Sign-to-Speech Technology: Enhancing Accessibility and Inclusion

Vaishali R. Shirkol[1], Monica K P[2], Thajunnisa N M[3], Sachin K[4]
[1]Postgraduate Student, [2]Professor, [3]Professor, [4]Professor, Department of Computer Science and Engineering, K.V.G. College of Engineering, Sullia, D.K, Karnataka, India

Abstract
Individuals with hearing and speech disabilities often encounter challenges when communicating with people who are unfamiliar with sign language. To address this issue, this project presents a real-time ***Sign Language to Speech Conversion System*** *that translates hand gestures into spoken words using computer vision and deep learning techniques. The proposed system utilizes a Raspberry Pi connected to a camera for capturing live hand gesture images. These images are analyzed using OpenCV for image processing and MediaPipe for efficient hand landmark detection and tracking.*
Before recognition, the captured frames undergo several preprocessing operations, including image scaling, normalization, and background refinement, to improve the quality of the input data. The extracted hand features are then provided to a Convolution Neural Network (CNN), which has been trained to identify various sign language gestures. Once a gesture is recognized, the corresponding text representation is generated and converted into speech using a text-to-speech engine. The generated audio is played through an attached speaker, enabling seamless communication between sign language users and non-sign language users.
The system is capable of operating in real time and supports the recognition of frequently used gestures with reliable accuracy. By combining embedded computing, computer vision, and machine learning, the proposed solution offers a compact, cost-effective, and user-friendly assistive technology. The developed framework demonstrates the potential of intelligent human-computer interaction systems to improve accessibility, promote social inclusion, and support independent communication for individuals with hearing and speech impairments in everyday environments.

Keywords- Sign Language Recognition, Raspberry Pi, Computer Vision, CNN, OpenCV, MediaPipe, Hand Gesture Detection, Text-to-Speech Conversion, Deep Learning, Assistive Communication System.

INTRODUCTION

Communication is essential for sharing ideas, information, and emotions. However, people with hearing and speech impairments often face difficulties in communicating with others because sign language is not widely understood by the general public. This communication gap can affect their participation in education, employment, healthcare, and social activities.

To address this challenge, this project proposes a Sign Language to Speech Conversion System that translates hand gestures into spoken words. The system uses a Raspberry Pi, camera module, OpenCV, and MediaPipe to capture and process hand gestures, while machine learning techniques are used for accurate gesture recognition. The identified gestures are then converted into speech using a text-to-speech engine.

By enabling real-time communication without the need for wearable devices or interpreters, the proposed system provides a practical, affordable, and user-friendly solution. It promotes accessibility, independence, and social inclusion for individuals with hearing and speech disabilities.

OBJECTIVES

- Develop a real-time sign language to speech conversion system using Raspberry Pi.
- Detect and recognize hand gestures using OpenCV, MediaPipe, and Python.
- Classify gestures accurately with a CNN-based machine learning model.
- Convert recognized signs into spoken words through text-to-speech technology.
- Provide a simple, contactless, and user-friendly communication solution.
- Enable future expansion with additional gestures and language support.
- Improve accessibility and communication for individuals with hearing and speech impairments.

