice since it does not draw to much attention to it, the weight was good it was not heavy on my head. The design was nice since it was made to fit all properly!

less haptic signals, less overwhelming

no, fine as it is.

Do you have any notes about the wearablity of the Vest or Gauntlet ?

3 responses

The vest was too small it did not fit properly on my chest thus the vibrations where not always properly touching my back, which caused issues.
The gauntlet was to big for my hand which had issues when trying to feel the vibrations.

Yes. the glove and the vest should be more tight, to better feel the vibration motors, both location and intensity. definitely needs a fashion designer to have a well fitting model in the future. For a prototype it was ok.

What did you like or dislike about the Vest/Sensor? Do you have suggestions for improvements ?

4 responses

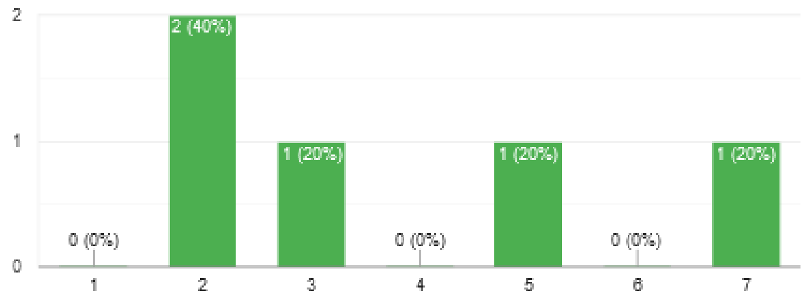
I liked the idea of vibrations, I would still prefer if it would vibrate only if there was an object and not if there is none. Since it was a bit of an overwhelming sensation when wearing the gauntlet as well. But it is still something you can kinda get used to so not a problem. I would prefer if I could feel the vibrations more thought because at times it was not reliable.

back sensors could be positioned further away, especially the columns

not tight enough, the 3 segments on the back are more intuitive than I expected.

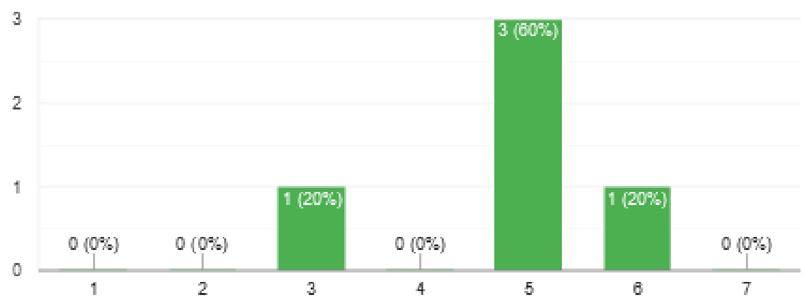
How would you rate the addition of the gauntlet to the vest

5 responses




How would you rate your experience of the device overall ?

5 responses




Which elements, positvely or negatively, influenced your experince of the device ? Were there any points of frustration ?

5 responses

I felt safe since the environment created was done nicely. I liked the idea of the vibrations on the vest but they needed to touch the back properly at all times something the right one did not do always. I preferred the vest over the gauntlet since the patterns where not always properly understood! Maybe that was due to lack of fit. Otherwise I was not caused frustration at any moment!

It was annoying that I did not feel able to discern the different vibrations of the vest; the vibration were changing so fast and I did not manage to make sense of the pattern

in chaotic environment it can be very confusing



I need more time to learn. Also, some improvement on the pattern (include rhythm) could be made.

During the test, were you able to imagine or understand your surroundings?

5 responses

Yes which was pretty incredible being able to move through objects just by following vibrations! The vest did a really good job with that!

