SMART GLASSES FOR THE BLIND

By

Abdul Maaieh (amaaieh2)

Ahmed Nahas (anahas2)

Siraj Khogeer (khogeer2)

Final Report for ECE 445, Senior Design, [Fall/Spring 201x]

TA: Sanjana Pingali

May 1st, 2024

Team No. 38

Abstract

Blindness and severe to moderate visual impairment affects about 43 million and 295 million people worldwide respectively [24]. Our goal was to create a method to help those with visual impairments see through sound. The solution includes sensors and cameras that sense the surroundings for the user and then relays that information using sound through earphones. The user is able to know if they are approaching an obstacle and the direction of the obstacle along with having their surroundings described to them upon request. The subsystems meshed together to create this ground-breaking device are described in detail throughout this document.

Table of Contents

1. Introduction

1.1. Problem

The underlying motive behind this project is the heart-wrenching fact that, with all the developments in science and technology, the visually impaired have been left with nothing but a simple white cane; a stick among today's scientific novelties. While the current solution, the cane, may help blind people walk, it is an outdated solution that only tells them there is an object if they touch it with the stick or possess exceptional hearing skill to detect a closeby obstacle. This causes the blind trouble and sometimes it could put them in harm's way if they don't place the cane in the right place. This is a major problem that blind people deal with on a daily basis. They also have no way of knowing what obstacle is in front of them or having the scene in front of them described without the help of another person. These lower the quality of life of blind people causing them more trouble throughout their life.

1.2. Solution

Our overarching goal was to create a wearable assistive device for the visually impaired by giving them an alternative way of "seeing" through sound. The completed solution revolves around a pair of glasses that allow the user to walk independently by detecting obstacles and notifying the user, creating a sense of vision through spatial awareness. The device maps the user's surroundings through a depth map and a normal camera, then converts them to audio that allows the user to perceive their surroundings.We are using a low-power I2C ToF imager to build a depth map of the user's surroundings as well as an SPI camera for ML features such as object recognition. The camera and imager are connected to our ESP32-S3 WROOM, which downsamples some of the input and offloads them to our web app for the heavier processing and ML algorithms. The web app then returns an audio signal that the user can listen to to understand what is in front of them.

1.3. High-Level Requirements

- ➔ The device allows the user to differentiate between an obstacle that is 0.5 m away vs an obstacle that is 2 m away [Spatial Awareness].
- ➔ The stereo audio output allows the user to successfully differentiate between obstacles up to 40° to the right and up to 40° to the left side of the user [Spatial Awareness].
- ➔ The device correctly identifies an object up to 1 meter in front of the user and communicates that to the user once prompted [Object Recognition].

2. Functional Overview and Design

2.1. Block Diagram

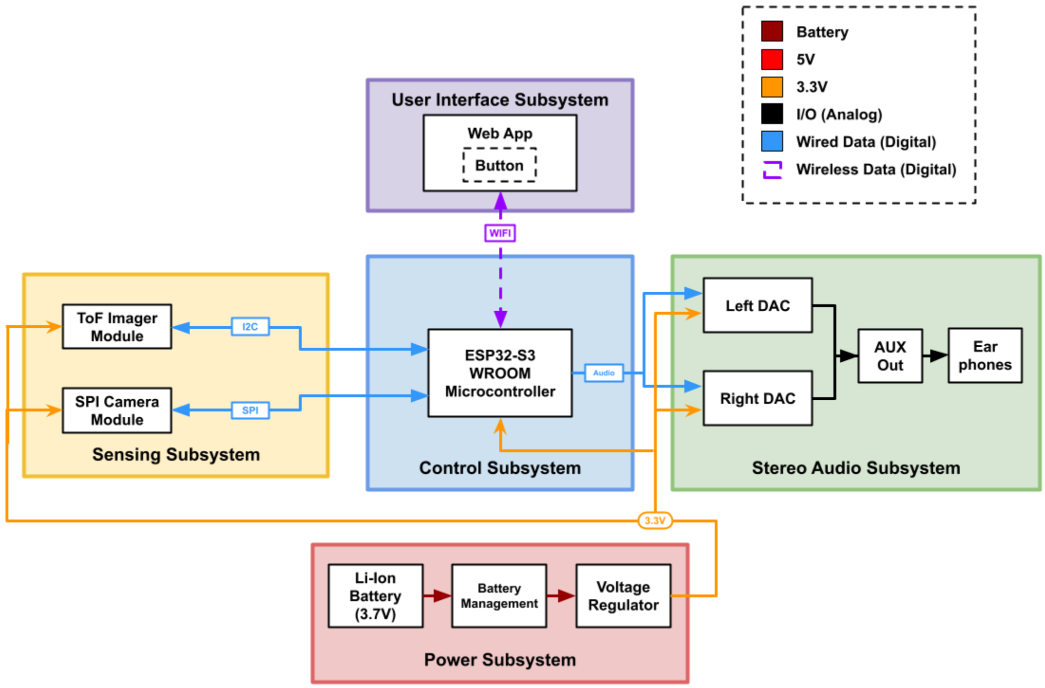


Figure 1: *Device Block Diagram*

Our block diagram, shown in Figure 1 above, includes all aspects of our device. It includes all of our subsystems: Power, User Interface, Control, Sensing, and Stereo Audio. The power subsystem manages the battery, and regulates the voltage to provide to the rest of the circuit. The sensor subsystem captures all the data from the sensors and sends it to the ESP32 through I2C and SPI. The control subsystem receives input from the sensing subsystem, performs some of the data processing, and offloads the data through WIFI to the User Interface subsystem, wherein the Web App performs most of the image processing and machine learning algorithms (pressing the button on the web app triggers the object recognition algorithm to produce an output), and sends back a corresponding spatial stereo audio output back to the ESP. The stereo audio system then takes audio from the ESP32 and plays it through the onboard AUX port. Our Printed Circuit Board (PCB) serves as the central hub for several subsystems in our device. It houses both the control and stereo audio subsystems entirely. The power subsystem is also integrated into the PCB, with the exception of the external battery, which connects via a cable. The sensing subsystem, while fully connected to the PCB through cables, is not physically attached to it. Finally, our User Interface subsystem is our only subsystem to not have any physical connection to the PCB; it connects wirelessly via WiFi.

