ECE 445 - Senior Design Lab

Project Design Document

# Smart Assistive Glasses for the Blind

TA Sanjana Pingali

February 22th, 2024

**Team Members**

Ahmed Nahas (anahas2)

Siraj Khogeer (khogeer2)

Abdul Maaieh (amaaieh2)

# Table of Contents

I. Introduction....... 3

A. Problem....... 3

B. Solution....... 3

C. Visual Aid....... 4

D. High Level Requirements....... 4

II. Functional Overview and Design Requirements....... 5

A. Block Diagram....... 5

C. Subsystem Overview and Requirements....... 7

Subsystem 1: Control....... 7

Subsystem 2: Sensing Subsystem....... 9

Subsystem 3: Stereo Audio Subsystem....... 12

Subsystem 4: User Interface and ML Subsystem....... 14

Subsystem 5: Power Subsystem....... 16

D. Tolerance Analysis....... 19

III. Software Design....... 22

A. Spatial Mapping Algorithm (Depthmap to Spatial Audio)....... 22

B. Object Recognition....... 24

IV. Cost and Schedule....... 29

A. Cost Analysis....... 29

B. Schedule....... 30

V. Safety and Ethics....... 32

A. Safety....... 32

B. Ethics....... 33

VI. References....... 34

# I. Introduction

## A. Problem

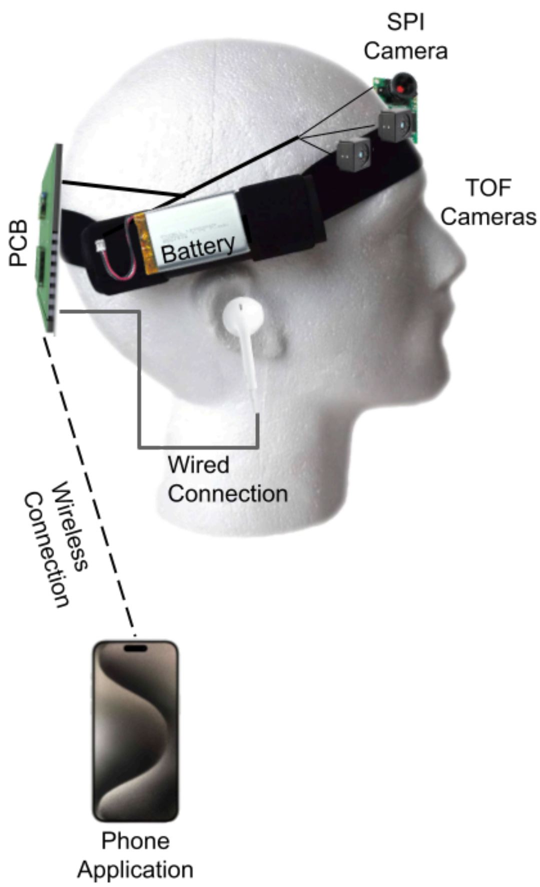
The underlying motive behind this project is the heart-wrenching fact that, with all the developments in science and technology, the visually impaired have been left with nothing but a simple white cane; a stick among today's scientific novelties. Our overarching goal is to create a wearable assistive device for the visually impaired by giving them an alternative way of "seeing" through sound. The idea revolves around glasses/headset that allow the user to walk independently by detecting obstacles and notifying the user, creating a sense of vision through spatial awareness.

## B. Solution

Our objective is to create smart glasses/headset that allow the visually impaired to 'see' through sound. The general idea is to map the user's surroundings through depth maps and a normal camera, then map both to audio that allows the user to perceive their surroundings.

We'll use two low-power I2C ToF imagers to build a depth map of the user's surroundings, as well as an SPI camera for ML features such as object recognition. These cameras/imagers will be connected to our ESP32-S3 WROOM, which downsampled some of the input and offloads them to our phone app/webpage for heavier processing and ML algorithms.