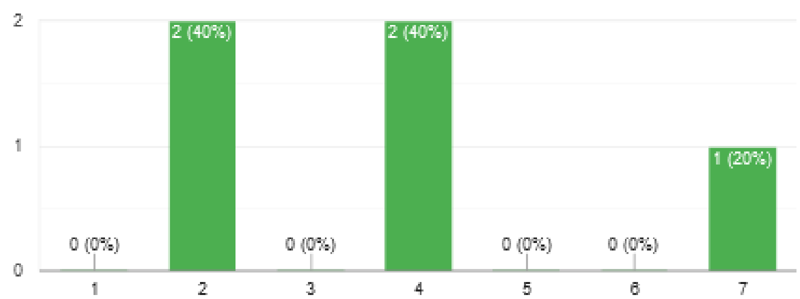
more or less

i was focusing too much on understanding the signals and their relative position to my position, completely forgetting about my absolute orientation

partially

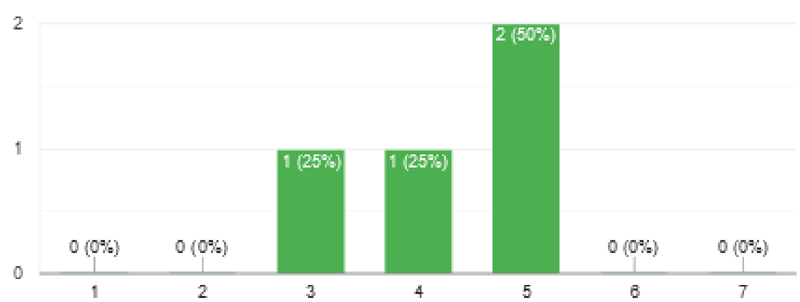
How would you rate the level of confidence you had in your surroundings during the test ?

5 responses



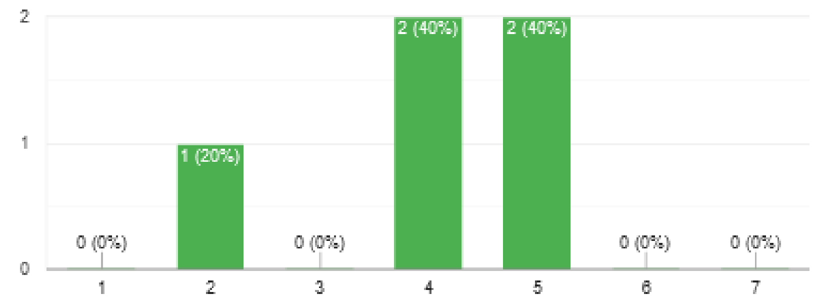

How would you rate your level of confidence in understanding and interpreting the device?

4 responses



How would you rate the clarity of the decice ?

5 responses




Which elements did you find difficult or easy to interpret?

5 responses

The only part that was difficult to properly interpret was the gauntlet

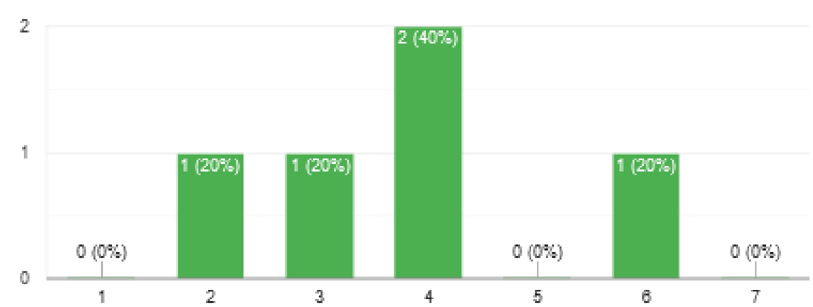
It was easy to interpret the signals of the gauntlet; but I was not able to understand my surrounding based on the signals of the vest

proximity and some of the object patterns (like door and stairs)

pattern on the hand, intensity on the back

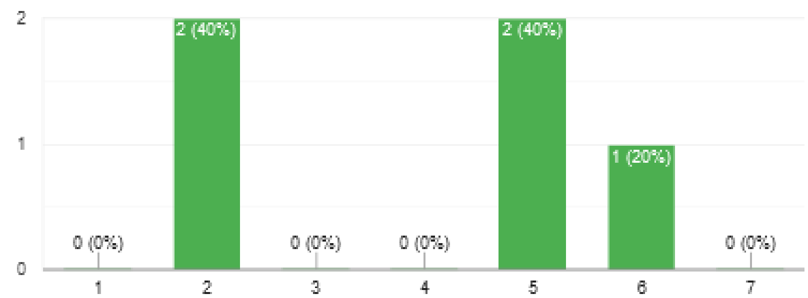
How well were you able to discern the distance of objects ?

