An intelligent Embedded System using Natural Language Processing for Deaf people

Dr. Mohamed Ahmed Hussein Ali
Computer Science Department
High Institute for Computers and Information Technology
AL-Shorouk Academy
Cairo – Egypt
email: dr.mohamed.hussein@sha.edu.eg

Abstract

Natural language processing (NLP) refers to the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

NLP combines computational linguistics rule-based modeling of human language with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process human language in the form of text or voice data and to 'understand' its full meaning, complete with the speaker or writer's intent and sentiment.

Deaf individuals and those who are hard of hearing often face challenges in completing basic daily tasks, such as interacting with others. While some solutions exist, they have significant limitations. For example, lip reading can only help understand about 30% of spoken words, and sign language interpreters are in short supply. These challenges contribute to high unemployment rates and mental health issues in the deaf community.

To address these challenges, this paper dealing with a design smart glasses that serve as assistive technology for people with hearing disabilities. The glasses provide real-time speech transcription and format it for display, allowing wearers to understand what is going on around them and interact with others. The glasses attach to regular prescription glasses and are highly effective in achieving their purpose.

This paper will have a significant impact on people with hearing disabilities, enabling effective communication with friends and society, raising awareness, improving education, and increasing productivity and participation in all fields. It will also encourage community members to engage with and participate in activities with people who are deaf or hard of hearing. We are grateful for the opportunity to contribute to this important and meaningful cause.

1.1 Introduction

Deafness is a problem that prevents an individual's auditory system from functioning or reduces an individual's ability to hear different sounds. Hearing impairment ranges in severity from simple and medium degrees that result in hearing impairment to very severe degrees that result in deafness. Over 1.5 billion people globally live with hearing loss. This number could rise to over 2.5 billion by 2050 [1]. The deaf around the world suffer in adaptation to society and difficulty with communication. Although there are sign language interpreters, they are few and this resulted in difficulty in communication results disadvantages in education, limited access to competitive employment, diminished social opportunities, and fewer financial and service resources to them.

3.3.1 Hardware Design

3.3.1.1 TTGO T-Display ESP32:

TTGO T-Display ESP32 is a small development board that is based on the ESP32 microcontroller. It features a 1.14-inch color TFT display, which makes it ideal for developing projects that require a graphical user interface. The board also includes a built-in Wi-Fi and Bluetooth module, which allows it to connect to the internet and other devices wirelessly.

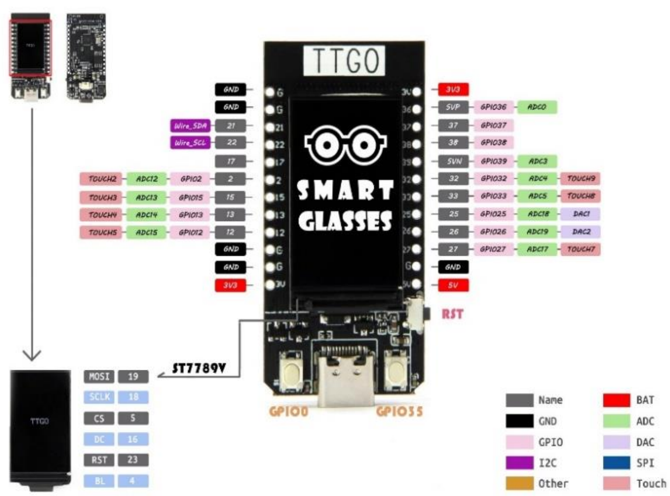

Figure 1 (ESP32 pin diagram).

3.3.1.2 Wave-Share Sound Sensor V2:

The Wave-Share Sound Sensor V2 is a type of audio sensor that can be used to capture sound waves and convert them into electrical signals. It is designed to be highly sensitive and is capable of detecting a wide range of sound frequencies. The sensor is based on Onboard audio power amplifier LM386, which is widely used in various audio applications due to its high sensitivity and low noise.

Figure 2 (Wave-Share Sound Sensor V2)

3.3.1.3 Rechargeable Battery:

- Lithium Polymer Rechargeable Battery.
- 3.7V.
- 1000mAh.