## C. Visual Aid



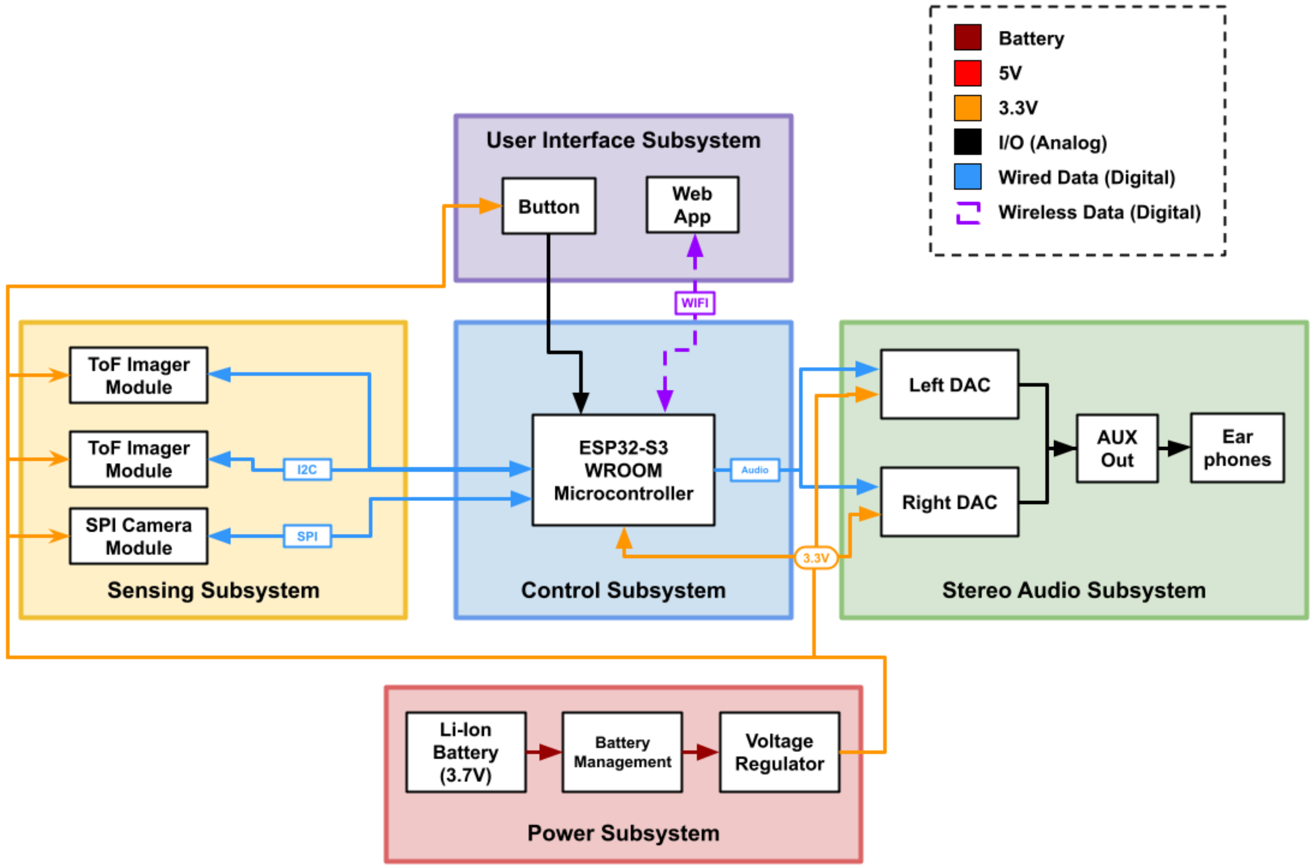

[1] [2] [3] [4] [5] [6] [7]

## D. High Level Requirements

- ➔ The device will allow the user to differentiate between an obstacle that is 1 meter away vs an obstacle that is 3 meters away [Spatial Awareness].
- ➔ The stereo audio output will allow the user to successfully differentiate between obstacles upto 40° to the right and upto 40° to the left side of the user [Spatial Awareness].
- ➔ The device will correctly identify an object upto 1 meter in front of the user and communicate that to the user once prompted [Object Recognition].

# II. Functional Overview and Design Requirements

## A. Block Diagram




The power subsytem 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 it through WIFI to the User Interface subsystem, wherein the Web App performs most of the image processing and machine learning algorithms (pressing the button 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.

## B. Physical Design

The images below depict a rough idea of how the sensors would be places on a set of sports goggles for initial tests, with the PCB attached to the back. If time permits, we will transition to a fully 3D printed set of glasses to enclose our design in a polished final product. [6][7][18]

## C. Subsystem Overview and Requirements

### Subsystem 1: Control

**Overview:**

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 will have two connections from it to the ESP32. First, the Tof Sensors will connect to it using the I2C bus to receive a 2x8x8 depth map array. Secondly, the camera will be connected via SPI, and when commanded by the ESP32, it will take a 2MP JPEG image and transmit it to the ESP32. The ESP32 will then use the data from these sensors, and send it via WIFI to the APP/web server. After processing it, the APP will send 8-bit unsigned stereo audio back to the ESP32 which it will then send to the stereo audio subsystem through 16 GPIO pins. The ESP32 is also in charge of programming the sensors if needed, and changing the control settings as needed. This will be done using the same SPI and I2C busses.

**Requirements and Verification:**

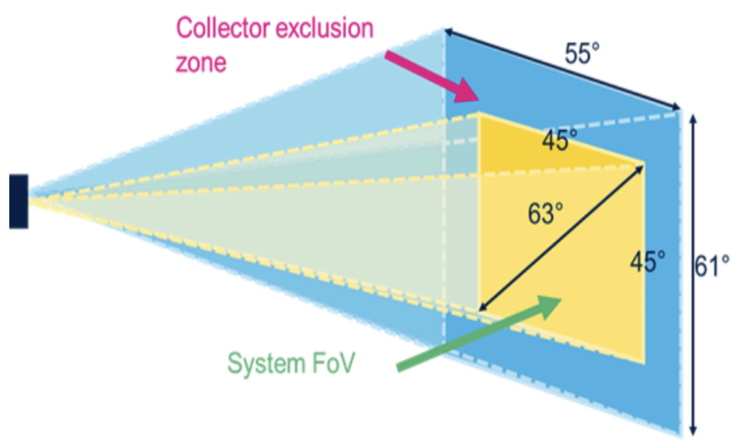
| **Requirements** | **Verification** |
|---|---|
| MCU successfully receives data from ToF sensor and sends it to the APP within 1s. | 1. Power on circuit using 3.7V lab power supply or battery<br>2. Run “control test #1” script on ESP32. Monitor depth map visualization on APP.<br>3. Place hand in front of sensor and move it around<br>4. Ensure that depth map/array changes according to hand movement<br>5. Ensure that delay between movement and change in APP doesnt exceed 1s by using stopwatch |
| MCU sends capture signal to 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 | 1. Power on circuit using 3.7V lab power supply or battery<br>2. Run “control test #2” script on ESP32<br>3. Press button<br>4. Monitor: script will print all steps taken by ESP32 (button pressed, image captured, image sent to APP) as well as compute and print the time starting from when the button was pressed to when it was sent successfully to the APP.<br>5. Confirm time printed on console is less than 2 seconds |
| MCU can successfully receive 8-bit audio from APP and sends it to the audio subsystem through its GPIO pins. | 1. Power on circuit using 3.7V lab power supply or battery<br>2. Run “control test #3” script on ESP32<br>3. Monitor: Script will connect to the APP and receive stereo audio data, to be outputted by the system, and will print steps taken (connected to app, received audio, outputting audio)<br>4. Confirm audio outputs correctly (should hear: “control test 3”) |

