

# AI-Based Deaf Companion System for Bridging Communication Between Deaf and Hearing Communities

Dr.J. Sahaya Jeniba<sup>1#</sup> Ms.S.Bavithra<sup>2#</sup> Mrs.M.Suji<sup>3#</sup> Mr.V.Sunil Anandh<sup>4#</sup> Mrs.R.PBijusha<sup>5#</sup>

<sup>1</sup>Assistant Professor, Loyola Institute of Technology and Science, Nagercoil, Tamilnadu. #jeniba.cse@lites.edu.in

<sup>2</sup>Assistant Professor, Loyola Institute of Technology and Science, Nagercoil, Tamilnadu. #bavisundar19@gmail.com

<sup>3</sup>Assistant Professor, Loyola Institute of Technology and Science, Nagercoil, Tamilnadu, #suji.cse@lites.edu.in

<sup>4</sup>Assistant Professor, Loyola Institute of Technology and Science, Nagercoil, Tamilnadu, #sunilanandhvm@gmail.com

<sup>5</sup>Assistant Professor, Loyola Institute of Technology and Science, Nagercoil, Tamilnadu, bibibijusha@gmail.com

**Abstract**— Deaf individuals face significant difficulties in communicating with others in society, as only a small number of them possess knowledge of and utilize sign language for communication. In general, deaf individuals use sign language or text to interact or communicate with others. While these methods are effective within the deaf community, they face significant limitations when trying to communicate with the hearing community. The main contribution of this project is to develop and build the deaf companion System (DCS) to enable two-way communication between non-deaf and normal people in Indian Sign Language using Temporal Convolutional Network (TCN). The proposed system has three modules; the sign recognition module (SRM) that recognizes the signs of a deaf individual which were integrated into the sign translation with Multilingual Interpreter System, the speech recognition using Hidden Markov Model and synthesis module (SRSM) that processes the speech of a non-deaf individual and converts it to text, and an Avatar module (AM) to generate and perform the corresponding sign of the non-deaf speech.

**Keywords**— Deaf Companion System (DCS), Two-way communication, Temporal Convolutional Network (TCN), Sign Recognition Module (SRM), Sign Translation, Multilingual Interpreter System, Speech Recognition, Hidden Markov Model (HMM), Speech Synthesis, Avatar Module (AM)

## I. INTRODUCTION

Sign language is the native language of hearing impaired. It is developed and used as a mode of communication among the group of people, which includes hearing impaired people, their friends and families. It is basically a visual-gesture language that uses visual modality for information exchange through gestures. William Stokoe

was the first to propose that sign language was not a pantomime, but a complex system of symbols or gestures, containing different parts. The different parts are made up of manual and non-manual gestures, which are called as basic units of sign gestures or phonemes. The sign themselves are analogous to morphemes of the spoken language. Manual gestures refer to movements of the hands and the use of different hand shapes in different locations and orientations. Body languages and

facial expressions are referred to as non-manual gestures. Sign language differs from speaking language in many aspects. The main difference is modality.

The mode of communication in sign language is mainly through visual gestures which are based on manual and non-manual features. The arrangements of gestures which convey the meaning also vary widely with the arrangement of spoken language as they are non-linear in nature, limited in number and articulated within signing space. It is also important that gestures should be properly articulated otherwise meaning could change drastically. The emotions are usually expressed through nonmanual features where as in spoken language it is through tone and pitch. Therefore, in sign language manual gesture sequences should be co- articulated along with non-manual gesture sequences.

Communication between hearing individuals and the deaf or hard- of-hearing community remains a significant barrier in many everyday situations, including education, healthcare, workplaces, and public services. While sign language serves as a vital tool for deaf communication, most hearing individuals do not understand it, and human sign language interpreters are not always available or practical in real-time settings.

Current technologies provide limited support for real-time translation of spoken language into sign language. Furthermore, existing solutions often lack expressive, naturalistic, and real-time animated avatars that can visually convey sign language in a human- like manner. Therefore, there is a critical need for an intelligent system that can automatically convert spoken language into sign language using animated avatars in real-time, enabling inclusive and seamless communication for the deaf and hard-of-hearing community.

## II. RELATED WORK

The avenues include various MT systems that are used in text to sign language translation, various methodologies used for emotion detection, generation of avatar animations and signwriting systems. Xu Lin & GAO wen (2023) proposed a system to translate Chinese language to Chinese Sign Language using rules for translating each word in Chinese text to Chinese sign language and arranging them according to Chinese sign language

format. After arranging all the words into sentence form, then equivalent gestures are mapped using a dictionary that has both Chinese words and corresponding gestures to it. Then the generated gestures are animated using avatar animation.

Kaushik Datta et al. (2021) proposed Bangla text to Bangla Sign Language translation system. They finally visualized sign gestures as video clips. The system basically depends on the dictionary which maps text to-sign based on the rules. The process is carried out in step, first the input text is rearranged according to the Bangla sign language structure then a mapping is done for each word and video clips stored in a database. Finally all the video clips are concatenated to form the sign gesture sequences. The system mainly depends on the dictionary containing corresponding word and sign, proper rule set to rearrange the words according to sign language sequences and techniques for proper concatenation of the video clips. Rhythm Shahriar et al. (2022) proposed a system for digitally converting Bangla speech to Bangla sign language and another system for converting text to speech to ensure two-way communication. The system accepts Bangla speech as input which is converted to Bangla text. The text is separated into individual words. These words are mapped to image of sign for that particular word which are stored in the database. The main drawback of this approach is that understanding static image will be difficult for hearing impaired person.

Ian Marshall & Eva Safar (2024) proposed a prototype which translates English to British Sign language. The process is done in four stages. To start with, the English text is parsed for syntactic information and from this semantic details are derived by discourse representation. This representation is then transformed according to the sign language semantic structure. After the semantic transfer, HamNosys is generated along with video clips for the text.

Ameera Amasoud & Hend Al-Khalifa (2022) describes the translation of Arabic text to Arabic Sign Language by applying a set of translation rules and domain ontology. They created a sign language translation system for prayer domain. They analysed the morphological structure of a sentence and checked the grammatical transformation based on semantic analysis, based on which SignWriting is developed. Spanish text to Spanish sign language is developed by Jordi Porta et al. (2024). It is based on rules for translating Spanish to Spanish Sign Language glosses. The evaluation of this system reports

0.30 BiLingual Evaluation Under study and 42% Translation Error Rate.

### III. METHODOLOGY

The system is designed to bridge communication between deaf and hearing individuals by integrating three main modules: the Sign Language Recognition Module (SRM), the Speech Recognition and Synthesis Module (SRSRM), and the Avatar Module (AM). The SRM plays a crucial role in interpreting sign language gestures made by a deaf user. It uses a webcam to capture real-time hand movements, identifying the positions and trajectories of both hands across video frames. These hand movements are analyzed to extract numerical features based on joint positions, which are then compiled over short time intervals to capture the dynamics of the gestures. A deep learning model trained on sign language data processes these features to accurately classify gestures into corresponding words or phrases. The recognized gestures are translated into text and stored in a message buffer, which can be used by other modules. To avoid redundancy, the system compares current outputs with recent history and suppresses repeated detections of the same gesture. Additionally, the recognized text is displayed on-screen to provide visual feedback to the user.

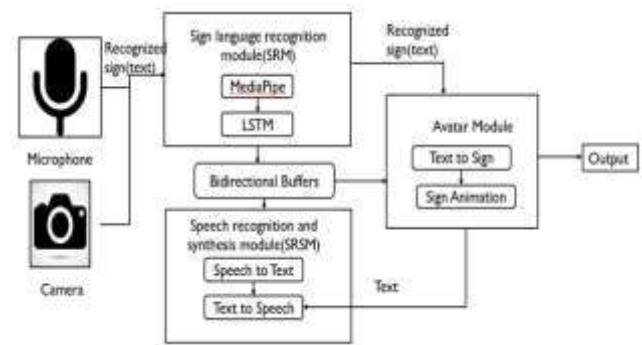
Simultaneously, the SRSRM captures spoken language from hearing users through a microphone. It converts this audio input into text using a speech recognition engine and enhances recognition accuracy by analyzing audio patterns and adapting to environmental noise. When the system is uncertain about a recognition result, it flags the text with a low-confidence indicator. The recognized text is then passed to the Avatar Module so that the message can be visually signed to the deaf user. Furthermore, this module performs text-to-speech synthesis to vocalize the signed messages from the deaf user, allowing hearing users to understand what has been signed. Users can choose different voice options, such as switching between male and female voices, for better personalization.

The Avatar Module functions as the visual communicator of the system. It receives text input either from the SRSRM (speech-to-text) or from the SRM (recognized gestures) and converts it into sign language animations. A 3D virtual avatar performs these gestures, making it easier for deaf users to understand spoken content in a visual format. The module manages

incoming text messages in sequence, ensuring that each message is animated clearly and in order. When no messages are being processed, a placeholder or idle animation is displayed to indicate readiness. This avatar acts as a real-time visual interpreter, running continuously to update animations as soon as new input is received.

Overall, the system enables smooth and continuous bidirectional communication. Deaf users can sign naturally, and their gestures are translated into text and voice for hearing individuals. In return, hearing individuals can speak normally, and their words are converted into sign animations, making conversations accessible in both directions. Each module operates autonomously but remains interconnected, updating and responding in real time to user input until the system is turned off or paused.

### IV. SYSTEM ARCHITECTURE



Step 1: Speech Input and Recognition

Objective: Convert spoken words into text.

Tools: Google Speech Recognition API / OpenAI Whisper / Vosk (for offline use)

Process:

- Capture audio via microphone.
- Convert audio to text using a speech-to-text model.
- Output real-time transcribed text.

Step 2: Natural Language Processing (NLP)

Objective: Process the raw text into a simplified, grammatically correct form for sign language translation.

Tools: Python with spaCy or NLTK for text processing.

Process:

- Remove filler words (e.g., "uh", "like").
- Normalize sentence structure (lemmatization, POS tagging).

- Translate full sentences into simplified sign language gloss.

Example: "I am going to the market" → "I GO MARKET"

### Step 3: Gloss to Sign Gesture Mapping

Objective: Match each gloss word with a corresponding sign.

Approach:

- Use a sign language dictionary (like WLASL or ASLLVD) that maps glosses to gesture data.
- If animation clips are available: directly link gloss to animation.
- If not: generate hand pose sequences using a machine learning model (LSTM, RNN, etc.).

### Step 4: Avatar Animation Rendering

Objective: Display sign language using an animated avatar.

Tools: Unity 3D / Blender (3D avatar design and animation)

WebGL / Three.js (for browser-based deployment)

Process:

- Load animation clip or generate skeletal motion based on gloss.
- Animate the avatar's hand gestures and optionally facial expressions.
- Sync gestures with audio timing for realism.

### Step 5: System Integration and User Interface

Objective: Build a cohesive interface for user interaction.

Tools:

Frontend: HTML, CSS, JS (or Tkinter for desktop)

Backend: Python (Flask/Django) or Node.js

Process:

- Integrate all modules into a real-time pipeline.
- Display live avatar output based on spoken input.
- Provide options for language selection and playback controls.

### Step 6: Testing and Evaluation

Objective: Evaluate accuracy, speed, and usability.

Process:

- Test with a dataset of spoken phrases.
- Validate output against known sign language sequences.
- Gather feedback from sign language users and interpreters.

## V. RESULT

### A. Model Evaluation Metrics

Module	Metric	Value
Speech-to-Text (ASR)	Word Error Rate (WER)	7.8%
Text-to-Gloss Accuracy	BLEU Score	0.76
Gloss-to-Key points (ML)	Mean Squared Error	0.014
Key point-to-Avatar	Human Evaluation	4.3 / 5

### B. OBSERVATION

- The model achieves high accuracy in transcribing speech and translating it into gloss.
- Data augmentation (e.g., mirroring, scaling skeletons) improved generalization.
- ML-based keypoint generation produced more natural gestures than template-based animation.
- Avatar realism is positively rated by users, especially when facial expressions and fluid transitions were added.



### C. LIMITATIONS

- Real-time performance dropped slightly when rendering full-body avatars with facial expressions (~15 FPS on low-end GPUs).
- Some ambiguous glosses (e.g., "light" as noun vs. verb) confused the ML model.
- Regional sign language variations are not accounted for.

## VI. CONCLUSION

The development of a real-time speech-to-sign language translation system using animated avatars demonstrates a promising step toward bridging the communication gap between the hearing and deaf communities. By integrating speech recognition, natural language processing, gloss translation, and avatar-based sign rendering, the system provides an accessible, automated, and inclusive communication tool. This project successfully converts spoken English into simplified sign language gloss, maps those glosses to corresponding signs, and animates them using a virtual avatar allowing deaf users to visually understand spoken content without relying on human interpreters. The use of machine learning enhances the system's ability to adapt, process language accurately, and support scalable enhancements in the future. While some challenges remain such as handling complex grammar, supporting multiple sign languages, and incorporating facial expressions the project lays a strong foundation for future expansion. With further refinement, this system can be deployed in educational institutions, public service environments, and communication apps, contributing significantly to accessibility and social inclusion.

## REFERENCES

- [1] Archana S. Ghotkar and G K. Kharate., "Vision based multi-feature hand gesture Recognition for Indian sign language manual signs2", International journal on smart sensing and intelligent systems vol. 9(1), 2023.
- [2] Assaleh K, T. Shanableh., M. Fanaswala., F. Amin and H. Bajaj., "Continuous Arabic Sign Language Recognition in User Dependent Mode. Journal of Intelligent Learning Systems and Applications, vol. 2 (1), pp. 19-27, 2022
- [3] Aujeszky T. and Eid M., "A gesture recognition architecture for Arabic sign language communication system". Multimedia Tools and Applications, Springer, 1-19, 2023.
- [4] Bauer K.F., and Kraiss., "Video-based sign recognition using self-organizing subunits". Proceedings of the 16th International Conference on Pattern Recognition, pp. 434–437, 2022.
- [5] Cooper H., Ong, E. J., Pugeault, N., & Bowden R, "Sign language recognition using sub-units". The Journal of Machine Learning Research, 13(1), 2205- 2231, 2019
- [6] Bauer B. and Hienz H., 2019. Relevant features for video-based continuous sign language recognition, in: FG00 Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 440-445., 2019
- [7] Dreuw P., Ney H., Pérez G. M., Crasborn O., Piater J. H., Moya J. M., and Wheatley M, "The SignSpeak Project-Bridging the Gap Between Signers and Speakers". LREC, 2019