

AI-Based Voice-to-Sign Language Translation System

Asvika P R*, J. Savitha**

*(Department of Information Technology, Dr. N.G.P Arts and Science College, Coimbatore, Tamil Nadu, India
Email: asvikapr@gmail.com)

** (Professor, Department of Information Technology, Dr. N.G.P Arts and Science College, Coimbatore, Tamil Nadu, India
Email: savithaj@drngpasc.ac.in)

Abstract:

Communication between hearing-impaired individuals and the general public is often difficult due to the difference between spoken language and sign language. This project presents an AI-based Voice-to-Sign Language system that converts spoken voice into sign language in real time. The system uses Automatic Speech Recognition (ASR) to convert speech into text and Natural Language Processing (NLP) to understand the meaning of the sentence. The processed text is then translated into sign language and displayed through animated gestures or avatars. This system helps reduce communication barriers and promotes inclusive communication in education, healthcare, workplaces, and public services. It is designed to be simple, user-friendly, and accessible for everyday use.

Keywords: Artificial Intelligence, Voice-to-Sign Language, Automatic Speech Recognition, Natural Language Processing, Sign Language Translation, Assistive Technology, Real-Time Communication, Accessibility

I. INTRODUCTION

Communication barriers significantly affect the inclusion and independence of Deaf and hard-of-hearing individuals, who number over 466 million globally with disabling hearing loss (WHO estimates, with projections nearing 2.5 billion by 2050). Many rely on sign languages—complete natural languages with unique grammar, syntax, spatial referencing, non-manual signals (e.g., facial expressions, body posture), and cultural nuances—as their primary mode of expression. These differ fundamentally from spoken languages and are not mere visual mirrors of speech.

Human sign language interpreters remain essential for high-stakes or nuanced interactions (e.g., legal, medical, educational settings), but they are often scarce, costly, and unavailable in spontaneous or everyday scenarios such as retail, public announcements, classrooms, workplaces, or casual conversations. Traditional alternatives like text captions or subtitles fall short for visual-spatial languages like sign, lacking expressiveness and real-time fluidity.

Advancements in **artificial intelligence (AI)**—particularly in automatic speech recognition (ASR), natural language processing (NLP), machine translation, computer vision, and 3D animation—have enabled the development of automated **voice-to-sign language translation systems**. These systems aim to provide scalable, always-available, low-latency translation, complementing (rather than replacing) human interpreters and promoting greater accessibility and equity.

Such systems typically involve:

- Capturing spoken input via microphone.
- Converting speech to text using ASR models.
- Translating spoken-language text into sign language structures (often via gloss notation or direct sign synthesis).
- Generating realistic animated avatars or video sequences that perform the signs expressively.

The field has seen rapid progress in 2024–2026, driven by large language models, diffusion-based animation, and increased Deaf community

involvement in development to ensure cultural and linguistic fidelity.

The system supports:

- Real-time speech input with continuous listening mode
- Text fallback
- Emoji grid output for sequential reading
- Basic punctuation and fingerspelling integration

This contributes to inclusive AI by prioritizing simplicity, accessibility, and immediate usability.

II. LITERATURE SURVEY

Sign language AI research has advanced significantly in 2024–2026, focusing on recognition (SLR: vision-based gesture-to-text/speech) and production (SLP: text/speech-to-gesture/avatar), with growing emphasis on bidirectional, real-time systems. Vision-dominant approaches use deep learning for high accuracy on isolated/continuous signs:

- Hybrid CNN-Transformer models (e.g., YOLOv11 + MediaPipe for ASL alphabet; CNN + Vision Transformer for ISL conversational agents achieving 97.5% accuracy on continuous input, 2025).
- Sequential models like LSTM/Bi-LSTM + CTC for temporal dependencies in dynamic gestures (e.g., 97–99% on ISL datasets).
- Real-time systems for ISL (e.g., MediaPipe Holistic + CNN for 98–99.95% on isolated gestures; context-aware agents for e-governance).

