Bridging the Communication Gap: A Sensor-Based System for Sign-to-Speech and Speech-to-Text Conversion

Yigithan OZER, Galip ERGUN, Goktug BAY, Hatice OKUMUS

Karadeniz Technical University, Turkiye, Department of Electrical and Electronics Engineering, Faculty of Engineering, 61080, Trabzon, Turkiye,

---***---

Abstract - *Hearing and speech-impaired individuals use country-specific sign languages to communicate effectively both with one another and with others in their daily lives. However, many people in society lack sufficient knowledge of sign language. This limitation makes it difficult for hearing-impaired individuals to build social relationships, advance in education, and succeed in professional life. Feeling excluded from society can also lead to a loss of self-confidence. These communication barriers restrict their ability to fully express themselves and actively participate in social life. In this study, a glove is designed that converts sign language gestures into speech and simultaneously converts spoken language from non-disabled individuals into text. The goal of this design is to enable real-time, two-way communication between hearing and speech-impaired individuals and those who do not know sign language, thereby supporting their more active inclusion in society. In the first phase, flex sensors placed on the fingers will be used to detect the hand movements of hearing and speech-impaired individuals and convert these gestures into speech. These sensors will accurately capture hand and finger movements, interpret sign language gestures through a microcontroller, and convert them into audible output via a speaker. In the second phase, the speech of non-disabled individuals will be detected using a microphone and speech recognition software, converted into text, and displayed on a screen. Through this study, hearing and speech-impaired individuals will not only be passive listeners but will also be able to actively participate in conversations and express their thoughts instantly.*

Key Words: Hearing impaired people; Sign Language; Arduino Nano; Flex Sensor; Gl*oves*

1. INTRODUCTION

Technological advancements not only make human life easier but also enhance the quality of life for individuals with disabilities, allowing them to integrate more effectively into society. In recent years, newly developed technological devices and software have made significant contributions to overcoming communication barriers faced by people with disabilities. Such technological solutions facilitate the daily lives of hearing-impaired individuals and enable them to participate more actively in social life. Various software tools and assistive hardware systems provide substantial benefits in reducing or even eliminating disabilities [1]. In the literature, many studies have been conducted to improve the lives of hearing-impaired individuals. Patil et al. designed a device that enables deaf and mute individuals to communicate with normal-hearing people by converting sign language expressions into audible commands. This device is a data glove mounted on the hand, equipped with five small, low-power, three-axis ±3 g accelerometer sensor units (ADXpL335) attached to the fingertips. Since the output of these sensors is in analog form, an Arduino Mega microcontroller was found to be suitable for interpreting the signals. The entire system was integrated and verified using the Arduino IDE [2]. Extensive experiments were carried out with numerous volunteers who could fluently perform multiple hand movements as specified in the Marathi Sign Language (MSL) for the Marathi alphabet. The real-time performance of the system was evaluated in terms of its accuracy and precision in recognizing Marathi alphabets, establishing a correlation between human-generated signs and those stored in the system [3]. Mathew et al. developed a device aimed at enhancing the mobility, independence, and safety of hearing-impaired individuals while crossing streets. The device detects approaching vehicles and provides vibration-based alerts to help users perceive cars coming from behind. An ultrasonic sensor and a camera are used to identify objects in the environment. In addition, the device can