## Subsystem 2: Sensing Subsystem

**Overview:**

The sensor subsystem is separated into two sections; the time-of-flight (Tof) sensors, and the SPI camera.

The Tof sensors are used to capture a depth map in front of the user. Two Tof sensors will be used to provide a wide field of view (FOV). Each sensor has a 8x8 array of pixels, and a IR transmitter which maps the distance of objects in front of it. This data allows a person to be able to tell the distance to obstacles around them, helping them navigate without sight. The depth map data is then sent to the ESP32 using I2C to be transformed into spatial audio. The Tof sensors operates at 15 Hz when at 8x8, and draw around 100mA. The data format of the sensors can differ, but minimally, its a 8x8 array of unsigned int, with some header information.

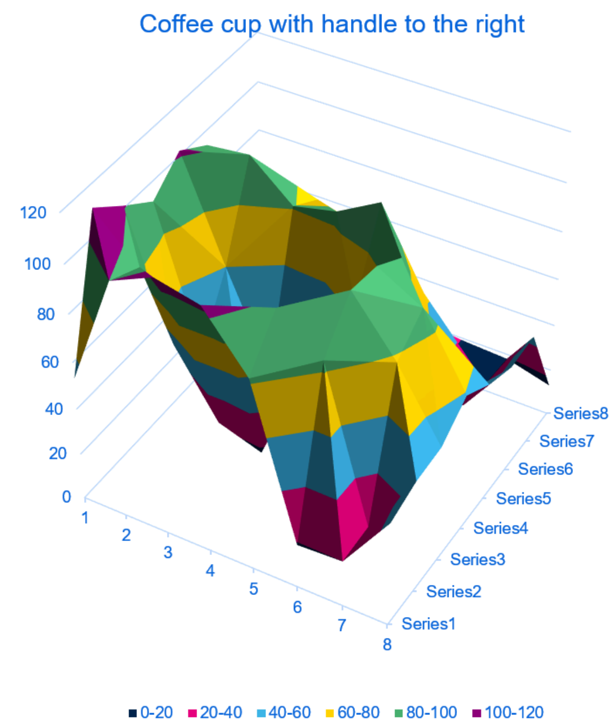



Tof Sensor FoV [15]

As for the SPI Camera, it will allow us to capture a colored image of the user's surroundings. A captured image will allow us to implement egocentric computer vision, processed on the app (more details highlighted in the user interface subsystem). This is exciting as having such an input will allow for other ML features/integrations that can be scaled drastically beyond this course. The camera will typically be idle, until a capture signal is sent by the ESP32, triggering it into capturing a image. Camera commands and image data are sent using the SPI bus. The camera has multiple formats, but in our case a 1920x1080 JPEG image will be sent to the ESP32

**Requirements and Verification:**

| **Requirements** | **Verification** |
|---|---|
| The Tof sensor successfully maps the surroundings into a 8x8 array and sends it to the ESP32. | 1. Power on the circuit using 3.7V and connect to it via WIFI<br>2. Run the "Tof test" script on the ESP32<br>3. Run the visualization program on the APP<br>4. Place hand infront of sensor and compare visualization program with reality. [See below for example]<br>5. Confirm that motion is captured by sensor by moving hand and seeing output change. |
| SPI camera captures a 1920x1080 image when a button is pressed, and sends the data to ESP32 within 2s | 1. Power on circuit using 3.7V lab power supply or battery<br>2. Run "Camera test" script on ESP32<br>3. Press button to capture image<br>4. Monitor: script will print all steps taken by ESP32 (button pressed, image captured) as well as compute and print the time starting from when the button was pressed to when the image is captured and received.<br>5. Confirm time printed on console is less than 2 seconds<br>6. Inspect image to ensure that there are no visual bugs |




Example of Tof Sensor Visualization [22]

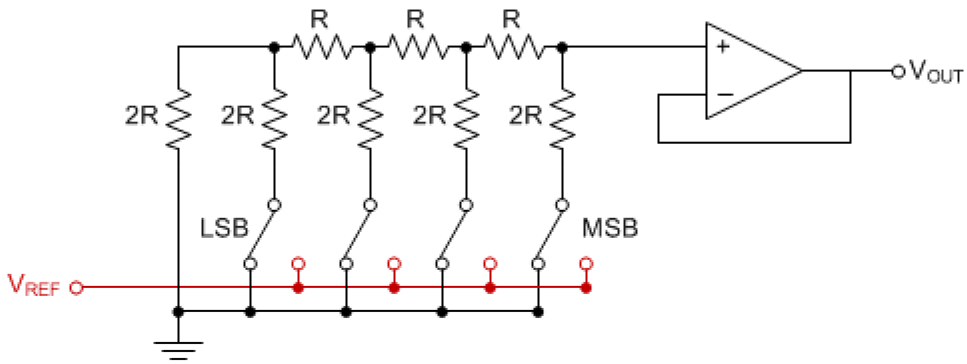
## Subsystem 3: Stereo Audio Subsystem

**Overview:**

The stereo audio sub-sytem will take in two channels of 8-bit audio from the ESP32 and convert them to an analog signal. This signal will be stepped down to audio safe levels and isolated using op-amps to be connected to a AUX. The user will then use the AUX port to connect earphones for onboard stereo audio. The signals will be a total of 16 bits from the GPIO pins of the ESP32. These represent a 8-bit unsigned number and are at standard sample rate of 44.1 Khz.

