Boston University
Electrical & Computer Engineering
EC463 Senior Design Project

First Semester Report

Visually Impaired AI Wearable

Student-Defined Project

By

Team 32
Mimir

Team Members
Louis Jimenez-Hernandez louisjh@bu.edu
Heather Li hli9753@bu.edu
Dylan Ramdhan dylram@bu.edu
Houjie Xiong xhj@bu.edu

Submitted: 12/08/24

Table of Contents

Executive Summary……………………………………………………………………………..1
1.0 Introduction……………………………………………………………………..……2
2.0 Concept Development…………………………………………………………………..4
3.0 System Description……………………………………………………………………..6
4.0 First Semester Progress…………………………………………………………………8
5.0 Technical Plan……………………………………………………………………..……9
6.0 Budget Estimate………………………………………………………………….……12
7.0 Attachments…………………………………………………………………….......13
7.1 Appendix 1 - Engineering Requirements………………………………………..…….13
7.2 Appendix 2 - Gantt Chart…………………………………………………….……..14
7.3 Appendix 3 - Other Appendices…………………………………………….......15

Executive Summary

Visually Impaired AI Wearable

Team 32 – Mimir

Cooking poses numerous challenges for the visually impaired such as operating touchscreen appliances, identifying ingredients, and doing measurements among others. To address these obstacles, we wish to propose Mimir, an AI bodycam with an accompanying application to provide real-time guidance in the kitchen. Leveraging computer vision and AI, Mimir will provide object recognition, text extraction, precise measurement, and recipe guidance to the user. Through the connected application, we will be able to utilize cloud-based computations for more complex tasks including web-scraping, image processing, and natural language processing. To assist with accessibility, we will incorporate voice command, and text-to-speech feedback delivering an intuitive and user-friendly experience. By combining cutting-edge AI technologies with an accessibility-focused design, Mimir aims to empower visually impaired individuals with independence, safety, and confidence in the kitchen.

Authored by Louis

1.0 Introduction

Cooking is an essential and enjoyable activity, but for the visually impaired, it poses a myriad of challenges. Tasks such as identifying ingredients, reading nutrition labels, taking measurements, operating touchscreen appliances, or following recipes all often have visual cues making them inaccessible to individuals with limited vision. As technology has gotten more modernized, many kitchens are relying more and more on complex and touchscreen heavy technology which makes them less visually impaired friendly. As it stands, the lack of accessibility makes it difficult for the visually impaired to be independent and safe as they cook without the use of accessibility technology.

Generally, navigating daily life and performing daily tasks without external assistance can present numerous challenges for the visually impaired, such as navigation, identifying key environmental features, or the aforementioned cooking. While there exists various products that aim to resolve these issues, many are quite cumbersome to utilize, limited in functionality, or expensive. Some assistive devices are quite bulky and uncomfortable creating a poor user experience. Other devices are fairly limited in their ability to detect and provide real-time feedback. In addition, many devices are also limited to more controlled environments or have very limited use cases causing them to not be adaptable to daily life. While more advanced devices with complex capabilities do exist, they are largely locked behind large price tags and are more focused on navigation. Thus, there exists the demand for an intuitive, unobtrusive, assistive device that has empowers the visually impaired as they focus on daily tasks, particularly cooking.

To address this issue, we are proposing Mimir, an AI bodycam with a connection application which serves to assist the visually impaired in the kitchen. With Mimir, our goal is to provide real-time guidance and feedback for various cooking tasks giving users more independence and confidence as they cook. Our device will feature a compact, simple and tactile design delivering an intuitive and unobtrusive user experience.

To that end, we aim to create a sleek and lightweight form allowing individuals to incorporate the device into their daily life without hassle. In addition, we will utilize verbal feedback allowing for real time and detailed information being provided to the user. With the use of computer vision and machine learning algorithms, we will provide our device with the ability to better process the information the device's camera takes in. By leveraging these technologies, Mimir will be able to identify text or ingredients, as well as assist with measurements such as depth, level or distance as the user cooks. In addition, we will utilize web-scraping to assist users be able to extract recipes from the web and have Mimir be able to read out and guide them through the recipe.

