



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 8, August 2025

Smart Translation for Physically Challenged People using Deep Learning Techniques

Deekshith TR¹, Naseerhusen. A²

PG Student, Dept. of MCA, City Engineering College, Bengaluru, Karnataka, India¹

Assistant professor, Dept. of MCA, City Engineering College, Bengaluru, Karnataka, India²

ABSTRACT: In the modern digital era, bridging the communication gap between hearing-impaired individuals and the hearing population has become a vital technological challenge. Sign language serves as the primary mode of communication for the deaf community, yet it is not widely understood outside it, creating barriers in education, employment, healthcare, and social integration. This project, Smart Communication Assistance using Deep Learning, is designed to address these challenges by creating a system capable of translating spoken language into Indian Sign Language (ISL) and vice versa, in real time. The system integrates speech recognition technology, Natural Language Processing (NLP), and a database of ISL signs (images and animations) to provide accurate and meaningful translations. It is lightweight, cost-effective, and can be deployed across various platforms, from personal devices to public information kiosks. Beyond daily conversation, the application has educational benefits, enabling children with hearing impairments to develop linguistic skills and improving inclusivity in classrooms.

KEYWORDS: Speech-to-Sign Translation, Indian Sign Language, Deep Learning, Natural Language Processing, Real-time Communication, Accessibility Tools, Sign Language Recognition, Camshift Algorithm, Google Speech API.

I. INTRODUCTION

Communication is the foundation of human interaction. For individuals with hearing impairments, sign language acts as the bridge to the hearing world, but its limited adoption among non-deaf individuals presents a persistent barrier. Indian Sign Language (ISL) consists of a complex combination of hand shapes, facial expressions, and body movements, making it rich in meaning but challenging for non-experts to interpret.

This project is motivated by the vision of an inclusive society where hearing-impaired individuals can interact freely in educational, professional, and personal settings without requiring a human interpreter. Leveraging advancements in deep learning and computer vision, the system processes voice inputs through the Google Speech API, applies NLP techniques for preprocessing, and generates corresponding ISL signs in image or animation form.

The proposed system also supports reverse translation — from ISL gestures to spoken words — using computer vision techniques such as the Camshift algorithm. This bidirectional approach enables not only accessible communication but also the possibility of integrating such systems into video conferencing platforms, educational content, and even emergency services.

In recent years, the demand for assistive communication technologies has grown significantly due to the increased emphasis on inclusivity and accessibility in both public and private sectors. While several applications exist for converting speech to text, very few extend their functionality to provide accurate sign language translation, especially in the Indian context where ISL is underrepresented in mainstream software solutions. This scarcity has left a considerable communication gap for millions of hearing-impaired individuals in India.

The implementation of deep learning models in this domain allows the system to not only recognize and translate common words but also adapt over time through continuous learning. By training on diverse datasets that capture variations in accents, pronunciations, and sign gestures, the system becomes more robust and context-aware. Additionally, incorporating Natural Language Processing ensures that the translated signs follow the grammatical structure of ISL rather than a direct word-for-word translation, resulting in more meaningful and natural communication.

Another key aspect of this project is its adaptability to different deployment environments. The lightweight architecture enables it to run on standard desktop systems, mobile devices, or even embedded platforms such as Raspberry Pi, making it feasible for schools, public service centers, healthcare institutions, and personal use. The system's ability to work with both live inputs and pre-recorded files adds further flexibility.

Ultimately, this project represents a step toward breaking down the communication barriers that isolate the deaf community from broader society. By providing a reliable, cost-effective, and scalable communication bridge, it not only enhances day-to-day interactions but also opens new opportunities in education, employment, and social participation for hearing-impaired individuals.

II. SYSTEM MODEL AND ASSUMPTIONS

The system is modeled as a modular, scalable application that integrates voice capture, processing, and ISL translation. It assumes users have access to devices with microphones, displays, and internet connectivity (for online API calls). However, offline mode can be incorporated for areas with low connectivity by preloading the ISL dictionary.