**Requirements and Verification:**

| **Requirements** | **Verification** |
|---|---|
| Converts 8-bit audio from ESP32 to analog signal at AUX out. | 1. Power on circuit using 3.7V lab power supply or battery<br>2. Run “audio test #1” script on ESP32, to play an audio signal from the APP through the ESP32<br>3. Probe the output of the DAC using an oscilloscope<br>4. Ensure that output signal looks similar to analog input from the app or AC. |
| Plays audio from ESP32 through AUX port using commercial earphones at normal and safe levels. | 1. Power on circuit using 3.7V lab power supply or battery<br>2. Run “audio test #2” script on ESP32, to play audio through the MCU either from the app or from storage<br>3. Connect commercial earphones to aux port.<br>4. Insert earphones and ensure that audio can be heard from both sides. |
| Stereo audio is clear and spatial distinction can be made from left vs right originating audio | 1. Power on circuit using 3.7V lab power supply or battery<br>2. Run “audio test #3”, will connect APP to MCU.<br>3. Ensure that sound only plays in the correct ear that corresponds the audio heard (“left” is heard on the left, “right” is heard on the right) and that noise is minimal. |




R-2R 4-Bit DAC [19]

**Hardware:**

To convert the 8-bit audio to analog a R-2R DAC will be used. The generalized circuit can be seen above. 8 bit audio was chosen since it provided the ability to distinguish speech while being as simple and low bandwidth as possible. To isolate and step down the voltage, a resistor divider and op-amp will be used to provide the power to drive the AUX port. The specific op-amp being used is OPA358 is able to be operated using a single power rail at the 3.3v available. The resistor values chosen should be high enough to not dissipate much power, but not too high such that the reading is off, any value higher than 10kohm should be sufficient.

## Subsystem 4: User Interface and ML Subsystem

The User Interface and ML subsystem will consist of a web app, as well as a button. Since we aim to implement one ML feature as a baseline for this project (one of: scene description or object recognition). This will only be given as feedback to the user once prompted by a button on the PCB: when the user clicks the button (along with the corresponding circuit) on the glasses/headset, they will hear a description of their surroundings. Therefore, we don't need real time object recognition, as opposed to a higher frame rate for the depth maps which do need lower latency.

The web app will carry all the heavy processing for (a) the spatial awareness algorithm as well as (b) the object recognition or scene description algorithms. We plan to use python to develop both (a) and (b), and use React Native to build a user-friendly interface (while this might not be used by the user, it will be helpful with testing). For (a), we will be building over research papers that effectively translate 2D scenes to "audioscapes", which allow the users to 'see' through sound, which is no simple feat, and no effective solution has yet been established. Our algorithm will build over these past approaches and introduce 3D depth maps to the scope, allowing for an exciting avenue to explore and innovate, which we do expect to be quite challenging and complex. As for (b), the Software Design section explains our approach.

**Requirements and Verification:**

| **Requirements** | **Verification** |
|---|---|
| Successfully connects to the Control Subsystem through WIFI and receives data packs at up to 2 seconds of latency as an upper limit for higher resolution images | 1. Power on circuit using 3.7V lab power supply or battery<br>2. Run python script "app test #1" on laptop<br>3. Script will connect to ESP32, prompt image to be captured, print out the image, resolution, and time taken to receive data packs (time from start of script to end of script)<br>4. Confirm that printed time is less than 2 seconds of latency |
| Sends back audio data successfully upon processing to the Control Subsystem to be outputted by the Stereo Audio Subsystem with 50ms of latency as an upper limit | 1. Power on circuit using 3.7V lab power supply or battery<br>2. Ensure circuit is on. Run "app test #2" script on ESP32<br>3. Script will connect to the APP and receive stereo audio data, to be outputted by the system<br>4. Script will compute time taken from start of running the script, to when audio starts outputting, and will print it to console<br>5. Confirm that printed time is less than 50ms of latency. |
| Pressing the button will successfully prompt the ML algorithm, the object upto 1 meter away is successfully identified, and communicated back to the user | 1. Power on circuit using 3.7V lab power supply or battery<br>2. Run "app test #3" script on the ESP32.<br>3. Place object 50cm in front of camera (e.g. mug) with limited to no noise (other objects) obstructing the view in front of the desired object.<br>4. Press button<br>5. Script will print all steps taken by ESP32 (button pressed, image captured, image sent to APP, audio received, audio outputting)<br>6. Confirm audio outputted (identified object by ML algorithm) is the correct object you had placed<br>7. Repeat steps 1-4 with three distinct objects (laptop, water bottle, chair) |

