First Semester Report Draft

Team 3: EchoView.AI

Hassan Hijazi CE’24. E-mail: hhijazi@bu.edu
Riya Deokar CE’24. E-mail: riyaa@bu.edu
Nicholas Hardy CE’24. E-mail: nckhrdy@bu.edu
Jazmyn Walker CE’24 E-mail: jazmynw@bu.edu
Marybel Boujaoude CE’24. E-mail: marybelb@bu.edu

Table of Contents

Executive Summary		**2**
1.0	**Introduction**	**2**
2.0	**Concept Development**	**3**
3.0	**System Description**	**4**
4.0	**First Semester Progress**	**5**
5.0	**Technical Plan**	**6**
6.0	**Budget Estimate**	**7**
7.0	**Attachments**	**7**
7.1	**Appendix 1 – Engineering Requirements**	**7**
7.2	**Appendix 2 – Gantt Chart**	**8**
7.3	**Appendix 3 – Other Appendices**	**8**

Executive Summary

The need for accessible, cost-effective, and immediate communication tools for the deaf and hard-of-hearing community is critical and urgent. Current available technologies are costly, complex, and unable to be seamlessly integrated into daily life. Our solution is to introduce a revolutionary wearable device: glasses that provide real-time speech-to-text transcription. By integrating advanced technology into a discrete, everyday accessory, we will be reducing communication barriers faced by individuals who are deaf or hard of hearing. The final deliverable is a pair of assistive glasses, equipped with an OLED display that will display transcribed text into the user's field of vision, a Teensy 4.1 microcontroller for processing, an ESP32-C3 for wireless communication, and a microphone for audio capture. Additionally, we will develop an iOS-compatible mobile application that enables the user to reference past conversations and customize the displayed text based on their preferences. Our technical approach leverages open-source speech recognition software with microcontroller technology, ensuring accurate transcription. Innovative features of our design include low power consumption, a customizable display, and adaptability since it utilizes open-source software that is enhanced by continuous, community-driven developments. Our end product delivers a new way for deaf and hard-of-hearing individuals to communicate, cultivating a more inclusive world where every voice and word can be heard and understood.

1.0 Introduction

Deaf and hard-of-hearing individuals face significant challenges in their daily lives when they wish to communicate with others. These challenges lead to social isolation and limited access to information, adversely affecting their quality of life. Current assistive technologies often fall short of addressing these issues comprehensively. These shortcomings, as well as the gravity of their challenges, underscore the critical need for innovative assistive technologies. While existing solutions often fail to provide a comprehensive solution, our project aims to bridge this gap by designing and building affordable smart glasses from scratch. These glasses will combine cutting-edge technology with affordability, empowering individuals with hearing difficulties to engage fully in conversations and access information effortlessly with added smartphone components. Designing our glasses from the ground up gives us the flexibility to incorporate microphone arrays and speech-to-text processing technologies that work well together. This enables our real-time transcription of spoken language to be highly accurate, ensuring that users can understand and participate in conversations. These glasses will also offer an ergonomic and customizable design, to prioritize user comfort and accessibility. Our user-friendly interface will allow individuals to personalize settings to ensure optimal readability. The end goal is for out glasses to seamlessly pair with a dedicated mobile application so that users can better control and store transcription and translation records. We are also actively exploring the integration of multi-language translation capabilities, which would render our product indispensable for other populations. With a focus on affordability, we are committed to making these glasses accessible to a wider audience. Our project is distinguished by several key highlights: affordability, accuracy, and ergonomics.

