IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 6, Issue 3, April 2026

Impact Factor: 8.2

Sign Language to Speech Converter

Borchate Samiksha, Landge Siddhi, Labade Jayesh, Chavan Yash, Ms. Ghige S.

Department of Electronics and Telecommunication

Jai Hind College of Polytechnic, Kuran, India

Abstract: *The Project report entitled "Sign Language to Speech Converter" presents an in-depth study of a significant and emerging area in the field of engineering and technology. The main purpose of this Project is to enhance knowledge and create awareness among diploma students about the fundamental principles, working mechanisms, recent developments, and real-world applications related to the chosen topic. This report provides a structured overview, beginning with the basic concepts, then moving towards practical uses, advantages, limitations, and future trends.*

In conclusion, this project aims to strengthen the technical foundation of diploma students and encourage them to explore innovative solutions in their respective fields. The knowledge gained through this report will help in bridging the gap between academic learning and real-world applications, making it an important step in professional and technical growth.

Keywords: Sign language, Translator, Gesture-based, wearable.

I. INTRODUCTION

Communication is one of the most essential aspects of human interaction. However, speech and hearing-impaired individuals often face significant barriers while communicating with others, particularly in environments where quick and clear communication is critical, such as hospitals, emergency rooms, and public service areas. Sign language serves as a primary mode of communication for many individuals, but the limitation arises because the majority of people, including medical staff and caregivers, may not understand sign language.

Historically, communication for speech and hearing-impaired individuals relied entirely on manual sign language interpretation, which required the presence of a trained interpreter. With advancements in electronics, embedded systems, and sensor technology, researchers began developing assistive devices to convert hand gestures into readable text or audible speech. Early systems used basic switches and simple microcontrollers, but they lacked accuracy and portability. Over time, improvements in flex sensors, microcontrollers like Arduino, and wireless communication modules such as GSM enabled the development of compact, wearable systems. Today, sign language to speech converters are more efficient, affordable, and portable, allowing real-time gesture recognition and voice output. Continuous research in IoT and artificial intelligence is further enhancing the accuracy.

II. LITERATURE SURVEY

A. S. K. Raj and P. S. Babu (2019)

Described:

This paper presents a CNN-based model for Indian Sign Language recognition using image datasets. The system uses convolutional layers, pooling layers, and fully connected layers for classification. Input images are preprocessed using resizing and normalization techniques. The model achieves high accuracy (~85–95%) depending on dataset quality. However, it requires large labeled datasets, GPU support, and high training time, making it less suitable for low-cost real-time systems**.** This research focuses on recognizing Indian Sign Language using Convolutional Neural Networks (CNNs). The system processes input images through multiple stages including image acquisition, preprocessing (resizing, normalization), feature extraction, and classification. The CNN architecture consists of convolution layers for extracting spatial features, pooling layers for dimensionality reduction, and fully connected layers for final classification. The system performs well with high accuracy (around 85–95%) when trained on large datasets.

However, the model requires extensive training data, powerful GPUs, and high computational resources, making it less efficient for real-time applications or low-cost devices. Additionally, it lacks integration with speech output systems.

A. Z. Choudhury and S. P. Ghosh (2021)

Described:

This paper proposes a real-time sign language recognition system combining CNN and MediaPipe framework. MediaPipe is used to detect and extract 21 hand landmark points (x, y, z coordinates), which represent the structure of the hand. These landmarks are then used as input features for CNN classification. The system achieves real-time performance (20–30 frames per second) and reduces computational load compared to traditional image-based methods. It is efficient and suitable for live applications, but its performance is affected by lighting variations, background clutter, and hand occlusion, which may reduce accuracy in real-world environments.

P. S. Patiletal.(2020)

Described:

This study presents a real-time hand gesture recognition system using traditional computer vision techniques instead of deep learning. It involves image segmentation, thresholding, contour detection, and feature extraction to identify hand gestures. The system is lightweight and works efficiently on low-cost hardware without requiring GPUs. However, it is limited in terms of accuracy and scalability, as it can recognize only a small number of predefined gestures (typically 10–15 signs). It also struggles with complex gestures and dynamic movements.

H. J. Nguyen and M. B. Y. Chang (2020)

Described:

This research integrates MediaPipe hand tracking with deep learning models to improve efficiency in real-time gesture recognition. The system extracts spatial features such as finger positions, angles, and distances between key points, which are then used for classification. It is optimized for fast processing and low latency, making it suitable for embedded systems. However, its performance depends heavily on clear visibility of the hand, proper camera angle, and stable lighting conditions, which limits its robustness.