2.2. Subsystem Overview and Design

2.2.1. Sensing Subsystem

The sensing subsystem consists of two main components: the time of flight sensor (ToF) and the SPI camera. These mesh together to receive an extensive understanding of the surroundings.

The ToF sensor is used to quickly receive a depth mapping of the space in front of the user. The sensor uses an IR laser to gather an 8x8 map of the distance of objects in front of it. The depth map data is sent to the ESP32 using I2C to be transformed into spatial audio.The specific sensor being used is STM VL53L7CX, chosen for its wide FOV and ease of use. The ToF sensors operate at 15 Hz when at 8x8, and draw around 100mA.

As for the SPI Camera, it allows us to capture a colored JPEG image of the user's surroundings. The captured image allows us to implement egocentric computer vision, processed on the web app (described in detail in the User Interface subsystem section 2.2.2 below). The camera is typically idle, until a capture signal is received by the ESP32 from the web app, triggering an image capture. Camera commands and image data are sent using the SPI bus. The camera captures a 640x480 JPEG image and is sent to the ESP32 which then will forward the image to the web app for processing (described in detail in the control subsystem section 2.2.3 below). The camera being used is the “ARDUCAM MEGA 3MP”.

2.2.2. User Interface

The User Interface and ML subsystem consists of a web app including a button. This subsystem is the main processing unit for our data collected by the sensing subsystem. The User Interface Subsystem has the ability to receive two types of data to process from the control subsystem, an 8x8 array of depth map information and a colored JPEG image.

The 8x8 array is processed at a quick latency for a consistent stream of critical information delivered to the user. This information is processed into a form of beeps ranging in speeds and volume to alert the user when approaching an obstacle (similar to a car parking sensor). This is the most crucial processing of our device as the safety of the user is dependent on receiving timely warnings if they are approaching an obstacle. Once this 8x8 array is processed into an audio signal, the user interface subsystem sends the signal back to the control subsystem via WiFi.

The JPEG image is a secondary form of data received from the control subsystem. This data collection is triggered by a button on the web app which sends a capture signal to the control subsystem. Once the image is captured and then received by the web app, a series of ML processing and image processing are run to get an understanding of the user's surroundings. This processing creates an audio signal that describes what objects are seen in front of the user. This audio signal is sent via WiFi to the control subsystem.

2.2.3. Control Subsystem

The control subsystem consists of an ESP32-S3 WROOM that acts as the main hub for the other subsystems to connect to. The sensor subsystem has two connections from it to the ESP32. First, the Tof Sensor connects to it using the I2C bus to receive the 8x8 depth map array. Secondly, the camera is connected via SPI, and when commanded by the ESP32, it takes a 640x480 JPEG image and transmits it to the ESP32. The ESP32 then uses the data from these sensors, and sends it via WIFI to the web app. After processing it, the APP sends a 16-bit stereo audio stream, which is then sent to the audio subsystem to be played. The ESP32 is also in charge of programming the sensors if needed, and changing the control settings as needed. This is all done using the same SPI and I2C buses. TCP is used for the WiFi protocol since it offers reliability which removes the chance for corruption with the tradeoff of slower speeds. We also decided to go with WiFi instead of bluetooth due to latency and bandwidth concerns.

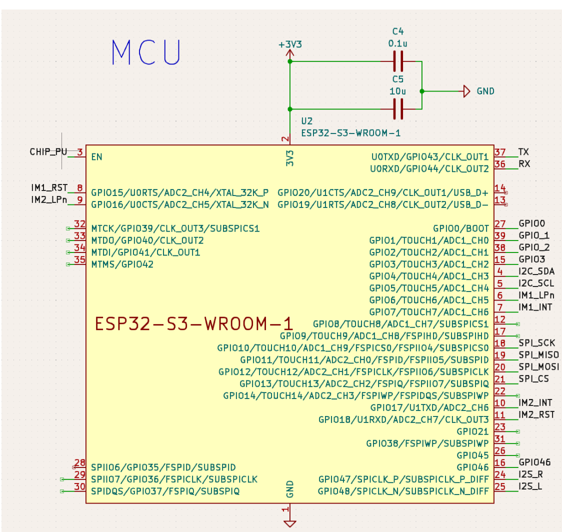


Figure 2: ESP32 Microcontroller

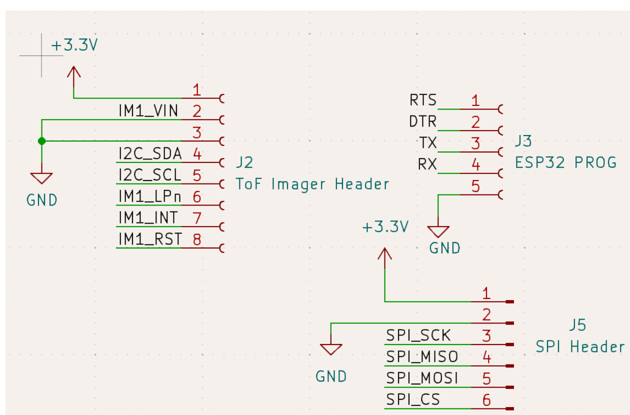


Figure 3: Input and Output headers

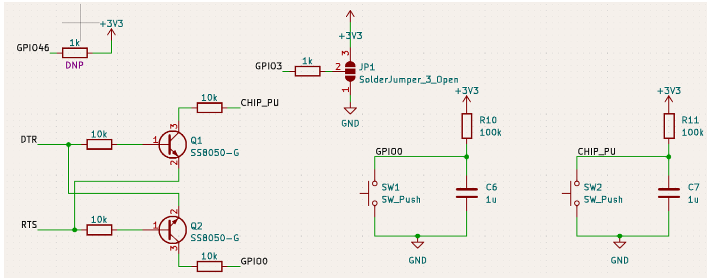
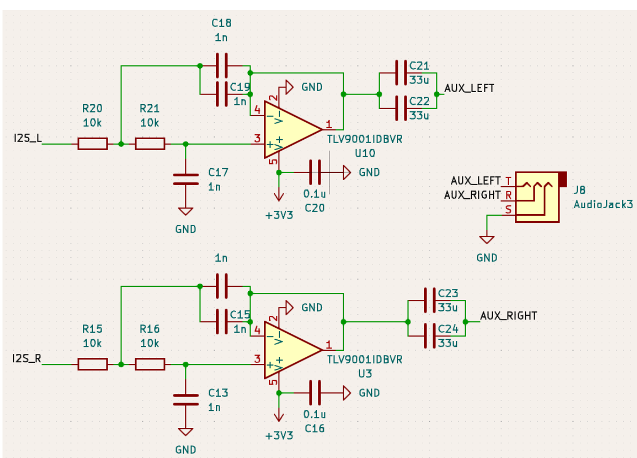


