iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.80246

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Gesture Driven Speech Synthesis

G. Sravan Kumar[1], Sathvik Dendi[2], K. Jayadeep Reddy[3]
[1, 2, 3]Electronics and Communication Department, MVSR Engineering College (Autonomous)

Abstract: Communication plays an important role in human interaction and daily life. However, people with speech impairment face many problems in expressing their feelings and emotions in their social life. Although sign language provides a medium for them to communicate with others, it is not understood by everyone and most people often fail to interpret it correctly. Some existing methods in this area propose a system that converts hand gestures into messages but the major concern with these methods is the bulkiness of the system. Our project proposes an easier solution by designing a system that is lighter and capable of producing human speech based on gestures performed rather than displaying the messages.
Keywords: Gesture Recognition, Speech Synthesis, Wearable System, Embedded System, Assistive Technology.

I. INTRODUCTION

Human interaction is incomplete without communication, as it is a fundamental aspect that enables individuals to express their thoughts, needs, and emotions. Many people with speech impairment find it difficult to express their thoughts, ideas, and emotions. Although sign language provides an alternative mode of communication for them, it is not universally understood and often requires trained interpreters, making it difficult to use in real-world scenarios.

With recent advancements in technology, many gesture-based communication systems have been designed to address these challenges. However, the problem with these systems is their complexity and bulkiness, which decreases their practical usage in day-to-day life. To overcome these limitations, this project focuses on designing and developing a compact wearable system that converts hand gestures into audible speech. The simplicity and portability of this design provide a better solution for individuals with speech impairment.

II. EXISTING METHOD

Several gesture-based communication systems have been developed to assist people with speech impairment. Most of them utilize sensor-based smart gloves connected to microcontrollers such as Arduino or ESP. These systems detect different hand gestures and convert them into text, which is displayed on a screen. Some other approaches make use of wireless communication technologies such as Bluetooth or Wi-Fi to transmit gestures to a mobile device, where the corresponding text is displayed.

However, these approaches are limited by their bulkiness, complexity, and dependence on mobile devices and external displays. To overcome these issues, a more compact and efficient system is required that can directly produce speech instead of text messages.

III. PROPOSED WORK

The proposed system is an embedded system designed as a standalone solution that operates independently without the need for a mobile device. The system consists of a microcontroller, flex sensors, a gyroscope and an accelerometer, a battery with a charging module, an MP3 module, and a speaker. The proposed system serves as an assistive technology designed to support individuals with speech impairment by enabling effective communication through gesture-based speech output.

The microcontroller selected for this design is the ESP32, due to its faster processing power and integrated features. Since the primary objective is to make the system lighter and more compact, we are using MPU6050 instead of separate accelerometer and gyroscope, thereby reducing the number of components. The MPU6050 is a 6-axis motion tracking device that can detect changes in motion, acceleration and rotation. It has an integrated 3-axis gyroscope and 3-axis accelerometer within a single module.

Four flex sensors are used to detect the bending of the index, middle, ring and little fingers. Each flex sensor is connected to a resistor to form a voltage divider circuit, which converts the change in resistance of flex sensor into a corresponding voltage signal that can be read by ESP32.

The entire system is powered by a 3.7V, 1000mAh rechargeable battery. The battery is connected through a TP4056 charging module, which enables convenient recharging of the system. The ESP32 runs at 3.3V but requires a stable 5V input for reliable performance. Therefore, an MT3608 DC-DC boost converter is used to step up the 3.7V battery voltage to 5V.

A DFPlayer Mini is used as the MP3 module for generating the speech output. It is a compact module capable of playing pre-recorded audio files stored in a memory card. It can be easily interfaced with ESP32, which controls the selection of audio files based on the recognized gesture. A speaker connected to the DFPlayer Mini is used to produce speech output. The memory inserted into the DFPlayer Mini contains all the necessary pre-recorded audio files required for the system. Since the DFPlayer follows a strict naming format, all the audio files in the memory card are named in a sequential manner, such as 0001.mp3, 0002.mp3, and so on.

