EchoView.ai: Augmented Reality Glasses with Real-Time Speech-to-Text Capability

Hassan Hijazi, Riya Deokar, Nicholas Hardy, Jazmyn Walker, Marybel Boujaoude

Abstract— The development of Augmented Reality glasses equipped with real-time transcription capabilities will introduce groundbreaking innovation within the field of assistive technology and assist in bridging the gap between the Deaf and hearing community. By seamlessly converting spoken language into text in real time, this technology will empower hard of hearing and Deaf individuals to engage fully with the hearing-centric world. This device will include speech recognition software and a built-in display system that projects text into the user's field of vision. Additionally, a mobile application will store transcription history, allowing users to revisit and reference past conversations. With a user-friendly interface and simple compatibility with mobile devices, these AR glasses will break down communication barriers and create a more inclusive society.

Index Terms—Artificial, augmented, and virtual realities, User Interfaces, Natural Language Processing, Speech recognition and synthesis, Assistive technologies for persons with disabilities, Affective computing applications, Education, Health care

1 Need for this Project

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 currently 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 market for AR headsets and glasses 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 that will benefit from improved accessibility at an affordable cost. Hospitals can have clearer communication 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 Problem Statement & Deliverables

2.1 Problem Statement

The Deaf and hard of hearing community has continuously faced several challenges in a world where the main form of communication is using spoken language. While alternative assistive technologies exist, they are very expensive, and often fall short of providing real-time, accessible solutions to communication barriers. Standard methods such as the reliance on sign language interpreters and conventional speech-to-text solutions can be costly, require prior planning, and limited in availability. Our proposed solution seeks to address this issue by developing Augmented Reality (AR) glasses equipped with real-time speech-to-text capabilities. By processing spoken language and displaying it as readable text, we aim to bridge the communication gap and allow users to easily interact in various environments including hospitals, educational institutions, court rooms, and more. Our design will feature a sleek and comfortable wearable device that is equipped with a high-quality microphone array, allowing for accurate transcription in loud or crowded environments. Intended for extended use, our glasses will provide up to 2 hours of

- *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*

battery life, making them ideal for daily wear. We will develop an accompanying mobile application with a user-friendly interface that will store previous transcriptions allowing the user to revisit past conversations. It will also provide options for customization of text size, font, and display preferences. Our device will create a more inclusive world that transforms how the Deaf and hard of hearing community is able to communicate by offering an immediate and accessible solution.

2.2 Deliverables

Augmented Reality glasses capable of:

1. Real-time speech-to-text transcription
2. Projecting transcribed text into user's field of vision
3. Providing a long-lasting battery life (2-3 hours)
4. Offering comfortable wear with an ergonomic design
5. Wireless pairing with a mobile application for enhanced control

A mobile application capable of:

1. Seamlessly pairing with the AR glasses
2. Storing transcription to allow users to reference previous conversations
3. Offering a user-friendly interface, enabling effortless control over the settings of the AR glasses
4. Ensuring compatibility with popular platforms (iOS and Android)

3 Visualization

Fig. 1. This is a representation of our final product design. The glasses will be engineered to offer real-time speech-to-text capabilities while integrating a sleek and sophisticated design to optimize comfort and style.

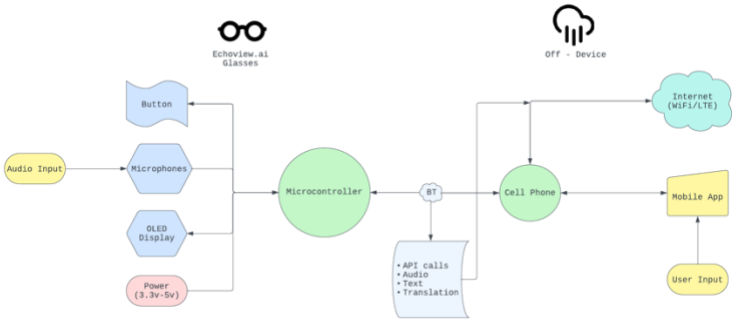

Fig. 2. This diagram offers a comprehensive schematic of our Echoview.ai glasses. Audio input is captured through a microphone array and an OLED display will provide real-time visual feedback to the user. For off-device communication, a microcontroller will interface with a cell phone via a Bluetooth connection. The device will also work using WiFi or LTE granting the system access to cloud resources and services. The mobile app provides a simple user interface for customization and further interaction with the device. Overall, this diagram displays how we will seamlessly integrate hardware and software components to create Augmented Reality (AR) glasses.

