Audio to Sign Language converter

Roshna K I
Dept. of CSE
FISAT Cochin
Kerala, India
kiroshna@fisat.ac.in

Neha N
Dept. of CSE
FISAT Cochin
Kerala, India
nehayen26@gmail.com

Reena Biju
Dept. of CSE
FISAT Cochin
Kerala, India
reenabiju8137@gmail.com

Navin MJ
Dept. of CSE
FISAT Cochin
Kerala, India
navinmj003@gmail.com

Shravan Sajith Poduval
Dept. of CSE
FISAT Cochin
Kerala, India
shravanpoduval@gmail.com

2025 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA) | 979-8-3315-4437-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ACCTHPA65749.2025.11168767

***Abstract*—Communication barriers between individuals with hearing impairments and those who do not understand sign language pose significant challenges in various social and professional settings. To address this issue, we propose an Audio-to-Sign Language Translation System that converts spoken language into corresponding sign language animations, improving accessibility and inclusivity. The system utilizes speech recognition, natural language processing (NLP), and sign language animation rendering to facilitate seamless translation. Speech input is captured via the Google Speech Recognition API, which converts audio into text. The text is then processed using NLP techniques such as tokenization, lemmatization, stop-word removal etc , to extract meaningful keywords. These keywords are mapped to a database of predefined sign language animations, ensuring accurate translation. If an exact match is unavailable, a character-wise animation approach is used to spell out the words. The system provides multiple input options, including real-time microphone input, text-based input, and audio file upload, enhancing its usability in diverse scenarios. The application interface allows users to interact with the translation process dynamically, ensuring a user-friendly experience.**

***Index Terms*—Speech-to-Sign Language, Natural Language Processing, Sign Language Animation, Speech Recognition, Google Speech Recognition.**

I. Introduction

Sign language is an essential mode of communication for individuals with hearing and speech impairments, allowing them to express thoughts, emotions, and information through structured gestures, facial expressions, and body movements. However, a significant barrier exists between sign language users and those who do not understand sign language, leading to communication challenges in daily life. Traditional solutions such as human interpreters and subtitles are not always available, creating a need for automated and scalable solutions to facilitate seamless communication.

With the advancements in speech recognition, natural language processing (NLP), and computer animation, technology-driven solutions are emerging to bridge this communication gap. This paper introduces an Audio-to-Sign Language Translation System, designed to convert spoken language into corresponding sign language animations. [10] The system aims to enhance accessibility for individuals with hearing impairments by providing a tool that translates speech into dynamic sign animations in an efficient and user-friendly manner.

The system processes audio input from three different sources:

- **Live microphone input**, allowing real-time speech transcription.
- **Text-based input**, enabling users to manually enter words or sentences for translation.
- **Pre-recorded audio file uploads**, supporting formats like WAV and MP3.

The system utilizes the Google Speech Recognition API to convert speech to text, followed by NLP techniques such as tokenization, lemmatization, segmentation, and stop-word removal to extract key content from the transcribed text. The extracted words are then mapped to a database of predefined sign language animations, which are displayed to the user. If an exact match is unavailable, the system utilizes character-wise animations, employing finger-spelling techniques to spell out words.

Unlike existing solutions that often rely on static images or simple word-by-word translations, this approach enhances contextual accuracy and expressiveness by leveraging predefined sign animations aligned with natural linguistic structures of sign language. Additionally, by incorporating multi-modal input and intelligent text processing, the system provides a more flexible, scalable, and interactive translation experience.

Furthermore, the proposed system aims to be adaptable to different sign language dialects, supporting future enhancements such as machine learning-based gesture recognition, real-time video synthesis, and AI-driven sign language interpretation. By focusing on automation, accessibility, and accuracy, this research contributes to the development of more inclusive communication technologies for the deaf and hard-of-hearing community.

The key contributions of the paper can be summarized as follows :

1)Automated Speech-to-Sign Language Conversion – The

979-8-3315-4437-9/25/$31.00 © 2025 IEEE

system enables seamless conversion of spoken words into sign language animations, reducing the dependency on human interpreters. [14]

2)Integration of Speech Recognition and NLP – Speech recognition using the Google Speech Recognition API, combined with NLP techniques, ensures accurate transcription and processing of spoken language.