2.0 Concept Development

The core problem for our customers who are deaf or hard-of-hearing is the inability to access auditory information in real-time, which creates a communication barrier and hinders their interactions throughout their daily lives. The solution to this problem is to engineer a device that captures spoken words, processes them into text accurately, and presents the transcription in a readily accessible format. This requires a system that is portable, discreet, comfortable, and capable of operating in real-time with high accuracy under various environmental conditions. First of all, the system must be able to process audio in real-time. To do so, it must capture sound effectively via a microphone and then convert this audio data into text format accurately, and with minimal delay. Secondly, the device must be able to work in a variety of settings, ranging from quiet rooms to noisy outdoor areas. It therefore requires advanced noise-cancellation and speech recognition algorithms capable of adapting to different conditions. Similarly, the user interface and display are critical components. The transcribed text must be visible and presented in an easily readable manner. We must consider the display's size, lighting conditions, and any potential pre-existing visual impairments of the user. Additionally, it must be portable and wearable. The device should be lightweight so it does not slip off the user's face and must be comfortable for daily use. We will also consider power efficiency as the user will most likely want to wear the device for a long period. Therefore our product must be designed to minimize energy consumption and reduce the frequency of recharging. To combat these challenges, we have chosen to develop glasses, integrated with an OLED display for real-time speech-to-text transcription. We chose glasses because they are a daily accessory that many people already wear which will make the device practical and discrete. The OLED will ensure that users can read the transcribed text even when looking at others, so that they can communicate naturally.. Through an iOS mobile application, we will also include the option to customize the text being displayed, including the size, brightness, and font. We will also incorporate a Teensy 4.1 microcontroller and an ESP32-C3 module that will allow for sophisticated audio processing and wireless communication. This will enable efficient and accurate real-time transcription. Finally, a microphone is necessary to capture audio to send to the microcontroller. Overall, this approach addresses the core engineering challenges, and offers a practical, user-friendly, and innovative solution.

3.0 System Description

The EchoView.ai glasses will transcribe speech onto a pass through display in real time. The glasses will be designed to minimize costs while maximizing the computational abilities.

The product itself fits together multiple components into a lightweight, unobtrusive and fashionable device. The glasses main microcontroller is the Teensy 4.1 *(or maybe Rpi 2w)* paired with an ESP32-C3. We chose the Teensy 4.1 because of its optimal power efficiency and the number of I2C ports, as well as its 7 UART communication pins. The Teensy 4.1 allows us the flexibility to build on top of existing Open-Source libraries. Since the Teensy 4.1 does not have any Bluetooth or WiFi capabilities, we are using the RX/TX pins to communicate with the ESP32-C3. We selected this model of ESP32 because of its exceptionally small form factor as well as its BLE and WiFi capabilities.

With both the ESP32-C3 and the Teensy 4.1 tied together with the microphones and powered by a 3v battery, the audio will be compressed and sent wirelessly to a mobile application running on the user's device. Once the wireless connection is established, the audio channels will constantly report to the mobile app and compute the live transcriptions using Apple's Siri frameworks. This allows for faster computation times as we are using a native transcribing tool on the device, as well as utilizing the faster processors from the mobile device.

Finally, once the transcription has occurred using Apple's frameworks, the text will be wirelessly transmitted back to the glasses via BLE connection, and displayed on the translucent OLED display over an I2C connection.