To assist these capabilities, we will have a companion application which will serve as an intermediary between our device and cloud servers. Through the incorporation of cloud servers, we can provide our device greater processing abilities, thus allowing more advanced functionality without compromising our small form factor. Within the application, the user will be able to insert online recipes for the aforementioned web-scraping. In conjunction with edge ai on our device, the user will be able to maintain a database of recipes. Overall, our design by combining our device and application will provide a practical and accessible design enabling easier and more confident navigation for visually impaired individuals.

Authored by Louis

2.0 Concept Development

This student-defined project began as a device that would help visually impaired individuals navigate their surroundings by using AI. However, through discussions with Former Assistive Technology Director of the Carroll Center for the Blind Brian Charlson, we found that the scope of our original goals for facial recognition, environment navigation, and object recognition were not areas, if we were to make a device with these specificities in mind, that would be the most beneficial for our target consumer. In the Youtube video *Recipes for Accessibility: Cooking Demo by Brian Charlson*, Brian Charlson demonstrates his recipes for fajitas, showcasing the inconveniences faced by the visually impaired when cooking as well as how he has learned to navigate through his situation. In addition to celebrating the skills and experiences of disabled cooks, the ‘Recipes for Accessibility’ series aims to spark conversations and a deeper understanding of how ableism, as well as adaptability, function in food spaces.

After deciding a purpose to guide the direction of our project, our team identified four main problems that we aim to solve by engineering such a device: ingredient identification, recipe accessibility, measurement accuracy, and appliance usability. All of these issues require the device to be able to perform basic capabilities, including the ability to capture images through a camera and to produce an audio output to relay feedback to the user. This audio output would also include the conditions of the device in use, such as battery life and potentially the quality of the images received as there may be blurring due to circumstantial conditions. We would like to implement a rechargeable battery, and although a long battery life of over three hours may not be necessary as the device would most likely only be used within the house with access to charging, there would be greater ease of convenience for charging not to be necessary after every use. It is also important for the device itself to be lightweight and able to withstand movement that may occur when one is cooking.

A basic translation of our problem statement into engineering terms would involve designing a device that meets the following criteria, the first of which is to be able to interpret visual data. This data can be received in the form of text, objects, and spatial information from a camera. Then, computer vision algorithms would not only be needed to be able to accurately recognize what it is but also process the information to extract only the necessary details since within an image, much non-essential information can be captured.

Another important aspect is real-time feedback. For the visually impaired, it is undoubtedly helpful for the feedback for what the individual is currently doing to be relayed as instantaneously as possible so that they have a clear understanding of their immediate surroundings and actions. Moreover, this communication of information needs to be clear and concise to not prevent any misinterpretation as it may lead to serious consequences working in the kitchen, such as burns or cuts. In addition, our design should be user-centric and also have the function of scalability and integration. This would include the ability to integrate seamlessly with external platforms, such as cloud computing and enhanced processing capabilities.

The conceptual approach of our proposed solution, Mimir, centers around leveraging machine learning to assist users with real-time visual interpretation and verbal feedback. With the key features of our device in mind, we considered many alternatives before being confident in our decision to create a wearable. It may be a more obvious choice to choose to create a pair of standalone smart glasses as the camera would then be placed at eye level, which may be more accurate in interpreting information as if someone was seeing it. Moreover, the speakers could be placed closer to the user's ears, which may be beneficial in the event that they are in a noisy environment. However, there were a few drawbacks that made us ultimately move away from the idea. Many devices on the market currently that are able to perform at our desired level cost upwards of $300, and for many of the blind, this is a cost that may be difficult for them to spend. The unemployment rate for people with visual impairments is higher than the general population, and a significant number of blind people live in poverty. It is estimated that the unemployment rate for blind people is 70%, which is 11.5 to 16.7 times higher than the general population rate, and we do not feel that the ability to perform such a daily task should be inaccessible. In addition to the expensive components and limited battery life, we disliked the potential discomfort that the user would experience due to the bulkiness of the components. For those who wore glasses within our team alone, we had discussed issues with slipping and the pressure that it can put on the ears and nose. Ultimately, the high cost and suboptimal user experience made this option less practical.