4 Competing Technologies

With the rapidly increasing need for assistive technologies, the emergence of competing ventures has grown. There are several competing AR glasses currently on the market from companies such as Google, XRAI, TranscribeGlass, and Xander.

4.1 Google Translation Glass

Google has begun development of augmented reality glasses that are capable of live transcription and live translation. The glasses also feature a camera which can interpret sign language and output the text on the lens in real time. The product is still in development, and there is no timetable for their release, or expected price.

4.2 XRAI Glass

XRAI Glass is an app developed to provide real time transcription and translation when paired with AR glasses which start at 379.99$. The application has a free version which limits the user to only transcription and one day chat history, or more costly monthly subscriptions which allow for translation and AI assistant technology. The monthly subscription cost is 19.99 for the premium version, and 49.99 for the ultimate version.

4.3 Transcribe Glass

Similar to our project, TranscribeGlass provides real-time subtitles using speech-to-text recognition, addressing the need for immediate transcription. It has an ease of an attachment, because it can be attached to any existing eyeglass, making it a convenient and cost-effective option for those who already wear glasses. Users can integrate their preferred speech-to-text recognition software, adjust font size, select languages, and store transcriptions, offering a personalized experience. Additionally, there is day-long battery life on a single charge.

4.4 Xander Glass by Vuzix

Xander glass offers real-time captions through speech-to-text technology. They excel in user friendliness by easily attaching to existing glassware. Customizability is a key feature, allowing users to tailor font size, choose languages, and store transcriptions, delivering a personalized user experience. Furthermore, Xander glasses have a day-long battery life on a single charge.

5 Engineering Requirements

This section describes the engineering requirements based on the specific subsection.

5.1 Hardware & Mechanics

1. Must be designed for comfortable wear made for extended periods of time of up to 4 hours.
2. Glasses must not exceed a weight of 300 grams to ensure comfort and keep up with competitors [5].

5.2 Speech-to-Text Processing

1. The AR glasses ust incorporate real-time speech-to-text processing capabilities with a target response time of 10 milliseconds for local processing, allowing users to view captions as conversations unfold.
2. Noise Specification – The device will work in environments with noise levels of up to 90 dBA. We must specify the ambient noise conditions for the system accurately, considering both local and remote processing scenarios, and explore the potential use of phone apps to estimate background noise levels (measured in dBAs).

5.3 Mobile Application

1. The application must be compatible with both iOS and Android platforms.
2. The system must maintain a stable connection to WiFi and LTE, ensuring access to cloud services.

5.4 Power & Cost

1. The AR glasses must remain operating for 2-3 hours on a single charge.
2. The total component cost should not exceed more than $500 to ensure a competitive market price.

Acknowledgment

The authors wish to thank Professor Alan Pisano, Professor Osama Alshkh, Professor Michael Hirsch, and Professor Thomas Little for their continuous support throughout the project.

References

[1] J. Wirth, A. Hall, and L. Jorgenson, "Deafness and hearing loss statistics," Forbes, https://www.forbes.com/health/hearing-aids/deafness-statistics/#:~:text=Approximately%2010.7%20million%20people%20ages,U.S.%20have%20profound%20hearing%20loss [accessed Oct. 12, 2023].

[2] "How to pay for Translation Services," Monitor on Psychology, https://www.apa.org/monitor/2020/01/career-other-words-sidebar.html [accessed Oct. 12, 2023].

[3] "Assistive Technology Market - Global Industry Assessment & Forecast," Vantage Market Research, https://www.vantagemarketresearch.com/industry-report/assistive-technology-market-1786 [accessed Oct. 12, 2023].

[4] T. Alsop, "Topic: AR headsets and Glasses," Statista, https://www.statista.com/topics/10134/ar-glasses/#topicOverview [accessed Oct. 10, 2023].

[5] T. Alsop, "AR headsets and glasses comparison by weight 2023," Statista, https://www.statista.com/statistics/1337293/ar-glasses-comparison-by-weight/ [accessed Oct. 12, 2023].