connect to smartphones via Bluetooth and employs a piezoelectric transducer to convert sound signals into vibrations, thereby improving the efficiency of alarm and communication systems [4]. Navaitthiporn et al. designed a glove that converts sign language gestures into both speech and text, addressing communication challenges faced by hearing-impaired individuals. In this project, flex sensors measure the bending of fingers, while the GY-521 module detects the orientation and movement of the hand. The data from the sensors are transmitted to the Arduino IDE, converted into alphabetic text, and then synthesized into speech [5]. Similarly, Rewari et al. used flex sensors and gyroscopes to detect finger movements and hand angles, converting these gestures into sound through a microcontroller [6]. Chin et al. developed a wearable assistive device that enables hearing-impaired individuals to recognize warning sounds from vehicles on the road [7]. For this purpose, an EfficientNet-based and fuzzy-ranking ensemble model was proposed and integrated into an Arduino Nano 33 BLE Sense development board. The sound files were taken from the CREMA-D dataset [8] and a large-scale sound dataset containing sirens from emergency vehicles [9], comprising a total of 8,756 audio samples. These include four types of vocalizations and three types of siren sounds. The sound signals were converted into spectrograms using short-time Fourier transform (STFT) for feature extraction. When one of the three siren sounds is detected, the wearable device provides a vibration alert and displays a corresponding message on an OLED panel. In the device developed by Yağanoğlu and Köse, sound data collected by a microphone are processed in a Raspberry Pi environment through a USB sound card. Once received, the sounds are classified using various techniques to determine their types. The system identifies environmental sounds such as doorbells, alarms, telephone rings, horns, brakes, dogs, human voices, and other noises. For each sound, a unique vibration pattern is designed and transmitted to the user via a vibration motor, providing tactile feedback that can be perceived through touch [10]. Asakura developed an augmented reality-based system that converts household sounds into visual information for hearing-impaired individuals. In this system, environmental sounds are continuously captured using a microphone. The sound pressure waveform recorded by an omnidirectional microphone connected to a sound level meter is transferred to a laptop through an audio interface for processing. These sound waveforms are then analyzed by a machine-learning-based classifier and simultaneously converted into spectrograms. Through augmented reality glasses, users can view not only the classification results but also real-time visual representations of the detected sounds [11]. WHK et al. developed an application designed to support hearing-impaired elementary school students in Sri Lanka. The study consists of three main components: a sound recognition and classification system, Android software that translates gestures in Sinhala Sign Language (SSL) into text, and a mobile application that performs emotion recognition and text-to-speech conversion in Sinhala. The sound recognition system rapidly detects potential hazards to ensure the safety of hearing-impaired children at home. The Android software captures SSL gestures through the device's camera and represents them as text, while the mobile application identifies emotions from facial expressions using Convolutional Neural Networks (CNNs) [12]. Baltacıoğlu et al. designed a wristband for parents who are hearing-impaired or have hearing loss. The main objective of this study is to analyze and differentiate between silent environments, speech sounds, and baby cries by using sound intensity distribution to create a warning system. The sound data were collected and analyzed to develop a vibration-based alert mechanism. The data from the baby's room were transmitted via radio frequencies to the parent's wristband. An analysis of 20-second sound samples revealed that speech sounds increased by approximately 75% compared to silent environments, while baby cries showed a 102.5%