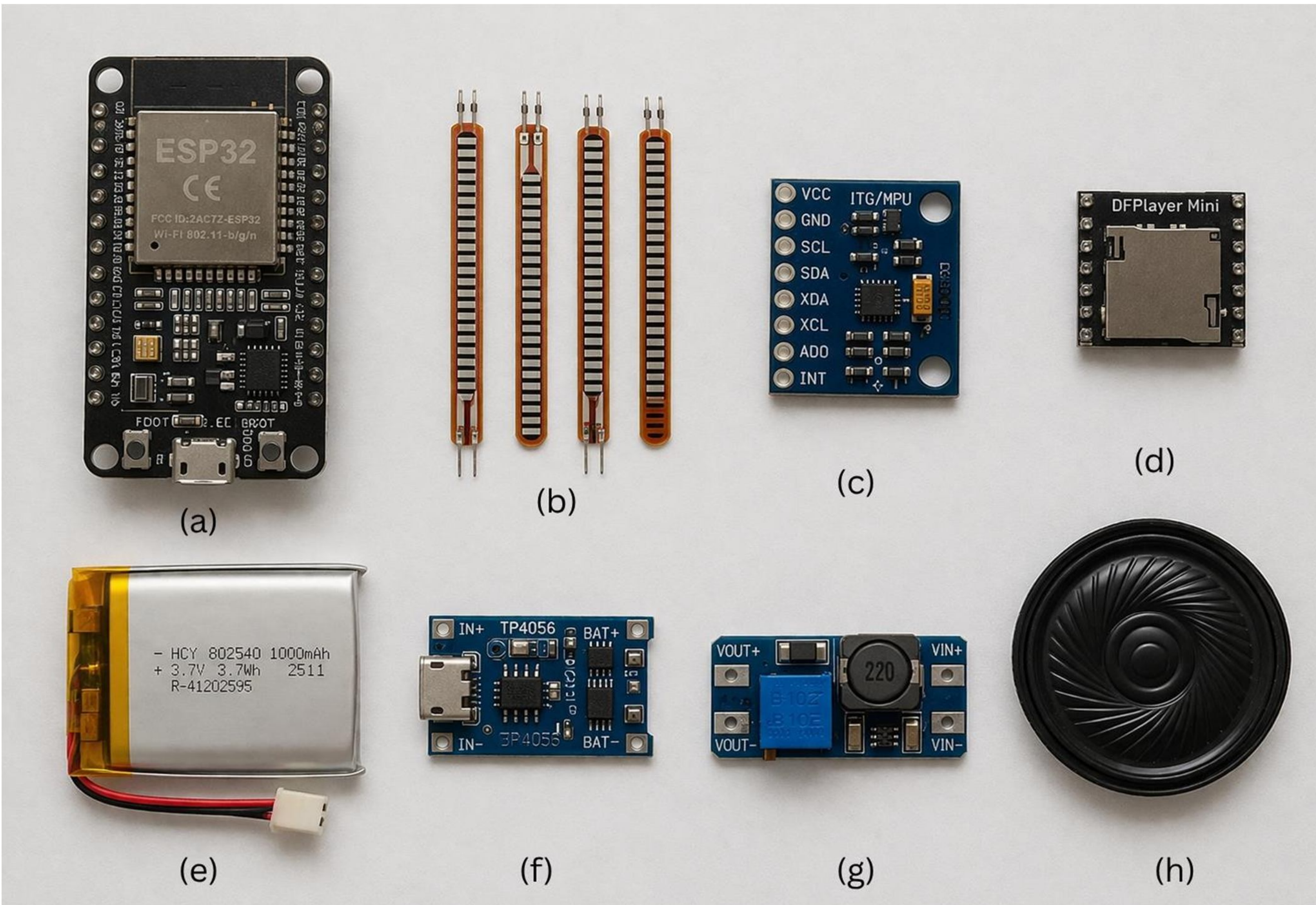

Figure 1: Components used in the system — (a) ESP32, (b) Flex Sensors, (c) MPU6050, (d) DFPlayer Mini, (e) Battery, (f) TP4056 Charging Module, (g) MT3608 Booster, (h) Speaker