Figure 4: *ESP32 Programming Circuit*

2.2.4. Stereo Audio Subsystem

The audio subsystem takes in stereo audio from the MCU and converts it to analog to play it through the AUX port. This subsystem ensures that each audio channel is distinct, and audio is legible and has minimal noise. An initial approach of using a binary DAC was used, however we found that it took too much space and resources, and wouldn't be feasible for good quality audio. Instead an active filter was used utilizing the Pulse Density Modulation (PDM) capability of the ESP32-S3. We also decided to go with an audio resolution of 16-bits, with a 22.05kHz sampling rate to allow for legible good quality audio, while minimizing data.



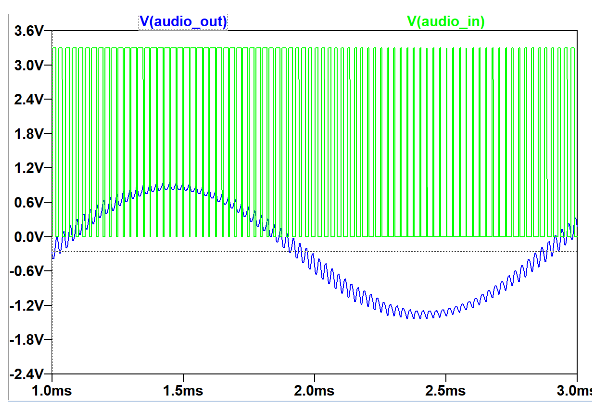


Figure 5: *Audio Subsystem Schematic*

Figure 6: *Audio Subsystem*

simulation

A PDM signal encodes an analog signal in the width of a digital signal. To decode this signal we utilized an active filter using the Sallen-Key topology. This was chosen to ensure minimal volume loss and better isolation. An initial cutoff frequency of 10khz was chosen, being half of what the human ear could hear, however we changed it to 3.56kHz to reduce noise from interference.

$$\omega_0 = 2\pi f_0 = \frac{1}{\sqrt{R_1 R_2 C_1 C_2}}$$

R1 = 100k ohm, R2 = 10k ohm, C1 = 2nF, C2 = 1nF;

Figure 7: *Cut-off frequency of Sallen-Key Filter [23]*

After the PDM signal is decoded we pass it through a high-pass filter to remove the DC component which is comprised of a 66uF capacitor and a series 47 ohm resistor. Since the cut-off frequency of a RC filter is $fc = [2\pi * RC]^{-1}$, a cutoff of 38Hz is achieved. The series resistor also serves to drop down the voltage from the filter, where:

$Vmax = 3.3/2 * (\Omega_{earphones}/(\Omega_{earphones} + 47) = 0.42V$; for a 16 ohm earphone.

Finally the analog signal is connected to the aux port to be played through earphones. The system has two filters, one for each channel. Lastly, the volume is controlled by the microcontroller, allowing the subsystem to work with a large range of earphones.

2.2.5. Power Subsystem

The power subsystem manages a Li-Ion battery, and provides power to the rest of the subsystems excluding the Web-App. To accomplish this, it uses two linear regulators to convert the 3.7V of the battery to 3.3V. Two regulators were used due to temperature concerns and to allow for modular replacement of sensors. This subsystem also contains three protections: short circuit, reverse polarity, and undervoltage protection. This also included a 3D printed enclosure to safely contain the battery.

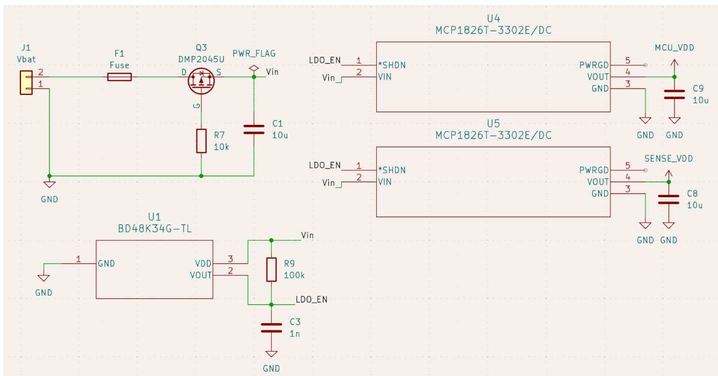


Figure 8: *Power Subsystem Schematic*

The battery voltage is first connected to a 1.25A fuse. This is because our max current draw estimate was ~800mA, and a general rule of thumb is to choose a fuse rating equal to max_current/0.75.[25]. This ensures that if a short occurs, the fuse pops and protects the circuit. The second stage is a P-Channel MOSFET which ensures reverse polarity protection. When the battery is connected correctly, the MOSFET gate voltage is low, which turns the MOSFET on allowing current to flow. When the battery is reversed, the gate voltage is high, turning the MOSFET off, stopping the flow of current. The third stage utilizes a 3.4V undervoltage detector (BD48K34G). This detector disables the LDOs when the voltage falls too low, indicating the battery is too discharged. A value of 3.4V was chosen after analyzing the drop-out voltage of the LDOs, which was found to be around 0.1mV.

$$Vcutoff = 3.3V + Vdropout, \qquad Vcutoff = 3.4V$$

The final stage is two (MCP1826T) LDOs which supply 3.3V to the sensing subsystem, and the microcontroller separately. This LDO was chosen due to its high 1A peak current draw, low dropout voltage, and disable pin. Finally, capacitors and resistors were included for stability and protection reasons.

