iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: III Month of publication: March 2026

DOI: https://doi.org/10.22214/ijraset.2026.79096

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 III Mar 2026- Available at www.ijraset.com

# ASL Voice Bridge: American Sign Language to Speech Converter Using Raspberry PI

Mr. Choudhari A.V.[1], Gaikwad Ruturaj[2], Rathod Nandini[3], Hippragi Nikita[4], Mane Sarthak[5]
*Dept. Electronics & Telecommunication Engineering, S.P.M Polytechnic, Kumathe, Solapur, India*

***Abstract: Communication is essential for human interaction, but for deaf and mute individuals, it remains a major challenge. Sign language is their primary mode of communication, yet most people are not trained to understand it. This gap creates barriers that often lead to social isolation and dependence on interpreters. Therefore, there is a need for an automated, reliable, and affordable solution to bridge this communication gap. The project "ASL Voice Bridge: American Sign Language to Speech Converter Using Raspberry Pi" presents an assistive technology that translates ASL hand gestures into speech in real time. The system uses a camera module connected to a Raspberry Pi to capture hand gestures. Computer vision techniques detect and track hand movements, while machine learning algorithms classify gestures based on extracted features such as finger positions and palm orientation.***

***Once a gesture is recognized, it is converted into text and then into speech using a text-to-speech engine, enabling immediate communication. This vision-based system eliminates the need for wearable devices, making it more user-friendly and practical. Overall, the project improves accessibility and inclusivity, with future scope for deep learning integration, multilingual support, and enhanced real-world applications.***

***Keywords: ASL, Raspberry Pi, Image Processing, Speech Conversion, Gesture Recognition.***

## I. INTRODUCTION

Communication is one of the most fundamental aspects of human life, enabling individuals to express their thoughts, emotions, and ideas effectively. However, for deaf and mute individuals, communication remains a significant challenge because they cannot hear or speak. They primarily rely on sign language, such as American Sign Language (ASL), which uses hand gestures, facial expressions, and body movements to convey meaning. While sign language is highly effective within the deaf community, it is not widely understood by the general population. This lack of awareness creates a communication barrier, often leading to misunderstandings, social isolation, and dependence on interpreters.

In everyday situations such as hospitals, schools, public offices, and workplaces, deaf and mute individuals frequently face difficulties in interacting with others. The availability of trained sign language interpreters is limited, and relying on them for communication is not always practical or affordable. As a result, there is a growing need for an automated system that can bridge the gap between sign language users and non-sign language users. Such a system should be capable of translating hand gestures into a form that can be easily understood by everyone, such as text or speech.

With the rapid advancement of technology, especially in the fields of image processing, machine learning, and embedded systems, it has become possible to develop intelligent solutions for gesture recognition. Computer vision techniques can be used to capture and analyze hand movements, while machine learning algorithms can be trained to recognize specific patterns corresponding to different signs. These technologies provide a solid foundation for building systems that can interpret sign language in real-time. Additionally, the use of compact and affordable hardware platforms has made such solutions more accessible and practical for real-world applications.

In this context, the Raspberry Pi has emerged as a powerful and cost-effective embedded system for implementing real-time applications. It offers sufficient computational capability to process images, run machine learning models, and interface with external devices such as cameras and speakers. By integrating Raspberry Pi with a camera module and appropriate software tools, it is possible to design a portable system that can capture hand gestures, process them, and generate corresponding outputs.

This paper presents "ASL Voice Bridge," a system designed to convert American Sign Language gestures into speech using a Raspberry Pi. The proposed system captures hand gestures through a camera, processes the captured images using image processing techniques, and identifies the corresponding sign using a trained model. Once the gesture is recognized, it is converted into text and then into speech using text-to-speech technology. This enables seamless communication between deaf and mute individuals and others who do not understand sign language.