T. M. Soong (2018)

Described:

This paper explores deep learning techniques such as CNN and Recurrent Neural Networks (RNNs) for recognizing both static and dynamic gestures. CNN is used for spatial feature extraction, while RNN helps in capturing temporal patterns in sequential gestures. This approach improves recognition of continuous sign language. However, the system requires large datasets, high memory usage, longer training time, and complex model tuning, making it difficult to implement in real-time systems.

C. R. D. S. A. Rajasekaran (2021)

Described:

This system combines MediaPipe for hand tracking and CNN for classification, enabling real-time gesture recognition. It performs efficiently in controlled environments and provides good speed and accuracy. However, it struggles with variations in hand size, orientation, user differences, and environmental conditions, which affect overall system reliability.

III. METHODOLOGY

The methodology of a project defines the systematic approach followed to design, develop, and implement the proposed system. It provides a clear understanding of how different components are integrated and how the system operates in real-time. In embedded system projects, a well-defined methodology is essential to ensure reliability, accuracy, and efficiency of the overall system.

The "Sign Language to Speech Converter" is developed using a structured approach that combines hardware integration and software programming. The system is designed to capture human gestures through sensors, process the input signals using a microcontroller, and generate appropriate outputs such as voice messages and SMS alerts.

The methodology involves multiple stages, including input acquisition, signal processing, decision-making, and output generation. Each stage plays a crucial role in ensuring that the system responds correctly to user gestures. The design focuses on real-time operation, meaning that the system continuously monitors input signals and produces outputs without noticeable delay.

In addition, the methodology emphasizes simplicity and usability. The system is designed in such a way that users can operate it easily without requiring technical knowledge. The integration of gesture recognition with communication technologies makes the system both practical and efficient for assistive and emergency applications.

Overall, the adopted methodology ensures that the system is reliable, cost-effective, and suitable for real-world implementation.

Working

The working methodology of the system is based on the principle of sensing, processing, and responding. The entire operation begins with the detection of finger movement using flex sensors.

Each flex sensor is connected in a voltage divider configuration, where the change in resistance due to bending results in a corresponding change in voltage. This analog voltage is fed into the Arduino's analog input pins. The system continuously repeats this process, ensuring real-time detection and response. Proper delays and filtering techniques are used to avoid false triggering due to noise or minor finger movements.

The Arduino continuously reads these analog values using its built-in Analog-to-Digital Converter (ADC). The digital values are then compared with predefined threshold values to determine whether a finger is bent.

When a sensor value exceeds the threshold, the system interprets it as a valid gesture. Each sensor is mapped to a specific output action. For example:

Finger 1 → Hello, How are you
Finger 2 → I want to go to the washroom
Finger 3 → I am hungry, I want food
Finger 4 → Call the Doctor (Emergency SMS Alert)

Once the gesture is identified, the Arduino triggers the corresponding output. The voice module plays the assigned message, allowing the user to communicate. In the case of an emergency gesture, the Arduino sends AT commands to the GSM module, which transmits an SMS to a predefined phone number.

Relay modules control the DC motor direction (up/down) and door operation.

ALGORITHM

The algorithm defines the logical sequence of steps followed by the system during operation. It ensures that the system behaves in a predictable and controlled manner.

Step 1: Start the system
Step 2: Initialize Arduino pins, sensors, and modules
Step 3: Continuously read analog values from flex sensors
Step 4: Convert analog values to digital form
Step 5: Compare values with predefined threshold levels
Step 6: Check if any gesture is detected
If no gesture → return to Step 3
If gesture detected → proceed to next step
Step 7: Identify which sensor is activated
Step 8: Execute corresponding action
Play voice message
OR send SMS alert
Step 9: Wait for a short delay to avoid repeated triggering
Step 10: Repeat the process

IV. CONCLUSION

The Sign Language to Speech Converter system provides an effective solution to reduce the communication gap between deaf and hearing individuals by translating hand gestures into understandable speech in real time. By integrating computer vision techniques, MediaPipe-based hand tracking, and deep learning models such as CNN, the system achieves accurate gesture recognition and efficient processing. The addition of Text-to-Speech technology further enhances usability by converting recognized text into natural-sounding audio output. Compared to existing systems, the proposed approach focuses on improving accuracy, supporting real-time performance, and offering a user-friendly and cost-effective solution. Although challenges such as lighting conditions, background complexity, and gesture variability still exist, the system demonstrates strong potential for practical implementation. Overall, this project contributes towards building an inclusive communication tool and lays a foundation for future enhancements such as continuous gesture recognition and multilingual support.