Working of the system: The system operates by detecting hand gestures using a combination of flex sensors and an MPU6050 sensor. The overall working process can be explained as follows:

1) Gesture Detection: When a user performs a hand gesture, the flex sensors detect the bending of fingers by producing varying resistance values and the MPU6050 sensor measures the orientation and motion of the hand using its built-in accelerometer and gyroscope.
2) Signal Acquisition: The analog signals from the flex sensors and digital data from the MPU6050 are continuously read by the ESP32.
3) Data Processing: The ESP32 processes the incoming sensor data and compares it with predefined threshold values and conditions to identify specific gestures.
4) Gesture Recognition: Based on the processed data, the system determines the corresponding gesture from the programmed set of gestures.
5) Audio Output: Once a gesture is recognized, the ESP32 sends a command to the DFPlayer Mini module via serial communication. The module then plays the appropriate pre-recorded audio file through the speaker.

The system is designed to operate in a sequential manner, where each stage contributes to the overall functionality of gesture recognition and speech generation. The sensor data collected from the flex sensors and MPU6050 is continuously monitored and processed to ensure accurate detection of gestures. The integration of sensing, processing, and audio output modules enables efficient communication within the system, resulting in reliable performance.

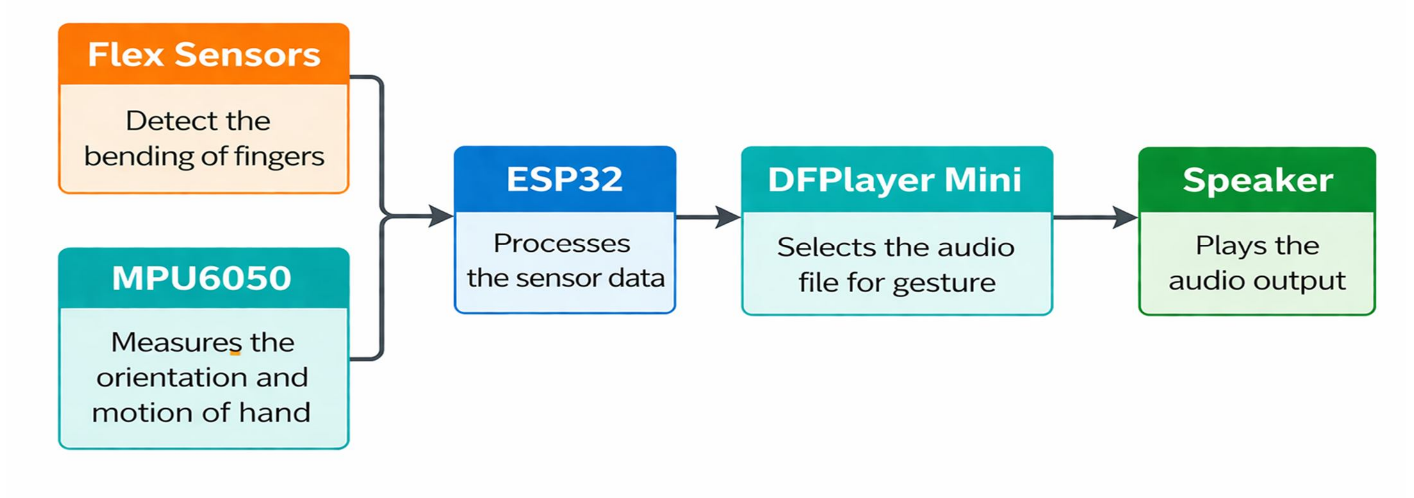

Figure 2: Block diagram illustrating the workflow of the system

The system is programmed using the Arduino IDE, which provides an easy development environment and is compatible with the ESP32. Gesture recognition is implemented using threshold-based logic to identify predefined gesture patterns. The processed data is then used to trigger the corresponding audio output through the DFPlayer Mini module.

IV. RESULTS

The system was successfully implemented and tested for different hand gestures. It was able to detect finger bending, orientation and motion of hand accurately using the flex sensors and MPU6050. Based on the acquired sensor data, the ESP32 identified the gestures using the predefined threshold values and triggered the appropriate audio output through the DFPlayer Mini module. The speech output from the connected speaker was clear and audible, showing the effectiveness of the system.