2.3. Software Design

Our software design can be divided into three parts: a) Spatial Awareness, b) Object Detection, and c) WebApp and Multithreading. Let's delve into each of the above:

a) Spatial Awareness:

Our spatial awareness algorithm takes the near real-time depth map data as input, and outputs stereo audio. The goal was to encode the data in such a way as to allow the user to 'see' through sound. To do this, and to simplify our algorithm for the purpose of giving a general overview here, we wrote an algorithm that takes each "box" or value in the 8x8 array of our depth map, and encoded the volume, left and right panning, as well as top and bottom panning. As for volume, we wanted to replicate the mechanics of how a human ear perceives depth, in almost an inverse square manner. Starting with a simple linear encoding (where closer objects are louder) After hours of experimentation and testing, we landed at this encoding for volume:

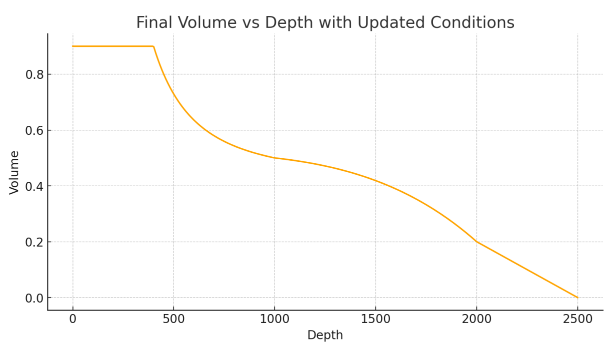


Figure 9: *Volume vs Depth Encoding*

As seen in Figure 9 above, we mapped the initial phase (Around 1000mm to 2000mm) to be almost square, such that a wall or distant object is perceived to be slightly closer, then an inverse square mapping (from 500mm to 1000mm) such that at close distances, smaller differences (like door frames, or couches) are more pronounced.

As for top and bottom mapping, we made it such that each row (starting from the bottom) increases by 25 Hz from the row below it, with the base frequency being set to 180Hz (we landed at this frequency as it allowed the differences to be more pronounced, without the sound being irritating during long usage). Hence, the top rows sound more high pitched than lower rows. As for right and left, we have them such that each column from left to right increases the respective stereo sound output (i.e. if an object is detected in the leftmost column, then the amplitude of the same resultant wave is increased to the left channel. If an object is seen near the middle, an equal amount of amplitude is set to both the right and left channels). Note that we landed at this encoding after experimenting with lots of different encodings, before tweaking the parameters of each encoding, and to also allow for multiple objects to be perceived by the user without wave overlap or interference that would confuse the user.

b) Object detection:

The second key feature involves recognizing and classifying objects around the user, such as tables, laptops, or mugs, and relaying this information via text-to-speech, also using spatial audio to map the location of sounds corresponding to the objects' positions.

We achieved two different levels of complexity in implementing object detection:

- Pretrained Model: We started with off-the-shelf models for basic object recognition and spatial audio mapping. Here, we used YOLOv8 with some minimal tweaks to the input and output, pretrained on the COCO dataset. This allowed for upto 60 object classes that can be identified. This worked surprisingly well; however, accuracy came as a tradeoff to the inference’s high speed.
- Model Retraining and Fine-tuning (Level II): To enhance consistency and accuracy, we retrained the YOLOv8 model and its architecture using multiple datasets (including the large

ImageNetv7 dataset) that better reflect our users' environments. This step included fine-tuning the model to ensure that it’s able to classify more objects at a higher accuracy. Our results were somewhat promising: we now had a wider range of upto 600 classes of objects that can be identified, with lower inference speed (an average of 320 ms) as the tradeoff.

We mainly used YOLOv8 for its flexibility in training, its fast inference speeds, and its ability to expand to fit different use cases: for example, we incorporated semantic segmentation (masking identified objects in an image), which will enhance and enable future features like Radar (discussed in Next Steps section). Starting with basic models allowed us to quickly prototype, whilst the retrained and fine-tuned model allowed us to achieve more consistent results. We used consistency and number of identified classes as our main objectives here, and consistency was greatly improved with our Level II model as compared to the pretrained Level I model. This iterative process of model development and testing enhanced our system's adaptability, accuracy, and user interaction quality. Figure 10 displays an example of our inference and semantic segmentation:

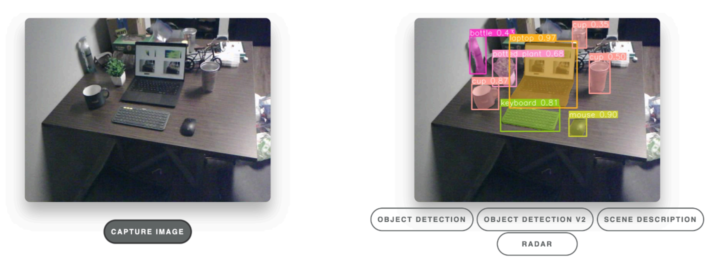


Figure 10: *Example of captured image (left) vs objects identified (right)*

c) Web app and Multithreading

Communication between our python scripts and our ESP was done through WIFI’s TCP protocol. To encapsulate all of our python scripts (object detection, spatial awareness as well as

WIFI communication with the ESP), we used Flask for its compatibility and support for python, and its integrate with HTML and Javascript (which allowed us to design and display different plots and input controls on our webpage). We had to overcome all the obstacles that arose from using Flask (such as its security measures, CORS, which gave rise to lots of issues when we used sockets to enable communication to and from the ESP).

Once we overcame the issues that Flask gave rise to, we used multithreading and spin locks to ensure that all features run with minimal overlap and delay. On the MCU, we had four threads: ToF Thread (constantly receives data from sensor), Camera Thread (to receive data from the camera once prompted), Spatial Audio Thread (to constantly receive spatial audio from the web app and play it), and ML Audio Thread (to output audio once ML is prompted). We have four sockets, one for each thread, such that no data is sent on the same WIFI thread.

The web app had two threads: ToF Thread (to constantly await and receive ToF data from the ESP), and an Audio Thread (that compiles the spatial audio for both the spatial awareness and the object detection, if available, and sends it back to the ESP).