Another solution that we considered was to create voice-activated smart appliances, which involved leveraging the capabilities of pre-existing appliances with voice-controlled interfaces. Some drawbacks of this idea included narrow functionality due to the limitations of specific appliances, our budget to purchase these devices, and upgrades that would be needed to specifically tailor their systems for full kitchen coverage. Compared to starting from scratch, the work required would not be comparable to its functionality as determined by our project requirements. This option presents limited scalability and an inability to address broader challenges such as ingredient recognition.

By combining machine learning, auditory feedback, and app integration, our conceptual approach is able to provide solutions for a wide range of challenges that an individual who is visually impaired may face when navigating the kitchen. Our device also boasts adaptability as its modular design allows for future enhancements. Although we only aim to target specific problems in relation to cooking, it then can be expanded to be implemented to other daily tasks outside of navigating the kitchen. Improved AI models or cloud-based capabilities also ensure long-term value for users.

Authored by Heather

3.0 System Development

Regarding the hardware of our project, the camera module is a key component. We aim for the camera to be high-definition and equipped with low-light enhancement capabilities so that the device can be used in most lighting situations and different environments. The camera's primary purpose is to capture real-time visual data, which is essential for processing tasks such as text recognition, ingredient identification, and spatial measurements. The device itself would be able to handle initial image processing locally, ensuring efficient operations. For more complex tasks, the processed data can be offloaded to cloud servers, enhancing the device's overall functionality and responsiveness.

The processor would be able to support real-time execution of lightweight AI models. For our prototype, we chose to use the Raspberry Pi 5 for its significantly improved processing power, a dedicated hardware image signal processor built into the chip, and latest Bluetooth technology. However, we have plans to transfer all device capabilities to run on a custom PCB as described in more detail later in the report. These capabilities include tasks such as optical character recognition (OCR) and object detection algorithms, enabling efficient and accurate performance. Additionally, the processor is optimized for low power consumption, ensuring energy efficiency and contributing to an extended battery life for the device.

The power system is driven by a 2500mAh rechargeable battery, which provides 6-8 hours of continuous operation. This ensures reliable performance for extended usage. The device is equipped with a USB-C fast charging interface, allowing it to restore a full charge within 2 hours, minimizing downtime and enhancing user convenience. For audio output, the device incorporates miniature speakers. This component delivers clear verbal feedback directly to the user without obstructing external sound awareness, and the design ensures that instructions are communicated effectively while maintaining safety and situational awareness. Furthermore, the ergonomics and build of the device prioritize user comfort and durability. With dimensions of approximately 50 x 75 x 25 mm and a total weight under 200 grams, the device should be compact and portable. Additionally, we want to offer a versatile mounting option that can attach to the user regardless of the collar or style of the piece of clothing.

The functional features of the system are designed to enhance usability and efficiency, particularly in the kitchen environment. One of the core functionalities is ingredient identification. Our device should achieve 95% recognition accuracy across diverse lighting conditions, ensuring reliability in various environments. The system would process an image within 2 seconds and provide a verbal announcement of the result, making it both quick and user-friendly.

Another key feature is recipe guidance, which employs natural language processing to parse recipes downloaded via the associated app into sequential instructions. This functionality enables the system to deliver step-by-step verbal prompts, allowing users to follow recipes effectively without needing to look at a screen.