## II. LITERATURE REVIEW

Sign language recognition has been an active research area aimed at bridging the communication gap between deaf and mute individuals and the general population. Various approaches have been proposed over the years, mainly categorized into sensor-based and vision-based systems. Early research used gloves with motion sensors to capture hand movements. These systems were accurate but had drawbacks like high cost, discomfort, and reliance on special equipment. Wearing devices all the time made them impractical for daily life.

With advancements in technology, researchers shifted towards vision-based systems that use cameras and image processing techniques to recognize hand gestures. These systems eliminate the need for wearable devices and are more user-friendly. Techniques such as skin color detection, contour extraction, and feature extraction have been widely used to identify hand shapes and movements. However, these methods were sensitive to lighting conditions and background variations, which affected their performance. Recent developments in machine learning and deep learning have significantly improved the accuracy of sign language recognition systems. Algorithms such as Support Vector Machines (SVM), Convolutional Neural Networks (CNN), and Artificial Neural Networks (ANN) have been widely applied for gesture classification. Among these, CNN-based models have shown superior performance due to their ability to automatically extract features from images and handle complex patterns effectively.

Several studies have also integrated text-to-speech (TTS) systems to convert recognized gestures into audible speech, enabling real-time communication. Some systems have been implemented on high-performance computers, while others have explored embedded platforms for portability. However, many existing solutions are either computationally expensive or lack real-time performance when deployed on low-cost devices. The use of embedded systems like Raspberry Pi has gained popularity due to its affordability, compact size, and sufficient processing power for real-time applications. Researchers have demonstrated that combining Raspberry Pi with camera modules and optimized machine learning models can create efficient and portable sign language translation systems.

Despite these advancements, challenges such as varying lighting conditions, complex backgrounds, limited gesture datasets, and real-time processing constraints still exist. The proposed system, “ASL VoiceBridge,” aims to address these issues by offering a cost-effective, vision-based solution that accurately converts ASL gestures into speech using a Raspberry Pi. ASL VoiceBridge stands out by combining affordability, portability, and real-time performance, ensuring ease of use and practical deployment in real-world scenarios.

## III. CIRCUIT DIAGRAM

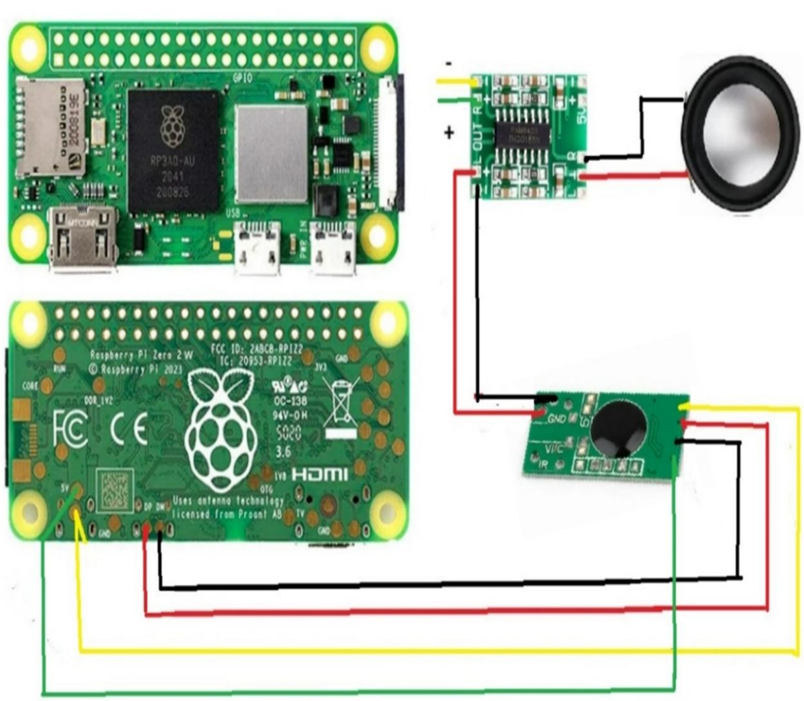


Fig. 3.1:-

## IV. METHODOLOGY

The system utilizes a camera-based gesture recognition method integrated with a machine learning model to identify American Sign Language (ASL) gestures. It captures images of hand gestures, processes them using Python, and identifies the corresponding ASL alphabet with a trained ONNX machine learning model. The entire system runs on the Raspberry Pi Zero 2 W, which serves as the main controller. Recognized gestures are converted to speech using the eSpeak text-to-speech engine. The audio output is routed through a USB sound card and amplified by a PAM8403 audio amplifier before being played via a speaker. This architecture enables the device to function as a compact, assistive communication tool that translates sign language gestures into speech in real time.