Figure 3 (Lithium Polymer Rechargeable Battery)

3.3.1.4 Lens:

The human eye can only focus an object at a distance of -ve 25cm. And all what we needed is the formula (1/f) = (1/o) + (1/i) where f is focal length of the lens, o is object distance to the lens and i is the distance of the virtual image.

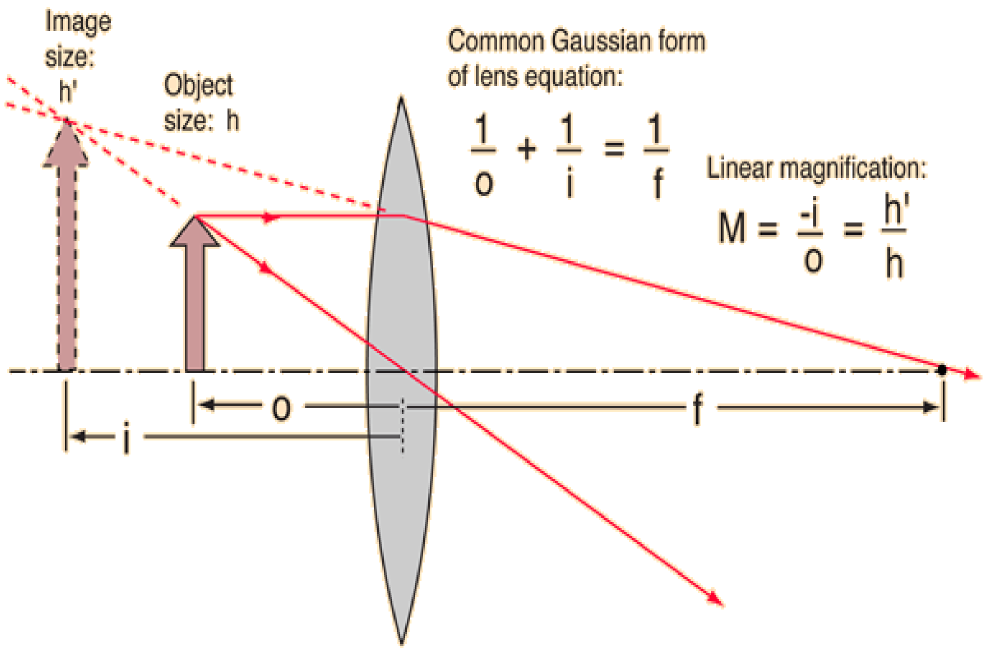

Figure 4 (Lens equation).

3.3.1.5 Mirror:

A mirror piece that reflects the received data, which is displayed on the T-Display ESP32, towards the lens.

3.3.1.6 Reflective Sheet:

The user will see the text on the small reflective sheet.

Figure 5 (Reflective sheet).

3.3.1.7 Holder Body:

The body of the smart glasses will be made using 3D printing and the material will be plastic.

Holder Weight: 45 grams.

The body of the smart glasses consists of five parts:

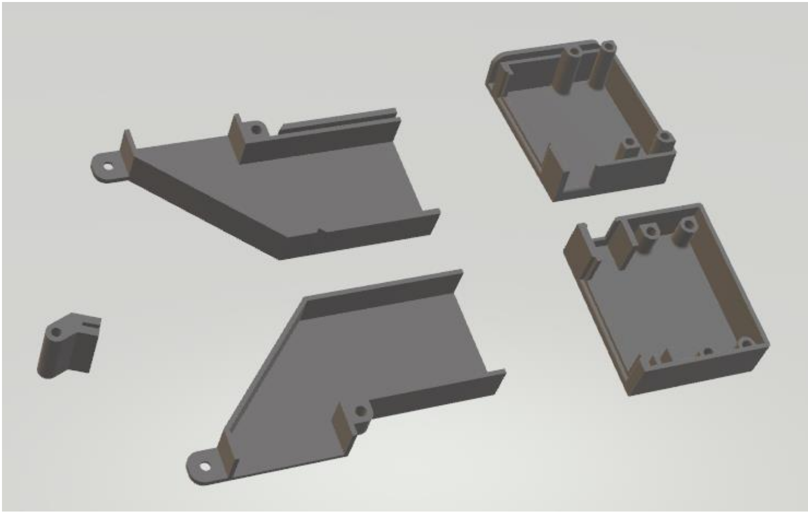

Figure 6 (Holder body 3D design).