Additionally, the device offers measurement assistance through the use of stereo depth estimation algorithms. This capability allows it to calculate distances and volumes with precision, addressing a range of measurement needs. For instance, the system can guide users in pouring a liquid until a specified volume is reached by providing real-time auditory updates. This feature ensures accurate measurements while simplifying complex tasks.

Authored by Heather

4.0 First Semester Progress

Since the beginning of the semester, our team has made substantial progress in developing and refining a prototype aimed at assisting visually impaired individuals with essential daily tasks, such as cooking. The project has been driven by a strong commitment to creating a functional and intuitive design, with the team dedicating significant effort to consistent iterations and collaborative problem-solving. Early stages of development focused on brainstorming and testing potential concepts, while the latter half of the semester saw rapid advancements as feedback was analyzed and applied to optimize functionality and usability. By refining each iteration based on real-world considerations, our team has worked to ensure the prototype not only meets but exceeds the needs of its intended users. Following the iterations, our team focused on prototype development.

In the first iteration of the prototype, the team worked on implementing the usage of a microcontroller, specifically the Raspberry Pi 5 (RPi 5), allowing our device a slim form factor. As seen in the System Design Prototype in *Appendix 7.3.1*, the system diagram demonstrates the overall model of our design. The device itself consists of a camera, microphone, processors (CPUs and I/O), speakers, and edge AI to handle running AI models. The server will be where we handle the majority of computation by running libraries such as Meta's LLaMa or the various image processing and OCR libraries we plan to utilize. Lastly, we have the Android application which will serve as an intermediary between our device and the cloud which will be achieved with bluetooth.

Through testing our prototype, it became obvious that there needs to be a higher computing power. Per *Appendix 7.3.2*, our First Deliverable Testing showed that we took about an average of 4 minutes to do image analysis which was accurate but far too slow. This introduced the notion of customizing a PCB that can handle the processing power and the utilization of running LLMs such as the Meta's LLama AI model. As seen in *Appendix 7.3.3*, an ECE PhD candidate has provided background and detailed explanation of how PCBs are verified/validated prior to the final rounds of scheduling to be printed. However, this calls for consistent iterations to perfect its outcome in being close to 99.99% accuracy.

As in moving towards future implementations and and progressions, the team has received criticism in the ideas of creating newer methods of the server by possibly viewing other more liable cloud computing power, such as Google Cloud Platform (GCP), which can store user data and be referred to whenever the user asks to be recalled. As for recent updates on the first progression, the team is currently refining the Test-to-Speech (TTS) model, finding reliable, affordable cloud servers and experimenting with various image processing libraries. Over the next semester, we will continue to refine our design.

Authored by Dylan

5.0 Technical Plan

Task 1. Design PCB *Expected end date: December*

A custom PCB shall be designed, implemented, and tested to integrate Mimir's core hardware components into a compact and efficient layout. The PCB will manage key functionalities, including the camera module, microcontroller, power management systems, and wireless communication interfaces.The design will be thoroughly tested to ensure compatibility with the device's enclosure and consistent functionality during extended use.

Lead: Dylan

Task 2. Design and Development of Mimir App *Expected end date: January*

A new mock-up of the Mimir app shall be created to refine the user interface (UI) and user experience (UX), ensuring accessibility and simplicity for visually impaired users. The app will be coded to serve as an intermediary between the bodycam and the cloud server, handling tasks such as user inputs, recipe management, and communication with the device. Special attention will be given to integrating voice commands, text-to-speech feedback, and seamless Bluetooth connectivity. The app will be rigorously tested to validate its functionality, responsiveness, and compatibility with the device and cloud infrastructure.