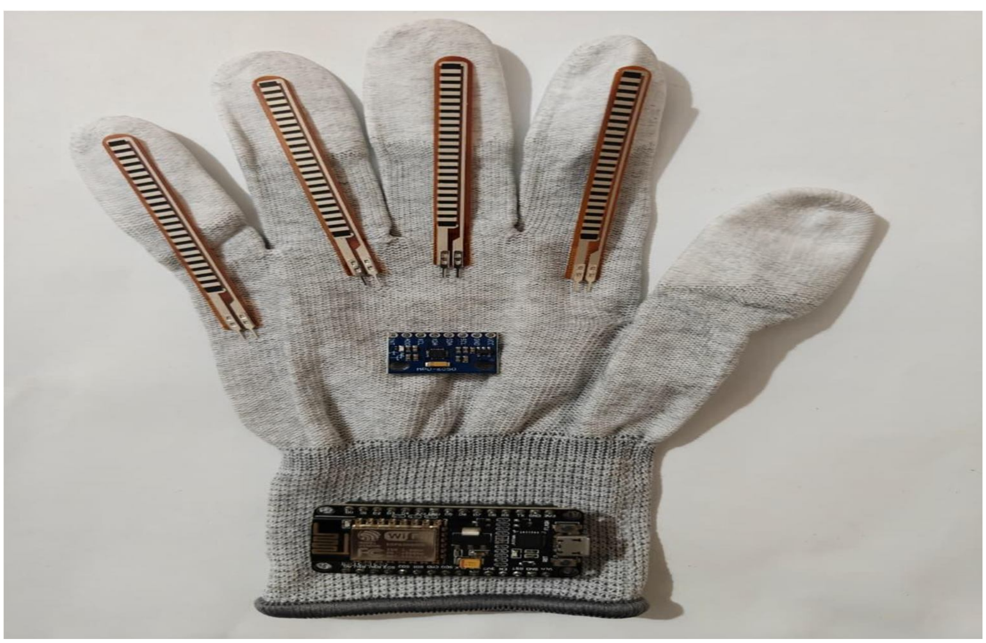

Figure 3: Wearable glove prototype with ESP32, Flex sensors and MPU6050

The prototype was tested with different hand gestures to verify its functionality. The system successfully detected variations in finger bending and hand movement, enabling accurate gesture recognition. The appropriate audio output was produced through the speaker, confirming the proper working of system.

