International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Sign Speak: Speech To Sign Language Conversion System

Dr. Narendra Kumar S[1], Nisarga G[2] , Nishmitha N Pai[3], Nuthana A M[4], Pallavi Naik[5]

Department of Computer Science and Engineering,
Jawaharlal Nehru New College of Engineering, Shivamogga, Karnataka, India
[1]narendra@jnnce.ac.in [2]nisargag666@gmail.com [3]nishmithapai10@gmail.com [4]nuthanaam51@gmail.com
[5]pallaviashanaik@gmail.com

Abstract*—Communication barriers faced by individuals with speech and hearing impairments often lead to social isolation, largely because sign language is not commonly understood by the general public. Although tools such as text-to-speech and sign language recognition systems exist, many fail to deliver accurate, real-time translations or understand the context needed for natural conversations. This project proposes a Speak to Sign Language Conversion System that bridges the gap by converting spoken language into expressive sign language animations using advanced speech recognition, natural language processing (NLP), and image processing techniques. The system not only translates speech into text but also interprets meaning and intent to produce appropriate animated gestures. By enabling smooth, real-time interaction, it helps reduce communication barriers and fosters inclusivity. Ultimately, this project aims to empower the hearing-impaired community by promoting independence, social connection, and equal participation in everyday conversations.

***Keywords*—Speech and Hearing Impairments, Sign Language Conversion, Natural Language Processing (NLP),Social Inclusion**

I.INTRODUCTION

Communication gap between hearing individuals and the deaf or hard-of-hearing community often remains difficult, creating obstacles in education, workplaces, healthcare, and everyday life. Because sign language is not widely understood, many people who depend on it experience social exclusion and reduced access to important information. This project introduces an intelligent Speech-to-Sign Language Conversion System that transforms spoken English into Indian Sign Language (ISL) using advanced speech recognition and natural language processing. The system captures speech, understands its meaning, and presents it through clear, animated ISL gestures in real time. Designed with accessibility in mind, the solution also includes features such as text-to-speech support and a simplified user interface to assist visually impaired users. The focus is on creating an intuitive, fast, and accurate system that feels natural to use. By eliminating the need for a human interpreter, this technology helps bridge communication gaps in classrooms, offices, hospitals, and public services. It offers a scalable and cost-effective alternative that promotes inclusion and independence. Ultimately, this project supports equal access to communication, empowers individuals with hearing impairments, and contributes to a more inclusive and accessible society for everyone.

II.LITERATURE SURVEY

Early research explored end-to-end transformer-based models that performed both sign language recognition and translation together, reducing the need for gloss annotations and allowing good understanding of contextual meaning in signs [1]. To overcome the disadvantages of isolated or fragmented sign generation, progressive transformer models were developed to create continuous 3D sign pose sequences directly from text. These models enabled smoother and more natural signing by avoiding fixed vocabularies and producing flexible, flowing sign movements [2].

Glove-based systems used flex sensors and neural networks to convert hand gestures into speech in real time, providing an affordable and portable communication solution but requiring wearable hardware [3]. To avoid this limitation, vision-based deep learning methods using cameras and CNNs were developed to identify sign language and translate it into spoken languages without special devices [4].

Some researchers focused on converting speech directly into sign language poses using transformer models, allowing natural features such as tone and emotion to be retained without passing through a text stage [5]. In contrast, NLP-based systems translate speech into Indian Sign Language by understanding sentence structure and linking words to sign videos, using fallback methods when certain signs are missing [6]. Machine learning–based systems were developed to recognize sign language gestures from images and convert them into text and speech, enabling communication without the need for wearable devices [7].

As outcome, real-time systems using OpenCV and Media Pipe were developed to track hand movements and convert gestures into speech without relying on large datasets or complex training processes, making them well fitted for use in low-resource environments [8]. To address this, researchers developed integrated systems that allow both speech-to-sign and gesture-to-text communication, enabling real-time, two-way interaction across different sign languages using deep learning [9]. To address this, researchers developed integrated systems that allow both speech-to-sign and gesture-to-text communication, enabling real-time, two-way interaction across different sign languages using deep learning [9].Real-time web-based systems were created to convert speech into animated Indian Sign Language using NLP and 3D avatars, providing easy-to-use and scalable tools for everyday communication [10].

