International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

AI Tool/Mobile App for Indian Sign Language (ISL) Generator from Audio-Visual Content in English/Hindi to ISL Content And Vice-Versa

Dr. Harsha R. Vyawahare[1], Sukhada Shripad Tare[2], Ashwini Nitin Shingane[3], Shreya Sunil Shinde[4], Bhavika Suraj Jain[5]

[1]Associate Professor, Department of Computer Science And Engineering , Sipna College of Engineering and Technology , Amravati
[2]Student , Department of Computer Science And Engineering , Sipna College of Engineering and Technology , Amravati

Abstract- This paper presents a practical and lightweight bidirectional communication system that translates between speech/text and Indian Sign Language (ISL) using machine learning and computer vision techniques. The system supports two modes: Speech-to-ISL and ISL-to-Text/Speech. In Speech Mode, spoken input is converted into text using speech recognition, then mapped to corresponding ISL alphabet images. In Camera Mode, hand gestures are captured using a webcam and classified using a Convolutional Neural Network (CNN) model to generate text and voice output. The system is implemented using Streamlit for the user interface, OpenCV for image processing, TensorFlow/Keras for gesture recognition, and pyttsx3 for speech synthesis. The proposed system provides a simple, real-time, and cost-effective solution to improve communication accessibility for the Deaf and Hard-of-Hearing (DHH) community.

Keywords – Indian Sign Language, Speech Recognition, Gesture Recognition, CNN, OpenCV, Accessibility.

I. INTRODUCTION

Communication between hearing individuals and the Deaf and Hard-of-Hearing (DHH) community remains a significant challenge in everyday life. While Indian Sign Language (ISL) serves as a primary mode of communication for many DHH individuals, a large portion of the population is not trained in understanding or using ISL. This lack of mutual understanding creates a communication barrier that affects education, healthcare access, employment opportunities, and social interaction.

To address this issue, this project proposes a bidirectional communication system that enables seamless interaction between both groups. The system is designed with two core functionalities:

1. **Conversion of speech into ISL images**

Spoken language is first converted into text using speech recognition techniques. The text is then broken down into individual characters, which are mapped to corresponding ISL alphabet images. These images are displayed sequentially, allowing DHH users to understand the message visually.

2. **Recognition of ISL gestures into text and speech**

Hand gestures performed by users are captured using a webcam and processed through a Convolutional Neural Network (CNN) model. The model classifies each gesture into its corresponding alphabet, which is then combined to form meaningful words and sentences. The output is displayed as text and also converted into speech using a text-to-speech engine, enabling hearing individuals to understand the message.

Unlike complex research systems that rely on computationally expensive 3D avatars, large-scale datasets, and advanced deep learning architectures, this system focuses on a lightweight, efficient, and practical implementation. By using simple image mapping for ISL representation and a CNN-based approach for gesture recognition, the system achieves real-time performance with lower computational requirements. This makes it more accessible, easier to deploy, and suitable for use on standard devices, thereby providing a cost-effective solution for improving communication between hearing and DHH individuals.

II. LITERATURE REVIEW

Speech-to-ISL Systems

The IJRASET AI Tool/Mobile App (2025) uses Google ASR with translation into English for mapping vocabulary to ISL GIF animations, showing effective real-time speech capture

and sign rendering ai-tool-mobile-app-for-indian [5]. Similarly, the Python-based Hindi speech recognition system uploaded by the user employs Google ASR, deep translation, and phrase-to-GIF mapping for sign output Speech_Recognition_Hindi[1][6].

Text-to-Sign Research (3D Sign Generation)
The IJACSA 2024 paper introduces a 3D-model–based gesture generation system with ISL grammar correction, achieving high accuracy (99.21%) for English-to-ISL conversion Paper_114-Harnessing_AI_to_Gene…. Their analysis of ISL grammar transformations (SOV, tense markers, spatial modifiers) informs the rule-based grammar engine in our model[3].

Gesture Recognition and Sign Interpretation
Deep learning models using CNNs, LSTMs, OpenPose/MediaPipe, and transformer architectures have been used extensively in gesture recognition systems. Related works highlight challenges such as occlusion, lighting variation, motion blur, and signer diversity [4].

3D Avatars and Animation Techniques
Virtual avatar approaches across multiple studies emphasize motion capture, HamNoSys-to-SiGML mapping, and 3D rigging issues—particularly around facial expressions and fluidity of movement [6].

Identified Gaps
Across the reviewed studies, several shared limitations are evident.