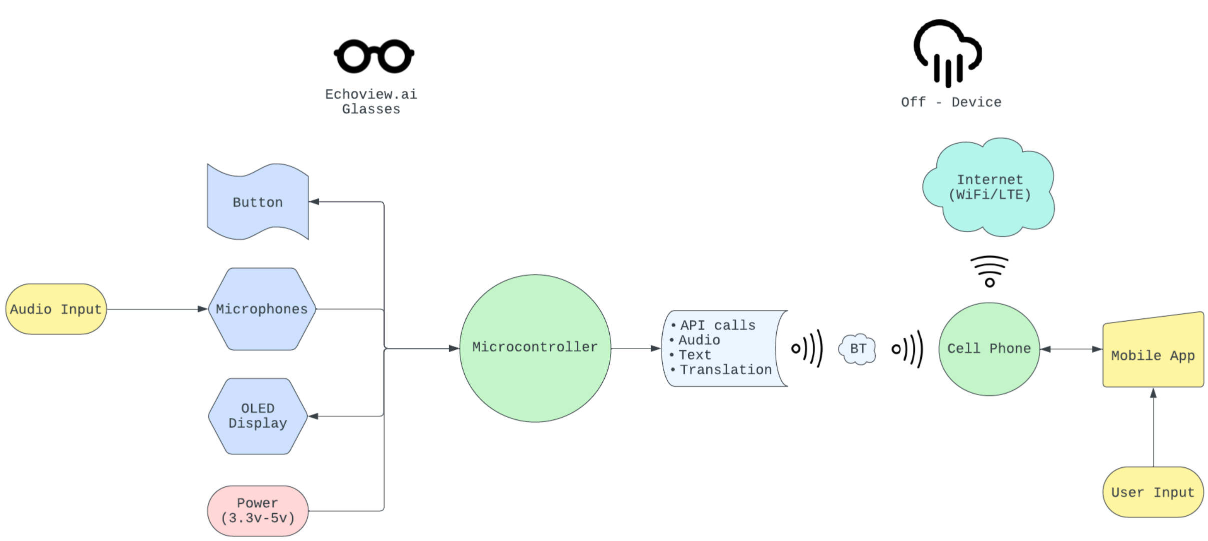


Fig. 1

4.0 First Semester Progress

In the first semester of our EchoView.AI project, we aimed to create affordable glasses from scratch while maintaining functionality. We made substantial technical progress by sourcing hardware, designing custom glasses components, and exploring AI-driven speech transcription.

4.1 Project Milestones and Goals

In the first semester of our project, Team 3 – EchoView.AI set out to achieve several key milestones and goals. Our primary objective was to develop affordable glasses from scratch, emphasizing cost-effectiveness without compromising on functionality.

4.2 Technical Achievements

During this semester, we made technical progress in the development of our glasses. We successfully obtained the hardware components from the microcenter, focusing on affordability and performance. Our technical achievements include:

- Beginning design and fabrication of custom glasses hardware
- Development of a prototype battery power supply with a focus on reusability and weight optimization.
- Initial design and testing of the optical display components, including OLED screens.
- Exploratory research into AI-driven speech transcription and translation for real-time communication.
- Coding of Swift app development for speech-to-text capabilities, a crucial component of our glasses.
- Integrated the innovative translucent OLED display and Teensy microcontroller and completed the pinout configuration and wiring
- Through comprehensive testing and validation, we confirmed that our system can effectively provide accurate transcriptions achieving an average accuracy rate of 95.4% in ideal noisy conditions and 99.4% in a quiet room.

4.3 Goals for Future Development

While our team has made significant progress in various aspects of the project during this semester, there are still key goals we aim to achieve in the upcoming phases of development. These goals include:

- Conducting comprehensive testing of the glasses in various indoor and potentially noisy environments to ensure reliability and functionality in real-world scenarios.
- Implementing advanced features, such as the ability to respond to transcribed conversations, as part of our stretch goals.
- Exploring and integrating AI-driven language processing capabilities to improve the accuracy of transcription and translation, including handling idiomatic expressions.
- Further refining the Swift app development for seamless speech-to-text functionality, ensuring it meets the needs of our target users.

- Continuously optimizing power management techniques to achieve our goal of at least 4 hours of battery life.
- Conducting detailed load-time testing for transcription/translation computations performed on and off the device to determine efficiency and responsiveness.
- Staying updated with developments in competing technologies and incorporating valuable insights into our project.

4.4 Non-Technical Progress

In parallel with our technical efforts, we have also made non-technical progress. This includes refining our project management approach, identifying key stakeholders, and initiating discussions with potential advisors.

5.0 Technical Plan

Task 1: Transparent OLED Display Selection

Select and procure transparent OLED displays for integration into the glasses. OLED displays should meet specifications for size, resolution, and anti-reflectivity. Compatibility with the chosen microcontroller should also be ensured. The deliverable is a set of transparent OLED displays meeting the project's technical requirements.

