Boston University
Electrical & Computer Engineering
EC464 Senior Design Project

User Manual

EchoView.ai

by

Team 3
EchoView.AI

Team Members

Nicholas Hardy nckhrdy@bu.edu
Riya Deokar riyaa@bu.edu
Jazmyn Walker jazmynw@bu.edu
Marybel Boujaoude marybelb@bu.edu
Hassan Hijazi hhijazi@bu.edu

User Manual

Table of Contents

Executive Summary

EchoView.AI
Team 03

The need for accessible, cost-effective, and immediate communication tools for the deaf and hard-of-hearing community is a critical issue. 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 proof of concept equipped with an OLED display that will display transcribed text into the user's field of vision, an ESP32-C3 for wireless communication, and two microphones 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 continuously improvable open-source speech recognition software, a practical optical design, and an efficient microcontroller. 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. Introduction

For the Deaf and hard-of-hearing community, seamless communication in a largely hearing world can be complicated and in some cases impossible. Despite continuous advancements in assistive technology, a significant gap persists in providing real-time, accessible, affordable, and convenient solutions. Approximately 1.5 billion individuals globally are affected by hearing loss in one or both ears, with an estimated 430 million people worldwide requiring rehabilitation for severe hearing loss. This number is expected to increase to 700 million people being affected by hearing loss by 2050 [1]. The effects of this communication gap are profound and have damaging consequences. Deaf and hard-of-hearing individuals often find themselves isolated from spoken conversations and missing out on everyday interactions. While current solutions exist, they are costly, lack immediate access, and are overall ill-suited for fast-paced real-world environments. Sign language interpreters can cost up to $150 per hour, not including additional travel costs, and often take prior planning to schedule [2]. Given these challenges, there is a clear and urgent need for an improved solution. The Global Assistive Technology Market was valued at 21.95 billion U.S. dollars in 2022 and is projected to reach a value of 31.22 billion U.S. dollars by 2030 [3]. More specifically, the domestic AR headsets and glasses market was valued at 2.4 billion U.S. dollars [4], and 31.12 billion dollars globally in 2023. AR and VR is a growing market with an expected growth of +366.91 percent by 2027, so entering this market with a more affordable option could accelerate the adoption of AR/VR and grant Echoview.ai a large share of the consumer market. Our proposed Augmented Reality (AR) glasses, with real-time speech-to-text capabilities and a mobile application, aim to address the communication barriers faced by the Deaf and hard-of-hearing community. Through a business-to-business (B2B) and business-to-consumer (B2C) model, we will introduce these Augmented Reality (AR) glasses to consumers who will benefit from improved accessibility at an affordable cost. Hospitals can communicate clearer between patients and staff, educational centers can integrate these glasses into their classrooms to promote an inclusive learning environment, and courtrooms can be sure that legal proceedings are accessible to all participants. Our device will enhance personal interactions, foster inclusivity, and provide a more cost-effective solution to existing assistive technologies.

2. System Overview and Installation

The EchoView.AI glasses system has evolved to utilize the ESP32-C3 microcontroller and the Siri framework for real-time speech-to-text transcription, enhancing accessibility for individuals who are deaf or hard of hearing. This system integrates Bluetooth microphones for audio capture, ensuring clear speech input in diverse environments. The ESP32-C3 microcontroller serves as the processing unit, executing the transcription model that converts speech into text. The text is then displayed on the OLED screen, strategically positioned for easy viewing without obstructing the user's field of vision. This setup ensures users who are deaf or hard of hearing can easily access spoken information.

Figure 2.1 illustrates the system workflow, where speech triggers the transcription module within the ESP32-C3 microcontroller. The text output is promptly displayed on the OLED screen, maintaining minimal latency and eliminating the need for an internet connection. Additionally, the system seamlessly connects to the EchoView.ai mobile app via Bluetooth, providing users with enhanced functionalities such as transcription management and device settings customization. Installation of the system entails setting up the ESP32-C3 microcontroller, pairing the Bluetooth microphones, and connecting the OLED display. Detailed instructions are provided to streamline the installation process, ensuring a user-friendly experience from setup to operation.