3.3.2 Software Design

3.3.2.1 Speech-To-Text:

Speech-to-Text is commonly used to transcribe spoken language into written text by identifying the phonetic units, words, and sentences in the speech, which are then displayed on the deaf person's smart glasses. This allows the deaf person to read what the speaker is saying in real time, enabling them to have a conversation without requiring an interpreter or lip-reading skills.

We use the Jonatasgrosman/wav2vec2-large-xlsr-53-english model which is a state-of-the-art speech recognition tool that has proven to be highly effective in accurately transcribing spoken language. In the context of the Smart Glasses for Deaf project, this model plays a crucial role in enabling users to communicate with others effectively.

One of the key benefits of using the Jonatasgrosman/wav2vec2-large-xlsr-53-english model is its accuracy and reliability. The model has been trained on a large dataset of English speech, making it highly effective in recognizing and transcribing spoken language. Additionally, the model is highly customizable, allowing developers to fine-tune its settings to suit the specific needs of the application and its users.

Overall, the Jonatasgrosman/wav2vec2-large-xlsr-53-english model is a critical component of the Smart Glasses for Deaf project, enabling effective communication and interaction with others. Its accuracy, reliability, and customizability make it a valuable tool for our project.

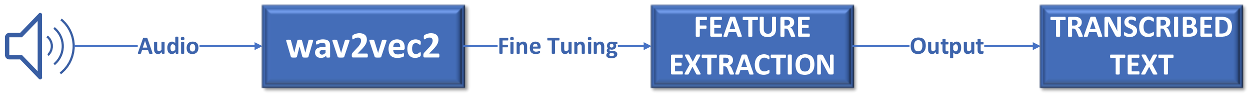

Figure 7 (Speech-to-Text Model)

3.3.2.2 Smart Glasses Server:

The Smart Glasses Server is a vital software component that serves as a communication bridge between the ESP32 and the server. Its primary role is to retrieve sound files from the sound sensor of the ESP32 and facilitate speech recognition to convert the audio to text. This text is then transmitted back to the ESP32 for display on the smart glasses, enabling effective communication and interaction with others.

The sound files retrieved from the ESP32 are processed through advanced speech recognition algorithms, which analyze the audio and convert it into accurate text representations. This process involves multiple complex steps, including noise reduction, feature extraction, and model training, all of which are optimized to ensure the highest possible accuracy levels.

The Smart Glasses Server is designed to be highly scalable and reliable, ensuring that it can handle large volumes of audio data and deliver fast and accurate results. Its advanced architecture and robust feature set make it a critical component of the Smart Glasses for Deaf project, enabling users to communicate and interact with others effectively.

3.3.2.3 Mobile Application:

The mobile application is designed to assist the deaf and hard of hearing to communicate more effectively.

Figure 8 (App Icon)

The app works by converting audio input from the user's surroundings, such as a conversation with another person, into a text format. This is done using automatic speech recognition technology, which analyses the speech signal and converts it into text by identifying the phonetic units, words, and sentences in the speech.

Once the audio is transcribed into text, the app sends the text to the smart glasses using Bluetooth, where it is displayed in real-time on the user's field of vision. This allows the user to read what the speaker is saying, enabling them to participate in conversation without relying on lip-reading or an interpreter.

The Smart Glasses for Deaf mobile app is user-friendly and can be easily customized to meet the user's specific needs, such as connecting the ESP32 to Wi-Fi or turning on or off the smart glasses.

Another important aspect of the mobile application is that it allows anyone to connect with the app and help a deaf person in need. Using the app, a hearing person can speak into their phone's microphone, and the app will transcribe their speech into text, which is then displayed on the smart glasses wearable device worn by the deaf person. This feature makes Smart Glasses for the deaf a powerful tool for fostering communication between people of different abilities. With the app, anyone can easily communicate with a deaf person, regardless of whether they know sign language or not.