Task 2: Microphone Procurement

Procure high-quality microphones suitable for speech recognition and transcription. Microphones should meet sensitivity, noise reduction, and compatibility specifications with the chosen microcontroller. The deliverable is high-quality microphones that adhere to project specifications.

Task 3: Battery Power Supply Design

Design and fabricate a battery power supply system for the glasses. The system should support recharging via an external connector, meet weight limitations, and manage heat dissipation effectively. Test the design using a dummy load of 900 ohms. The deliverable is a functional battery power supply system with the ability to recharge and meet all technical specifications.

Task 4: Transparent Lens Procurement

Procure transparent lenses suitable for the glasses. Lenses should meet specifications for magnification and compatibility with the OLED displays. Considering StickTite Lens products for magnification. The deliverable is transparent lenses that fit the glasses and provide the desired magnification.

Task 5: Testing Framework Development
Develop an emulator framework for testing the glasses' functionality. Ensure the framework can simulate various scenarios, including real-world conditions and conversations. The deliverable is a functional testing framework for glasses with the ability to simulate different usage scenarios.

6.0 Budget Estimate

Item	Description	Cost
1	StickTite Bifocal Lenses 2.0x Magnification	$39.98
2	StickTite Bifocal Lenses 3.0x Magnification	$39.98
3	Transparent OLEDs’	$71.97
1	Sandisk 64GB	$19.99
1	Teensy 4.1 Development Board	$31.99
3	Adafruit ESP32	$29.97
	Total Cost	$151.93

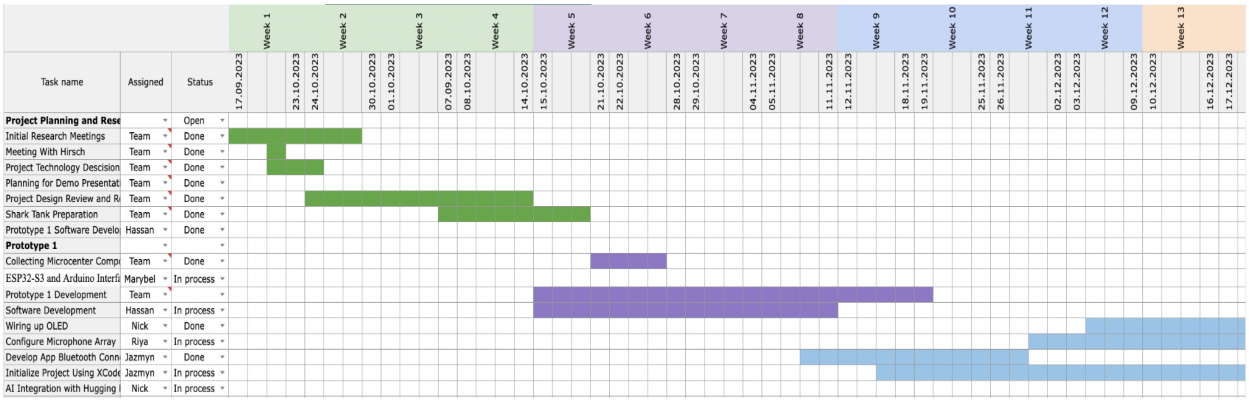
7.0 Attachments

7.1 Appendix 1 – Engineering Requirements

Team #___3___ Team Name: EchoView.AI

Requirement	Value, range, tolerance, units
Weight	<300 grams
Response Time	~10 ms
Noise Specification	Work with noise levels of up to 90 dBA
Compatibility	iOS compatibility
Connection	Connects to WiFi and LTE
Battery Life	2-3 hours operation time
Cost	<$300
Durability	Must withstand drops from 1.5 meters
Audio Input range	2 meters distance from source
Customization	Adjustable text size and contrast on display