BLOCK DIAGRAM REPRESENTATION

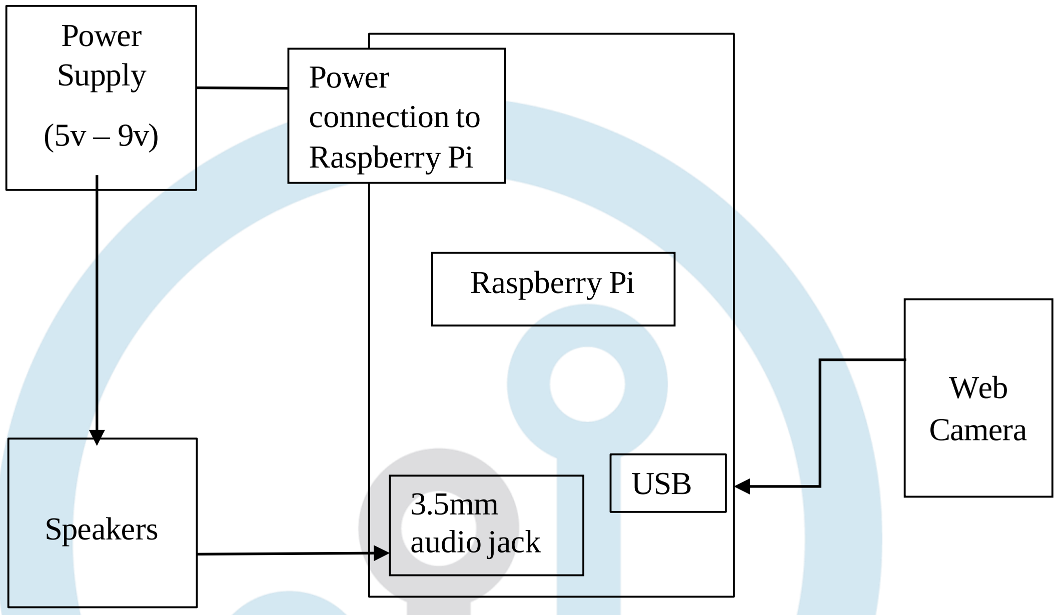

***Figure 1:** Block diagram.*

The proposed system converts sign language gestures into speech in real time. It is built using a Raspberry Pi, a USB camera, and a speaker. The Raspberry Pi acts as the main processing unit of the system. When a user performs a hand gesture in front of the camera, the camera captures live video frames. These frames are processed using OpenCV for image acquisition and preprocessing. MediaPipe is then used to detect and track the user's hand. MediaPipe identifies 21 hand landmark points that represent the position of fingers and the palm. The extracted landmark data is used as input for the gesture recognition model. A CNN-based classification model analyzes these features and determines the corresponding sign language gesture. Once the gesture is recognized, the system maps it to a predefined word or phrase. The recognized text is then passed to a text-to-speech module. The text-to-speech engine converts the text into audible speech. The generated speech is played through a connected speaker, allowing others to understand the user's message. The entire process takes place in real time, providing fast and effective communication. The system is compact, affordable, and easy to use. It does not require gloves, sensors, or any wearable devices. The vision-based approach allows users to communicate naturally using hand gestures. This makes the system suitable for educational institutions, workplaces, healthcare environments, and public spaces. The project provides an efficient assistive solution for improving communication and accessibility for individuals with hearing and speech impairments.

COMPONENTS

Raspberry Pi: The Raspberry Pi 4 is a small and efficient single-board computer used for embedded and intelligent applications. It features a quad-core processor, built-in Wi-Fi, Bluetooth, and supports Linux-based operating systems. Powered by a 5V, 3A USB-C supply, it performs image processing, gesture recognition, and speech generation in this project. Its compact design and reliable performance make it suitable for real-time sign language to speech conversion.

The Raspberry Pi serves as the core controller of the system. It processes images captured by the camera, recognizes hand gestures using OpenCV, MediaPipe, and CNN, and converts the detected signs into speech output through a speaker. Its compact size and efficient performance make it ideal for real-time operation.

Figure 2: **Raspberry Pi**

Camera Module: The Logitech webcam serves as the image acquisition device in the system. It continuously records hand movements and transfers the video stream to the Raspberry Pi. The captured frames are processed using computer vision techniques to identify hand positions and gestures. With its reliable performance, good image clarity, and simple USB interface, the webcam is an effective component for real-time sign language recognition.

Figure 3: Web Camera

Speaker / Earphone: The speaker or earphone functions as the output device of the system. It delivers the voice generated from recognized sign language gestures, allowing users to hear the converted message instantly. Connected through the Raspberry Pi, it provides clear audio feedback and supports effective real-time communication.