Additionally, the Smart Glasses for Deaf mobile app can be used in a variety of settings, such as classrooms, meetings, or public spaces, where it may be difficult for a deaf person to follow along with spoken conversation. By enabling real-time transcription, the app empowers deaf individuals to fully participate in these settings and engage with others on an equal footing.

We use Speech to Text Flutter package which is a powerful tool that enables us to integrate speech recognition capabilities into our mobile application. In the context of the Smart Glasses for Deaf application, this package plays a vital role in enabling users to communicate with others effectively.

One of the key benefits of using the Speech to Text Flutter package is its accuracy and reliability. The package offers robust speech recognition capabilities, even in noisy environments, which is essential for users with hearing disabilities.

Overall, the Speech to Text Flutter package is a critical component of the Smart Glasses for Deaf application, enabling effective accuracy, reliability, and customizability make it a valuable to use.

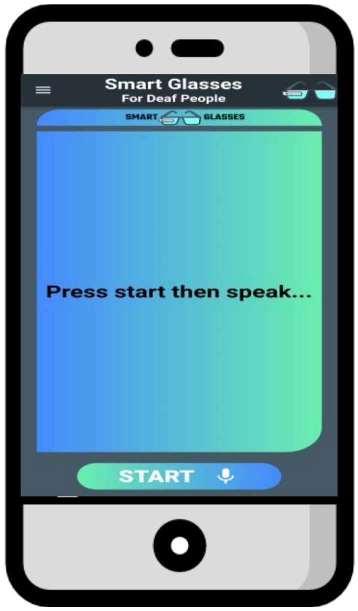

Figure 9 (Smart Glasses for Deaf People Mobile App)

3.4 Testing Phase

3.4.1 Hardware Testing

Hardware testing is an essential part of any project that involves physical components. In our project, we used an SPH0645 I2S microphone for audio input. However, during testing, we found that the microphone was not working as expected and produced noise. To address this problem, the team decided to switch to a Wave-Share Sound Sensor V2, which is known for its high sensitivity and noise-canceling capabilities. After testing the new sensor, we found that it performed much better than the previous microphone and was able to capture high-quality audio without any noise. This experience highlights the importance of thorough hardware testing during the development process and the need to be flexible and adaptable when faced with unexpected issues. By being proactive and testing multiple components, the team was able to find a suitable replacement and ensure the success of our project.

3.4.2 Software Testing

Software testing is an essential aspect of our project, and it becomes even more critical when dealing with complex technologies like speech recognition. First, we used Facebook's Wav2Vec2--960h for sound to text transcription, and testing revealed that the dataset was not perfect for the task. Despite its reputation as a high-quality dataset, it did not perform as well as expected.

However, the team was able to identify the issue through rigorous testing, and quickly pivoted to a new solution. We are going to another dataset called Jonatasgrosman/wav2vec2-large-xlsr-53-english, which proved to be a better fit for the project's needs. This highlights the importance of testing, as it allowed the team to identify and address the issue before it could cause any significant problems. By taking a proactive approach to testing, the team was able to ensure that the final product met expectations and delivered the desired results.