7.2 Appendix 2 – Gantt Chart



7.3 Appendix 3 – Other Appendices

Other typical attachments that are added to bolster the competitiveness of your proposal:

- Technical references (in proper bibliographic form) including key URLs.
- Your drawings and schematics (rather than embedding in text)
- Team information sheet (Biographical paragraph on each member; phone numbers and e-mail, history of team and company)

– Nick Hardy (Microphone Array Configuration and Transparent OLED Display Lead):
Nick Hardy leads the configuration of the directional microphone array, ensuring it effectively cancels noise and synchronizes seamlessly with the microcontroller output. Additionally, Nick takes charge of sourcing a suitable OLED display for integration, testing readability in various lighting conditions, and working on a magnification solution. His dual role ensures that both the microphone array and OLED display meet the project's requirements.
CE’24, e-mail: nckhrdy@bu.edu, phone number: 617-610-0539

– Hassan Hijazi (Xcode/Swift/Backend Code and Testing Lead):
Hassan Hijazi is responsible for the Xcode, Swift, and backend code development. He initiates the iOS app project using Xcode, designs the app architecture, and builds user interface components for the app's backend functionality. Hassan's expertise ensures the smooth operation of the app and its seamless integration with the hardware. In addition to his app development

role, Hassan played a crucial part in the testing phase. He was the lead for conducting thorough tests in both quiet and noisy rooms to assess the performance of the real-time transcription capabilities.
CE’24, e-mail: hhijazi@bu.edu, phone number: 224-283-2675

– Jazmyn Walker (App Development Frontend Lead):
Jazmyn Walker leads the frontend development of the app, focusing on creating a user-friendly and accessible interface using IOS frameworks. Her role involves designing and building user interface components. Jazmyn's work ensures that the app provides real-time transcription capabilities. Jazmyn's technical proficiency extends to leveraging iOS frameworks for live transcription. This involves implementing advanced algorithms and speech-to-text libraries to achieve real-time transcription accuracy. Her knowledge of Swift, coupled with the power of iOS frameworks, ensures that the app operates efficiently while delivering high-quality transcription results.
CE’24, e-mail: jazmynw@bu.edu, phone number: 585-709-1460

– Marybel Boujaoude (Microcontroller Interface and Glasses Manufacturing Lead):
Marybel Boujaoude plays a pivotal role within the team, taking on responsibilities in two critical areas. As the Microcontroller Interface Lead, she is responsible for programming both the ESP32-S3 and Arduino Teensy microcontrollers, ensuring seamless real-time data processing. Her role involves guaranteeing connectivity between the microcontrollers and the various hardware components, such as the OLED display and microphone array. Collaborating closely with Nick on data reception, Marybel ensures that data flows efficiently and accurately within the system.
CE’24, e-mail: marybelb@bu.edu, phone number: 603-952-8495

– Riya Deokar (Machine Learning Integration Lead):
Riya Deokar continues her role as the Machine Learning Integration Lead, focusing on researching and selecting speech-to-text libraries from Hugging Face. She actively plans the integration of AI models with both the mobile app and hardware components, aiming to enhance the accuracy and real-time capabilities of our transcription system. Riya defines and monitors metrics to evaluate the quality of our real-time transcription, ensuring it meets the highest standards.
CE’24, e-mail: riyaa@bu.edu, phone number: 510-953-1814