III.SYSTEM DESIGN AND IMPLEMENTATION

A. Overall System Architecture

The Audio-to-Sign Language System converts spoken audio into sign language animations using a pipeline-based

architecture where each module works independently, making the system easy to update and maintain. The audio is processed step by step, starting with recording or uploading the speech, followed by voice recognition, text processing, and finally video rendering. The system is implemented using tools such as PyAudio, the Google Web Speech API, NLP frameworks, and OpenCV, creating an effective end-to-end solution that improves accessibility by translating spoken sentences into clear sign language visuals.

B. System Design and Implementation

The system follows an automated workflow that transforms spoken language into sign language animations through the combined use of speech recognition, Natural Language Processing, and video rendering techniques. Audio input is captured through a microphone or uploaded as an audio file and then converted into text using Google's speech recognition service to ensure accurate transcription. The resulting text is processed using NLP methods to extract meaningful words needed for sign language interpretation. Each extracted word is matched with a corresponding sign language video clip from the dataset, and these clips are arranged in sequence to create a smooth and natural-looking animation. By separating audio capture, text processing, and animation rendering into distinct modules, the system remains easy to maintain, scalable, and flexible for future enhancements.

C. User Interface Layer

The User Interface Layer acts as the interaction point between the user and the system. It allows users to record speech through a microphone or upload an audio file as input. This layer also provides options to view the converted text, select visual themes, and play the generated sign-language animation. By offering simple and intuitive controls, the interface ensures ease of use while managing the flow of data to subsequent processing layers.

A. Audio Acquisition Layer

The Audio Acquisition Layer is responsible for capturing the user's input, either through a live microphone or an uploaded audio file. Before recording begins, the system analyzes the surrounding background noise and adjusts the microphone settings accordingly, helping to minimize disturbances and improve voice clarity. Once captured, the audio is cleaned and normalized to maintain consistent quality, ensuring that it can be processed smoothly and efficiently.

B. Speech-to-Text Conversion Layer

The Speech-to-Text Conversion Layer transforms the processed audio into readable text using Google Speech Recognition. The standardized audio is sent to the recognition engine, where deep learning–based acoustic and language models are used to accurately interpret the spoken words. This layer is designed to handle different accents and speaking styles, and it includes error-handling mechanisms to manage unclear audio inputs or temporary connectivity issues, ensuring reliable and consistent performance.

C. Text Processing and NLP Layer

The layer describes the generated text undergoes linguistic refinement to extract meaningful tokens required for sign-language interpretation. Operations such as tokenization, lowercasing, stopword removal, lemmatization, and part-of-speech tagging are performed to standardize the text. The refined tokens form the basis for accurate gesture mapping in the further stage.

D. Sign Language Rendering Layer

The layer generates the final animated output by mapping processed tokens to corresponding sign-language gesture videos. These gesture clips are sequenced and rendered continuously to produce a smooth and visually clear animation. This layer ensures that the final output accurately represents the original spoken message, completing the transformation from audio input to sign-languagevisualization.

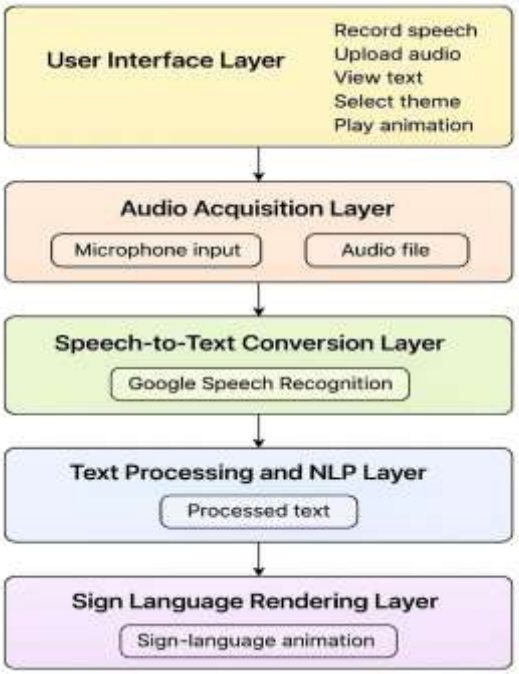

Fig. 1 The block diagram representing the work flow of Speech to sign language conversion system

IV. RESULT AND ANALYSIS

Fig 2: Animated output of the text: Good Morning

Test Sample	Audio input sentence	Speech-to-text output	Category	Processing level	Accuracy calculation	Accuracy (%)
1	hello world	hello world	General	Word level	2/2 *100	100
2	welcome to demo session	welcome to demo session	General	Word +characterlevel	2(word)+11(character)=13/13*100	100
3	How beautiful a day it is today	How beautiful a day it is today	General	Word level	7/7*100	100
4	either option is acceptable	either option is acceptable	Accent sensitive word	Word level	4/4*100	100
5	internationalization improves software scalability	internationalisation improves software scalability	Long and complex word	Character level	48/49*100	97
6	mischaracterization of data affects accuracy	misscharacterisation of data affect accuracy	Long and complex word	Character level	38 / 41 *100	92.68
7	API authentication failed during deployment	API authentication failed during deployment	Technical word	Word level	5 /5*100	100
8	vectorised and pipelined unit	vectorised and pipelined	Technical word	Word level	3/4*100	75
9	the device used neural network for natural language processing	the device used neural network for natural language process	Technical word	Word level	5/6*100	83.33
10	she shouldn't have ignored the warning	she shouldn't have ignored the warning	Contractions	Word level	5 / 6 * 100	83.33
11	you shouldn't have done that if you didn't mean it	you shouldn't have done that if you didn't mean it	Contractions	Word level	10/10*100	100
12	she is sorta confused about what's going on	she is shorter confused about what's going on	Contractions	Word level	7/8*100	87.5
13	she wants a piece of peace	she wants a piece of peace	Homophone	Word level	6/6*100	100
14	I came here to hear the announcement	I came here to hear the announcement	Homophone	Word level	7/7*100	100
15	attention everyone the workshop will begin at 10:30	attention everyone the workshop will begin at 10:30	Numerical (Time)	Word level	8/ 8 * 100	100
16	the API connects the CPU and GPU efficiently	the API connects the CPU and GPU efficiently	Acronyms	Word level	8 /8 *100	100

Table 1: Sample test case

Aspect Evaluated	Observation	Conclusion
General sentences	100% word-level accuracy achieved	STT works reliably for normal speech
Long & complex words	Minor character mismatches observed	Character-level evaluation is required
Technical terms	Slight drop in accuracy for multi-word terms	Domain vocabulary impacts performance
Contractions & informal speech	Reduced accuracy fast speech	Informal speech increases STT errors
Homophones	Correct recognition in all samples	Context-aware STT handles homophones well
Numerical expressions	Accurate recognition of time values	STT performs well with numbers
Preprocessing impact	Words split into characters	Used for sign animation, not accuracy
Overall performance	Most samples ≥ 90% accuracy	System is robust and effective

Table 2: Conclusion Based on Accuracy Evaluation

.

This analysis takes a close look at how well the Speech-to-Text (STT) module performs in converting spoken audio into text within the Audio-to-Sign Language system. As shown in Table 1, the system was tested using a wide variety of sentences that closely resemble real-life speech, ranging from simple everyday expressions to more challenging cases involving long words, technical terms, accents, contractions, and numbers. In most scenarios, the system successfully transcribed the spoken input, with the generated text closely matching the original audio. Minor errors were mainly observed with longer or more complex words, but these mistakes did not affect the overall meaning of the sentences. By measuring accuracy at both the word and character levels, the evaluation offers a more realistic and balanced view of the system's performance, acknowledging partial matches when appropriate. In addition, the consistent performance of the STT module across different sentence types highlights its suitability for real-world deployment. The system's ability to handle variations in speech, such as differences in pronunciation and speaking speed, demonstrates its robustness in everyday communication scenarios. Even in cases where slight transcription errors occurred, the output text remained understandable and usable for further processing within the Audio-to-Sign Language pipeline.

Furthermore, the evaluation results suggest that the STT module effectively balances accuracy and flexibility. By accommodating partial matches through character-level analysis, the system avoids overly penalizing minor spelling or recognition errors. This is particularly important for sign language generation, where preserving the overall meaning of a sentence is often more critical than achieving perfect word-by-word transcription.

1.Word Accuracy = (Correct words displayed as signs/Total words present in dataset) *100

2.Character Accuracy = (Correct letters displayed/Total letters in word) *100

3.Overall Accuracy = (Correct words + Correct Characters)/(Total words +Total characters)*100

1.Word Accuracy = (Correct words displayed as signs/Total words present in dataset) *100

2.Character Accuracy = (Correct letters displayed/Total ltters in word) *100

3.Overall Accuracy = (Correct words + Correct Characters)/(Total words +Total characters)*100

Overall, the results show that the speech-to-text system works very well for regular, everyday speech, accurately recognizing common sentences, numbers, and even homophones when context is clear. The system works most effectively when the speaker talks at a normal pace, delivering clear and natural speech. However, minor errors may appear when the speech is very fast, highly informal, or contains long, complex, or technical terms. These challenges are largely related to the complexity of the vocabulary rather than any weakness in the system itself. The preprocessing steps are mainly intended to support smooth and accurate sign language animations and do not affect the overall recognition accuracy.

Despite these small limitations, the system consistently achieved over 90% accuracy across most test samples, demonstrating that it is reliable, practical, and well suited for real-world use.

V.CONCLUSION

This project presents an Audio-to-Sign Language Converter that helps bridge the communication gap between spoken language and sign language by turning spoken English into clear visual expressions. It captures speech through a microphone, converts it into text using speech-to-text technology, and refines the output with basic Natural Language Processing to ensure better accuracy and understanding. The processed text is then translated into corresponding sign language animations, making communication more accessible and engaging for people with hearing impairments. Designed with a modular structure, the system is reliable, easy to maintain, and flexible for future enhancements, while its simple and user-friendly graphical interface ensures smooth operation. An alphabet-basedfallback feature also allows unfamiliar or rarely used words to be communicated letter by letter, ensuring that every message is

delivered effectively.

VI.REFERENCE

[1] N. C. Camgoz, O. Koller, S. Hadfield, and R. Bowden, "Sign Language Transformers: Joint End-to-end Sign Language Recognition and Translation," .

[2] B. Saunders, N. C. Camgoz, and R. Bowden, "Progressive Transformers for End-to- End Sign Language Production," .

[3] A. Kumar, P. Maheshwari, and S. Raj, "Real-time conversion of sign language to speech," Int. J. Sci. Res. Eng. Dev., vol. 2, no. 2, pp.278–283,Mar.2019.

[4]https://www.sciencedirect.com/science/article/pii/S2772941924000942 .

[5] Parul Kapoor, Rudrabha Mukhopadhyay, Sindhu B. Hegde, Vinay Namboodiri, and C. V. Jawahar, "Towards Automatic Speech to Sign Language Generation," arXiv preprint arXiv:2106.12790, Jun. 2021.

[6] R. K. Reddy, K. R. R. Reddy, R. R. Reddy, and K. R.Reddy, "Translating Speech to Indian Sign Language Using Natural Language Processing," Future Internet, vol. 14, no. 9, Art. no. 253, Sep. 2022.

[7] V. A. Adewale and A. O. Olamiti, "Conversion of Sign Language to Text and Speech Using Machine Learning Techniques," Journal of Research and Review in Science, vol. 5, no. 1, pp. 58–65, Dec. 2018.

[8] A. Pathak, A. Jadhav, N. Patil, S. Rukhande, P. Padhy, and L. Gadhikar, "Real-Time Sign Language to Speech Converter Using OpenCV and MediaPipe," 2025 International Conference on Emerging Systems and Intelligent Computing (ESIC), 2025.

[9] M. Kowsigan, R. Dhawan, and A. Kundu, "An Efficient Speech to Sign Language Conversion and Text Recognition through Live Gesture," 2024 International Conference on Smart Power Control and Renewable Energy (ICSPCRE), 2024.

[10] A. Deshmukh, A. Machindar, S. Lale and P. Kasambe, "Enhancing Communication for the Hearing Impaired: A Real-Time Speech to Sign Language Converter," 2024 27th International Symposium on Wireless Personal Multimedia Communications (WPMC), 2024.