3.5 System Implementation Phase

3.5.1 Hardware Implementation

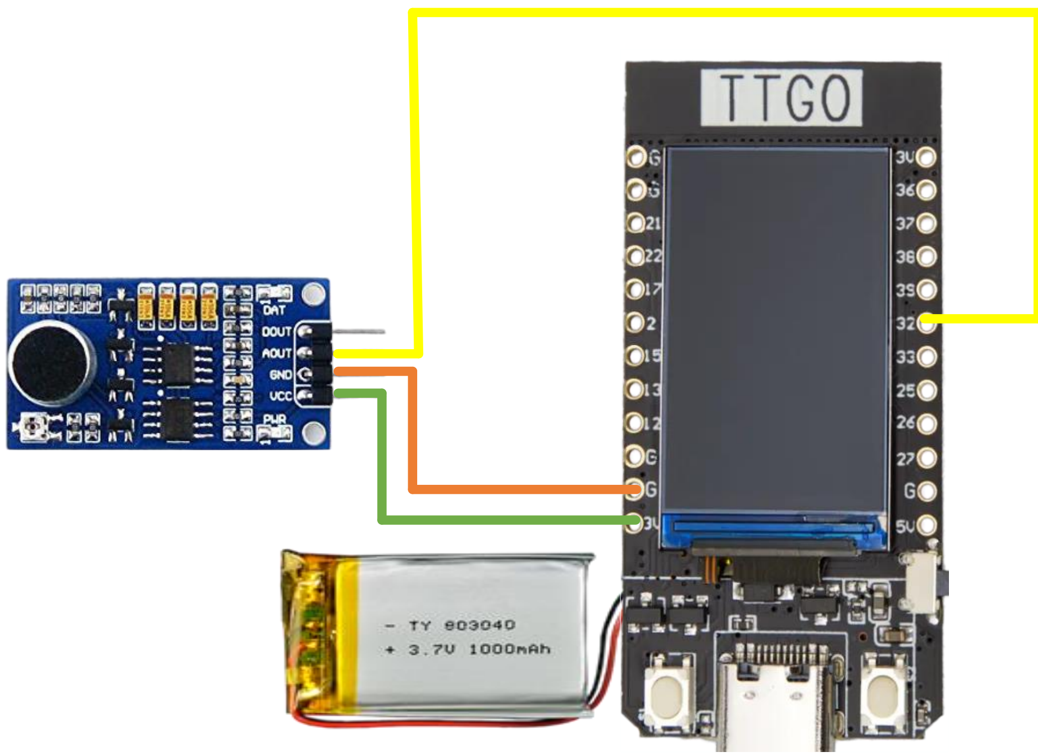

Figure 10 (circuit diagram).

3.5.1.1 Components Wiring:

The wiring between components plays a crucial role in ensuring the proper functioning and seamless communication of various hardware elements. To establish the connection, the microphone's output pin, which carries the analog audio signal, should be connected to the appropriate analog input pin on the ESP microcontroller. Moreover, microphone’s communication pins are connected to specific ESP pins in order to ensure proper interaction.

3.5.1.2 Holder 3D Printing:

In the world of technology, 3D printing is a fascinating process that brings ideas to life. Using this technique, a holder is created to hold an ESP and a microphone. The printer carefully builds each layer to make sure the holder fits the components perfectly according to the calculated dimensions.

3.5.1.3 Final Product:

Putting the components in the designed and printed holder, the final product is ready to perform its functionality.