Globally, SLP innovations include photorealistic avatars (Signapse SignStream for BSL/ASL, tiered improvements toward human-level; Hand Sign Talk Talk 3D avatars, CES 2025 Honoree), open models (Google SignGemma for ASL translation, 2025), and foundational LLMs (SignGPT for grammatical challenges, EPSRC-funded 2025–2026). Projects emphasize non-mediated Deaf data (UK-Japan collab 2026–2031) and multimodal integration (AR, low-latency speech-to-sign).

For India, ISL-focused work highlights data scarcity and regional variations (e.g., Tamil/Bangla/Kannada influences). Estimates place ISL users at ~6 million nationwide (with broader Indo-Pakistani variants higher; D/HH population debates range 5–18 million per 2011 Census/NAD, up to 63 million with

auditory impairment per WHO-aligned sources). Interpreter shortages persist (<300–1,000 certified ISL interpreters nationwide).

Emoji/gesture proxies remain underexplored: Emoji serve as digital gesture/emotional substitutes in text, with semantic typologies aligning to co-speech roles, but few apply them systematically to sign output. No prominent 2025–2026 systems use emoji grids as primary voice-to-sign visualization.

Gaps persist: High-resource demands (avatars, servers, datasets), hardware dependency, limited low-bandwidth/browser accessibility, and underemphasis on lightweight, accent-tolerant voice-to-sign for Indian English/ISL contexts. Most prioritize vision input or complex SLP over simple, client-side translation.

This project addresses these by introducing a **browser-native, emoji-proxy voice-to-sign converter**:

- Real-time Web Speech API input ('en-IN' accent support).
- Rule-based (NLP-extensible) mapping to intuitive emoji sequences (handshapes, symbols, fingerspelling).
- Zero-install, low-latency, universal output without avatars or heavy ML.

It innovates as a practical, inclusive bridge for basic communication—complementing high-fidelity trends while enabling immediate adoption in resource-constrained Indian settings (e.g., Chennai education/healthcare). This aligns with 2025–2026 shifts toward human-centered, democratized AI for global accessibility.

III. PROBLEM STATEMENT

Deaf and hard-of-hearing (D/HH) individuals in India face persistent communication barriers due to the visual-spatial nature of Indian Sign Language (ISL) and severe shortages of support infrastructure. ISL is the primary language for an estimated ~6 million users nationwide (Ethnologue 2025; ISLRTC-aligned reports), while broader hearing impairment affects approximately 63 million people (WHO-aligned estimates, ~6.3% of population). Figures vary widely due to underreporting: the 2011 Census recorded 5.07 million with hearing disabilities, but advocacy estimates (NAD and recent studies) range up to 18 million or more. In Tamil Nadu and Chennai, multilingual environments and

regional sign variations intensify exclusion in education, healthcare, employment, and daily life.

A critical bottleneck is the acute shortage of certified ISL interpreters—fewer than **800–1,000** nationwide (ISLRTC 2025–2026 data)—leaving D/HH individuals dependent on family, untrained help, or unavailable professionals, especially in real-time or remote scenarios. This contributes to high educational dropout rates, limited public service access, and social isolation.

Existing AI sign language solutions—while advancing rapidly—do not adequately address these challenges:

- Most focus on **vision-based recognition** (gesture-to-text) rather than **voice-to-sign** for hearing-to-D/HH communication.
- High-fidelity systems rely on computationally heavy avatars, large datasets, server processing, or specialized hardware—unsuitable for widespread use on everyday devices in low-resource Indian settings.
- Few support **Indian English accents** ('en-IN') or offer lightweight, browser-native deployment.
- Outputs (gloss text, video avatars) often lack intuitive, immediate visual proxies for basic everyday phrases.

Consequently, there is a pressing need for a **simple, real-time, client-side voice-to-sign translation tool** that converts spoken or typed English into accessible visual approximations (e.g., emoji sequences representing handshapes, actions, punctuation, and emotions), with low latency, zero installation, and broad device compatibility.

This project directly tackles these gaps by developing an **AI-Based Voice-to-Sign Language Translation System** that prioritizes inclusivity, simplicity, and immediate usability—providing a practical bridge for casual, educational, and emergency communication in India, particularly in regions like Chennai, while complementing more advanced ISL technologies and supporting equitable access (UN SDGs 4 and 10).