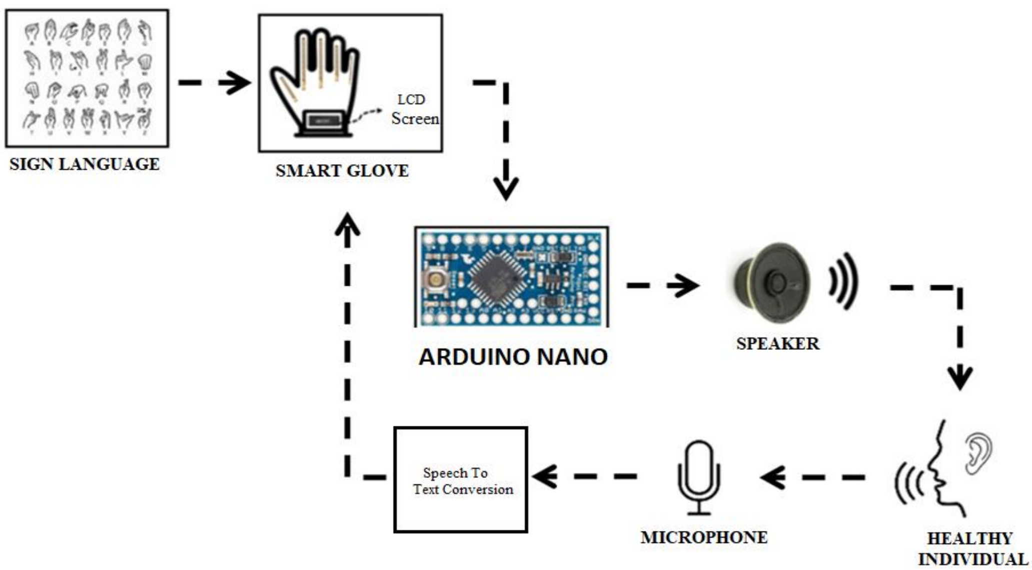

Fig- 1. The general block diagram of the system increase. These results demonstrate that the system can effectively function as a wearable technology [13]. Quiapi et al., in their project, designed a system that converts captured American Sign Language gestures into simple English sentences using a sensor-based glove [14]. When examining previous studies, it is observed that most of them offer one-sided solutions. However, the proposed study aims to design a glove that enables two-way communication by both converting sign language into speech and transforming the speech of non-disabled individuals into text. In the first phase of the glove's design, flex sensors attached to each finger detect the hand movements of hearing- and speech-impaired users. The analog signals from these sensors are processed by an Arduino Nano microcontroller, which interprets the gestures and converts them into corresponding audio outputs. In the second phase, the speech of non-disabled individuals is transformed into written text using smartphone-based Speech-to-Text (STT) technology. This text is then wirelessly transmitted to the Arduino-connected display, where it is instantly shown in real time.

2. PROPOSED SYSTEM

In this study, a smart glove that detects the hand movements of hearing and speech-impaired individuals and converts them into both auditory and visual outputs was developed. The system employs flex sensors to measure finger movements. The analog data received from the flex sensors were processed using an Arduino Nano microcontroller. Each finger movement was defined based on specific threshold values, and the system instantly recognized the corresponding words associated with these movements.

The identified words were played as audio files through a speaker and simultaneously displayed on a TFT screen without delay. The speech of non-disabled individuals was converted into text using the Speech-to-Text (STT) technology commonly employed in smartphones. This text data was transmitted wirelessly to the Arduino via the HC-06 Bluetooth module. The Arduino processed the received text data in real time and displayed it on the TFT LCD screen mounted on the glove. In this way, spoken information could be instantly followed in text form. The general block diagram of the system is presented in Fig- 1.

Each piece of equipment and component used in the development of this study is given in detail in Table-1. Arduino Nano was chosen due to its low cost and compact design, which allowed the overall size of the project to be reduced and portability to be achieved. If larger boards (such as the Arduino Uno or Mega) had been used instead, the project would have required more space and power. Additionally, these boards are more expensive and less suitable in terms of portability. A Li-Po battery was used to meet the project's power requirements. Li-Po batteries are preferred power sources in portable projects due to their high energy density and lightweight structure. The selected

Table -1: List of the equipment and components used in the development of the glove.

Component	Functionality
Arduino Nano	To process data received from sensors, control operations, and provide appropriate feedback.
Flex Sensor	To detect finger movements.
Speaker	To audibly transmit data obtained from sensors and provide instant feedback to the user.
LCD TFT Display	To display the text form of the speech from a non-disabled individual.
Li-Po Battery	To supply power to the system.
Micro SD Card Module	To interface the SD card with the circuit.
SD Card	To store audio data.
HC-06 Bluetooth Module	To provide wireless (Bluetooth) communication between the Arduino and other devices.
PAM8403 Audio Amplifier	To increase the output sound level of the speaker.

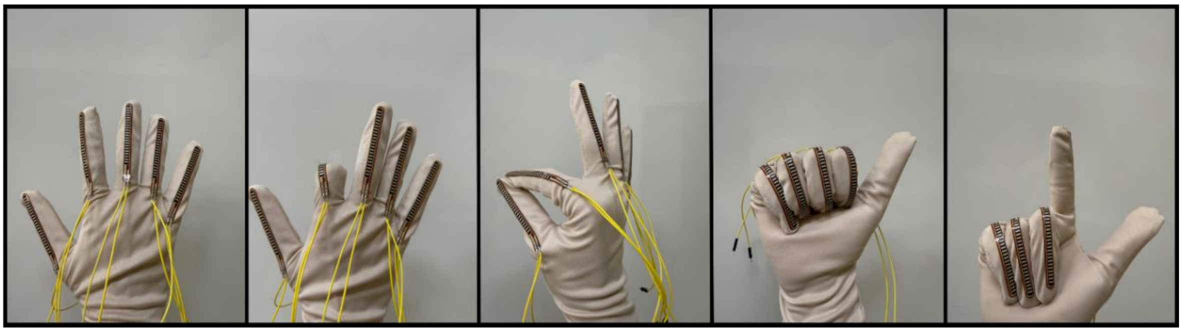

Fig- 2. Representation of Finger Positions Corresponding to Words battery has sufficient capacity to ensure continuous operation of the system. Moreover, the rechargeable nature of Li-Po batteries allows for long-term use of the project. Flex sensors were employed to detect finger movements. These sensors measure the degree of finger bending, enabling accurate detection of the user's gestures. Their sensitivity allows the system to provide precise and reliable feedback.