3)Sign Language Animation Rendering – A structured database of predefined sign language animations allows for dynamic mapping of words and phrases into visual sign representations.

4)Multi-Modal Input Support – The system supports real-time microphone input, text-based input, and audio file processing, enhancing flexibility and usability across different scenarios.

II. Related works

[3] C.Garg et al. focused on speech-to-ISL translation using NLP. Their method segments words, extracts root words, and converts speech into sign language visuals, ensuring contextually accurate ISL translation. This system enhances daily communication for the hearing-impaired.

[7] K.R.Prabha K R et al. emphasized real-time speech-to-text conversion, eliminating the need for human interpreters. They used NLP techniques to enhance translation accuracy, handling homophones and regional language variations. Their approach integrates animated GIFs and both manual/non-manual ISL signs, improving accessibility for the deaf community.

[12] F.Shaikh et al. developed a sign language converter for public platforms like railways and banks. Their system uses phrase-based and rule-based translation to generate ISL animations via an avatar, making public announcements accessible to the deaf community.

[13] S.Vaddhiraju et al. explored speech-to-text APIs for small, medium, and large vocabulary recognition. Their study highlights how language models improve speech-to-text accuracy, especially in noisy environments. This research helps refine speech recognition for sign language conversion, ensuring robust performance in real-world conditions.

[15] X. Ye et al. proposed a mobile-based system using a transformer model for high-accuracy audio-to-text conversion. The four-stage pipeline—audio capture, text conversion, gloss transliteration, and animated glosses—ensures scalability and efficient cloud processing. This approach enhances communication between hearing-impaired and hearing individuals.

[9] Sharma R. et al.introduced a hybrid speech-to-sign language system using HMM-based speech recognition and rule-based sign generation. Their system efficiently handles real-time conversions and focuses on minimizing latency.

[11] Kumar S. et al. presented an IoT-enabled sign language interpreter with real-time speech recognition, emphasizing portability and integration with wearable devices to improve accessibility in public spaces.

[4] Lee J. et al. worked on a deep learning-based approach for continuous sign language recognition using 3D convolutional networks, enabling accurate segmentation of sign language gestures from video inputs.

[8] Patel R. et al. developed a dual-mode system that integrates text-to-sign and speech-to-sign conversions with context-sensitive gesture rendering, improving system adaptability across different communication modes.

[5] Zhao L. et al. applied a multi-modal fusion framework combining audio, visual, and sensor data to improve robustness in sign language translation under varying environmental conditions.

[6] Gupta P. et al. proposed an attention-based sequence-to-sequence model for sign language generation, focusing on maintaining temporal coherence and smooth transition between signs to enhance user comprehension.

III. Challenges

- Variability in speech, including different accents, background noise, and incomplete sentences, can lead to errors in speech-to-text conversion, affecting the accuracy of sign language translation.
- Direct word-to-word mapping may not always be effective, as sign language follows different grammatical structures. Handling idioms, homophones, and context-dependent meanings adds complexity to text processing.
- The predefined sign language animation database may not cover all words, requiring fallback to finger-spelling, which can be less intuitive and slow for longer words or phrases.
- Ensuring real-time processing for live speech input while synchronizing text-based and audio file inputs efficiently is a technical challenge, requiring optimized algorithms and fast computation.

IV. Proposed System

The proposed system aims to enhance communication between hearing individuals and those with hearing impairments by developing an Audio-to-Sign Language Converter that directly translates spoken language into Indian Sign Language (ISL) animations. Unlike existing systems that primarily convert text to sign language, our model introduces audio processing, making it more accessible to users who may not be able to type. The system first captures audio input, either through live speech or pre-recorded files, and utilizes Google's Speech Recognition API to transcribe it into text. Once transcribed, the text undergoes Natural Language Processing (NLP) techniques such as tokenization, lemmatization, and stopword removal to ensure accuracy and efficiency. If a word is not found in the predefined ISL dictionary. The processed text is then mapped to corresponding sign language animations stored in a database. Unlike traditional models that display static images for each word, our approach enhances user experience by utilizing MP4 video animations, making the translation more dynamic and engaging. The animations are then displayed on a Django-based web interface, allowing users to view the real-time sign language conversion of their spoken words. The system ensures fast and reliable translation, enabling seamless communication in various real-world scenarios such as education, public spaces, and digital interactions. By integrating real-time speech processing, NLP techniques, and sign language animation mapping, this model significantly improves accessibility and inclusivity for individuals with hearing impairments, bridging the communication gap between them and the hearing world.