At the core of the system is the app, running a transcription model that converts speech captured by the microphones into text. These microphones are designed to pick up clear audio in various environments, ensuring reliable transcription. The text is then displayed on the OLED display, positioned to allow users to read without hindering their view.

Figure 2.1 shows the system workflow. Speech triggers the transcription module, where the ESP processes the audio and displays the resulting text almost instantaneously on the OLED screen. This process is designed to minimize latency and does not require an internet connection.

The system also connects to the EchoView.ai mobile app via Bluetooth, offering additional functionalities like transcription management and device settings customization.

Installation involves running the Swift and ESP code, connecting the microphones to Bluetooth and OLED display, and pairing the device with the mobile app. The process is streamlined for ease of setup, with detailed instructions provided for a hassle-free user experience.

2.1. Overview block diagram

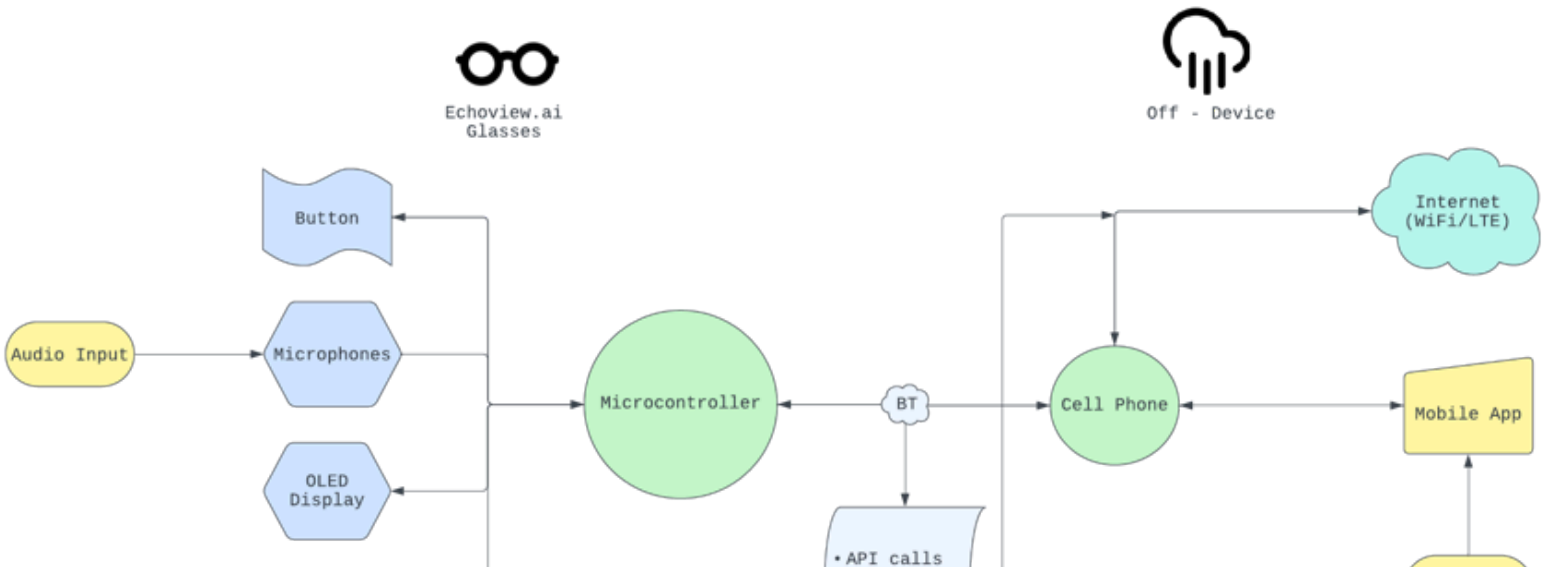


Figure 2.1: System Overview of the EchoView.AI Glasses

2.2. User interface.

The user interface for the EchoView.ai app is designed to be intuitive and user-friendly, focusing on seamless interaction between the user and the app's core functionalities. Activation of the app's features is straightforward: users can start the transcription process with a simple tap, eliminating the need for complex gestures or commands.