## Subsystem 5: Power Subsystem

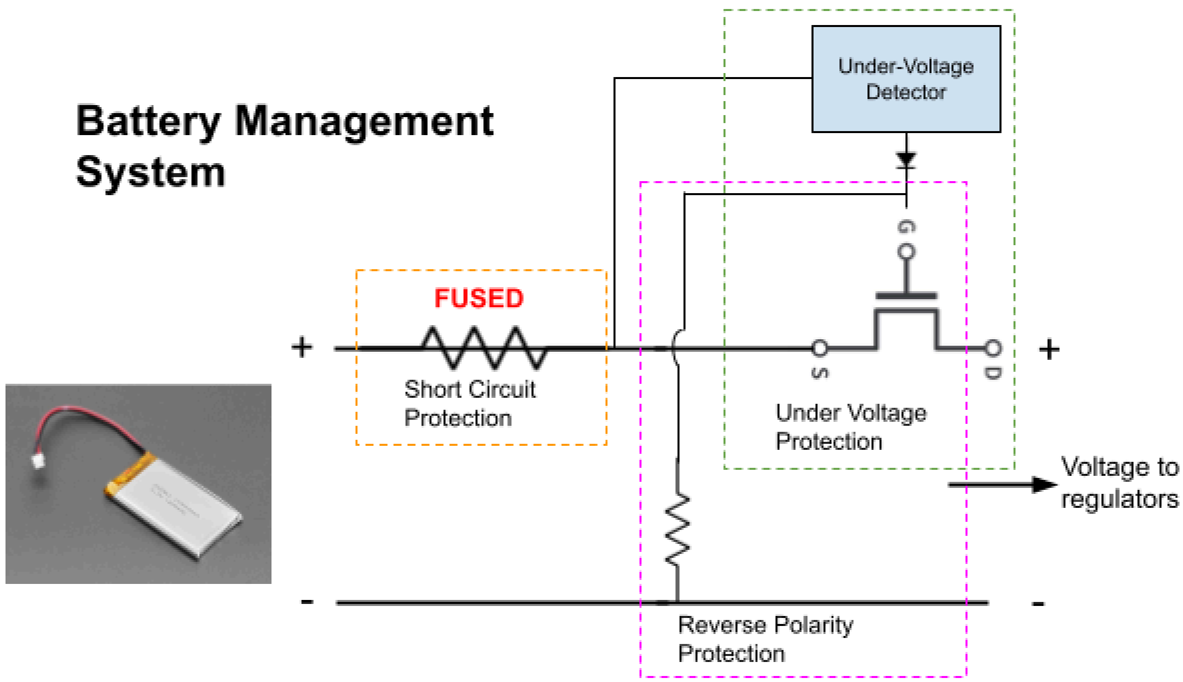
**Overview:**

This sub-system is will supply power to all other subsytems except the app. The power subsystem will contain a rechargeable lithium ion battery pack that will be mounted alongside the rest of the circuits. The battery is regulated to the working voltage of 3.3V and a battery management system will be included to ensure short circuit and undervoltage protection. Battery will be encased to allow for easy removal and charging. Two linear regulators (LDO) will be used to separately. One for the MCU and Stereo Audio subsystems, and one for the Sensor subsystems. This is to reduce heat dissipation, and allow for modular placement of sensors.

**Requirements and Verification:**

| **Requirements** | **Verification** |
| --- | --- |
| Regulate 4.2-3V Li-Ion battery to 3.3V +/- 5% at 1A max while ensuring under-voltage protection | 1. Connect lab power supply to input of power subsystem in place of battery.<br>2. [Turn on power supply at 4.2V and use multimeter to probe power rails to ensure 3.3V +/-5%.<br>3. Slowly lower voltage to 3.5 while ensuring voltage reading is within bounds<br>4. Make the sensors and device work by listening to the audio and capturing an image.<br>5. Lower power supply below 3.4V at which point the circuit should shut-off, drawing less than 2 mA. |
| Ensure that short circuit of battery is stopped to ensure safety of user | 1. Connect lab power supply to input of power subsystem<br>2. Turn on power supply to 3.7V<br>3. Using tweezers connect a 1 ohm resistor to the power rails to draw 3.3A<br>4. Check if 1A fuse is blown or circuit protection worked by reading power supply current reading to be below 10mA. |

| | |
|---|---|
| Ensure that reverse polarity connection is stopped to ensure safety of circuit | 1. Connect lab power supply to input of power subsystem<br>2. Turn on power supply to -3.7V<br>3. Check the current reading to be below 5mA |



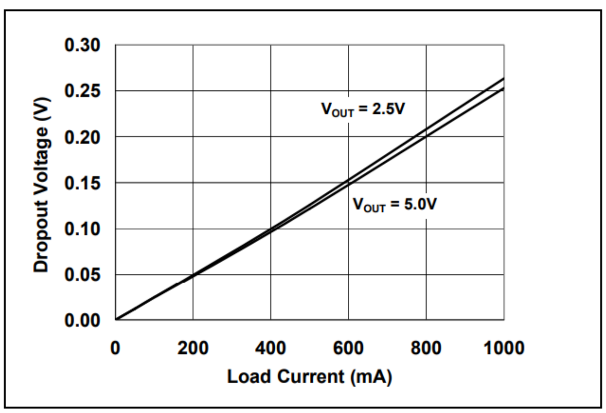

**Hardware:**

The battery being used is a Li-Ion battery from adafruit with a capacity of 2500mAh at a nominal voltage of 3.7, going from 4.2-3v from max to discharged[14]. The battery will use a JST-PH connection to ensure secure connection and protect against reverse connection.