- Limited ISL datasets compared to ASL/BSL [1]
- Non-standardized gloss representation [3]
- Incomplete models lacking bidirectional translation [1][2]
- Inability to handle continuous signing [6]
- Lack of smooth, human-like 3D animation [5]
- This motivates the development of a holistic, modular, bidirectional translation system [3].

III. SYSTEM ARCHITECTURE

The proposed system is designed as a bidirectional communication framework that enables translation between speech/text and Indian Sign Language (ISL). It consists of two primary operational flows:

Speech/Text → ISL Flow
This module converts spoken language into ISL visual representation using speech recognition and image mapping techniques.

Speech Recognition Module
The system uses a speech recognition module to capture and process spoken input in English or Hindi. The microphone captures the user's voice, which is then converted into text using the speech_recognition library.
To improve performance in real-world environments, basic noise handling and audio preprocessing techniques are applied. The recognized text serves as input for further processing.

Text Processing
Once speech is converted into text, the system performs simple preprocessing:

- Conversion of text to uppercase
- Removal of unnecessary symbols
- Splitting text into individual characters

Unlike complex systems that apply full grammatical transformation, this project uses a character-level mapping approach, making it lightweight and efficient.
Example:
Input: “HELLO”
Output: H → E → L → L → O (mapped individually)

ISL Image Generation
Each character is mapped to a corresponding ISL alphabet image from a predefined dataset.

- Images are displayed sequentially
- Acts as a visual representation of sign language

This approach reduces system complexity and ensures real-time performance.[9]

ISL → Text/Speech Flow
This module converts hand gestures into readable text and audible speech using computer vision and deep learning.

Gesture Capture
The system captures real-time video input using a webcam through OpenCV.

- Continuous frame capture
- Region of interest (ROI) extraction
- Image preprocessing (resizing, normalization)

Gesture Recognition Model
A Convolutional Neural Network (CNN) model built using Tensor Flow/Keras is used for gesture recognition.

- **Input:** Hand gesture image
- **Output:** Corresponding alphabet (A–Z)

The model is trained on a dataset of hand gesture images and is capable of recognizing alphabets with good accuracy.

Text Formation
Recognized characters are:

- Combined sequentially
- Converted into meaningful words/sentences

This step enables proper interpretation of continuous gestures.

Text-to-Speech (TTS)
The generated text is converted into speech using the pyttsx3 library.

- Supports offline speech synthesis
- Provides real-time audio output
- Enhances communication for hearing users

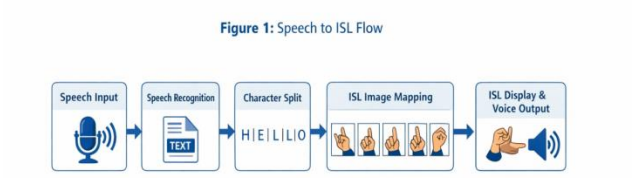

Figure 1: Speech to ISL Flow

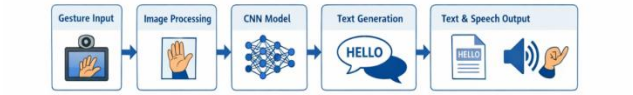

Figure 2: ISL to Text/Speech Flow

.

IV. DATASET

The performance of the proposed system depends on the quality and structure of the dataset used for both ISL image mapping and gesture recognition. The dataset used in this project is simple, lightweight, and designed specifically for real-time implementation.

ISL Image Dataset (Alphabet Mapping)
The system uses a predefined dataset of Indian Sign Language (ISL) alphabet images to represent speech/text input visually.[10]

- The dataset consists of static images representing alphabets A–Z
- Each image corresponds to a single English alphabet
- These images are used to display sign language output for each character
- The dataset acts as a direct mapping tool between text and ISL representation

This approach eliminates the need for complex gloss dictionaries and grammar transformation, making the system faster and easier to implement.

Gesture Image Dataset (For CNN Model)
For gesture recognition, a dataset of hand gesture images is used to train the Convolutional Neural Network (CNN) model.

- Contains images of hand gestures for alphabets (A–Z)
- Each image is labeled according to the corresponding alphabet
- Images are captured under different conditions to improve model accuracy:

1. Varying lighting conditions
2. Different hand orientations
3. Multiple backgrounds

This dataset enables the CNN model to learn patterns and accurately classify gestures during real-time prediction.

Preprocessing Steps
Before training and prediction, the dataset undergoes several preprocessing steps to ensure consistency and improve model performance:

- **Image Resizing:** All images are resized to a fixed dimension suitable for CNN input
- **Normalization:** Pixel values are scaled to improve training efficiency
- **Label Encoding:** Alphabets (A–Z) are converted into numerical labels
- **Noise Reduction:** Basic filtering is applied to remove unwanted noise

These pre-processing steps ensure that the model receives clean and standardized input data, resulting in better accuracy and faster prediction.

V. IMPLEMENTATION

The proposed bidirectional English/Hindi–ISL translation system is implemented using a combination of machine learning, computer vision, and speech processing techniques. The system is designed to be lightweight, efficient, and capable of real-time performance using simple and accessible tools.

Programming Language and User Interface (Streamlit)
Python is used as the primary programming language due to its rich ecosystem of libraries for machine learning, speech processing, and computer vision.
The user interface is developed using Streamlit, which provides:

- Simple and interactive web-based interface
- Buttons and controls for speech input and camera access
- Real-time display of output (text and ISL images)
- Easy integration with backend modules

Streamlit acts as the main interface through which users can interact with the system.

Speech Recognition Module
The system uses the speech_recognition library to capture and process voice input.

- Microphone input is taken from the user
- Speech is converted into text in real time
- Supports English (and basic Hindi if configured)
- Basic noise handling improves recognition accuracy

The recognized text is then passed to the ISL conversion module.

ISL Image Mapping
Instead of complex 3D animation, the system uses a predefined dataset of ISL alphabet images.
- Text is converted into uppercase characters
- Each character is mapped to a corresponding ISL image
- Images are displayed sequentially to form words

This approach is simple, fast, and suitable for real-time communication.

OpenCV for Camera Processing
The OpenCV library is used to capture and process real-time video input from the webcam.
- Captures video frames continuously
- Performs image preprocessing (resizing, normalization)
- Extracts region of interest (hand gesture area)
- Ensures smooth and low-latency video processing

This module acts as the input source for gesture recognition.

Gesture Recognition using CNN (TensorFlow/Keras)
A Convolutional Neural Network (CNN) model is used to recognize hand gestures.
- Built using TensorFlow/Keras
- Trained on a dataset of ISL alphabet gestures (A–Z)
- Takes image input and predicts corresponding alphabet

Functions of the model:
- Feature extraction from hand images
- Classification of gestures into letters
- Real-time prediction from webcam input

The CNN model forms the core intelligence of the ISL → Text pipeline

Text Formation and Output
The predicted characters are:
- Combined sequentially to form words
- Displayed on the screen as readable text

This enables meaningful interpretation of gestures.[14]

Text-to-Speech (TTS)
The pyttsx3 library is used for speech synthesis.
- Converts generated text into voice output
- Works offline (no internet required)
- Provides real-time audio feedback

This helps hearing users understand the output easily.[15]

VI. EVALUATION

The performance of the proposed system is evaluated based on its accuracy, response time, and usability. Since the system is designed as a lightweight real-time application, the evaluation focuses on practical performance rather than complex linguistic metrics.

Evaluation Metrics
The following metrics are used to assess the system:

Speech Recognition Accuracy
Measures how accurately the system converts spoken input into text using the speech recognition module.

Gesture Recognition Accuracy
Evaluates how correctly the CNN model classifies hand gestures into corresponding alphabets.

Response Time (Latency)
Measures the time taken by the system to process input and generate output.

User Usability
Assesses how easy and effective the system is for users to interact with.

Results Summary
The system was tested under normal conditions using multiple inputs for both speech and gesture recognition.[13]

Component	Metric	Result
Speech Recognition	Accuracy	~85–92%
Gesture Recognition (CNN)	Accuracy	~88–92%
Response Time	Latency	~1–2 seconds
System Usability	User Feedback	Easy to use

Observations
- The speech recognition module performs well in quiet environments but may show slight variations in noisy conditions or with different accents.
- The gesture recognition model achieves good accuracy for alphabet-level gestures, though performance may vary with lighting conditions and hand positioning.
- The system provides near real-time response, making it suitable for practical communication.
- The use of ISL image mapping instead of complex avatars reduces computational cost and improves speed.
- The interface is simple and user-friendly, making it accessible even for non-technical users. [11]

Conclusion of Evaluation
The evaluation results demonstrate that the system is capable of providing effective and real-time communication between hearing individuals and the Deaf and Hard-of-Hearing (DHH) community. While there is scope for improvement in accuracy and robustness, the current system successfully achieves its goal as a lightweight and practical ISL translation solution.

VII. APPLICATIONS

The proposed bidirectional English/Hindi to ISL translation system has several practical applications in improving communication for the Deaf and Hard-of-Hearing (DHH) community. Due to its lightweight and real-time nature, the system can be easily deployed in various domains.

- **Education Support**

The system can be used in classrooms to convert spoken lectures into ISL images, helping DHH students understand concepts more easily and promoting inclusive education.