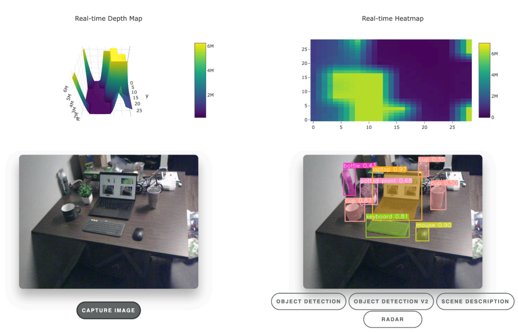


Figure 11: *A screenshot of our web app, with imager plots (top) and object detection (bottom)*

3. Design Verification

All requirements and verification were passed, please see appendix for full list of instructions and results.

3.1. User Interface

The User Interface subsystem passed all requirements. We were able to obtain depth and image data from the MCU. We were able to capture an image from the PCB to the APP in 1.86s which is below our 2s requirement. We were also able to process and send depth audio in 43ms, faster than the frequency of the sensor. Finally, the subsystem was able to use the captured image and identify the objects and relay it to the user using text-to-speech.

3.2. Control

The microcontroller subsystem was successful in all tests. It communicated with the ToF imager and the SPI camera to send their data to the APP through wifi. MCU was also able to play audio from the app with minimal delay. All timing constraints were met.

3.3. Sensing

The sensor subsystem worked fully. The ToF imager captured depth data at 15hz and relayed it to the MCU. The camera captured a 640x480 image and sent it to the MCU within 2 seconds.

3.4. Stereo Audio

The Stereo Audio subsystem received PDM audio from the MCU and successfully demodulated the signal. Audio was played through earphones and was clear and left and right channels were distinct. While all requirements passed, there was unwanted noise due to interference.

3.5. Power

The power subsystem used the Li-Ion battery and regulated the voltage within 5% of 3.3V. Fuse was confirmed to be working, protecting circuit in case of short, drawing no current. Reverse polarity protection was also tested, showing < 1mA of current draw, indicating no power draw. Finally, when the battery went below 3.5V, the power subsystem turned off to stop the battery from over discharging.

4. Costs and Conclusion

4.1. Costs

The total component cost for the project before shipping is \$71.51. With taxes and shipping an approximate cost is \$83. For a complete breakdown please check the appendix. The largest costs were the sensors, and the battery. We can expect a salary of $\$40/\text{hr} \times 2.5\ \text{hr} \times 60 = \$6000$ per team member. Hence, for our team of three, we get $\$6000 \times 3 = \$18{,}000$ in labor cost. This comes out to be a total cost of \$18,109.

Work was split up between team members, and weekly meetings were made. Consistent progress was made throughout the semester, detail can be seen in Appendix C.

4.2. Summary

Our device was a total success. Our device is able to accurately direct users down a hallway or through their day to day journeys. But just directing them through their journeys does not enhance their lives to the fullest. So we were able to go even further and enhance their lives by giving them the ability to have the scene in front of them described by objects along with their direction to them. Our goal was to enhance the lives of the blind and visually impaired and we were able to accomplish just that.

4.3. Accomplishments

We have a long list of accomplishments we were able to complete. To begin, we were able to successfully achieve all of our high level requirements. This gives our device the ability to assist the blind in their daily lives. Our device correctly identifies and relays the difference between obstacles that are 0.5 m and 2 m away to the user. It also correctly identifies the difference between obstacles 40° to the left and right and relays their direction to the user via ambisonic audio. Lastly, our device correctly identifies objects ahead of the user and relays that information to the user along with their direction (via ambisonic audio) when prompted.

But to get these high level requirements to be successes, we needed many parts of our design to work in tandem to achieve them. First, our PCB works successfully and is used in our headset. This was necessary in order to get our device to actually function and be able to do all of the required data collection and transfers quickly. Second, our sensing subsystem works properly in collecting an 8x8 array depth map of the space ahead of the user and delivers it to the control unit along with collecting a JPEG image and delivers it to the control as well. Third, our control subsystem quickly directs all data traffic to the correct destination in order to keep the device functioning properly. Lastly, our power subsystem works properly to power the entire device and the audio subsystem delivers the ambisonic audio through an AUX port.

These minor successes all are necessary but tying them all together was a huge accomplishment as well. There are many little parts to our design that all must function properly and efficiently in order for our device to be considered successful. After tying it all together, correctly mounting everything to our physical headset in a sleek and professional design was also a great accomplishment.

4.4. Uncertainties

During verification, voltage measurements had an uncertainty of +/- 0.1mV, and the current readings an uncertainty of +/- 1mA, which depend on the multimeter and supply. Since the project depends on user interpretation, the uncertainty of audio and depth can't be measured.

4.5. Ethical Considerations

Our device is created as a way to increase the safety, awareness, and lifestyle of the blind. Our device is relied on by people who need it every day. With that in mind, it was created and adapted to be the most accurate it can possibly be. Inaccuracies can potentially cause harm to those who are relied on. The IEEE Code of Ethics states "To uphold the highest standards of integrity, responsible behavior, and ethical conduct in professional activities. To hold paramount the safety, health, and welfare of the public." This holds us to guarantee the safety of those using and relying on our device for guidance within their daily lives.

The ACM Code of ethics also states "An essential aim of computing professionals is to minimize negative consequences of computing, including threats to health, safety, personal security." This includes the head mounting of batteries and other electronic components. We have ensured that all components follow the ACM Code safety measures and have fully inspected any technical components that have the potential to cause harm thoroughly in order to guarantee the safety of anyone using our device. Our components have been enclosed in order to protect the user as an additional security measure.

Our device keeps all data off of any network and is completely processed on the device's system. This helps protect their data and also can help prevent any malicious attacks. The ACM code of ethics states that honesty and trustworthiness are critical for development of a product and we hold that highly. We are completely open and straightforward with our users in any circumstances that might require the proper information. We also want them to know what is happening with their data and that it is secure.

4.6. Impact and Next Steps

Our project truly has the capacity to enhance the lives of visually impaired individuals for the better. We have also seen that there are many research papers that aim to develop Machine Learning or Egocentric Computer Vision features for the visually impaired, but lack the hardware to do so on scale. Since our team ensured that we design our prototype to be modular, in the sense that we have fixed input (camera and imager data) and a fixed output (stereo audio), but the software and web app remain flexible. Therefore, one potential route is to develop an open-source set of glasses that maintain this modularity and allow researchers to continue to develop and test features with the visually impaired community. Hence, this would not only enable visually impaired individuals, but also the research that advances technology for them.