Figure 4: Speaker / Earphone

Software Required and Description

OpenCV / cv2 (Library): OpenCV is an open-source image processing library used to acquire and analyze video frames from the camera. It enhances the captured images through preprocessing operations, ensuring better hand detection and tracking. In this project, OpenCV serves as the foundation for gesture recognition by supplying optimized image data for further analysis. ***Code:*** *import cv2*

MediaPipe (Framework): MediaPipe is a machine learning framework that enables real-time hand detection and tracking. It identifies key points on the hand, including finger tips and joints, and provides detailed landmark information for gesture analysis. By using its built-in pre-trained models, the framework accurately tracks hand movements without requiring additional training. In this project, MediaPipe supplies the essential hand landmark data used to recognize sign language gestures efficiently.

Code: **import mediapipe as mp**

WORKING

Figure 5: Working System

The system is processed using Python and OpenCV to improve image quality and prepare them for analysis. MediaPipe detects the hand and extracts 21 landmark points that represent the positions of fingers and joints. These points are then analyzed by a CNN model to identify the performed gesture. After recognizing the gesture, the Raspberry Pi converts it into text and uses a text-to-speech module to generate voice output through a speaker. This enables sign language users to communicate more easily with others.

RESULT

The system tracks the key points on each finger as needs to be, and these detects the points of thumb, index, middle, ring and little finger, overall 21 landmarks are detected and the results for different hand points are as follows:

1. If all the fingers are up, that is Thumb == 1; IF ==1; MF == 1; RF == 1; LF == 1; then the landmarks captured and detected as bellows:

Fig. 6.1 (a) Landmarks of all fingers up

2. For the condition Thumb == 0; IF ==1; MF == 1; RF == 1; LF == 1; landmark are as follows:

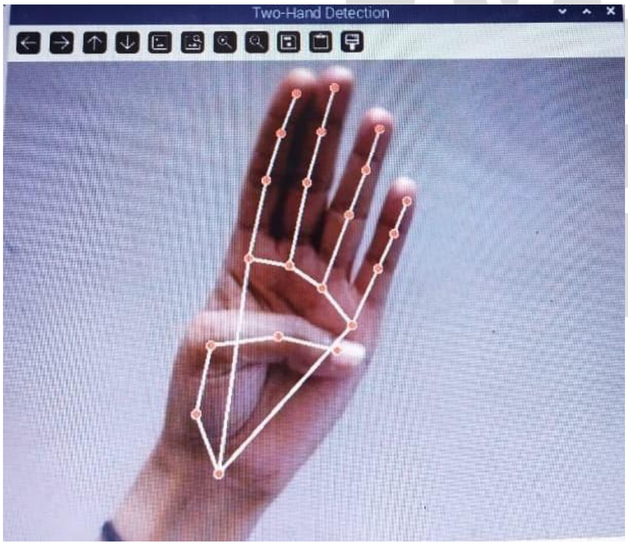

Fig. 6.1 (b) Landmarks of four fingers up

3 When only three fingers are up the conditions are as follows Thumb == 1; IF ==1; MF ==0; RF ==0; LF ==0;

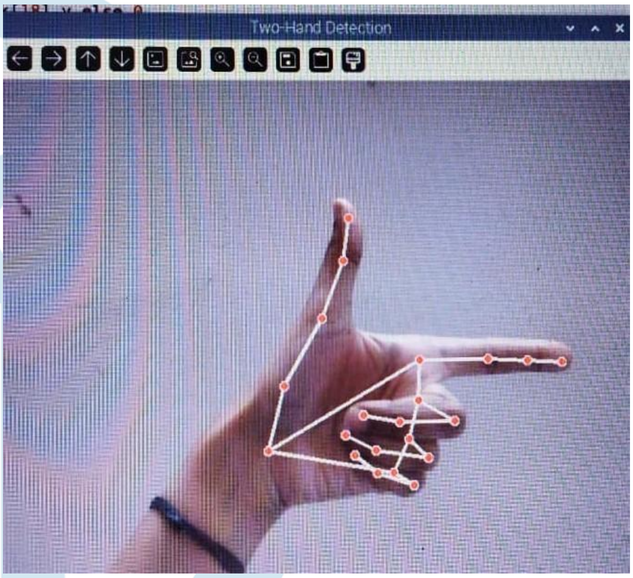

Fig. 6.1(c) Landmarks of thumb & index finger up

4 When all the fingers are closed landmarks detected are as follows Thumb == 0; IF ==0; MF==0; RF ==0; LF == 0;

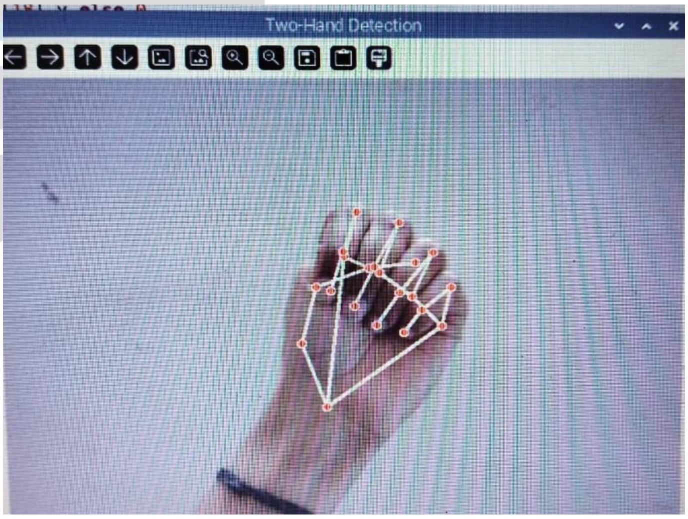

Fig. 6.1 (d) Landmarks of all fingers down

5. When Thumb == 1; IF ==0; MF==0; RF ==0; LF == 0;

Fig. 6.1 (e) Landmarks of thumb finger up

6. If Thumb == 1; IF ==1; MF==1; RF ==0; LF == 0; the points and fingers are detected as follow:

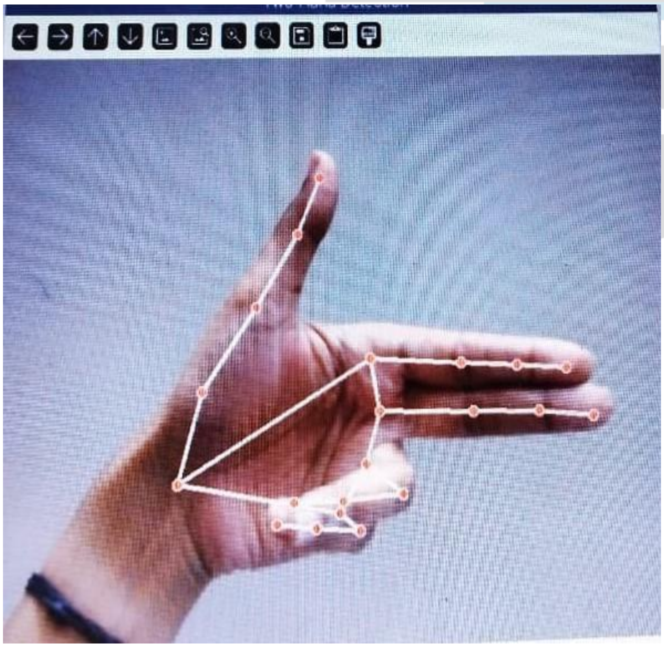

Fig. 6.1 (f) Landmarks of thumb, index & middle finger up

7. If the conditions are Thumb == 0; IF ==1; MF==1; RF ==1; LF == 0;

Fig. 6.1 (g) Landmarks of middle, index & ringer finger up

8. When Thumb == 0; IF ==1; MF==1; RF ==0; LF == 0;

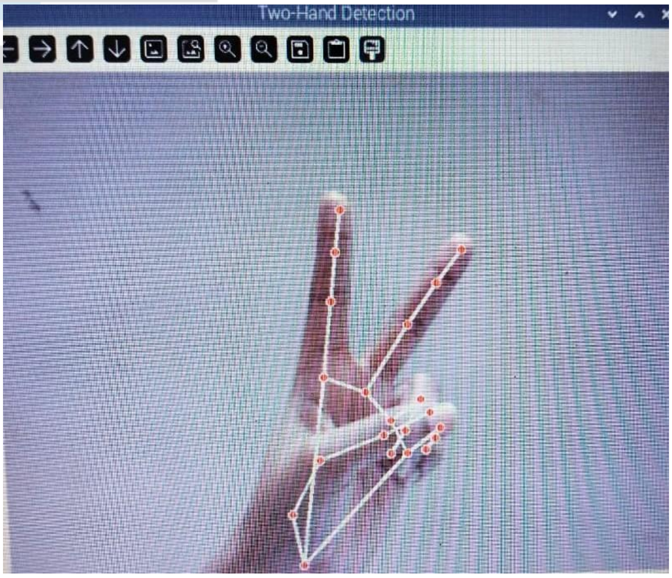

Fig. 6.1 (h) Landmarks of index & middle finger up

The Sign Language to Speech Conversion system was tested using a Raspberry Pi 4, a webcam, and a speaker. The camera captured hand gestures in real time, while MediaPipe extracted hand landmark points for gesture recognition. These landmarks were processed by a CNN model to identify the correct gesture, and the output was converted into speech using a text-to-speech module.

The system achieved an accuracy of approximately 90% and responded within 40–50 milliseconds per frame, providing near real-time performance. It successfully recognized common gestures and converted them into spoken sentences. Custom gestures could also be assigned to specific messages, such as generating an emergency alert when a particular hand sign was detected.

Some errors occurred when gestures looked very similar or when lighting conditions were poor. Despite these challenges, the system performed reliably and demonstrated its effectiveness as a communication aid for sign language users.

FUTURE SCOPE

This system can be enhanced by increasing the range of sign language gestures it can recognize. Future versions may support complete phrases and continuous conversations, making communication more natural and effective.