Lead: Louis;

Assistant: Houjie

Task 3. 3D Modeling and Printing Chassis *Expected end date: February*

A custom chassis shall be 3D modeled and printed to house the bodycam and its internal components. The design will prioritize a lightweight, compact, and ergonomic form factor to ensure comfort and usability for the wearer. The chassis must also provide adequate ventilation and structural support to accommodate the device’s PCB, camera, and power systems. Once modeled, the design will be prototyped through 3D printing and iteratively tested for durability, fit, and compatibility with the device's hardware. This process will ensure the final chassis meets both functional and aesthetic requirements.

Lead: Heather

Task 4. Set Up Cloud Server and Connection to App *Expected end date: February*

A cloud server shall be designed, implemented, and tested to enable communication with the app and connected devices. It shall store and manage built-in recipes, images, and text input/output for the Llama AI system while meeting specifications for storage capacity and latency. The server must be thoroughly tested to ensure functionality and reliability, including communication with the app and devices.

Lead: Houjie; Assisting: Louis

Task 5. Web-Scraping *Expected end date: March*

A Python script shall be developed to summarize recipes from given URLs, ensuring the output is concise, reliable, and includes actionable steps. The script will extract key information such as ingredients, measurements, and step-by-step instructions, and format it into a summarized recipe. It will be tested with URLs from three target websites: **www.allrecipes.com**, **www.tasteofhome.com**, and **www.delish.com**, ensuring compatibility and accuracy. The deliverable is a fully functional script capable of producing summarized recipes for various websites.

Lead: Louis; Assisting: Heather

Task 6. Image Processing *Expected end date: March*

A main program shall be developed and deployed on the server to handle image processing and decision-making. The program will use computer vision techniques and machine learning algorithms for tasks such as identifying ingredients, reading text from labels or screens, and assessing depth, level, and distance. This will involve using OCR for text extraction or CV for image analysis. Through these, we will be able to provide real-time feedback to the user in the aforementioned tasks.

Lead: Louis; Assisting: Houjie

Task 7. Local Database *Expected end date: March*

A local database will be developed on our device to store a predefined number of recipes directly on the device. By utilizing edge AI, the system will enable offline access to essential recipe information, allowing users to retrieve and follow recipes without relying on constant cloud connectivity. Testing will ensure that the local database operates seamlessly alongside other device functions, providing reliable support even in offline scenarios.

Lead: Houjie

Task 7. LiDAR Touchscreen Implementation *Expected end date: March*

A LiDAR-based touchscreen assistance feature will be developed to help users interact with touchscreen devices such as microwaves. By leveraging LiDAR cameras on phones, the system will identify the position of the user's finger on the touchscreen and provide real-time feedback on the button or number being touched. This feature will enable visually impaired users to accurately select options or settings on devices with minimal visual input.

Lead: Heather; Assisting: Louis

Task 9. Optimizing Speed of Image and Text Analysis *Expected end date: April*

The image processing and text recognition systems will be improved to make Mimir faster and more responsive. This involves finding ways to process tasks like identifying objects, reading text, and measuring depth more quickly without losing accuracy. The goal is to ensure users receive real-time feedback while cooking, even in different environments or lighting conditions. Testing will confirm that the improvements make the device reliable and easy to use in everyday situations

Lead: Houjie; Assisting: Dylan

Task 10. Tuning Guidance Features *Expected end date: April*

The web-scraping, recipe guidance, and measurement features will be fine-tuned to ensure accuracy, reliability, and ease of use. Web-scraping will be adjusted to extract recipes cleanly from a variety of websites, while recipe guidance will be optimized to provide clear, step-by-step instructions. Measurement features, such as assessing depth, level, and distance, will be refined to deliver precise and helpful feedback to users. These improvements will be tested thoroughly to ensure Mimir provides consistent and valuable assistance in the kitchen.