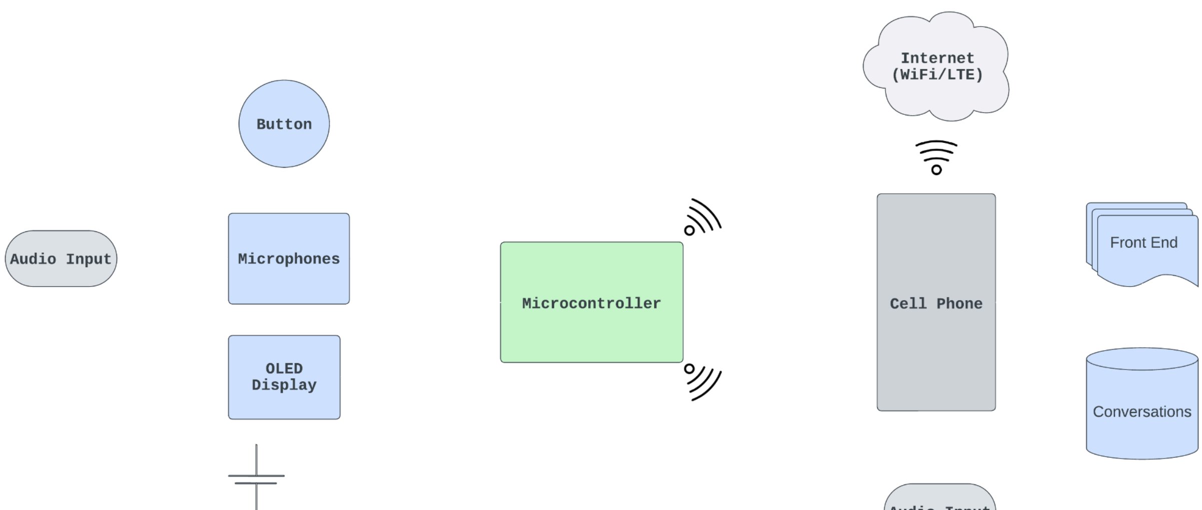

Internet
(WiFi/LTE)
Button
Audio Input
Microphones
Microcontroller
Cell Phone
Front End
OLED
Display
Conversations
Audio Input

Fig. 2

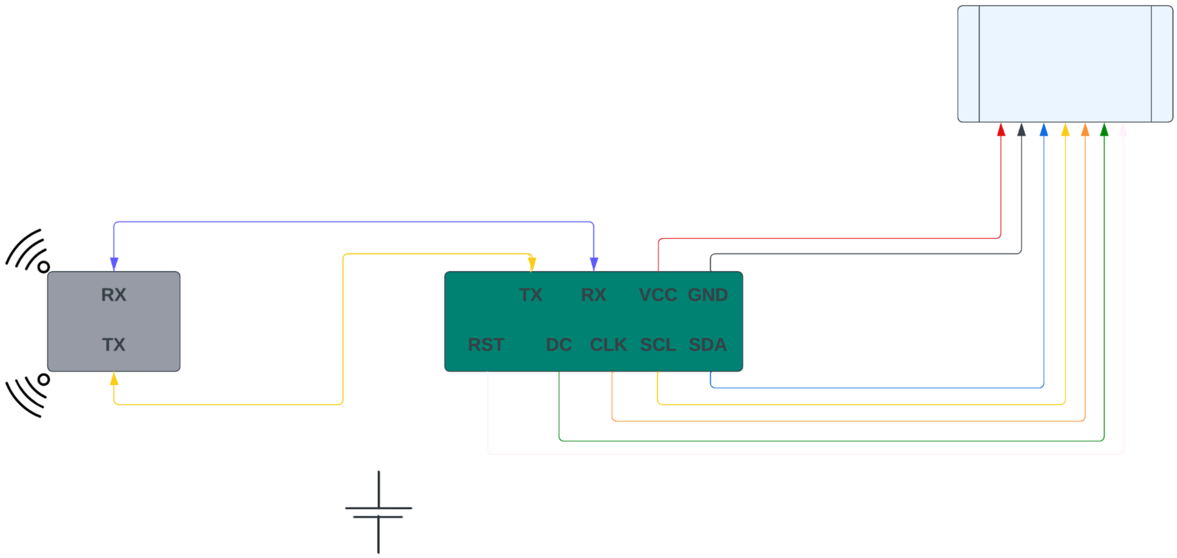


Fig. 3