The performance of the model can be improved by training it with more diverse data collected under different lighting conditions and environments. Support for multiple languages can also be added so that the generated speech can be understood by a wider audience.

The system may be adapted for smart phones, wearable devices, or other portable platforms to improve accessibility. Further research and testing can help increase recognition accuracy and make the solution suitable for everyday use in schools, healthcare centers, and public places.

CONCLUSION

The Sign Language to Speech Conversion system was designed to help people communicate more easily by translating hand gestures into spoken words. By combining a Raspberry Pi, camera, MediaPipe, and a CNN-based recognition model, the system was able to detect gestures and produce speech output in real time.

The project showed that reliable gesture recognition can be achieved using affordable hardware and open-source software tools. The generated speech output allows users to convey messages quickly and effectively without requiring manual text input.

Although certain environmental factors, such as poor lighting and closely related gestures, may affect performance, the system demonstrated consistent and satisfactory results. Overall, this work highlights the potential of technology to improve accessibility and support better communication for individuals who use sign language.

ADVANATAGES & APPLICATIONS

Advantages

- **Affordable Solution**

  The system uses low-cost components such as a Raspberry Pi, webcam, and speaker, making it economical for users and organizations.

- **Fast Response**

  Hand gestures are recognized quickly, and speech is generated almost instantly, enabling smooth communication.

- **Portable Design**

  The compact setup can be easily carried and used in different locations without requiring bulky equipment.

- **Easy to Use**

  Users only need to perform gestures in front of the camera, and the system automatically converts them into speech.

- **Improved Accessibility**

  It helps bridge the communication gap between sign language users and people who do not understand sign language.

- **Low Power Consumption**

  Since the system runs on a Raspberry Pi, it consumes less energy while maintaining reliable performance.

- **Future Expandability**

  New gestures, languages, and advanced features can be added without major changes to the system.