Lead: Heather; Assisting: Louis

Authored by Houjie

6.0 Budget Estimate

Item	Description	Cost
1	Raspberry Pi 5 Kit	$160
2	RPi Camera	$15
3	Custom PCB	$16
4	Cloud Server	$50
5	Speaker	$10
6	Microphone	$25
7	LiDAR Camera	$100
	Total Cost	$376

In our budget estimate, the Raspberry Pi 5 Kit is the largest expense and was to assist in creating our prototype. It included components such as fans, heat skins, speakers, SD cards, connecting cables, USB converters, power cables. The other major expense is the LiDAR camera which we may not end up committing to due to it raising the cost of our device and one of our original goals was to create a cheap accessible device. For now, we are trying to get similar benefits using ML and computer vision, but if that isn't effective, we will utilize the LiDAR camera.

Authored by Dylan

7.0 Attachments

7.1 Appendix 1 – Engineering Requirements

Team #32 Team Name: Visually Impaired AI Wearable
Project Name: Mimir

Requirement	Value, range, tolerance, units
Case Dimensions	Within 50x75x25mm
Device Weight	Less than 200g
Visual Recognition Accuracy	Text, objects, and spatial relationships can be recognized with at least 95% accuracy
Latency	Image is processed to produce auditory feedback within 2 seconds
Battery Life	Single charge will last a minimum of 6 hours
Durability	Device can withstand movement, minor impacts, and expose to the kitchen environment (e.g. steam, splashes)
Affordability	Target production cost of $250