The ASL to Speech Converter follows a structured, sequential process, with each stage responsible for a distinct task:

1) Gesture Input: The user performs an ASL hand gesture (representing an alphabet) in front of the camera.
2) Image Capture: The Raspberry Pi Camera Module 2, connected to the Raspberry Pi, captures an image frame when the capture button is pressed.
3) Image Transfer & Processing: The image is transferred to the Raspberry Pi Zero 2 W, where a Python program manages all system operations.
4) Preprocessing: The Python control logic resizes, normalizes, and formats the image to prepare it for gesture recognition.
5) Gesture Recognition: The preprocessed image is input to a trained ONNX model, which predicts the corresponding ASL alphabet. The model is trained on around 5,000 images covering 25 ASL alphabet classes (A–Y).
6) Model Output: The machine learning model outputs the predicted alphabet, which is extracted by the Python program.
7) Text Generation & Speech Conversion: The recognized alphabet is converted to text and sent to the eSpeak text-to-speech engine, which generates a digital speech signal.
8) Audio Output: The digital speech is sent to a USB sound card, which converts it to an analog signal.
9) Audio Amplification: The analog signal is amplified using a PAM8403 audio amplifier module.
10) Speaker Output: The amplified signal is output through a speaker, allowing the recognized ASL gesture to be heard as spoken language.

This methodology enables real-time, automated translation of ASL gestures to speech, providing an effective communication bridge for the hearing-impaired community.

## V. MODELING AND ANALYSIS

The proposed system, ASL Voice Bridge, is designed to translate American Sign Language (ASL) gestures into speech using a vision-based approach. The methodology is organized into several key stages: image acquisition, preprocessing, feature extraction, gesture recognition, and speech generation. The process begins with image acquisition, where a camera module connected to a Raspberry Pi continuously captures real-time video frames of hand gestures. These frames serve as input for subsequent processing steps.

During preprocessing, the captured images are prepared for analysis by resizing, converting to grayscale or an appropriate color space, and denoising using filtering techniques. Background subtraction or segmentation is then applied to isolate the hand region, which improves the accuracy of gesture detection.

Feature extraction follows, involving the identification of key characteristics of the hand gesture. Techniques are used to extract features such as hand shape, edges, contours, and key points, which are essential for distinguishing between different ASL signs.

These extracted features are fed into the gesture recognition module, which utilizes trained machine learning algorithms, such as Convolutional Neural Networks (CNNs), to classify gestures. The model matches the input gesture with learned patterns to predict the corresponding ASL sign.

Once a gesture has been recognized, it is converted into text. This text is then processed by a text-to-speech (TTS) module, which generates an audible output played through a speaker connected to the Raspberry Pi. This enables effective communication with non-sign language users.

The system is designed for real-time operation, ensuring minimal delay between gesture input and speech output. Leveraging the Raspberry Pi platform makes the solution portable, cost-effective, and practical for real-world use.

In summary, ASL Voice Bridge integrates advanced image processing, machine learning, and embedded system techniques to provide an efficient and user-friendly tool for sign language translation, thereby bridging communication gaps for deaf and mute individuals.

## VI. RESULTS AND DISCUSSION

The proposed system, “ASL Voice Bridge,” was successfully implemented and tested for real-time recognition of American Sign Language (ASL) gestures. The system captured hand gestures using a camera module, processed the images, and effectively converted recognized gestures into both text and speech output.