A. Recognition of Sign Language Gestures

In the recognition of sign language gestures, five flex sensors were used to measure the degree of finger bending. Each flex sensor generates analog signal values corresponding to finger movements, and these values are calibrated at specific intervals. The finger positions corresponding to each word were modeled using the minimum and maximum threshold values obtained from the sensor readings. Some example finger positions corresponding to specific words are shown in Fig-2.

For instance the system recognizes the gesture given in Fig- 3 as the "Good Morning" command when the real-time sensor readings fall within the predefined threshold ranges given below. Once the gesture is detected, the system triggers the audio file named gunaydin.wav stored on the SD card, providing auditory feedback through the speaker.

Fig-3. Recognition of the word "Good Morning"

Minimum Threshold Values: {160, -1, 160, 160, 160}

Maximum Threshold Values: {500, 159, 500, 500, 500}

In the code structure, the measurement range of each of the five flex sensors was defined separately for each word, and the incoming sensor values were checked to determine whether they fell within these ranges. For example, for the word "Hello", all sensor values must be within the range of 160 to 250, whereas for the word "Hi", the first sensor value is expected to be within the range of 0–159, while the other sensors operate within higher ranges. A threshold value of -1 indicates that there is no minimum limit for that particular sensor, and only the maximum threshold is evaluated. This threshold-based approach numerically defines the unique finger positions corresponding to each word. The real-time data obtained from the flex sensors are compared with these predefined value ranges. If the sensor readings fall within the specified intervals, the system correctly identifies the corresponding word. Thus, the analog data provided by the flex sensors serve as an effective measure for accurately and precisely distinguishing detailed finger movements.

B. Speech Detection and Conversion to Text

In this study, the speech-to-text conversion process was performed using a pre-existing Speech-to-Text (STT) application operating on a mobile phone. This application captures and analyzes the user's spoken commands through the microphone of the mobile device and converts them into written text in real time. The converted text is transmitted wirelessly to the Arduino via the HC-06 Bluetooth module and instantly displayed on the TFT LCD screen integrated into the system. In this way, spoken information is transformed into written form, providing hearing-impaired individuals with a means of visual communication.

One of the main advantages of the mobile STT application used in this system is its high accuracy in online operation, as well as its ability to store previously converted text on the phone's screen. This feature allows users to view not only the current text but also previous conversation content, enabling easy access to the communication history. While new text appears on the screen, earlier data remain stored in memory, allowing users to revisit past information when needed and ensuring a clearer and more consistent communication process. The STT application interface used in the system is shown in Fig-4.

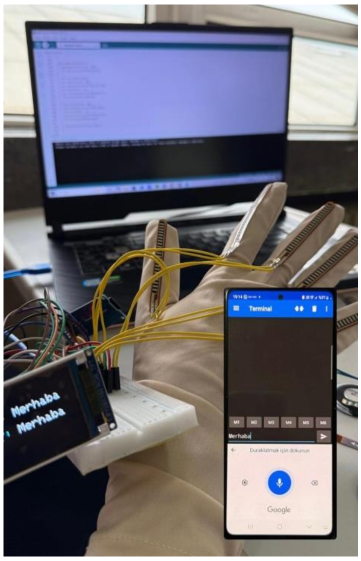

Fig- 4. Mobile Speech-to-Text application

With its user-friendly and portable design, the system can be conveniently used in any environment. Thanks to the practicality of the mobile application, users can quickly and effectively convert their speech input into text without requiring technical expertise, and directly communicate this text to hearing-impaired individuals.

2. CONCLUSIONS

In this study, a real-time, bidirectional communication system is developed to bridge the communication gap between hearing and speech-impaired individuals and those unfamiliar with sign language, using flex sensors to translate hand gestures into speech and speech recognition technology to convert spoken words into text. During the conversion of sign language gestures into speech, the predefined flex sensor threshold values for different hand movements were accurately detected, and the corresponding audio files were correctly triggered. The tests demonstrated high accuracy rates in recognizing the defined sign language gestures. In the speech-to-text phase, the system exhibited excellent performance under low ambient noise conditions. However, when environmental noise increased, a noticeable decrease in accuracy was observed. This effect was particularly evident in areas with intense background sounds, where the speech recognition process was negatively influenced. According to user feedback, the text displayed on the screen was generally clear and comprehensible, although occasional distortions in sentence structure were reported.