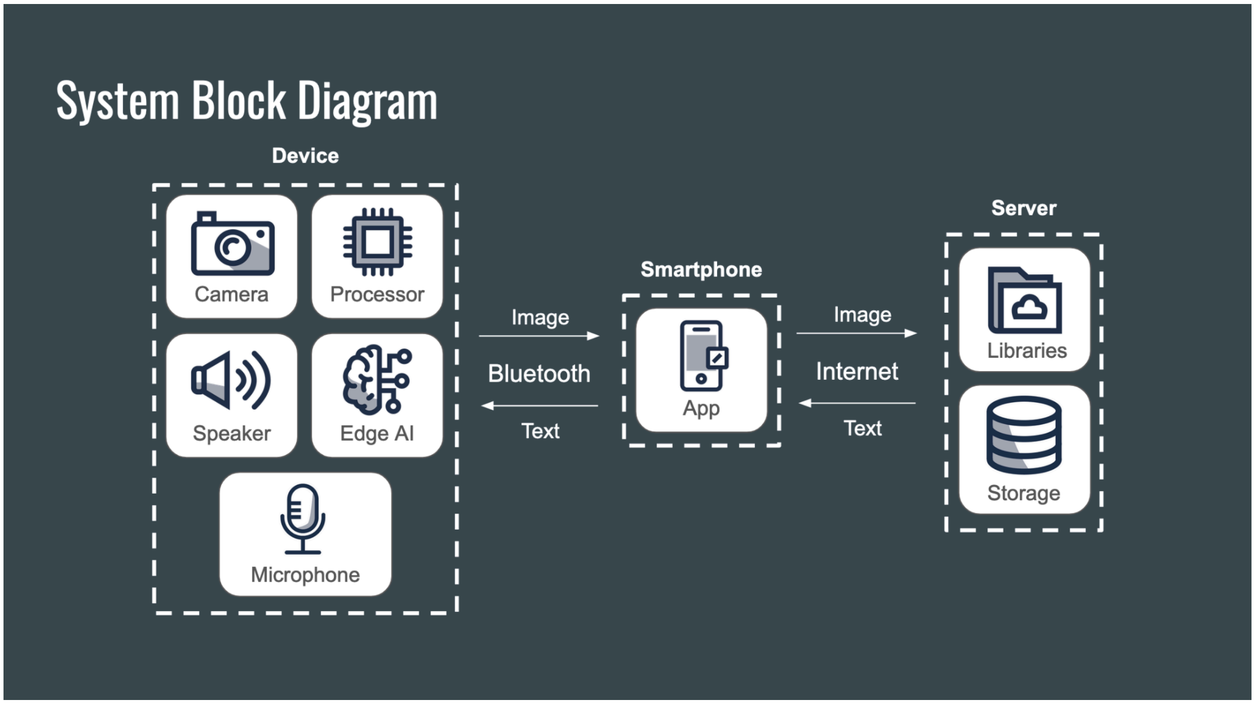
7.2 Appendix 2 – Gantt Chart

Senior Design Team 32

Louis Jimenez-Hernandez, Heather Li, Dylan Ramdhan, Houjie Xiong

TASK	ASSIGNED TO	PROGRESS	September	October	November	December	January	February	March	April	May
Prototype											
OLLaMa Analysis		100%									
Mock Application		100%									
Connect Camera and take images		100%									
3D Print Chassis		100%									
Connect Speaker		100%									
TTS		100%									
Automatically Push and Pull		100%									
Application											
Create New Mock Up		0%									
Set Up Cloud Server and connection to App		0%									
Web-Scraping		0%									
Image Processing		0%									
Device											
Design New PCB		0%									
Vocal Recognition		0%									
3D Model New Chassis		0%									
Print Chassis		0%									
Local Database		0%									
LiDAR Touchscreen Implementation		0%									
Optimization											
Speed Up Object Recogniztion		0%									
Speed Up Text Recogniton		0%									
Tuning Measurement		0%									
Tuning Web-scraping		0%									
Tuning Recipe Guidance		0%									

7.3 Appendix 3 – Other Appendices

1. 

System Block Diagram

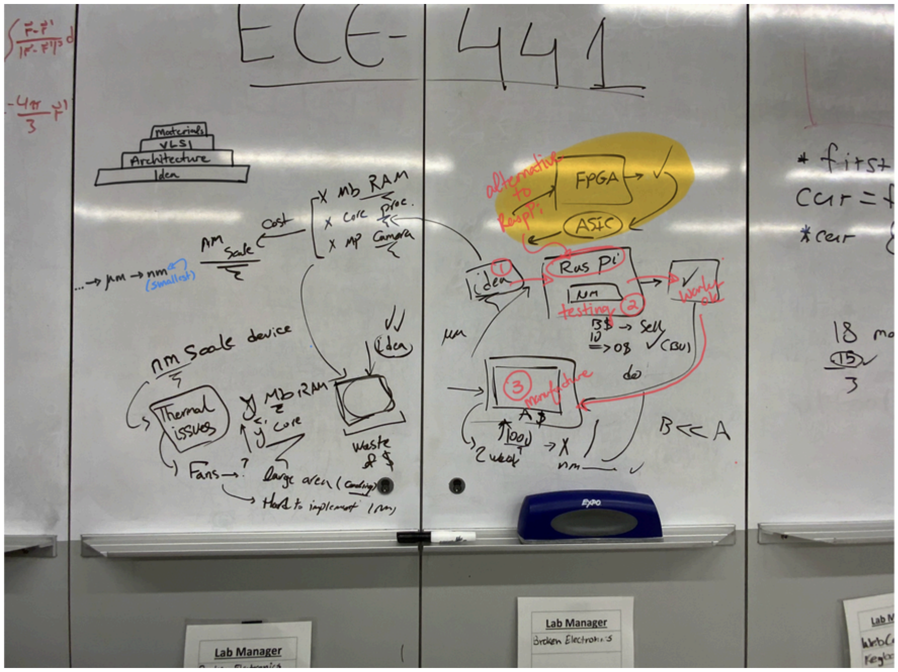
2.

Test Results

Object	Category	Correct	Time
Pasta Label	Food Label	Yes	3:50
Madeline Label	Food Label	Yes	4:13
Foundation Bottle	Bottle	Yes	4:24
Onion Powder	Bottle	Yes	3:41
Check	Document	Yes	3:55
Jury Duty	Document	Yes	4:02

Average Time: 4:01 Minutes

Prototype Test Results

3. 

Consultation of PCB Design by ECE PhD Student