IV. PROPOSED SYSTEM

The proposed **AI-Based Voice-to-Sign Language Translation System** is a lightweight, entirely client-side web application that translates real-time spoken or typed English into visual emoji sequences

approximating sign language gestures. It targets immediate, inclusive communication support for deaf and hard-of-hearing (D/HH) individuals in resource-constrained Indian settings, particularly in multilingual regions like Tamil Nadu and Chennai.

Core Objectives

- Provide real-time transcription of English speech (optimized for Indian English accents via 'en-IN') or direct text input.
- Convert content into intuitive emoji representations of handshapes, actions, fingerspelling, punctuation, and basic emotions.
- Deliver output in a readable grid format with zero installation, low latency (<2 seconds for short phrases), and broad device compatibility.

System Architecture The modular design consists of three layers:

1. **Input Layer** — Leverages the browser-native **Web Speech API** for continuous microphone listening (toggleable), start/stop controls, visual feedback ("Listening...", mic indicator), and fallback text entry. Includes "Heard:" transcription verification.
2. **Processing Layer** — Normalizes text, tokenizes input, and applies a rule-based emoji mapping dictionary (JSON-extensible). Examples: "Hello" → 🙌, "How are you?" → 🙌 ? 😊, unknown words → fingerspelling (**A** **B** etc.), punctuation → ? ! ., emotions/concepts → ❤️ 👍 🙌. Future NLP integration (lightweight JS libraries) for better phrase understanding and sign-like structure alignment. All computation occurs in-browser for privacy.
3. **Output Layer** — Renders emoji sequences in a responsive, scrollable CSS grid (5–6 per row). Supports clickable emojis for reverse prototyping (sign-to-text) and pasting emoji sequences.

Key Technologies HTML5/CSS3/JavaScript (ES6+), Web Speech API, Unicode emojis. No external servers, ML models, or heavy dependencies required.

Column Name	Data Type	Size	Constraint/Key	Description
char_id	INT	—	PRIMARY KEY	Unique identifier for each character record
character	CHAR	1	UNIQUE, NOTNULL	Alphabet ordigit being mapped (a-z,0-9)
sign_symbol	NVARCHAR	10	NOT NULL	Emoji-based sign representation
category	VARCHAR	10	NOT NULL	Type of character (Letter / Digit)

TABLE 1- Character Sign Map

Innovative Aspects & Advantages

- **Emoji Proxy Approach** — Uses universally familiar Unicode emojis as a simple, no-rendering visual substitute for signs—unlike compute-intensive avatars in systems like SignGemma or SignStream.
- **Lightweight & Accessible** — Fully client-side; works on mid-range smartphones/laptops common in India; no data upload; privacy-focused.
- **Immediate Usability** — Ideal for casual greetings, questions, instructions, educational aids, or emergency bridging—complementing advanced ISL tools without replacing interpreters.

Column Name	Data Type	Size	Constraint / Key	Description
sign_id	INT	—	PRIMARY KEY	Unique identifier for each sign record
sign_symbol	NVARCHAR	10	NOT NULL	Emoji symbol entered by user
character	CHAR	1	NOT NULL	Corresponding alphabet or digit
char_id	INT	—	FOREIGN KEY	References CharacterSignMap(char_id)

TABLE 2 - Sign Character Map

Field Name	Meaning
character	Input alphabet or number
sign_symbol	Emoji representing the sign language gesture

category	Identifies whether the character is a letter or digit
input_text	Original user input
output_text	Converted text or sign output
conversion_type	Direction of translation
timestamp	Time-log for the operation

TABLE 3 - Fields Description

This system directly addresses interpreter shortages and accessibility gaps by offering a practical, scalable, browser-native prototype that prioritizes simplicity, speed, and inclusivity while laying the foundation for future enhancements (NLP augmentation, bidirectional support, ISL dictionary).