The system's response time was also examined under conditions where both communication directions operated simultaneously. On average, the time from a sign language gesture to audible output through the speaker was measured as 1.2 seconds, while the conversion from speech to text display on the LCD screen took approximately 1.5 seconds. These durations were found to be short enough not to affect user experience negatively, confirming that the system achieved its goal of real-time communication.

In conclusion, the developed smart glove prototype enabled effective, fast, and bidirectional communication between hearing and speech-impaired individuals and non-disabled users. The system has the potential to contribute significantly to the active participation of hearing-impaired individuals in various fields such as education, public services, and social interactions.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to the Scientific and Technological Research Council of Türkiye (TÜBİTAK) for their financial support under the BİDEB 2209-A Research Projects Support Program for Undergraduate Students, and also to the Karadeniz Technical University Scientific Research Projects Coordination Unit (Project Code: FLÖ-2024-16241).

REFERENCES

[1] M. Güngör, A. Bolat, H. Cengiz & N. Aslan, "Özürlülere yönelik teknolojik düzenlemeler', Sektörel Araştırma ve Stratejiler Dairesi Başkanlığı, 2011.

[2] The Ministry of Home Affairs, Access Date: October 2 2024.https://www.ijcaonline.org/archives/volume118/number13/20802-3485

[3] Indian Sign Language Portal, Access Date: October 2 2024. https://indiansignlanguage.org/

[4] J. Mathew, A. S. Philip, A. Thomas, A. A. Raju & A. Tennyson, A., "IoT based Alerting Device for Hearing Impaired Population", 7th International Conference on Trends in Electronics and Informatics (ICOEI), 11-13 April, 2023, Tirunelveli, India, pp. 497-501.

[5] N. Navaitthiporn, P. Rithcharung, P. Hattapath & C. Pintavirooj, "Intelligent glove for sign language communication", 12th Biomedical Engineering International Conference (BMEiCON), 19-22 November, 2019, Ubon Ratchathani, Thailand, pp. 1-4.

[6] H. Rewari, V. Dixit, D. Batra & N. Hema "Automated sign language interpreter", 11th international conference on contemporary computing (IC3), 2-4 August, 2018, Noida, India, pp. 1-5.

[7] L. Chin, C. C. Lin, J. W. Wang, W. C. Chin, Y. H. Chen, S. W. Chang, ... & S. H. Liu, "A Wearable Assistant Device for the Hearing Impaired to Recognize Emergency Vehicle Sirens with Edge Computing", Sensors, 23(17), 7454, 2023.

[8] CREMA-D Dataset. Access Date: October 2 2024.https://paperswithcode.com/dataset/crema-d.

[9] M. Asif, M. Usaid, M. Rashid, T. Rajab, S. Hussain, S. Wasi, "Large-scale audio dataset for emergency vehicle sirens and road noises", Sci. Data, 9, 599, 2022.

[10] M. Yağanoğlu & C. Köse, "Real-time detection of important sounds with a wearable vibration based device for hearing-impaired people", Electronics, 7(4), 50, 2018.

[11] T. Asakura, "Augmented-reality presentation of household sounds for deaf and hard-of-hearing people", Sensors, 23(17), 7616, 2023.

[12] W. H. K. Fernando, P. L. D. N. Perera, M. I. N. Fernando, P. T. E. W. Gunasekara, H. Fernando & R. De Zoysa, "Empowering Deaf Children with Sinhala Sign Language, Emotion Detection, and Sound Recognition", 5th International Conference on Advancements in Computing (ICAC), 07-08 December, 2023, Colombo, Sri Lanka, pp. 388-393.

[13] K. Baltacıoğlu, M. T. Başar, V. Akdoğan, M. Karaaslan & E. A. İlçigen, "Baby cry-sensitive armband design for parents with hearing loss", European Mechanical Science, 7(4), 253-258, 2023.

[14] E. A. Quiapo & K. N. M. Ramos, "Development of a sign language translator using simplified tilt, flex and contact sensor modules", IEEE region 10 conference (TENCON), 22-25 November, 2016, Singapore, pp. 1759-1763.