To ensure we continue on our current trajectory, we want to start developing our features around the users. Thus, for next steps, we're looking forward to testing our prototype with a number of visually impaired students here on campus (once we get the necessary approvals)!

Our modular design was essential in allowing us to continually change and test our features and algorithms, and opened the door for constant iteration and updates. As for features, there are several features that we want to implement and test (some of which we have already started working on), and these include:

- **Scene description:** Instead of simply implementing object detection, a potentially useful feature would be using AI to describe the captured scene with detail, communicating full sentences back to the user. We have implemented a small example of this as a proof of concept, but this holds a lot of potential given the recent rise of LLMs that can be leveraged.

- **Radar Algorithm:** Instead of having the depth maps translated into near real-time spatial audio, a different method would be to encode a captured image (at 1 frame-per-second), into a radar-like frequency encoding.
- **Color recognition:** One major issue that visually impaired individuals face is the ability to deduce color, whether it be while picking out clothes or when choosing the correct money note.

In addition to the above software features, there is a lot that can be improved when it comes to hardware:

- **ToF Imager:** Our current imager did not work as well as we had anticipated, as it only allowed for a range of upto 2.5 meters with our current configuration, and did not perform well outdoors due to ambient lighting, nor was it able to 'see' certain objects that were less reflective, darker, or transparent. Therefore, we want to explore different options that encode depth maps more reliably at longer ranges than our current sensor.
- **Communication Protocols:** While we're currently using TCP to communicate back and forth between the ESP and the web app, we would like to explore different methods like UDP that might allow for faster communication. The goal is to stream video at a very low resolution in the background for potential integration with the ToF imager to enhance our depth map algorithms (i.e. we might be able to identify important landmarks like doors and stairs, and add that encoding with a unique frequency to each to be 'heard' by the user when using the spatial awareness feature).

5. References

[1] “Pov headband holder Elastic Head Mount for TACTACAM 4K solo 5.0 6.0 wide WIFI,” Stuntcams, https://stuntcams.com/tactacam-4k-5-0-wide-wifi-pov-headband-holder-elastic-head-mount.html.

[2] Phunk, “BLM PCB side view 2,” MIDIbox Forum, http://midibox.org/forums/gallery/image/350-blm-pcb-side-view-2/.

[3] “Verizon Apple iphone 15 pro max 512GB natural titanium,” Walmart.com, https://www.walmart.com/ip/Verizon-Apple-iPhone-15-Pro-Max-512GB-Natural-Titanium/5055309496.

[4] “EarPods (lightning connector),” Apple, https://www.apple.com/shop/product/MMTN2AM/A/earpods-lightning-connector.

[5] “PKCELL flat 3.7V 2000mah rechargeable lithium polymer 803860 battery W,” Rokland, https://store.rokland.com/products/pkcell-flat-3-7v-2000mah-rechargeable-lithium-polymer-803860-battery-with-jst-type-ph-2-0-plug?currency=USD&utm_source=google&utm_medium=cpc&utm_campaign=google%2Bshopping&gclid=CjwKCAiA8YyuBhBSEiwA5R3-Ey4whPgbrOMG6_z4ZGZ5wZIhXDMhlLOUQZO6LEmSkiu2qe-QqpAAVxoC6fcQAvD_BwE.

[6] “3D time of Flight Camera - DepthVista,” e, https://www.e-consystems.com/3d-depth-cameras/tof-time-of-flight-camera.asp.

[7] “ArduCam-mini OV5642 5mpx 2592x1944px 120fps SPI -,” BOTLAND, https://botland.store/arduino-rpi-cameras/6558-arducam-mini-ov5642-5mpx-2592x1944px-120fps-spi-camera-module-for-arduino-uno-mega2560-raspberry-pi-pico-5904422375133.html.

[8] “IEEE code of Ethics,” IEEE, https://www.ieee.org/about/corporate/governance/p7-8.html.

[9] Code of ethics, https://www.acm.org/code-of-ethics.

[10] E. I. S. Services, “Propeller-less multi-rotor,” :: ECE 445 - Senior Design Laboratory, https://courses.engr.illinois.edu/ece445/wiki/#/regulators/index.

[11] “ESP32-S3-WROOM-1-N8: Digi-Key Electronics,” Digi, https://www.digikey.com/en/products/detail/espressif-systems/ESP32-S3-WROOM-1-N8/15200089.

[12] “TLC27L1ACP last Time Buy,” TLC27L1A | Buy TI Parts | TI.com, https://www.ti.com/product/TLC27L1A/part-details/TLC27L1ACP.

[13] “AZ1117CD-3.3TRG1: Digi-key electronics,” Digi, https://www.digikey.com/en/products/detail/diodes-incorporated/AZ1117CD-3-3TRG1/4470979.

[14] A. Industries, “Lithium ion polymer battery - 3.7V 2500mah,” adafruit industries blog RSS, https://www.adafruit.com/product/328#technical-details (accessed Feb. 8, 2024).

[15] UM2887, https://www.st.com/resource/en/user_manual/um2887-a-software-integration-guide-to-implement-the-ultra-light-driver-of-the-vl53l5cx-timeofflight-8-x-8-multizone-ranging-sensor-with-wide-field-of-view-stmicroelectronics.pdf.

[16] “Arduino Camera: SPI camera module from Arducam,” Arducam, https://www.arducam.com/spi-arduino-camera/#:~:text=The%20Arducam%20camera’s%20SPI%20speed,recommended%20speed%20is%204Mbps~8Mbps.

[17] Pui Audio, https://www.digikey.com/Site/Global/Layouts/DownloadPdf.ashx?pdfUrl=36123F849FAA4FD9939A80185959CD4D.