While the battery includes some protections, due to the sensitive location the battery will be located at, which is the head more protections will be implemented to ensure the safety of the user. To protect against shorts, a 1A fuse will be used directly connected to the battery. This is because while the regulator has some short circuit protections, they can fail or cause the regulator to heat up too much. A fuse with resistance under 100 mohm will be used to minimize the voltage drop to the regulator.

For under-voltage and reverse polarity protection, a undervoltage detector such as LM810M3 and a P-MOSFET will be used. The PMOS will act as a switch, when the under-voltage is detected it will be shut off, furthermore if a reverse voltage is detected then it will also be turned off. This will circuit and battery safety.

During the regulation stage two MCP1826S linear regulators (LDO) will be used to supply the sensor subsystem, and the MCU plus the stereo subsystem respectively. While each regulator can supply over 1A each, the amount of heat dissipated by the LDO is a concern since it is close to the battery, but by using two regulators, the heat produced in each is cut in half. The efficiency of the LDO; $n = (Vbat - Vout)/Vbat$. In our case n = 89% which is high enough removing the need to use a more efficient buck/boost converter. MCP1826S was chosen specifically due to its low dropout voltage aswell as its low thermal resistance.




Dropout Voltage vs. Load Current[20] (At 400mA, Vdrop = 0.1V)

## D. Tolerance Analysis

One aspect of our design that might pose a risk to the project is the power draw of the circuits. There are many power hungry components which may be too much for the regulator to handle. All components work using 3.3V from the regulator. The MCU draws under 400mA, Tof sensors under 200mA, and Camera under 150mA. The audio subsystem draws a negligible amount of current. A liberal estimation for the max current draw is 800mA, which can be handled using a single high current regulator, but we chose to use two due to temperature concerns. The formula below can be used to estimate the temperature change of the linear regulator. [10]

$$T_j = i_{out}(v_{in} - v_{out})(\Theta_{jc} + \Theta_{ca}) + T_a$$

| **Parameter** | **Value** |
|---|---|
| $V_{IN}$ (V) | 3.7 V |
| $V_{IN}$max (V) | 4.2 V |
| $V_{OUT}$ (V) | 3.3 V |
| $T_j$ max (C) | 125C |
| $\Theta ja = \Theta jc + \Theta ca$ | 62 C/W |
| Current Draw (mA) | 800mA |

Calculation:

1 LDO $Typical\ T_j = 0.8 * (3.7 - 3.3) * 62 + 25C = 45\ C$

2 LDO $Typical\ T_j = 0.4 * (3.7 - 3.3) * 62 + 25C = 35C$

1 LDO $Max\ T_j = 0.8 * (4.2 - 3.3) * 62 + 25C = 70C$

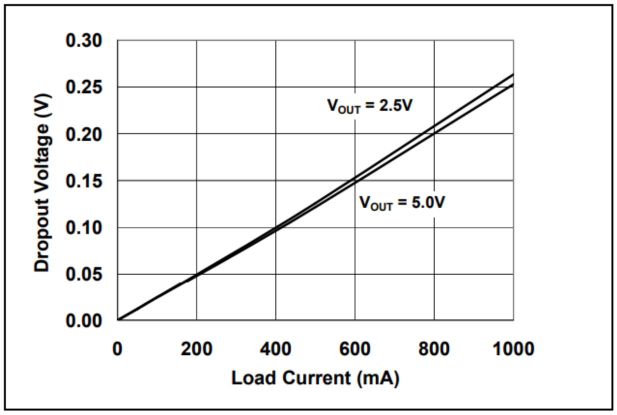
2 LDO $Max\ T_j = 0.4 * (4.2 - 3.3) * 62 + 25C = 48C$

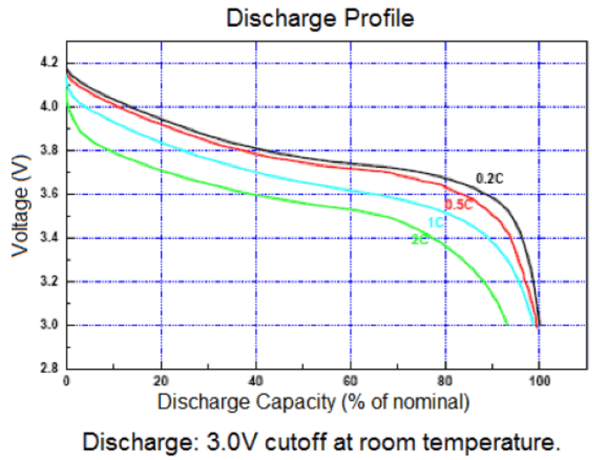
As seen above, while using 1 LDO is typically fine since Tj < 125C, at worst case the temperature can rise above 60C, which is the maximum recommended temperature for the Li-Ion battery. This is why using two LDOs would be better to ensure better heat management.

The second area of concern is the dropout voltage of the regulators. Linear regulators need to have an input voltage that is higher than the output by atleast the dropout voltage to work properly. This is the main reason we chose the MCP1826S since at 400mA each it has a Vdrop = 0.1V. Furthermore, the fuse and P-MOSFET used in the battery management system will have a small voltage drop that we need to take into account. This drop is around 0.1V at max current draw, this value was approximated using the on resistance of the MOSFET and the fuse.

$Vcutoff = Vout + Vfuse + Vmosfet + Vdropout, Vcutoff = 3.5V$

This means that the regulators will work consistently as long as the battery voltage doesnt go below 3.5V. As seen in the discharge diagram below, the Li-Ion battery would have been discharged by 80-90%, which is sufficient depending on the battery size. If the drop-out voltage was higher, then the cut-off voltage for the under-voltage detector would need to increase, cutting down the battery capacity.