The ISL dictionary forms the core knowledge base, containing thousands of predefined words and phrases, each represented by high-resolution images or GIFs. Synonym mapping ensures that even words not directly in the database are substituted with appropriate equivalents.

Key Assumptions: - The speech environment is relatively noise-free for better recognition. The system will have continuous updates to improve vocabulary and gesture accuracy. Gesture datasets for reverse translation will be collected and stored securely.

III. PROPOSED METHODOLOGY

The methodology follows a structured sequence of processing steps:

1. Audio Capture: Speech is collected through a microphone in real time.
2. Speech-to-Text Conversion: Google Speech API processes the input into a textual format.
3. NLP Preprocessing: Tokenization, stop-word removal, stemming, and synonym replacement are applied to the text.
4. Sign Mapping: The processed text is matched against the ISL dictionary. If no match exists, the closest synonym is used.
5. Output Rendering: The ISL image or animation is displayed to the user.
6. Reverse Communication: Using Camshift tracking, ISL gestures are identified, converted into text, and then synthesized into speech.

The system is designed to run efficiently on low-spec devices, with lightweight deep learning models optimized for faster inference. Additionally, modular coding ensures easy integration into other platforms like mobile apps and web-based conferencing tools.

IV. SYSTEM ARCHITECTURE AND EFFICIENT COMMUNICATION

The architecture is divided into multiple layers: Input Layer, Processing Layer, Translation Layer, and Output Layer. Communication between these layers is handled through internal APIs, ensuring modularity and scalability. - Input Layer: Captures voice or video-based sign language. - Processing Layer: Handles preprocessing using NLP for text and image filtering for signs. - Translation Layer: Maps input to output using deep learning models and ISL dictionaries. - Output Layer: Displays ISL signs or produces speech.

This architecture ensures efficient communication between modules, reducing latency to under 3 seconds per translation. This speed makes it feasible for live conversations, an essential requirement for effective accessibility tools.

V. SECURITY

Given the sensitive nature of personal communication, security is integrated into every stage of system development. All translations occur locally where possible, and any cloud-based processes are encrypted using industry standards (e.g., AES-256). Access controls restrict the sharing of stored vocabulary or usage logs to authorized users only.

VI. RESULT AND DISCUSSION

Testing involved 500 phrases from everyday conversations. The system achieved: - Accuracy: 94.8%- Response Time: 2.8 seconds average- User Feedback: 91% satisfaction rate among deaf students.

Feedback indicated that while the system performed exceptionally in quiet environments, its performance dropped slightly in noisy locations. This can be addressed by integrating noise-cancellation preprocessing into future iterations.

VII. CONCLUSION

The Smart Communication Assistance using Deep Learning project demonstrates a practical approach to breaking down communication barriers between hearing and non-hearing individuals. By incorporating speech recognition, NLP, and sign language rendering, it creates an effective and accessible tool for daily use. With further development, it has the potential to integrate seamlessly into education, public services, and online platforms.

REFERENCES

1. Mary Jane C. Samonte, 'Using Deep Learning in Sign Language Translation to Text,' 2022.
2. Purushottam Sharm, 'Indian Sign Language Generation using Natural Language Processing and Audio Speech,' 2022.
3. Shruti Chavan, Xinrui Yu, Jafar Saniie, 'Convolutional Neural Network Hand Gesture Recognition for American Sign Language,' 2021.
4. Onkar Bidkar, 'Speech/Audio to Sign Language Translator for Deaf People,' 2021.
5. Shagun Gupta, 'Sign Language Converter Using Hand Gestures,' 2020.
6. Akshay Goel, 'Sign Recognition and Speech Translation Using OpenCV,' 2020.
7. B. Sivaranjani, 'Audio to Sign Language Translator Using Python,' 2020.
8. Ashmi Katariya, Vaibhav Rumale, 'Voice to Indian Sign Language Conversion for Hearing Impaired People,' 2020.
9. Chaithraj, 'Sign Language Converter for Deaf and Dumb People in Two Way Communication for Regional Languages,' 2019.
10. Vishwas S, Hemanth Gowda, 'Sign Language Translator using Machine Learning,' 2019.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details