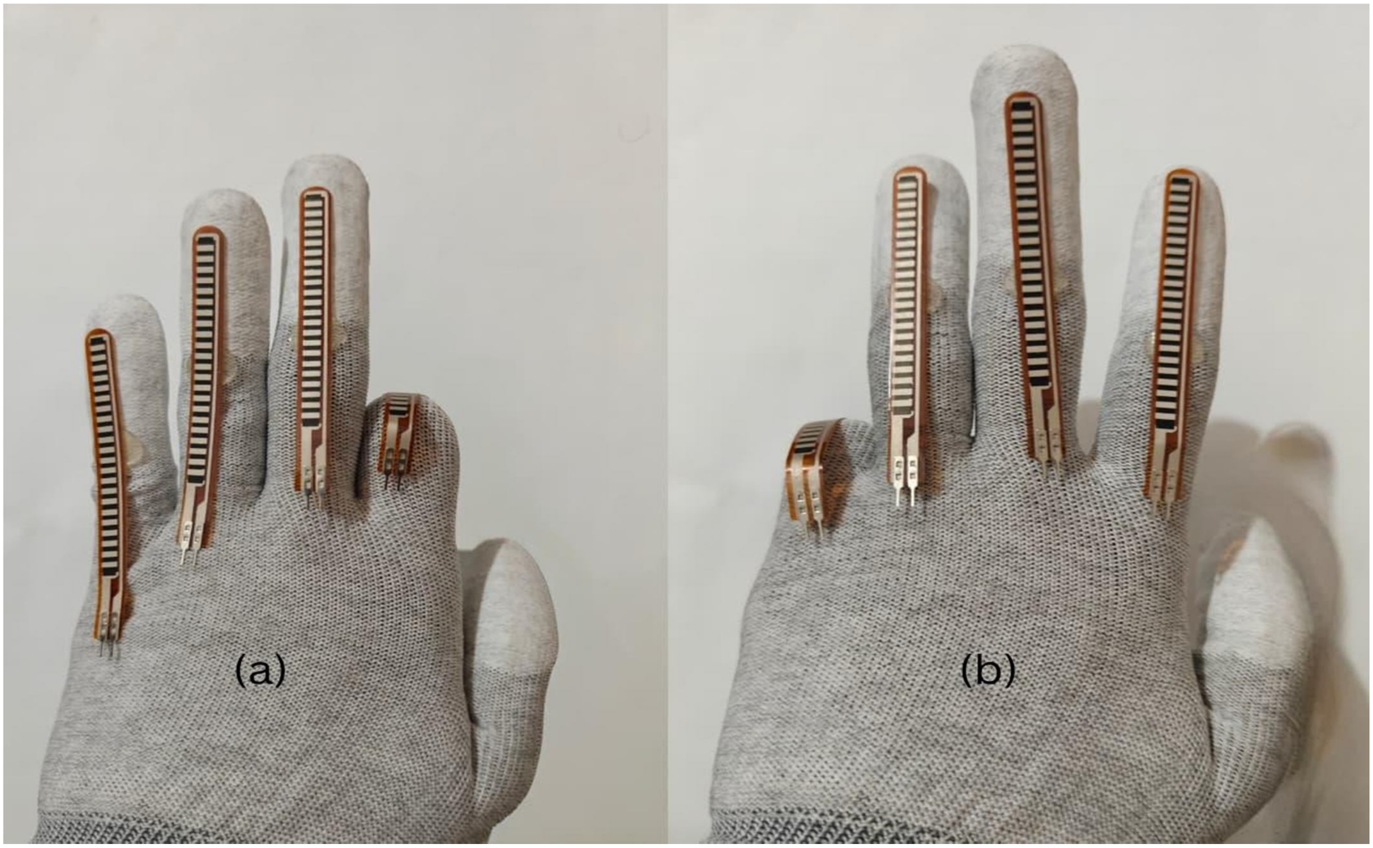

Figure 4: Demonstration of gesture recognition - (a) HELLO gesture, (b) HELP gesture

Some example gestures tested are shown in Figure 4. The gesture shown in Figure 4(a) represents “HELLO”, where the index finger is bent and the middle, ring and little finger are straight. The gesture shown in Figure 4(b) represents “HELP”, where the little finger is bent and the index, middle and ring finger are straight. The system also provides flexibility in defining gesture patterns based on user requirements. Different combinations of finger movements and hand orientations can be mapped to specific audio outputs by modifying the predefined threshold conditions. This allows the system to be easily customized for a wider range of gestures and user preferences.

V. CONCLUSION

In this work, a compact and efficient gesture-based speech synthesis system was developed to assist individuals with speech impairment. The system successfully converts hand gestures into audible speech using a wearable glove integrated with sensors and an ESP32 microcontroller. The use of a standalone design eliminates the dependency on external devices, making the system more practical and portable. The experimental results demonstrate that the system can accurately recognize predefined gestures and generate corresponding speech output effectively. This approach provides a simple and user-friendly solution to improve communication for people with speech disabilities. In future, the system can be extended to support a larger set of gestures and enhance recognition accuracy.

REFERENCES

[1] Rekha H, “MEMS-Based Wearable Smart Glove for Gesture Detection and Sign Language Classification,” International Journal of Innovative Science and Engineering Applications (IJISAE), 2024.

[2] Devyani Pravin Nagarale, “IoT-Based Sign-to-Speech Converter System,” International Journal for Research in Applied Science and Engineering Technology (IJRASET), 2024.

[3] Sangeeta Kurundkar, “Hand Gesture Recognizer Smart Glove Using ESP32 and MPU6050,” International Journal for Research in Applied Science and Engineering Technology (IJRASET), 2024.