In addition to the primary app interface, the team has developed a companion iOS application that allows users to monitor their transcription history and view real-time transcription status updates. The app's UI is crafted to automatically refresh and display the latest transcription data as soon as new speech is detected and processed, ensuring that users always have access to the most current information.

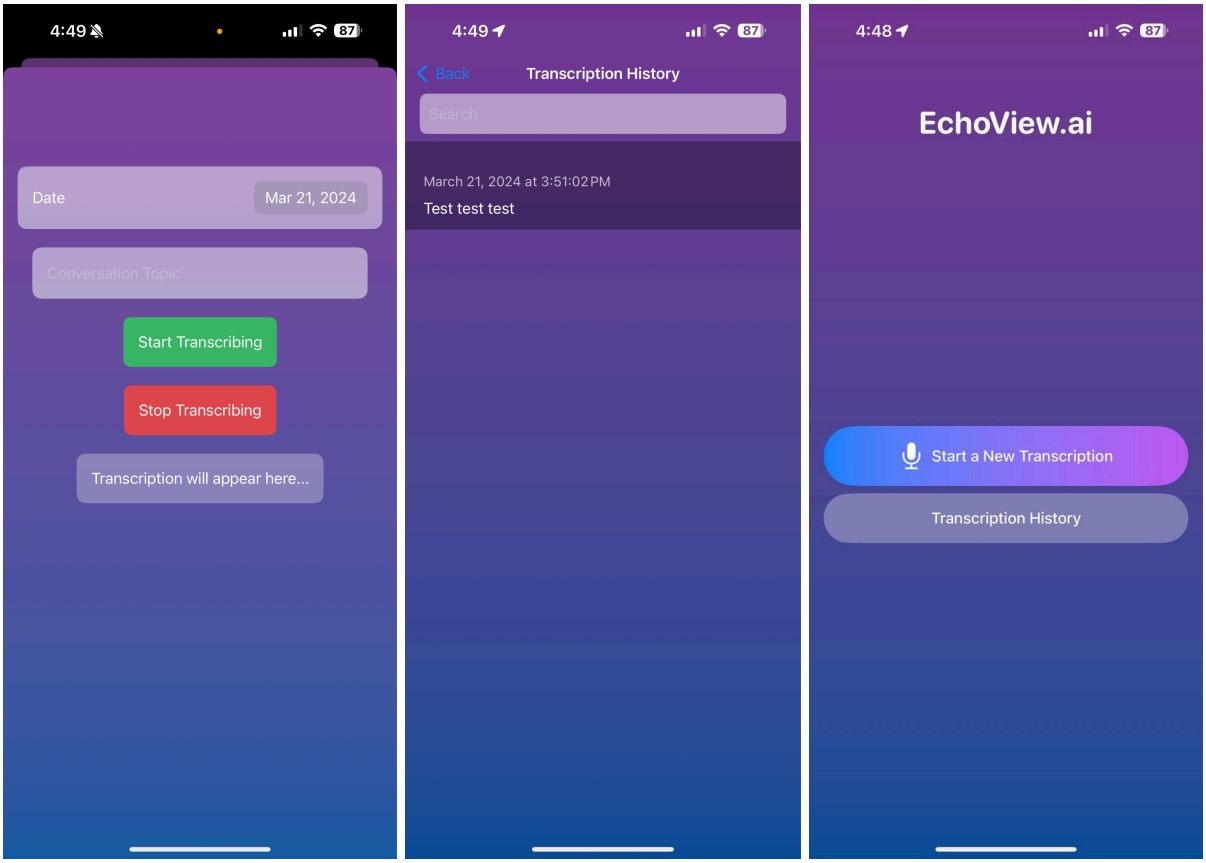
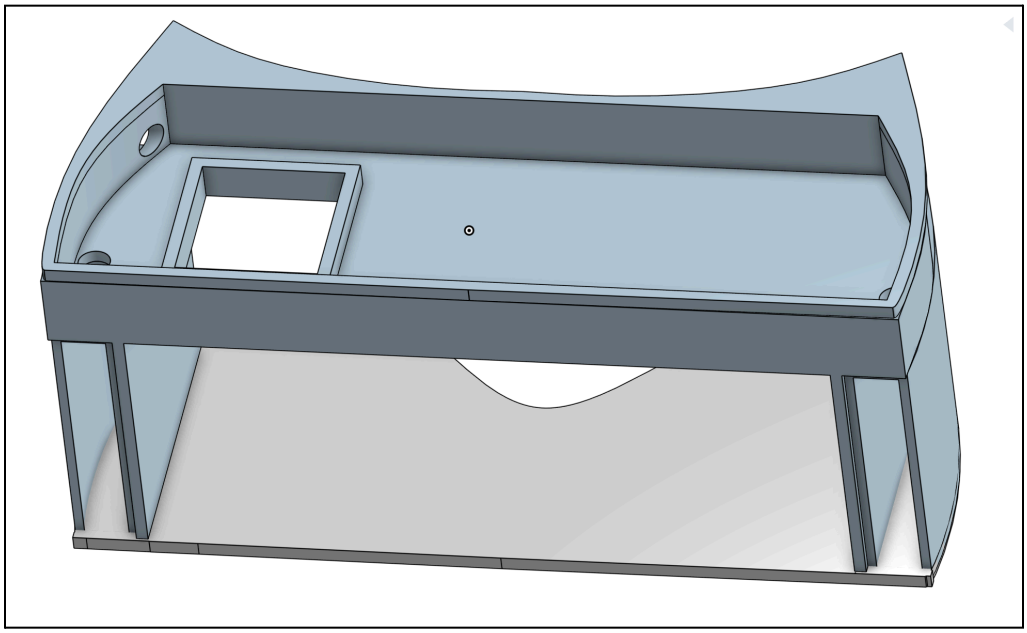