Dropout Voltage vs. Load Current[20]

Discharge Profile[21]

The third area of concern is the amount of data generated by the sensors, and if the ESP32 can store and transmit the data. The vl53l5cx packet size is 3.36KB, at a bandwidth of 50.909KB using the 8x8 15 hz mode [15]. Two of these sensors mean that the packet size is 6.72KB with a 101.82KB/s bandwidth. The SPI camera has a bandwidth of 4-8MB/s with a compressed 2MP image being 150KB [16]. This gives us a packet size of 157 KB and max bandwidth of 252 KB/s. And given that the ESP32-S3 has a 512KB SRAM, and 16MB PSRAM data should be limiting.

# III. Software Design

Detailed below are the two main software features that we aim to develop. The descriptions below highlight the general scope as the implementation details are to be finalized as we start development. All below ML/algorithm features will be developed using Python, and encapsulated in a React Native webApp to allow us to easily interface, monitor and debug features as we start testing later in the semester.

## A. Spatial Mapping Algorithm (Depthmap to Spatial Audio)

The key feature we aim to implement is our spatial mapping algorithm, which takes the depth map of the user's surroundings, and translated that into spatial audio, effectively allowing the user to 'see' through sound.

This is no trivial task since the translation from visuals to an audio form that proves effective in representing the surrounding scene in a way that is understandable by the user is a challenge that requires quite a bit of research into how we perceive audio, visuals, as well as testing various solutions with end users to see what works best. There are two approaches that we aim to implement and test:

1. **Spatial Audio:** This method transforms the depth information from our 16x8 depth maps —essentially how far away objects are in each of the 16x8 squares—into sound cues that users can hear through headphones. The process translates each square in this depth

map’s location and distance into a specific sound direction and intensity, allowing users to understand where objects are around them and how close they are.

- By leveraging stereo sound, the system mimics the way humans naturally use sound to locate and identify objects in their environment, making the technology intuitive and immediately useful for navigation and object identification.
- This approach not only enhances the autonomy and mobility of visually impaired users but also enriches their understanding of and engagement with their surroundings.

2. **Radar-like Mapping:** The approach, inspired by 3Blue1Brown's video on Hilbert's curves, adopts a unique method for translating visual data into audio information for the visually impaired. This technique employs a Hilbert curve, a continuous fractal space-filling curve, to convert images into soundscapes in a manner that preserves spatial relationships. By mapping the visual field onto a one-dimensional line that then gets translated into sound, it ensures that points which are close together visually result in sounds that are close in frequency, thereby maintaining the spatial integrity of the image.
   - This auditory representation allows users to interpret complex visual scenes through sound, enabling them to "see" their environment with sound. The user then hears audio that is similar to a radar scanning this new soundscape from left to right, providing a tone/frequency variation depending on depth and edges perceived, which is very unintuitive initially but takes several hours for the user to get trained on. But by leveraging the brain's ability to learn and adapt to new sensory inputs for navigation and object identification.

- Our algorithm would build over this general idea, but also include depth maps align with our captured image to highlight foreground from background and provide a more accurate input, but developing the algorithm to create that translation to soundscape will be quite complex and is a challenge that will require lots of experimentation and testing.

## B. Object Recognition

The second key feature will be allowing the glasses to recognize/classify objects around the user (i.e. identifying a table, a laptop, a mug, etc), and effectively communicate it back to them through text-to-speech. The goal is to first ensure that object recognition works effectively (for a large number of objects in different contexts, with high levels of confidence), then communicating it back in a way that is effective (how and when it is prompted to identify and object, then mapping the location of the object with spatial audio, such that if it is located to the bottom right, that is where the audio is mapped to).

There are various complexity levels to achieving object detection in a way that is more accurate and more tailored to our specific solution:

**Level 1: Pretrained Model Utilization**

- This would include implementing basic functionality using off-the-shelf pretrained models for both spatial audio mapping (based on depth information) and object recognition (using standard computer vision models like YOLO or MobileNet). This level involves minimal customization, focusing on integration and basic output through stereo audio.
- **Complexity:** Low. Relies on existing models and tools with straightforward implementation. Extra complexity would be achieved here by mapping the resulting objects to the spatial audio mapping algorithm.

**Level 2: Model Retraining and Fine-tuning**

- Enhance accuracy by retraining models with custom datasets that are more representative of the environments and objects the users will encounter. This could involve fine-tuning object recognition models for better performance on relevant objects and adjusting audio spatialization techniques based on user feedback.
- **Complexity:** Moderate. Requires data collection and model training but remains within established machine learning workflows.

**Level 3: Custom Model Development**

- Develop custom machine learning models tailored specifically for our application. This could involve creating new neural network architectures that better integrate depth imaging and HD camera data for spatial audio cues and object recognition, optimizing for both accuracy and computational efficiency.

- **Compelxty:** High. Involves significant research and development effort, including algorithm design and validation.

**Level 4: Context-Aware Processing**

- Integrate context-aware processing that dynamically adjusts the behavior of the system based on the environment and user activity. This might involve creating models that can distinguish between different types of environments (indoor, outdoor, crowded spaces) and adjust audio spatialization and object recognition priorities accordingly.
- **Complexity:** Very High. Requires sophisticated algorithm development, possibly involving multi-modal data integration and real-time adaptive learning.