[18] Amazon.com: Texsport Single Burner Propane Stove : Patio, lawn ..., https://www.amazon.com/Texsport-Compact-Lightweight-Propane-Backpacking/dp/B000P9CZY0/ref=pd_pss_dp_d_1_d_sccl_2_29/147-2861360-3310742?pd_rd_w=gTn0k&content-id=amzn1.sym.3dd6ff5f-e1b5-4c4b-9429-39c9f308776c&pf_rd_p=3dd6ff5f-e1b5-4c4b-9429-39c9f308776c&pf_rd_r=XSNJ0K2JDVKY5BRTS34Y&pd_rd_wg=qwqF3&pd_rd_r=9b49f15e-3086-4714-8713-fbde768dabb3&pd_rd_i=B000P9CZY0&psc=1 (accessed Feb. 23, 2024).

[19]"DAC Essentials: The resistor ladder - Analog - Technical articles - TI E2E support forums," e2e.ti.com, May 29, 2013. https://e2e.ti.com/blogs_/b/analogwire/posts/dac-essentials-the-resistor-ladder

[20] Microchip Technology, "1000 mA, Low-Voltage, Low Quiescent Current LDO Regulator", DS20002057C datasheet, Aug. 2007, https://www.mouser.com/datasheet/2/268/MCP1826_MCP1826S_1000_mA_2c_Low_Voltage_2c_Low_Qui-2325821.pdf

[21]"Li-Ion & LiPoly Batteries," *Adafruit Learning System*. https://learn.adafruit.com/li-ion-and-lipoly-batteries/voltages

[22]"Pololu - VL53L7CX Time-of-Flight 8×8-Zone Wide FOV Distance Sensor Carrier with Voltage Regulator, 350cm Max," www.pololu.com. https://www.pololu.com/product/3418/resources (accessed Feb. 23, 2024).

[23] Wikipedia Contributors, “Sallen–Key topology,” *Wikipedia*, Nov. 02, 2019. https://en.wikipedia.org/wiki/Sallen%E2%80%93Key_topology

[24]“New data shows 33 million people living with avoidable blindness,” Orbis, https://www.orbis.org/en/news/2021/new-global-blindness-data#:~:text=2021%20has%20seen%20the%20official,%2Dto%2Dsevere%20visual%20impairment (accessed May 1, 2024).

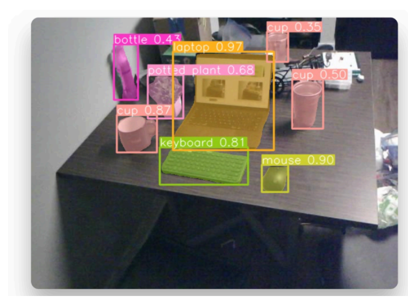
[25] P. Sagsveen, “Fuse Tutorial,” *DigiKey*, Oct. 03, 2017. https://www.digikey.com/en/articles/fuse-tutorial (accessed May 02, 2024).

[26] “The voice for Android,” Augmented Reality for the Blind? There’s an app for that: The vOICe for Android, https://www.seeingwithsound.com/android.htm (accessed May 1, 2024).

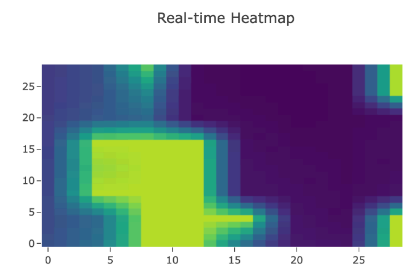
6. Appendices

Appendix A: Requirement and Verification [In-Depth Results]

User Interface

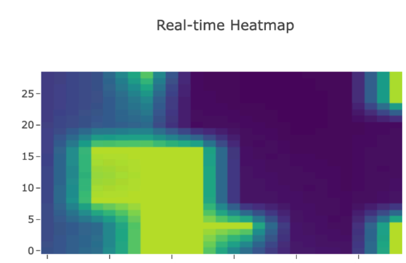
Requirements	Results
Successfully connects to the Control Subsystem through WIFI and receives image at up to 2 seconds of latency	Sending Capture Signal Recieving Data Time Taken: 1.861 seconds Image received and saved as 'received_image.jpg'
Sends back audio data successfully upon processing to the Control Subsystem to be outputted by the Stereo Audio Subsystem with 50 ms of latency as an upper limit	Recieved Depth Data Processing Data Sending Audio ... Total Time: 42.97 milliseconds
Pressing the button will successfully prompt the ML algorithm, the object upto 1 meter away is successfully identified, and communicated back to the user	Sending Capture Signal Recieved Data Performing ML ... Outputting Audio ...  Successfully hear "laptop, keyboard, mouse, cup …." etc. Objects on the left were heard on the left, while objects on the right were heard on the right.

Control

<table>
<tr><th>Requirements</th><th>Results</th></tr>
<tr><td>MCU successfully receives data from ToF sensor and sends it to the APP within 1s.</td><td>

Tof sensor is updated at a 15 Hz frequency and latency is below 500ms</td></tr>
<tr><td>MCU sends a capture signal to the camera when a button is pressed on the application. Receives signal from APP, Captures 2MP JPEG image through SPI bus and sends it to APP through WIFI, all within 2 seconds</td><td>Waiting for Capture Signal ...
Recieved Capture Signal!
Sending Image data ...
Time Taken: 1.635 seconds.</td></tr>
<tr><td>MCU can successfully receive 16-bit audio from APP and sends it to the audio subsystem through its GPIO pins.</td><td>Waiting for Audio Client ...
Audio Client Connected! Waiting for packets
Packet Recived, Writing to Audio (I2S)
“Control Test” was heard on audio output</td></tr>
</table>

Sensing

<table>
<tr><th>Requirements</th><th>Results</th></tr>
<tr><td>The Tof sensor successfully maps the surroundings into a 8x8 array and sends it to the ESP32.</td><td>

The ToF sensor recognized a hand, and motion had minimal latency.</td></tr>
</table>

SPI camera captures a 640x480 image when a button is pressed, and sends the data to ESP32 within 2s	`Waiting for Capture Signal ...`
	`Recieved Capture Signal!`
	`Sending Image data ...`
	`Time Taken: 1.635 seconds.`

Stereo Audio

Requirements	**Results**

Converts 16-bit audio from ESP32 to analog signal at AUX out.	`Running Audio Test`
	`Playing left channel only`
	`..........`
	`Playing right channel only`
	`..........`
	Audio was successfully played using earphones. Left and Right channels were distinct. Audio quality was understandable with some noise.