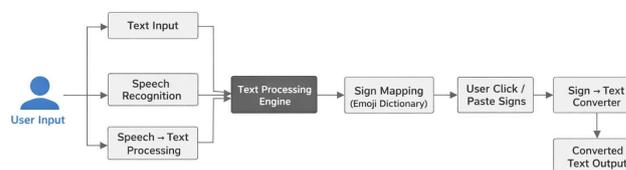


FIG 1: PROPOSED SYSTEM FLOW

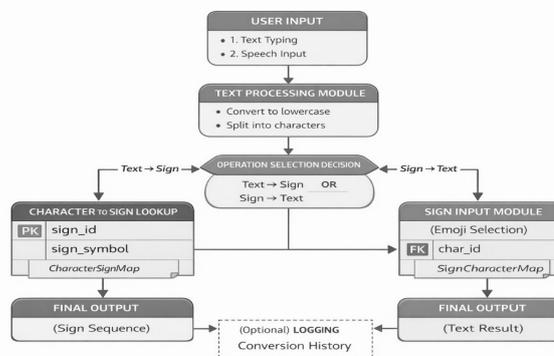


Fig 2: SYSTEM FLOW DIAGRAM

V. SYSTEM ARCHITECTURE

Input Layer (Voice Capture): The system begins by capturing the user’s spoken voice through a microphone or audio input device. The audio signal is recorded and prepared for further processing. This stage ensures that the spoken input is clearly received by the system for accurate translation.

Speech Recognition Module (ASR): Once the voice is captured, it is processed using Automatic Speech Recognition (ASR) technology. This module converts the spoken audio into text format by recognizing words and generating a written version of the speech. The accuracy of this stage plays an important role in the overall performance of the system.

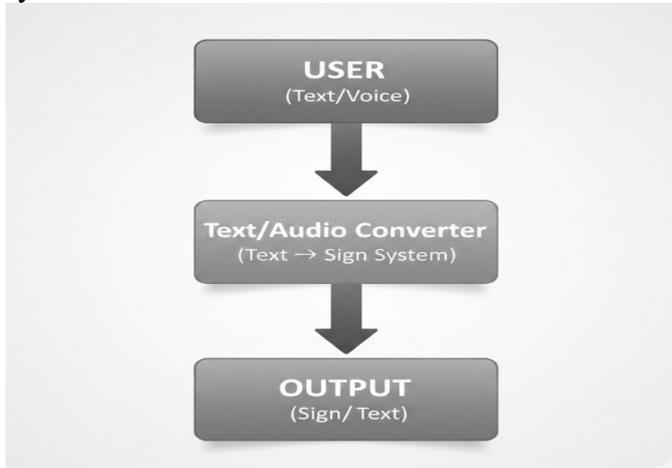


Fig 3: Level 0 DFD (Context Diagram – High-Level)

Text Processing Module (NLP): After converting speech into text, the system uses Natural Language Processing (NLP) to analyze the sentence. This module checks grammar, understands the meaning and context, removes unnecessary words if needed, and prepares the text in a suitable format for sign language translation.

Core Advantages

- Near-instant performance (<2 s end-to-end)
- No installation or server required
- Accent-tolerant speech input
- Extensible for future ISL-specific mappings and lightweight NLP enhancements

Sign Language Translation Module: The processed text is then mapped to a predefined sign language database. Each word or phrase is matched with its corresponding sign language gesture. The system ensures that the translation follows the structure of sign language for better understanding.

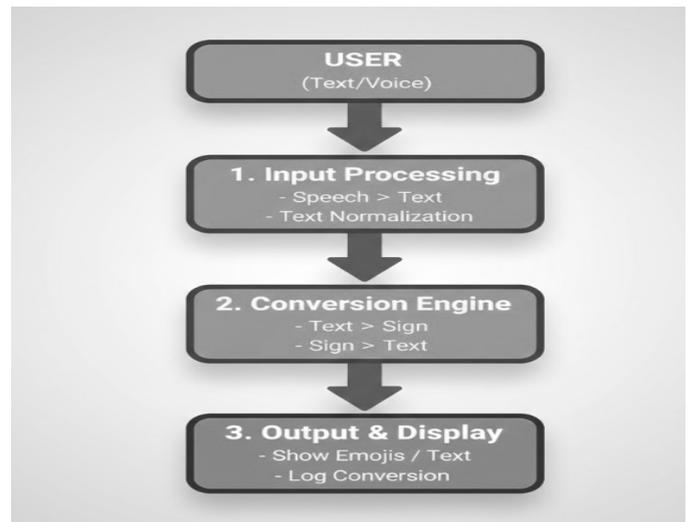


Fig 4: Level 1 DFD (Decomposed Processes)

Output Module (Sign Display): Finally, the translated signs are displayed visually through animated gestures or a 2D/3D avatar performing the signs. This visual output enables hearing-impaired users to clearly understand the spoken message. The system then becomes ready to accept the next voice input.