Figure 2.2: iOS Application UI

2.3. Physical description



2.4. Installation, setup, and support

Installation: Each pair of EchoView.AI glasses is designed for immediate use, requiring minimal setup by the user. Upon purchase, users simply need to ensure the glasses are charged. The device is built to be intuitive and user-friendly, eliminating complex assembly processes. For optimal performance, the EchoView.AI glasses require a stable internet connection. It's recommended to use them within a strong Wi-Fi signal range to maximize the functionality of their real-time speech-to-text transcription features.

Setup:

1. Download the EchoView.AI application from the App Store.
2. Create a user profile by registering with an email and password.
3. Pair the glasses with the app using a unique identification code provided with each pair.
4. Start experiencing real-time speech-to-text transcription.

Support: Recognizing the advanced nature of EchoView.AI glasses, we offer a comprehensive support system through an interactive iOS mobile application. This app is designed to enhance the user experience by offering clear visuals of transcribed speech, adjusting settings, and providing tips for optimal use. It also serves as a platform for troubleshooting and accessing customer support.

3. Operation of the Project

3.1. Operating Mode 1: Normal Operation

Activation

1. To activate EchoView.AI glasses for use, the wearer simply needs to turn them on.

   Notice: The current version is optimized for processing individual conversations in real-time. Overlapping or simultaneous speech may impact transcription accuracy.

2. The glasses are ready for immediate use once powered on, requiring no physical adjustments by the user.

iOS Application

The EchoView.AI iOS application serves as an interactive platform for users to customize settings, view transcription history, and access support.