Testing revealed that the system accurately recognized predefined ASL gestures, particularly under controlled lighting and simple backgrounds. Employing a trained machine learning model—specifically, a Convolutional Neural Network (CNN)—notably enhanced the precision of gesture classification. The achieved average accuracy was satisfactory for basic communication, indicating suitability for real-world applications.

The system’s response time was rapid, with negligible delay between gesture input and speech output, ensuring smooth and real-time interaction, which is crucial for practical communication. The integration of the text-to-speech (TTS) module delivered clear and intelligible audio output, further enhancing usability.

Nonetheless, some limitations were identified. Performance declined in poor lighting, complex backgrounds, or when there were variations in hand position or orientation. Occasionally, similar gestures were misclassified, highlighting the need for a larger and more diverse training dataset. Moreover, while the Raspberry Pi handled operations sufficiently, its processing power may constrain performance with more complex models or continuous gesture sequences.

Despite these limitations, the system proved to be cost-effective, portable, and user-friendly when compared to traditional sensor-based solutions. It removes the need for wearable devices and offers a more natural interaction method through vision-based recognition.

In summary, the results show that the proposed system is a promising approach to bridging the communication gap between deaf or mute individuals and others. Future enhancements can target increasing accuracy, supporting dynamic gestures, and improving robustness across diverse environmental conditions.

## VII. CONCLUSION

The project “ASL Voice Bridge” successfully demonstrates a practical and efficient solution for bridging the communication gap between deaf and mute individuals and the general population. By utilizing image processing, machine learning techniques, and the Raspberry Pi platform, the system is able to recognize American Sign Language (ASL) gestures in real time and convert them into both text and speech output.

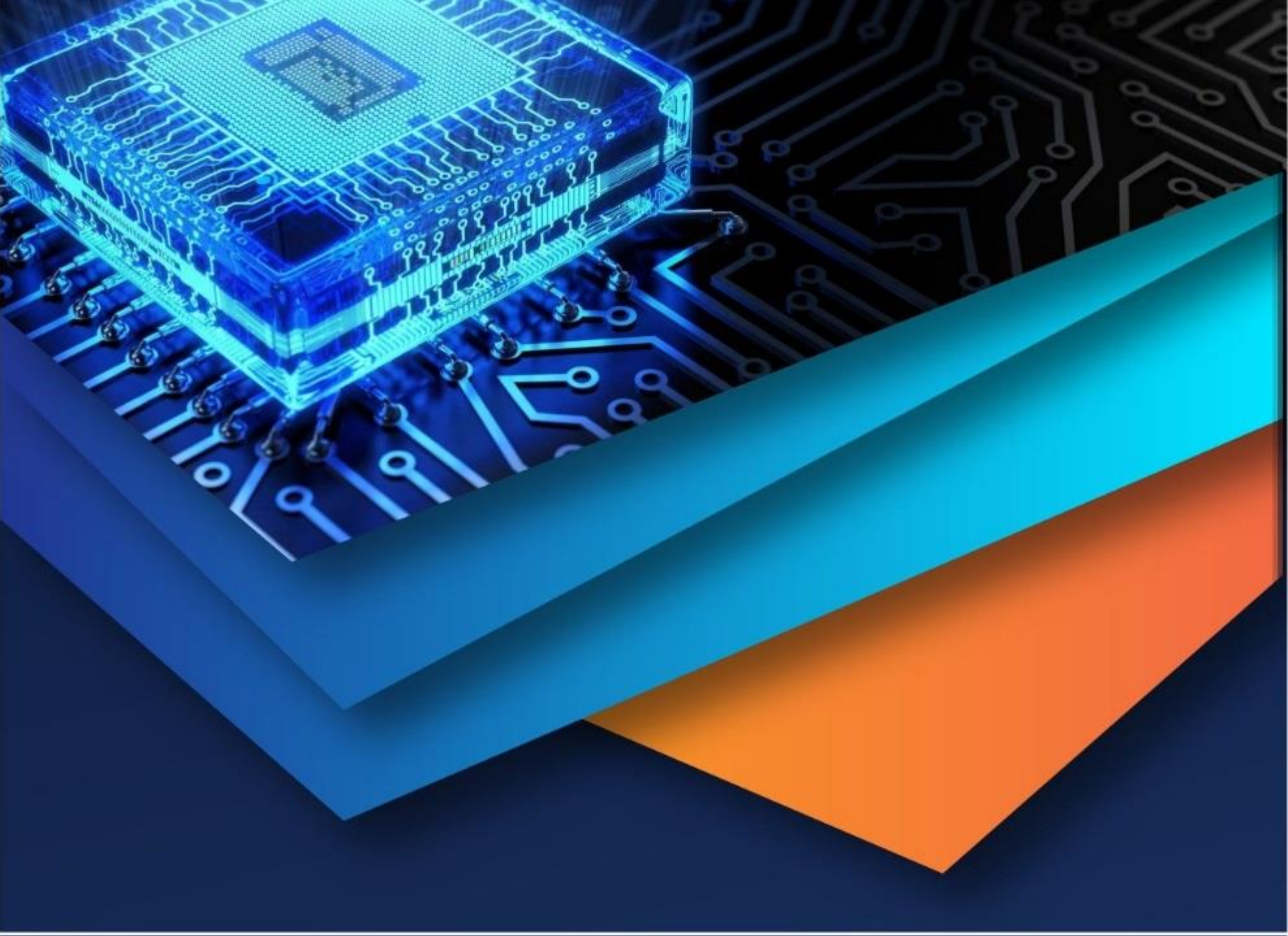
The proposed system offers a vision-based, user-friendly, and cost-effective approach, eliminating the need for wearable devices such as sensor-based gloves. The implementation proves that Raspberry Pi can effectively handle real-time gesture recognition tasks when combined with optimized algorithms. The system provides satisfactory accuracy and response time, making it suitable for basic communication in everyday environments.