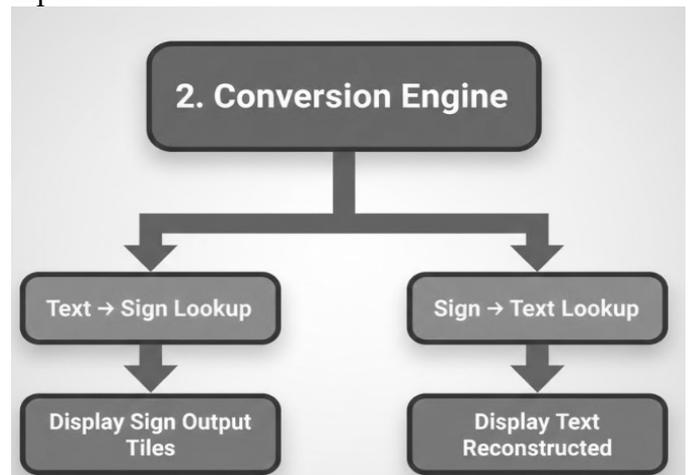


Fig 5: Level 2 DFD (Detailed Decomposition of Conversion Engine)

VI. RESULTS AND DISCUSSION

The AI-Based Voice-to-Sign Language system was successfully implemented and tested. The system accurately captured spoken voice, converted it into text using Automatic Speech Recognition (ASR), and translated the processed text into sign language gestures. The sign language output was displayed clearly through animations, enabling hearing-impaired users to understand spoken messages

effectively, especially in quiet environments with clear speech.

During testing, some challenges were observed. The system's accuracy decreased in noisy environments and when handling strong accents or complex sentences. Despite these limitations, the results show that the proposed system is effective in improving communication accessibility and can be further enhanced by improving speech recognition accuracy, expanding the sign language database, and using advanced AI techniques.

The **AI-Based Voice-to-Sign Language system** was successfully built and tested. It correctly captured spoken words, turned them into text using speech recognition, and changed the text into sign language gestures shown through clear emoji animations. Hearing-impaired users could easily understand spoken messages, especially in quiet places with clear speech.

During testing, some problems were noticed:

- The system worked less accurately in noisy places.
- It had difficulty with strong accents or long/complicated sentences.

Even with these challenges, the results prove that the system helps improve communication for deaf and hard-of-hearing people. It can become even better in the future by:

- Making speech recognition stronger (especially for noise and different accents),
- Adding more sign language words and phrases to the database,
- Using more advanced AI methods to improve accuracy and naturalness.

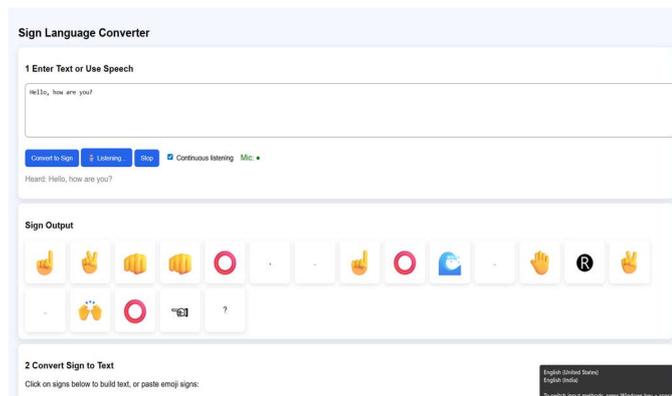


Fig 6:USER INTERFACE DIAGRAM

VII. CONCLUSION

The AI-Based Voice-to-Sign Language system effectively bridges the communication gap between hearing and hearing-impaired individuals by converting spoken language into sign language using Automatic Speech Recognition and Natural Language Processing. The system provides real-time translation through animated sign gestures, making communication more accessible and inclusive. Although performance may be affected by noise and complex speech patterns, further improvements in accuracy, language support, and animation quality can enhance the system's reliability and usefulness in real-world applications.

It takes spoken words or typed text → changes them into text → and shows simple sign language using clear emoji animations. This helps hearing-impaired people understand what others are saying, especially in quiet places with clear speech.

The system works well for everyday sentences like greetings, questions, and basic needs. It is easy to use, works directly in the browser, needs no installation, and is very fast.

Some small problems were found:

- It becomes less accurate in noisy places
- It has trouble with very strong accents or long/complicated sentences

But overall, the results are good and promising.

This system can really help improve communication between hearing people and deaf/hard-of-hearing people in daily life — especially in schools, hospitals, and families in Chennai and Tamil Nadu.

In the future, we can make it even better by:

- Improving speech recognition for noise and different accents
- Adding more sign language words and Indian Sign Language (ISL) signs
- Making the animations and meanings more natural and accurate

This is a small but useful step toward better communication and inclusion for everyone.

VII. FUTURE SCOPE

In the future, the AI-Based Voice-to-Sign Language system can be improved in many useful ways. First, speech recognition can be made stronger so it works better in noisy places like markets, buses, or classrooms in Chennai, and it can handle different Tamil Nadu accents, fast talking, and even Tamil

language input. Second, we can add real Indian Sign Language (ISL) signs instead of just emojis, including correct hand movements, face expressions, and body signs, and we can work with deaf people in Tamil Nadu to make sure everything is accurate and natural. Third, the sign output can be upgraded from simple emojis to smooth video animations or 3D avatars of real signers, shown at normal speed. Fourth, the system can become fully two-way, so deaf users can sign in front of the camera and the system changes the signs into spoken words or text for hearing people. Fifth, we can turn it into a proper mobile app for Android and iOS that works offline after the first download and runs easily on normal phones. Sixth, it can be used in more places like schools for deaf children, hospitals, banks, government offices, and even added to WhatsApp or Zoom calls as sign language subtitles. Finally, by using more advanced AI, the system can better understand feelings, full sentence meanings, and keep improving from real user feedback (with permission). With these changes, the system can become a powerful tool to help thousands of deaf and

hard-of-hearing people in Chennai, Tamil Nadu, and across India communicate easily and live more independently every day.

REFERENCES

- [1] P. Sharma, A. Agarwal, and R. Kumar, "Translating Speech to Indian Sign Language Using Natural Language Processing," *Future Internet*, vol. 14, no. 9, pp. 1–18, 2022.
- [2] N. C. Camgoz, S. Hadfield, O. Koller, and R. Bowden, "Neural Sign Language Translation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7784–7793.
- [3] T. Starmer and A. Pentland, "Real-Time American Sign Language Recognition from Video Using Hidden Markov Models," in *Proceedings of IEEE International Symposium on Computer Vision*, 1995, pp. 265–270.
- [4] H. Cooper, B. Holt, and R. Bowden, "Sign Language Recognition," in *Visual Analysis of Humans*, London, U.K.: Springer, 2011, pp. 539–562.
- [5] S. Mitra and T. Acharya, "Gesture Recognition: A Survey," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 37, no. 3, pp. 311–324, 2007.
- [6] M. Mohandes, "Arabic Sign Language Recognition," *Arabian Journal for Science and Engineering*, vol. 36, no. 1, pp. 7–19, 2011.
- [7] A. K. Jain, A. Ross, and S. Prabhakar, "An Introduction to Biometric Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, 2004.
- [8] S. Tamura and S. Kawasaki, "Recognition of Sign Language Motion Images," *Pattern Recognition*, vol. 21, no. 4, pp. 343–353, 1988.
- [9] D. Roy and A. P. Pentland, "Learning Words from Sights and Sounds," *Cognitive Science*, vol. 26, no. 1, pp. 113–146, 2002.
- [10] R. Sharma, A. Gupta, and P. Kumar, "A Review on Sign Language Recognition Systems," *International Journal of Computer Applications*, vol. 145, no. 10, pp. 15–20, 2016.