**Level 5: Predictive and Interactive Systems**

- Implement advanced predictive algorithms that not only recognize the current state but also anticipate future changes in the environment or object positions, enhancing user navigation and interaction. This could involve deep learning models that predict trajectories of moving objects and simulate potential future scenarios for more immersive spatial audio experiences.

- **Complexity:** Extremely High. Entails cutting-edge research, potentially developing new techniques in predictive analytics, and real-time interactive audio feedback systems.

Starting with pretrained models will allow us to rapidly prototype initially before advancing through the levels enhances the system's adaptability, accuracy, and user interaction quality. We aim to achieve an intersection between Level 2 and Level 3 by the end of this project, but this will come at notable complexity: most object recognition algorithms are trained and labelled for 2D images. Adding our depth-map array as a feature would definitely enhance the accuracy of our model for this specific use case, since depth information can aid in separating foreground objects from the background, making object detection algorithms more accurate and robust, and it may also contribute to a system's ability to understand the context of a scene in more advanced implementations, such as identifying whether an object is on a table or the floor, which is valuable for providing relevant information to visually impaired users.

Hence, we aim to learn from currently available models trained on 2D images, as well as models trained on images with RGB-D depth, and find develop a model that proves most accurate in our use case. Another big challenge here would be finding a large enough dataset with correctly labeled depth maps and objects classified: one potential solution would be to use RGB-D depth datasets (such as the NYU Depth Dataset V2, SUN RGB-D Dataset,  KITTI Vision Benchmark Suite, or Microsoft COCO Dataset), and play with the entire dataset to transform their labelled high-resultion depth maps into 16x8 resolution to reflect our depth-images, but also find ways to transform the corresponding labels for the model that we develop. All in all, this will result in significant software complexity with custom model development and dataset preparation, as well as testing with multiple different models to find one that works best for our use-case.

# IV. Cost and Schedule

## A. Cost Analysis

The total cost for parts, as listed below, before shipping is $94.79. An approximate 5% shipping cost adds $4.74 and 10% sales tax adds another $9.48. We can expect a salary of $40/hr×2.5 hr×60 = $6000 per team member. Hence, for our team of three, we get $6000× 3 = $18,000 in labor cost. This comes out to be a total cost of $18,109.

| Description | Manufacturer | Quantity | Unit Price | Total Price | Link |
|---|---|---|---|---|---|
| ESP32-S3-WROOM-1-N16 | Espressif Systems | 1 | 3.48 | 3.48 | https:/ |
| VL53L7CX Time-of-Flight 8×8-Zone Wide FOV Distance Sensor | Pololu | 2 | 19.95 | 39.9 | https:/ |
| Mega 3MP SPI Camera Module | ArduCam | 1 | 25.99 | 25.99 | https:/ |
| Lithium Ion Polymer Battery - 3.7v 2500mAh | Adafruit | 1 | 14.95 | 14.95 | https:/ |
| SF-0603S (1 A SMD Fuse ) | Bourns | 1 | 0.56 | 0.56 | https:/ |
| COMPLEMENTARY PAIR ENHANCEMENT MODE MOSFET | Diodes Incorporated | 1 | 0.45 | 0.45 | https:/ |
| STM809 3V Voltage Detector | STMicroelectronics | 1 | 0.87 | 0.87 | https:/ |
| LDO Voltage Regulators 1A CMOS LDO Vout 3.3V ETR | Microchip Technology | 2 | 0.91 | 1.82 | https:/ |
| 3V Single-Supply 80MHz High-Speed Op Amp in SC70 | Texas Instruments | 3 | 1.52 | 4.56 | https:/ |
| CONN JACK STEREO 3.5MM TH R/A | Kycon, Inc. | 1 | 0.76 | 0.76 | https:/ |
| CAP CER 4.7UF 25V X5R 0805 | Samsung Electro-Mechanics | 5 | 0.1 | 0.5 | https:/ |
| RES 10K OHM 5% 1/8W 0805 | Stackpole Electronics Inc | 25 | 0.017 | 0.425 | https:/ |
| RES 20K OHM 5% 1/8W 0805 | Stackpole Electronics Inc | 19 | 0.017 | 0.323 | https:/ |
| RES 100K OHM 5% 1/8W 0805 | Stackpole Electronics Inc | 2 | 0.1 | 0.2 | https:/ |
| Total Price | | | | 94.788 | |

## B. 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 webpage/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, recieving 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**<br>**Team Contract Fulfillment** | Everyone |
| **04/22** | Prepare Presentation, Film Demos | Everyone |
| | **Final Demos**<br>**Mock Presentations** | Everyone |
| **04/29** | **Final Presentations** | Everyone |

# V. Safety and Ethics

## A. Safety

1. This device is created as a way to increase the safety, awareness, and lifestyle of the blind. This device will be relied on by people who need it every day. With that in mind, it must be 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.

2. 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 will ensure that all components are up to the ACM Code with safety measures, such as insulating all batteries and other high temperature components, as well as ensure thorough safety testing, thermal and otherwise, once the project is complete. We will fully inspect any technical components that have the potential to cause harm thoroughly in order to guarantee the safety of anyone using our device. Our components will also be enclosed in order to protect the user as an additional security measure.

## B. Ethics

When it comes to ethics, our project is fairly straight forward. Protect our users with our full ability. Their data must be protected. Our vision is to have no data transmitted out of the system onto a network. This will help 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. While our ML algorithms will use visual data to identify the user's surroundings, none of that data will be stored beyond real-time processing. We will be 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.

## VI. 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).