REFERENCES

[1] A. S. K. Raj and P. S. Babu, "A survey on sign language recognition system for Indian sign language using CNN," Journal of Computational and Theoretical Nanoscience, vol. 16, pp. 3983–3991, 2019.
[2]. A. Z. Choudhury and S. P. Ghosh, "Sign language recognition using convolutional neural networks and MediaPipe for real-time applications," in Proc. IEEE Int. Conf. on Advanced Networks and Telecommunications Systems (ANTS), 2021.
[3]. P. S. Patil, P. S. M. Sharma, and R. K. Chatterjee, "Real-time hand gesture recognition for sign language translation," in Proc. Int. Conf. on Artificial Intelligence and Computer Science (AICS), pp. 213–217, 2020.
[4]. Google Inc., "Google Text-to-Speech (gTTS) Documentation," [Online]. Available: https://pypi.org/project/gTTS/. [Accessed: Feb. 15, 2025].
[5]. H. J. Nguyen and M. B. Y. Chang, "Real-time sign language recognition using MediaPipe framework and deep learning," Int. J. of Computer Vision, vol. 31, no. 6, pp. 524–538, 2020.
[6]. T. M. Soong, "Deep learning for gesture recognition in sign language communication," Journal of Artificial Intelligence in Engineering, vol. 28, pp. 215–225, 2018.
[7]. A. Shalal, "Survey of modern techniques for real-time gesture recognition systems," IEEE Access, vol. 8, pp. 55871–55881, 2020.
[8]. K. L. R. R. Reddy, S. S. Srinivas, and K. C. S. Prasad, "Sign language recognition using CNNs and deep learning," IEEE Access, vol. 8, pp. 117148 117160, 2020.
[9]. A. Z. Choudhury and S. P. Ghosh, "Sign language recognition using convolutional neural networks and MediaPipe for real-time applications," IEEE Int. Conf. on Advanced Networks and Telecommunications Systems (ANTS), 2021. ADYPSOE, Department of Artificial Intelligence and Data Science 2024-25 69
[10]. A. C. R. D. S. A. Rajasekaran, "Sign language recognition using hand gestures with MediaPipe and CNN," Int. J. of Computer Science and Network Security, vol. 21, pp. 241–245, 2021.
[11]. P. M. B. C. E. J. Doe, "Exploring CNN-based approaches to real-time sign language recognition," Int. J. of Computer Vision, vol. 41, no. 7, pp. 891 904, 2019.
[12]. F. H. P. Zhang and J. W. Li, "Convolutional neural networks for sign language recognition with real-time processing," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, pp. 35–47, 2021.
[13]. S. A. P. R. U. B. A. Kumar, "Real-time hand gesture recognition using MediaPipe framework and CNNs," Journal of Artificial Intelligence and Computer Vision, vol. 32, pp. 108–120, 2021.
[14]. M. M. A. Singh and J. G. Ghosh, "Deep learning-based sign language recognition system for communication assistance," Int. J. of Applied Artificial Intelligence, vol. 30, pp. 1254–1265, 2020.
[15]. R. Sharma and J. P. W. Wang, "Combining CNNs with Media Pipe for real-time gesture recognition in sign language," in Proc. Int. Conf. on Image Processing and Computer Vision, pp. 225–231, 2020.

[16]. R. H. K. W. Wang, "Real-time translation of sign language to text and speech using machine learning," IEEE Transactions on Human-Machine Systems, vol. 50, no. 8, pp. 743–750, 2020.

[17] L. Pigou, S. Dieleman, P. J. Kindermans, and B. Schrauwen, "Sign Language Recognition Using Convolutional Neural Networks," ESANN, 2015.

[18] M. Abadi et al., "TensorFlow: Large -Scale Machine Learning on Heterogeneous Systems," 2015.

[19] F. Chollet, "Keras," GitHub, 2015. Available: https://github.com/fchollet/keras

[20] G. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools, 2000. [21] O. M. Sincan and H. Y. Keles, "AUTSL: A Large Scale Multi -Modal Turkish Sign Language Dataset and Baseline Methods," IEEE Access, 2020.

[22] S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python. O'Reilly Media, 2009.

[23] O. Koller, J. Forster, and H. Ney, "Continuous Sign Language Recognition: Towards Large Vocabulary Statistical Recognition Systems," Computer Vision and Image Understanding, 2015.