5 responses



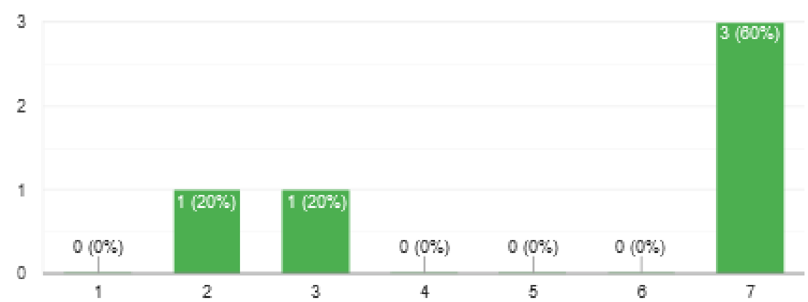
How well were you able to discern the position of objects ?

5 responses




How would you rate your percieved level of safety with the device, given that you were blindfolded ?

5 responses




Which elements did influence you perception of safety in either way ?

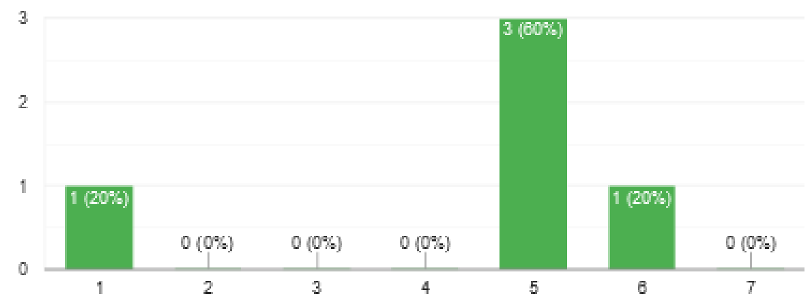
3 responses

I felt very safe at all moments of the experiments but that was because I was in a fixed environment. In a train station I would be a little more anxious!

you guys standing there and helping!

How would you rate the level of intuiton for the vest ?

5 responses




Did you asssociate the vibration of the vest with anything?

4 responses

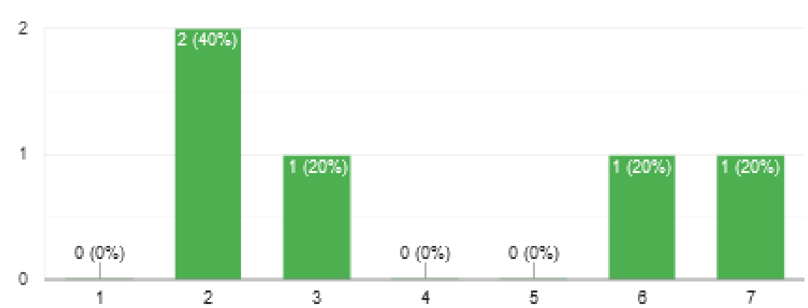
I don't understand the question
Long vibrations = no object
Double vibrations = object

It was more like a massage, or someone tapping my shoulder

no

How would you rate the level of intuition for the gauntlet ?

5 responses



Did you associate the vibration of the gauntlet with anything ?

3 responses

We where supposed to associate the pattern vibrations with objects ?

I think I understand the motivation for some of the pattern

To your knowledge, did you have any intuitive responses to any of the vibrations from the device ?

4 responses

No since I was taught before hand what everything meant

For example the three step vibration for a stairs was very intuitive

no

Do you have any notes about the intuitiveness of the device?

3 responses

It was okay, read above

I do not believe that any complex information can be transferred completely intuitively. The mapping of the grid, however is very easy to understand.

Do you have any further notes about the device, the test or the idea in general ?

3 responses

The device in a hole was good the ideas where nice as well as the testing.
The only changes I would make would be the size of the vest and gauntlet. As well as the power of the vibrations and maybe simplify the gauntlet

I am impressed.

If you could change/add one thing, what would you want to change/add?

3 responses

Read above!

better contact with vibration motors, small changes in pattern on the hand, more time to learn

Do you have any suggestions about the device, the test or the idea in general?

3 responses

Read above

this should be developed further