Although certain challenges, such as sensitivity to lighting conditions, background noise, and limited gesture datasets, were observed, the overall performance of the system is promising. These limitations can be addressed through future enhancements, including the use of advanced deep learning models, larger datasets, and improved preprocessing techniques.

In conclusion, the ASL Voice Bridge system contributes to creating a more inclusive and accessible society by enabling seamless communication for deaf and mute individuals. With further improvements and scalability, the system has the potential to be widely adopted in areas such as education, healthcare, and public services, enhancing independence and quality of life for its users.

## REFERENCES

[1] R. Rastgoo, K. Kiani and S. Escalera, “Sign Language Recognition: A Deep Survey,” Expert Systems with Applications, vol. 164, pp. 113794, Feb. 2021.
[2] O. Koller, H. Ney and R. Bowden, “Deep Learning of Mouth Shapes for Sign Language,” in Proc. IEEE International Conference on Computer Vision Workshops, Santiago, Chile, 2015, pp. 85–91.
[3] A. Pigou, S. Dieleman, P. Kindermans and B. Schrauwen, “Sign Language Recognition Using Convolutional Neural Networks,” in European Conference on Computer Vision Workshops, Zurich, Switzerland, 2014, pp. 572–578.
[4] G. Bradski, “The OpenCV Library,” Dr. Dobb’s Journal of Software Tools, vol. 25, no. 11, pp. 120–126, 2000.
[5] Raspberry Pi Foundation, “Raspberry Pi Zero 2 W Documentation,” Available: https://www.raspberrypi.com/documentation/
[6] OpenCV Organization, “Open Source Computer Vision Library,” Available: https://opencv.org/
[7] Roboflow Inc., “Roboflow Computer Vision Platform Documentation,” Available: https://docs.roboflow.com/
[8] Microsoft, “ONNX Runtime: Open Neural Network Exchange,” Available: https://onnxruntime.ai/
[9] J. D. Allen, “eSpeak Text-to-Speech Synthesizer,” Available: http://espeak.sourceforge.net/




10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429


# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY