International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 3, May-Jun-2026, ISSN (Online): 2395-566X

Vaani2mudra – Indian Sign Language (Isl) Translation For Deaf People

Khushboo Lokhande[1], Samruddhi Mahajan[2], Sayali Pawar[3], Janhavi Wankhede[4], Vijay More[5]

[1,2,3,4]Student, Dept. of Computer Engineering MET's Institute of Engineering Nashik, India

[5]Associate Professor, Dept. of Computer Engineering MET's Institute of Engineering Nashik, India

Abstract—Vaani2Mudra is an online assistive communication platform that converts spoken or written language into gestures representing Indian Sign Language (ISL). The platform utilizes a compact speech recognition model to process voice input and employs natural language processing methods to restructure spoken content into a format compatible with ISL. Through a rule-based linguistic framework, the system eliminates redundant grammatical elements and standardizes text for gesture mapping. For multilingual functionality, Marathi language input is first converted to English before further processing. The output is presented through a series of pre-established ISL gesture visuals shown on a web-based interface. This system prioritizes ease of use, instantaneous processing, and user accessibility, positioning it as an effective tool for learning environments and assistive communication applications.

***Keywords*—Indian Sign Language, Speech-to-Sign Conversion, Whisper Technology, Rule-Based Natural Language Processing, Assistive Communication Tools**

I. INTRODUCTION

Communication represents a fundamental human entitlement that encompasses educational access, social engagement, and independent living. Data from the World Health Organization indicates that over 466 million individuals across the globe experience debilitating hearing impairment. Within India, the National Sample Survey Office reports approximately 18 million people living with hearing or speech disabilities. For this population, Indian Sign Language serves as the most organic and effective communication medium.

However, despite ISL's vital importance to these individuals, insufficient public awareness results in substantial communication barriers. Although technological innovations exist for American Sign Language (ASL) and British Sign Language (BSL), comprehensive, expandable, and openly accessible systems for ISL translation remain in their early developmental stages within India.

This is done through an online real-time system that translates spoken language into ISL gestures using animated 3D avatars. The system integrates with the Google Speech-to Text API for accurate speech transcription, Natural Language Processing NLP for syntactic and semantic transformation into

ISL-compatible structures, and a curated ISL gesture library for efficient gesture mapping. Combining speech recognition, linguistic processing and gesture animation, Vaani2Mudra provides a seamless, efficient, inclusive, culturally relevant communication platform. This research-driven system is designed to improve one-to-one communication between hearing and deaf people. especially in classrooms, hospitals, public service places, and everyday conversations. As it continues to evolve, A more diverse vocabulary of gesture, advanced machine learning models, Vaani2Mudra has the potential to evolve into A full-featured assistive communication device used by millions of users across India.

This work compiles and discusses the existing research work available on the translation of the speech to Indian Sign Language. The proposed system and flow have been incorporated to provide a concept and prototype development based on the existing studies carried out.

A. Motivation And Objectives

To bridge the communication gap between hearing and speech-impaired individuals in India, the *Vaani2Mudra*

system focuses on Realtime ISL translation and accessibility. Objectives of work are:

1. Developed a Realtime translation system that converts spoken language into Indian Sign Language (ISL) gestures.

2. Utilize NLP and machine learning techniques for accurate sentence segmentation and gesture mapping.

3. Design an intuitive 3D animated avatar to visually represent ISL signs, ensuring cultural and linguistic accuracy.

4. facilitate inclusivity in education, healthcare and professional communication for deaf individual

5. Ensure that System is Lightweight, Scalable and Deployable on web and Mobile Platforms

II. LITERATURE REVIEW

Sign language translation system development has emerged as a prominent research domain in recent years, driven by advancements in artificial intelligence, natural language processing, and computer vision technologies. The majority of existing scholarly work focuses on American Sign Language (ASL) and British Sign Language (BSL). In contrast, Indian Sign Language (ISL) remains a relatively underexplored field. Recent research work carried out in this area states the need for linguistically adaptive systems for various sign languages with real-time constraints [7], [9], [11]. This section provides insight into the main contributions in the area of recognition of sign language systems, speech-to-sign conversion systems, avatar animation systems, and ISL systems.

A. Sign Language Recognition And Translation

Initially, research work for sign language recognition used image processing techniques and sensor techniques like glove-based systems which allowed for accurate gesture recognition but still had limitations in terms of usability and scalability [7]. But with the development of deep learning techniques, vision-based approaches with convolution networks and transformer networks for ASR models allowed for enhanced accuracy for vision-based models [4]. Zhou et al. introduced a model for the translation of sign languages with no need for glosses using the power of pre-trained models for visual languages [1]. Models based on the transformer allowed for advancements in multilingual models for sign language translation with better handling of gestures through temporal dependencies [9], [10]. Yet again, these models are primarily tuned with ASL/BSL corpora, pointing out the need for localized models for ISL [6].

B. Speech-To-Sign Conversion

Speech to sign translation constitutes the amalgamation of Automatic Speech Recognition and Natural Language Processing. Multimodal deep learning networks are used for direct speech to sign translation with quite encouraging results for aligning inputs with signs [11]. Large datasets like Sign Bank pave ways for multilingual translation in ISL by exploiting language models and their annotations [3]. However, current translation systems neglecting grammatical variations in spoken languages and ISL, which is structurally different from English and Hindi, are not effective. Current research signifies the importance of incorporating grammatical or rule-based mapping approaches for effective ISL translation, as discussed in [2], [4].

C. Avatar-Based Sign Language Animation

Avatar-based sign language rendering has appeared as a promising substitution for video-based rendering because of the scalability and storage advantages of the first solution over the second solution. Studies on sign language animation avatars have shown enhanced understandability and engagement of deaf users [5]. Neural machine translation with a generative network approach has been utilized to advocate realistic sign animation based on text data input [8]. Progressive transformer models continue to advance continuous motion and expressiveness of sign gestures in sign language production systems [10]. On the other hand, the challenge of synchronized animation of the input speech in real time continues to be a problem in developing an ISL animation system [5], [8].

D. Indian Sign Language (Isl) Systems

Although research work in Indian Sign Language has recently gained momentum, very little work has explored its integration with speech recognition. The Indian Sign Language Research and Training Centre (ISLRTC) has been involved in building lexical datasets and gesture resources. Approaches employing deep learning techniques in MediaPipe and hybrid CNN-LSTM models have demonstrated success in the Realtime recognition of Indian Sign Language gestures with high levels of accuracy [12].

Several scholarly works have highlighted the significance of regional linguistic variations, challenges related to limited datasets, and the need for comprehensive integration within speech processing frameworks [6], [13]. Previous research has addressed isolated gesture recognition systems and text-to-gesture conversion; however, speech recognition systems that integrate with ISL gesture generation remain largely unexplored. Recent studies have also explored deep learning–based frameworks for ISL recognition aimed at assisting deaf individuals [13].
This research gap has led to the requirement of developing tools like Vaani2Mudra.

III. RESEARCH GAPS

Based on the reviewed literature, several key research gaps have been identified:

Lack of ISL-Specific Datasets and Benchmarks: Existing models largely depend on ASL datasets ,with limited availability of large scale ,annotated Indian Sign Language(ISL)datasets for training and evaluation

Absence of Grammar-Aware ISL Translation Models: Current speech-to-sign translation systems follow English/Hindi grammatical structures However possesses distinct syntactic and morphological features that are not addressed by existing frameworks

1. Limited End-to-End Integration of ASR, NLP, and Gesture Generation: Most approaches treat automatic speech recognition, natural language processing, and sign animation as separate modules. There is lack of

unified ,real time pipeline tailored specifically for ISL

1. Inadequate Realistic and Real-Time Avatar Animation: Avatar based ISL System face challenges such as unnatural gesture transitions and limited synchronization

with continuous speech

1. Minimal Consideration of Regional Variations in ISL: ISL exhibits regional and dialectal differences across India, yet most existing system fail to incorporate these variations, reducing real-world applicability.

IV. SYSTEM ARCHITECTURE

The Vaani2Mudra system is developed around a modular * client–server architecture to allow for the evaluation of the spoken or the textual information into the gestures of Indian Sign Language. The This system can be divided into five different modules: Input Layer, Speech Processing Layer, NLP Processing Layer, Translation Layer, and Gesture Mapping and Visualization Layer. Each module is for a specific task in the whole translation process.

Input Layer: The input layer essentially acts as a user interface between user and the system. It receives audio either through browser-based audio recording or prerecorded media content. Audio recordings. In addition to audio input, the system also supports a manual text entry mode The level of complexity exhibited indicates that this software has to interact with the system in a different manner based on their requirements and available resources.

1. *Speech Processing Layer:* This is responsible for converting the input speech or audio into text format. The received audio is passed through the Whisper Tiny model, which is chosen based on its low latency and capability to support multiple languages. Before starting to transcribe, FFmpeg is employed to decode and pre-process the audio data so that it is an optimal manner for accurate speech recognition.

2. *NLP Processing Layer*: This layer performs linguistic analysis and text Normalization. The transcribed text or manually entered text is processed using SpaCy for Tokenization and Lemmatization. These processes aid in breaking the text into viable segments and reducing words to their base forms. To make the sentence structure compatible with Indian Sign Language grammar, rule-based NLP processing is applied.

3. *Translation Layer:* To accommodate multilingual users, the information to be Translation Layer is incorporated into the system. If the detected input language is Marathi, translation is performed using API-based translation services to translate the text into English. This translation process ensures a uniform processing flow, as well as facilitating the generation of accurate gesture mapping irrespective of the medium the original input language

4. *Gesture Mapping and Visualization Layer*: Within this layer, the tokens obtained as the result of NLP operations will be mapped to the pre-specified ISL gestures through the gesture dictionary. Each of these gestures will be presented as an image. These images will then be presented on the web interface through the HTML tags of images. With this visualization, the results obtained can be easily comprehended by the users.

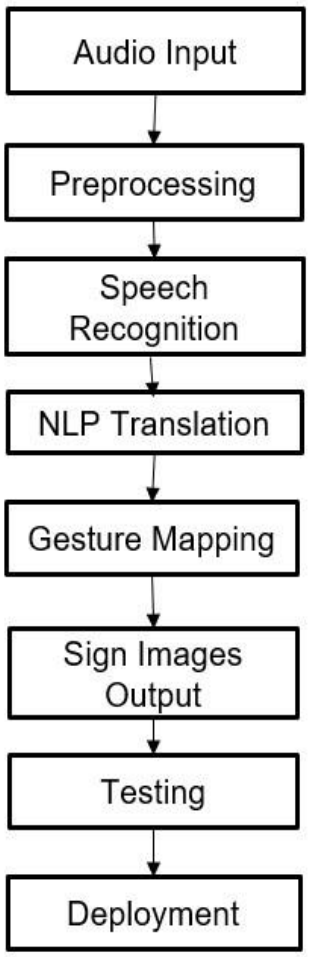


Fig. 1. Overall System Architecture of the Vaani2Mudra speech-to-ISL translation system

V. METHODOLOGY

The Vaani2Mudra system follows a multi-stage processing pipeline to convert spoken or textual input into Indian Sign Language (ISL) gestures. It consists of speech recognition, natural language processing, translation handling, gesture mapping and output visualizations. The overall workflow of the system is illustrated in Fig. 2.

A. *Speech Recognition*

The system performs speech-to-text conversion using the Whisper Tiny model. The Whisper Tiny model is selected due to its low latency and support for multiple languages, including English, Hindi, and Marathi. The output of this stage is textual representation of spoken input.

B. Text Processing And Nlp Transformation

The transcribed text or manually entered text undergoes Natural Language Processing to prepare it for ISL translation. SpaCy is used for tokenization and lemmatization, enabling the text to be broken into meaningful units and reduced to base word forms. In addition to this rule-based NLP techniques are applied to transform the sentence structure according to Indian Sign Language Grammar. These rules eliminate unnecessary linguistic elements such as stop word, pronouns, and connectors resulting in simplified ISL-complaint sentence structure

C. Translation Handling

To support multilingual input a translation mechanism is incorporated into system. If the detected input language is Marathi, API based translation services are use to translate text into English This step is performed before further NLP processing to maintain a uniform processing pipeline and ensure accurate gesture mapping across different languages.

D. Gesture Mapping

After Linguistic processing, each process token is mapped to its corresponding ISL gesture using a predefined gesture Dictionary. The gestures are stored as static images representing individual ISL signs. This mapping process ensures that each meaningful token is associated with a visually interpretable gesture.

E. Output Visualization

In the final stage, the mapped ISL gesture images are displayed sequentially on the web interface The overall methodology ensures lightweight, efficient and practical speech to ISL Translation system suitable for assistive communication and educational applications

F. *Backend Architecture*

The system is built on a lightweight and scalable stack consisting of:

Frontend: HTML, CSS, and JavaScript

Backend: Python 3.10 with FastAPI framework

Server: Uvicorn ASGI server

Workflow

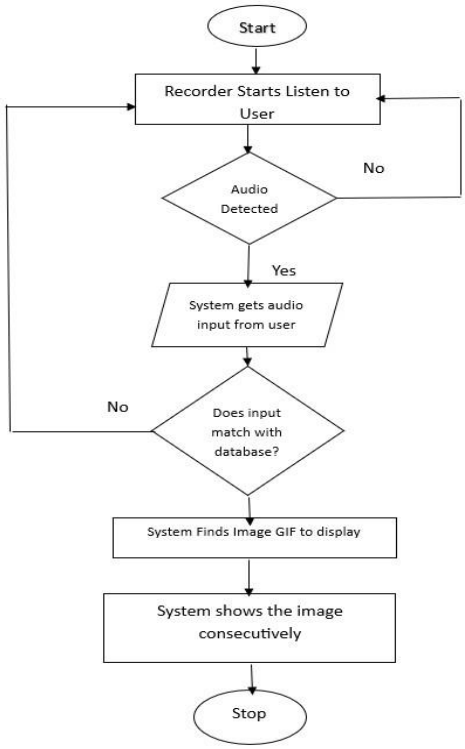


Fig. 2. Workflow of the Speech-to-ISL Translation Process.

VI. MATHEMATICAL MODEL

The functional behavior of the Vaani2Mudra system can be mathematically represented as a tuple:

S={I,F,O}

Where:

• I: Set of Inputs →{Ia,It,Disl}

• Ia: Audio input (speech signal)

• It: Text input (manually entered)

• Disl: ISL dataset containing gesture mappings

• F: Set of Functions →{F1,F2,F3}

• F1: Speech-to-Text Conversion

• F2: Text Preprocessing and NLP-based translation

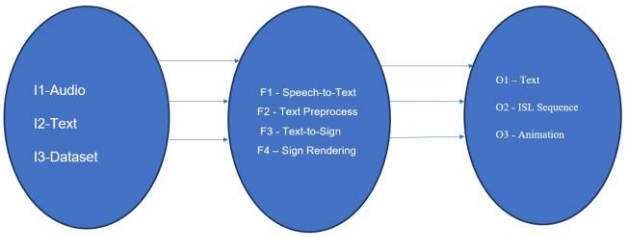
• F3: Text-to-Sign Gesture Mapping

• O: Set of Outputs →{O1,O2}

• O1: Recognized text from speech

• O2: ISL gesture sequence for visualization

Venn diagram of the system is shown in Fig.



VII. CHALLENGES AND RESEARCH GAPS

Although the proposed system provides an efficient translation workflow, certain issues remain open for future enhancements:

1. Dataset Scarcity: Limited availability of large-scale, annotated ISL datasets affects model generalization.

2. Continuous Signing: Generating smooth transitions for sentence-level signing remains challenging.

3. Performance on Edge Devices: Real-time translation on mobile hardware introduces latency constraints.

4. Cultural Localization: Variations in regional gestures require adaptive, context-aware learning.

VIII. SYSTEM EVALUATION

The Vaani2Mudra system was evaluated based on three parameters: Speech Recognition performance, effectiveness of gesture mapping, and system response time. The major observations obtained during system testing are summarized below.

A. Speech-To-Text Performance

The system demonstrated reliable speech-to-text conversion when the audio input was clear using the Whisper Tiny model. The Model was able to transcribe speech in English ,Hindi and Marathi with Acceptable accuracy for assistive communication scenarios.

B. *Gesture Mapping Effectiveness*

The Gesture mapping module successfully mapped on majority of processed tokens to their corresponding ISL gestures using predefined gesture dictionary.The effectiveness of the gesture mapping process depends on the availability of gesture images in the dictionary.

IX. SYSTEM RESPONSE TIME

The overall processing pipeline, which includes speech input handling, text processing, NLP transformation, and gesture mapping, returned responses in under a few seconds on average. This performance enables near real-time translation suitable for interactive use.

X. POTENTIAL IMPROVEMENTS

A number of enhancements have been proposed to further improve the system's accuracy, scalability, and linguistic coverage:

1. Vocabulary Expansion: Extend the gesture library coverage to more than 1,000 ISL signs, including phrases, classifiers, and region-specific variations.

2. Advanced Translation Models: Integrate transformer-based encoder–decoder architectures to better capture ISL grammar, gloss ordering, and contextual meaning.

3. Continuous Gesture Generation: Improve animation smoothness for multi-word and continuous gestures using advanced motion-interpolation algorithms.

XI. CONCLUSION

Vaani2Mudra represents a significant advancement toward inclusive communication technology in India—integrating speech recognition, natural language processing, and gesture animation to provide real-time speech-to-sign translation. The Deep learning integrated with skeleton-based recognition Systems may offer even further improvements in the accuracy of Indian Sign Language translation.

The proposed system is practically usable and provides a strong foundation for future expansion in the field of assistive communication technologies.

REFERENCES

1. B. Zhou, Z. Chen, A. Clapes, J. Wan, Y. Liang, S. Escalera, Z. Lei, and D. Zhang, "Gloss-Free Sign Language Translation: Improving from Visual-Language Pretraining," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023.

2. S. Sabharwal and P. Singla, "Translation of Indian Sign Language to Text: A Comprehensive Review," International Journal of Intelligent Systems and Applications in Engineering, vol. 12, no. 14s, pp. 309–319, 2024.

3. A. Moryossef and Z. Jiang, "SignBank+: Preparing a Multilingual Sign Language Dataset for Machine Translation Using Large Language Models," arXiv preprint arXiv:2309.11566, 2023.

4. "Machine Translation from Signed to Spoken Languages: State of the Art and Challenges," Universal Access in the Information Society, vol. 23, pp. 1305–1331, 2023.

5. M. Al-Ahmad, M. Al-Qudah, and O. Al-Jarrah, "A Real-Time Arabic Avatar for the Deaf–Mute Community Using Attention Mechanism," Neural Computing and Applications, vol. 35, pp. 21709–21723, 2023.

6. A. Kumar and N. Sharma, "Deep Learning Approaches for Indian Sign Language Recognition: Challenges and Future Directions," arXiv preprint arXiv:2303.11223, 2023.

7. D. Bragg, C. Huenerfauth, M. Koller, and T. Starner, "Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective," Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility, 2019.

8. S. Stoll, N. C. Camgoz, S. Hadfield, and R. Bowden, “Text2Sign: Towards Sign Language Production Using Neural Machine Translation and Generative Adversarial Networks,” International Journal of Computer Vision, vol. 128, pp. 891–908, 2020.

9. N. Camgoz, O. Koller, S. Hadfield, and R. Bowden, “Multilingual Sign Language Translation Using Transformers,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12089–12098, 2020.

10. B. Saunders, N. C. Camgoz, and R. Bowden, “Progressive Transformers for End-to-End Sign Language Production,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11836–11845, 2022.

11. Y. Li, T. Huang, and X. Han, “End-to-End Speech-to-Sign Translation Using Multimodal Neural Architectures,” IEEE Transactions on Multimedia, vol. 26, pp. 740–753, 2024.

12. B. Ojha and A. Mittal, “Indian Sign Language Recognition Using Media Pipe and Deep Learning for Real-Time Applications,” Journal of Imaging, vol. 10, no. 3, pp. 1–17, 2024.

13. K.A.Chandan and V.B More “Innovation in Indian Sign Language Recognition for Deaf and Dumb Individuals Using Deep Learning. In: Guru, D.S., Rajurkar, A.M., Vinay Kumar, N., Gudivada, V.N. (eds) Data Analytics and Learning. ICDAL 2024. Lecture Notes in Networks and Systems, vol 1540. Springer, Singapore,2026