1. The app allows users to adjust font size, display speed, and other settings to personalize their viewing experience.
2. It displays transcription accuracy and a history of conversations for reference.
3. Users can access visual statistics on their usage patterns, including frequency and duration of use.
4. The app provides notifications for software updates, ensuring the glasses operate with the latest features and improvements.

3.2. Operating Mode 2: Abnormal Operations

Hardware errors

1. If the glasses fail to power on or respond, a simple reset by turning them off and on may resolve the issue.
2. Should sensors not activate properly, restarting the device and checking for obstructions or damage is advised.

3.3. Safety Issues

1. EchoView.AI glasses are designed for indoor use and may require specific care to maintain functionality, such as avoiding exposure to water or extreme temperatures.
2. Users should handle the glasses with care, avoiding rough treatment that could damage the delicate sensors and electronics.
3. When not in use, storing the glasses in a safe, dry place will help preserve their condition and ensure longevity.

4. Technical Background

4.1. Hardware Component

- MEMS Microphones
    - MEMS (Microelectromechanical Systems) microphones are employed to capture ambient audio inputs. These microphones boast high sensitivity and low noise characteristics, ensuring accurate and reliable audio capture for subsequent processing.
    - The MEMS microphones are strategically positioned to optimize audio capture from the user's environment, facilitating robust speech recognition and transcription capabilities.
- OLED Display
    - An OLED (Organic Light Emitting Diode) display module is integrated into the EchoView.ai device to provide visual feedback to the user. The OLED display renders real-time transcribed text, enabling users to conveniently read incoming messages or conversation transcripts.
    - The OLED display offers crisp, high-contrast visuals and consumes minimal power, making it well-suited for battery-operated wearable devices.
- Optical Combiner:
    - The EchoView.ai device features an optical combiner, a critical component that overlays the OLED display output onto the user's field of vision. The optical combiner ensures that the transcribed text appears seamlessly integrated with the user's surroundings, enhancing readability and user experience.
    - By employing advanced optics and precise alignment techniques, the optical combiner achieves optimal transparency and clarity, minimizing visual distractions while maximizing legibility.

4.3. Software Component

- Real-Time Speech-to-Text Transcription
    - The iOS application developed with Swift 4 and Xcode features a speech-to-text transcription engine that converts spoken words captured by the device's microphones into text in real time. This feature enables deaf and hard-of-hearing users to easily understand spoken conversations by viewing the transcribed text on the device's display and storing conversations with generated titles.
- User Interface Design

    - The frontend application has a user-friendly interface designed to enhance accessibility and usability. It incorporates clear and readable visuals, and intuitive navigation to ensure a seamless user experience.
- Large Language Model Whisper
    - The C++ model facilitates accurate speech recognition and transcription, enhances natural language understanding, provides code generation assistance, and continuously improves over time. By leveraging its deep knowledge of C++ programming, it enables the EchoView.ai device to transcribe communication.
- Bluetooth Low Energy (BLE) Integration
    - The EchoView.ai system utilizes BLE technology for wireless communication between the wearable device and the companion mobile application. BLE enables efficient data transfer while conserving energy, facilitating seamless interaction between the glasses and the user's smartphone. Implemented using BLE libraries, the code establishes services and characteristics for bidirectional communication, handles connection events, and enhances the device's usability and accessibility.
- MEMS code
    - The microphone code facilitates audio capture for real-time speech recognition. It configures the MEMS microphones for optimal performance, capturing clear audio in various environments. Utilizing I2S communication protocol, the code reads audio samples from the microphones, processes them for transcription, and sends the data to the speech recognition module. Through efficient sampling and data processing techniques, the code ensures accurate transcription and seamless integration with the overall system.

5. Cost Breakdown

The cost breakdown approximates the costs to produce one EchoView.ai unit. It accounts for the hardware, software, and mechanical components of the EchoView.ai.

Item Description Cost/Unit Quantity Total Cost Hardware Components			
Quantity	Item	Individual	Total
1	Teensy 4.1 Development Board	$19.99	$19.99
1	Adafruit QT Py ESP32-C3 WiFi Dev Board with STEMMA QT 5405	$31.99	$31.99
2	Adafruit I2S MEMS Microphone Breakout - SPH0645LM4H	$6.95	$13.90
1	Monochrome 0.96" 128x64 OLED Graphic Display - STEMMA QT	$23.99	$23.99
1	Sandisk 64GB	$19.99	$19.99
Mechanical Components			
1	3D Printing Filament	$11.00	$11.00
	Total Cost		$175.85

Table 5.1. Production Cost Estimate for EchoView.ai

6. Engineering Standards

1. Use of Approved Components and Protocols: The project utilizes standard components like the Teensy 4.1 microcontroller and the ESP32-C3 module, which are likely designed in accordance with industry standards for electronic devices. The use of BLE (Bluetooth Low Energy) for communication is in line with the Bluetooth standard IEEE 802.15, ensuring interoperability and efficient power usage.
2. Safety and Efficiency: By focusing on a lightweight design (<300 grams) and ensuring the device can withstand drops from 1.5 meters, the project likely aligns with standards related to the safety

and durability of wearable devices. The emphasis on power efficiency, aiming for a battery life of at least 4 hours, also indicates an effort to comply with standards for energy-efficient electronic devices.

3. Software and Documentation: The development of an iOS-compatible mobile application and the use of open-source speech recognition software suggest adherence to software development and documentation standards. Proper software documentation is crucial for maintainability, scalability, and usability, aligning with the guidelines mentioned in the presentation.
4. Component Specifications: The project's attention to using components according to manufacturer specifications, such as the OLED display and microphones suitable for speech recognition, aligns with the standard compliance practices outlined in the PowerPoint. This ensures the reliability and optimal performance of each component within the system.

Appendix A - Specifications

Requirements	Value, range, tolerance, units
ESP32-C3 Dimensions	40mm x 27mm
Power Supply	Single 5V power supply with a minimum output of 2A for ESP32-C3
Bluetooth Microphone Compatibility	Compatible with Bluetooth microphones with a range of 50 meters (164 feet)
OLED Display	Size: 1.3 inches; Resolution: 128x64 pixels
Latency Transcription	No latency progressed from less than 4 second latency to realtime audio capture to display on OLED
Battery Life	Minimum battery life of 6 hours for wearable device functionality
Wearability	Comfortable fit for users with easy readability of OLED display

Appendix B - Team information

Team 3, or Team EchoView.ai, consists of Riya Deokar, Hassan Hijazi, Nicholas Hardy, Jazmyn Walker, and Marybel Boujaoude. All team members are majoring in Computer Engineering while Riya Deokar is additionally concentrating in Machine Learning and Nicholas Hardy and Marybel Boujaoude are concentrating in Technology Innovation.

Hassan Hijazi is a senior computer engineering student with a passion for creating innovative

solutions using emerging technologies. He has a keen interest in machine learning and its applications in assistive technology.

Nicholas Hardy is a senior computer engineering student with a focus on embedded systems and IoT devices. He enjoys tackling complex hardware challenges and optimizing system performance.

Riya Deokar is a senior computer engineering student with a background in software development and human-computer interaction. She is dedicated to designing user-friendly interfaces that enhance accessibility and usability.

Marybel Boujaoude is a senior computer engineering student with a strong foundation in system design and integration. She is committed to ensuring the seamless operation of all hardware and software components.

Jazmyn Walker is a senior computer engineering student with expertise in wireless communication and signal processing. She is passionate about developing reliable and efficient wireless systems for real-world applications.

Together, Team EchoView.AI is driven by a shared goal of creating a transformative assistive technology solution that improves the lives of individuals who are deaf or hard of hearing. Through collaboration, innovation, and perseverance, we are committed to delivering a product that makes a meaningful difference in the world.