Applications

- **Education**
  Supports communication for students with hearing or speech impairments in classrooms and learning environments.
- **Healthcare**
  Assists patients in communicating their needs and concerns to doctors, nurses, and healthcare staff.
- **Daily Communication**

  Helps users interact more easily in public places such as banks, shops, and government offices.

- **Workplaces**
  Encourages better communication and participation of sign language users in professional environments.
- **Smart Home Systems**

  Can be connected to smart devices, allowing gestures to trigger voice commands for home automation.

- **Public Services**

  Useful in locations such as airports, railway stations, and customer service centers where clear communication is important.

REFERENCES

- Kumar, DN Nagesh, et al. "Sign Language to Speech Conversion—An Assistive System for Speech Impaired." *2019 1st International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE)*. IEEE, 2019.
- Vijayalakshmi, P., and M. Aarthi. "Sign language to speech conversion." *2016 international conference on recent trends in information technology (ICRTIT)*. IEEE, 2016.
- Kaur, Binwant, et al. "Fostering inclusivity through effective communication: Real-time sign language to speech conversion system for the deaf and hard-of-hearing community." *Multimedia Tools and Applications* 83.15 (2024): 45859-45880.
- Shashidhar, R., et al. "Indian sign language to speech conversion using convolutional neural network." *2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon)*. IEEE, 2022.
- Yalabaka, Srikanth, et al. "Sign language to speech conversion using esp32." *2024 4th Asian Conference on Innovation in Technology (ASIANCON)*. IEEE, 2024.
- Hrithik, T. H., S. Rhethika, and Akhil KH. "ML Assisted Sign Language to Speech Conversion Gloves for the Differently Abled." *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*. Vol. 2. IEEE, 2024.
- Raman, Archith, and West Foster. "Sign2Speech: A Novel Approach to Real Time Sign Language to Speech Conversion with Convolutional Neural Networks." *2024 IEEE MIT Undergraduate Research Technology Conference (URTC)*. IEEE, 2024.

- An, Xiaochun, Hongwu Yang, and Zhenye Gan. "Towards realizing sign language-to-speech conversion by combining deep learning and statistical parametric speech synthesis." *International Conference of Young Computer Scientists, Engineers and Educators*. Singapore: Springer Singapore, 2016.

- Saikumar, K., et al. "RFID and IoT-Enabled Sign Language to Speech Conversion Using Deep Learning." *2024 IEEE International Conference on RFID Technology and Applications (RFID-TA)*. IEEE, 2024.

- Bhat, Abhijnya, et al. "Deep Learning Based Speech Synthesis with Emotion Overlay." *International Conference on Advances in Computing and Data Sciences*. Cham: Springer Nature Switzerland, 2023.

- Conlon, Clancy, Barbra Zupan, and Robyn Preston. "The confidence and competence of speech language pathologists in augmentative and alternative communication: a scoping review." *Augmentative and Alternative Communication* 40.4 (2024): 292-305.

- Skurr, Barbara. "Speech Discrimination in the Hard of Hearing." *Australian Journal of Human Communication Disorders* 4.1 (1976): 20-21.

- Moy, Patricia. "Another Look at "Over Protection" and the Child with a Speech Problem." *Journal of the Australian College of Speech Therapists* 20.2 (1970): 90-95.

- Mikelsons, Penny, and Trevor R. Parmenter. "Language Acquisition Programs for the Handicapped." *Australian Journal of Human Communication Disorders* 4.1 (1976): 65-72.

- Zubrick, Ann. "A National Language Policy: Implications for the Communicatively Handicapped." *Australian Journal of Human Communication Disorders* 11.2 (1983): 73-76.

- Ailwood, Peter. "The Intellectual Assessment of Children with Communication Disorders." *Australian Journal of Human Communication Disorders* 4.1 (1976): 61-64.

- Magnússon, Egill, et al. "Representation of the ICF in research of speech intelligibility: A systematic review of literature describing deaf and hard-of-hearing children." *International Journal of Speech-Language Pathology* (2024): 1-14.