Stereo audio is clear and spatial distinction can be made from left vs right originating audio	
Plays audio from ESP32 through AUX port using commercial earphones at normal and safe levels.	Audio signal was found to be below 0.4V, matching requirements.

Power

<table>
<tr><th>Requirements</th><th>Results</th></tr>
<tr><td>Regulate 4.2-3V Li-Ion battery to 3.3V +/- 5% at 1A max while ensuring under-voltage protection</td><td><table>
<tr><td>Vin [V]</td><td>MCU_VDD [V]</td><td>SENSE_VDD [V]</td></tr>
<tr><td>3.7V</td><td>3.29V</td><td>3.28V</td></tr>
<tr><td>3.55V</td><td>3.29V</td><td>3.28V</td></tr>
<tr><td>3.4V</td><td>0.04V</td><td>0.03V</td></tr>
</table>Current reading at Vin = 3.4V: 1mA</td></tr>
<tr><td>Ensure that short circuit of battery is stopped to ensure safety of user</td><td>Current before manual short: 438mA
Current after short: 1mA</td></tr>
<tr><td>Ensure that reverse polarity connection is stopped to ensure safety of circuit</td><td>Current in normal operation: 438mA
Current in reverse connection: 1mA</td></tr>
</table>

Note: 1mA is the minimum reading for current using the lab power supply.

Appendix B: Components Costs Breakdown

Components	Footprint	Quantity	Total Price	Source
1 kOhms Chip Resistor	Metric 0805	2	-	Electronic Service Shop
10 kOhms Chip Resistor	Metric 0805	9	-	Electronic Service Shop
100 kOhms Chip Resistor	Metric 0805	3	-	Electronic Service Shop
SF-0603S125-2 (1.25A Fuse)	Metric 0603	1	0.59	Digi-Key
1000 pF Ceramic Capacitor	Metric 0805	7	-	Electronic Service Shop
0.1 µF Ceramic Capacitor	Metric 0805	3	-	Electronic Service Shop
1 µF Ceramic Capacitor	Metric 0805	2	-	Electronic Service Shop
10 µF Ceramic Capacitor	Metric 0805	4	-	Electronic Service Shop
33 uF Ceramic Capacitor	Metric 0805	4		Electronic Service Shop
ESP32-S3-WROOM-1-N16	41-SMD Module	1	-	Electronic Service Shop
DMP2045U-7	SOT23	1	-	Electronic Service Shop
SS8050-G	SOT-23-3	2	-	Electronic Service Shop
MCP1826T-3302E/DC (Regulator)	SOT-223-5	2	1.82	Digi-Key
BD48K34G-TL (3.4V Detector)	3-SSOP	1	0.48	Digi-Key
TLV9001IDBVR (Op-Amp)	SOT-23-5	2	0.94	Digi-Key
VL53L7CX Time-of-Flight 8×8-Zone Wide FOV Distance Sensor	Module	1	19.95	Pololu
Mega 3MP SPI Camera Module	Module	1	25.99	Arducam
Lithium Ion Polymer Battery - 3.7v 2500mAh	JST-PH	1	14.95	Adafruit
SJ-3523-SMT-TR (AUX Port)	3.5mm SMD	1	1.01	Digi-Key
JST PH Conn 1x2	S2B-PH-K-S	1	0.17	Digi-Key
PPTC021LFBN-RC (2 Pin Female Header)	1x2 Header	1	0.33	Digi-Key
4 Position Female Dupont	1x4 Header	4	-	Electronic Service Shop
3 Position Female Dupont	1x3 Header	3	-	Electronic Service Shop
Button	6x3.5mm	2	0.28	Digi-Key
Jumper Wires + Miscellaneous	-	-	5	N/A
Total Components Costs: $71.51				

Appendix C: Schedule

Week of	Tasks	Assigned To
02/19 (current)	Finalize Circuit Designs	Everyone
	Order parts for prototyping	Everyone
	Design Document	Everyone
02/26	Finalize PCB Designs, pass audit	Siraj, Ahmed
	Establish communication, Imagers and Camera with ESP	Abdul, Ahmed
	PCB Review	Everyone
	Design Review	Everyone
03/04	Start developing pre-trained Object Recognition model	Ahmed
	Order parts for PCB	Everyone
	Prototype stereo audio circuit on breadboard	Siraj
	Develop and test spatial audio algorithm on ESP	Abdul, Ahmed
	1st Round PCBway	Everyone
03/11	Continue developing, test object detection algorithm	Ahmed
	Continue developing spatial audio ESP algorithm, test with breadboard circuit	Siraj, Abdul
	Establish ESP WIFI connection, connect to and create web page/webapp	Ahmed, Abdul
	Spring Break	Everyone
03/18	Assemble PCB	Siraj
	Test each PCB subsystem, make revisions	Siraj, Abdul
	Start developing Spatial Mapping Depth algorithm	Ahmed
	Research and decide on effective audio outputs	Everyone
	2nd Round PCBway	Everyone
03/25	Test different configurations, finalize ESP Imager and Camera Code	Abdul

	Design 3D printed enclosure	Everyone
	Data prep for Level 2 Object Recognition model	Ahmed
	Continue working on Spatial mapping algorithm	Ahmed, Siraj
	Continue testing PCB subsystems, make revisions	Siraj
	3rd Round PCBway	Everyone
04/01	Continue testing and configuring Object Recognition model	Ahmed
	Continue testing and configuring Spatial Mapping algorithm	Siraj
	Finalize ESP-Code (communication with imagers, camera, WIFI, receiving and outputting spatial audio, and protocols)	Abdul
	Print and test 3D printed enclosure	Everyone
	4th Round PCBway (if needed)	Everyone
04/08	Finalize webapp	Ahmed, Abdul
	Conduct Requirements and Verification for all subsystems	Everyone
	Make revisions to PCB and enclosure	Siraj
	5th Round PCBway (if needed)	Everyone
04/15	Finalize software, test thoroughly	Everyone
	Test with visually impaired students at UIUC	Everyone
	Mock Demos **Team Contract Fulfillment**	Everyone
04/22	Prepare Presentation, Film Demos	Everyone
	Final Demos **Mock Presentations**	Everyone
04/29	**Final Presentations**	Everyone