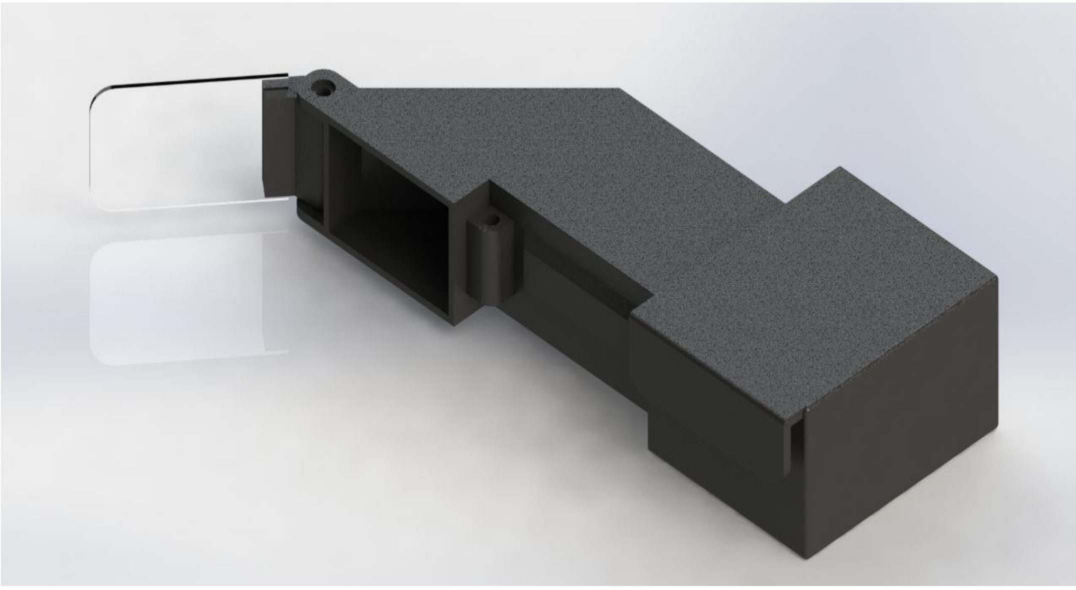

Figure 11 (Final Product)

3.5.2 Software Implementation

3.5.2.1 ESP Configuration:

To connect a microphone to an ESP TTGO setup, a specific code is required to establish the necessary communication and enable audio processing capabilities. The code begins by initializing the required libraries and configuring the necessary pins on the TTGO board. It then sets up the microphone as an input device, configuring its parameters such as sampling rate and resolution.

5.1 Conclusion

In conclusion, the smart glasses system developed using the ESP32 TTGO microcontroller and sound sensor successfully enables speech-to-text conversion for deaf individuals. The compact and user-friendly design incorporates a built-in display to relay converted text, enhancing communication and accessibility. With its ability to convert spoken language into readable text, the smart glasses provide a valuable tool for deaf individuals to overcome communication barriers. Future iterations of the system can focus on refining speech recognition algorithms, exploring augmented reality displays, and optimizing user experience and ergonomics. By continuously improving the technology, these smart glasses have the potential to significantly improve the quality of life for the deaf community, fostering greater independence and inclusivity."

5.2 Summary of archived results

1. Developed a functional smart glasses system: Successfully created a smart glasses system using the ESP32 TTGO microcontroller and sound sensor. The system is capable of converting spoken language into text and displaying it on the built-in display of the glasses.
2. Enabled speech-to-text conversion: Implemented a robust speech recognition algorithm that accurately converts spoken language into readable text. This functionality allows deaf individuals to easily understand and communicate with others.
3. Improved communication and accessibility: By displaying the converted text on the built-in glasses display, the project enhances communication and accessibility for deaf individuals. It provides them with a real-time visual representation of spoken language, enabling them to participate in conversations more effectively.
4. Compact and user-friendly design: Designed the smart glasses to be compact and user-friendly, ensuring comfort and ease of use for the wearer. The integration of hardware components, such as the microphone, lens, mirror, acrylic reflector sheet, and holder body, results in a wearable device that is practical and convenient.
5. Lay the foundation for future advancements: Through extensive testing and experimentation, the project establishes a solid foundation for future enhancements. This includes potential improvements in speech recognition algorithms, exploration of augmented reality displays for a more immersive experience, and optimization of user

experience and ergonomics. These advancements aim to further refine the smart glasses system and make it even more effective and user-friendly for the deaf community.

5.3 Future work

1. Improved microphone sensitivity: Enhancing the sensitivity of the microphone can lead to better speech recognition accuracy, especially in noisy environments. This improvement can involve exploring advanced microphone technologies, noise reduction algorithms, or signal processing techniques to capture clearer and more precise audio input.
2. Real-Time Translation: Integrating real-time translation capabilities into the smart glasses system would enable deaf individuals to communicate with others who speak different languages. This expansion would require incorporating additional language processing modules and databases.
3. Expand Display Options: Consider incorporating augmented reality (AR) technologies to provide a more immersive and intuitive display of the converted text. AR overlays could be used to superimpose the text directly onto the user's field of view, eliminating the need for them to look down at the glasses' display.
4. User Experience and Ergonomics: Conducting user studies and gathering feedback to refine the design, comfort, and usability aspects of the smart glasses would ensure that the device meets the specific needs and preferences of deaf individuals. Iterative improvements based on user input can enhance the overall user experience.