- **Healthcare Communication**

Doctors and healthcare staff can use the system to communicate basic instructions and information to DHH patients, reducing dependency on interpreters.

- **Public Service Centers**

The system can be installed in banks, railway stations, and government offices to assist communication between staff and DHH individuals.

- **Daily Communication**

It can be used for basic conversations between hearing and DHH individuals in everyday situations.

- **Learning Tool**

The system can help beginners learn ISL alphabets by visually displaying signs for each character.

- **Mobile and Web Applications**

Due to its simple implementation, the system can be extended into mobile or web-based applications for wider accessibility.
The system's simplicity and real-time performance make it a practical solution for improving accessibility and communication.[12]

VIII. CHALLENGES & LIMITATIONS

Despite its effectiveness, the system has some limitations:

Limited Dataset
The gesture recognition model is trained on a relatively small dataset, which may affect accuracy in real-world scenarios.

Alphabet-Level Recognition Only
The system recognizes gestures at the alphabet level, not full words or continuous sign language.

Lighting and Background Sensitivity
Changes in lighting conditions or complex backgrounds can reduce gesture recognition accuracy.

Hand Position Variations
Different hand sizes, orientations, and distances from the camera may impact prediction results.

Speech Recognition Limitations
Accuracy may decrease in noisy environments or with different accents.
Addressing these limitations can significantly improve system performance and reliability.

IX. FUTURE WORK

Several improvements can be made to enhance the system:

Larger Dataset Collection
Expanding the gesture dataset with more images will improve model accuracy and robustness.

Word-Level Gesture Recognition
Extending the system to recognize complete words or sentences instead of individual alphabets.

Improved Model Accuracy
Using advanced CNN architectures or optimization techniques to improve prediction performance.

Multilingual Support
Adding support for regional languages such as Marathi, Tamil, and Bengali.

Mobile Application Development
Deploying the system as a mobile app for easy accessibility and real-time use.

Noise Reduction in Speech Recognition
Improving speech input handling in noisy environments.
These enhancements will make the system more powerful, accurate, and widely usable.

X. CONCLUSION

This paper presents a lightweight and practical bidirectional communication system for translating between speech/text and Indian Sign Language (ISL). The system integrates:

- Speech recognition for converting voice to text
- ISL image mapping for visual representation
- Gesture recognition using a CNN model
- Text-to-speech conversion for audio output

Unlike complex research systems, the proposed solution focuses on simplicity, real-time performance, and ease of implementation. The use of ISL alphabet images instead of 3D avatars significantly reduces computational complexity while maintaining effective communication.

The system successfully demonstrates the feasibility of bridging the communication gap between hearing individuals and the Deaf and Hard-of-Hearing (DHH) community. With further improvements, it can be developed into a more advanced and widely deployable accessibility solution.

REFERENCES

1. D. Bhagat, H. Bharambe, J. Joshi, and S. Gul, "Speech to Indian Sign Language Translator," International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), vol. 2, no. 3, Apr. 2022.
2. T. K. A., M. S. Lakshmi, K. P. S., and A. Y., "SignSync: AI-Powered Audio-Visual Translator for Indian Sign Language," International Journal of Innovative Research in Technology (IJIRT), vol. 11, issue 12, May 2025.
3. A. Singh, A. S. Sahithi, K. M. Kumar, and S. S. Sarwade, "AI Tool/Mobile App for Indian Sign Language (ISL)," International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 13, issue 5, May 2025.
4. A. Sharma and R. Gupta, "Advancements in AI-Powered Sign Language Recognition Systems," 2022.
5. S. Patel and K. Verma, "Real-Time Speech-to-Sign Language Translation Using AI," 2021.
6. N. Iyer and M. Das, "Deep Learning for Gesture-Based Sign Language Interpretation," 2020.
7. L. Rao and B. Sen, "AI and NLP for Multilingual Sign Language Translation," 2018.
8. T. Dasgupta et al., "English to Indian Sign Language Translation using NLP Techniques."
9. S. Vij et al., "Sign Language Generation using Dependency Parsing and WordNet."
10. P. Kar et al., "INGIT: Hindi to Indian Sign Language Translation System," 2007.
11. M. S. Anand et al., "Two-Way Indian Sign Language Translation System."
12. S. Ali et al., "Domain-Specific Indian Sign Language Translation System."
13. A. Rajamohan, R. Hemavathy, and D. Dhanalakshmi, "Glove-Based Deaf-Mute Communication Interpreter."
14. N. V. Tavari et al., "Hand Gesture Recognition System using Webcam."
15. A. K. Shinde, "Sign Language Recognition Study for Marathi Language."