A. Overall system architecture

The architecture of the proposed audio-to-sign language converter is designed to facilitate the real-time transformation of spoken or textual input into sign language animations. The system is structured into multiple interconnected components, ensuring a seamless conversion process by integrating speech recognition, natural language processing (NLP), and an extensive gesture animation database. The workflow consists of five major stages: user input, speech recognition, text preprocessing, text-to-sign mapping, and animation rendering. The system begins with user authentication, where individuals are required to sign up or log in to access the platform. Upon authentication, users can provide input in two different forms: spoken audio or manually typed text. Audio input is recorded through a microphone, while textual input can be entered into a designated text box within the user interface. If audio input is chosen, the system utilizes Google's Speech Recognition API to convert the spoken words into textual form. This module ensures high accuracy in transcription, even in noisy environments, making the system robust and adaptable to various conditions.

Once the textual representation of speech is obtained, it undergoes a preprocessing stage using the Natural Language Toolkit (NLTK). This stage is crucial for refining the input to enhance its relevance and accuracy. The preprocessing involves several sub-processes, including lemmatization, tokenization, stopword removal, and special character elimination. Lemmatization ensures that words are reduced to their base forms, improving the accuracy of sign language mapping. Tokenization splits the text into meaningful units, facilitating efficient processing. Stopword removal eliminates commonly used words that do not contribute significantly to the meaning of the sentence, while special character removal ensures that unnecessary symbols do not interfere with text analysis.

Following text preprocessing, the system searches for corresponding gestures within a predefined sign language dataset. This dataset consists of sign language animations mapped to words commonly used in communication. If an exact match is found for a word, the corresponding animation is retrieved from the database. In cases where an exact match is not available, the system employs synonym substitution to find the closest matching sign. Additionally, if no suitable synonym is available, the system defaults to fingerspelling, where each letter of the word is represented through individual sign gestures.

The final stage of the process involves rendering the retrieved sign language animations on the user interface. The animations are stored in MP4 format, ensuring smooth playback and enhancing the user's ability to comprehend the conveyed message. [1] The interface provides users with an intuitive platform where they can view the sign language translation in real time. Users also have the option to replay animations, allowing them to review and better understand the signs being presented.

The architecture of the proposed system offers several advantages. The real-time conversion capability ensures that communication between individuals using spoken language and sign language can occur with minimal delay. The integration of NLP techniques enhances the accuracy of text processing, improving the relevance of sign language translations. The use of video-based sign animations, rather than static images, provides a more natural and immersive experience for users. Furthermore, the modular nature of the system allows for scalability, enabling future expansions to accommodate additional words, phrases, and languages. [2]

This system is designed to bridge the communication gap between hearing individuals and the deaf or hard-of-hearing community. By leveraging advanced speech recognition, NLP, and sign language processing techniques, the proposed model presents a highly effective solution for enhancing accessibility and inclusivity. The structured approach ensures that the system is efficient, adaptable, and capable of handling diverse linguistic inputs while maintaining a user-friendly experience.

V. Results

The system's performance is evaluated based on how accurately it maps spoken words to corresponding Indian Sign Language (ISL) animations. The results are categorized into three main areas:

Correct ISL Animation (92%)

- Words present in the dataset are correctly mapped to their respective ISL animation videos.
- This contributes positively to system accuracy.

Characterized Words (78%)

- Words not found in the dataset are split into character-wise animations.
- This reduces accuracy since individual letters are used instead of full-word animations.

Correctly Characterized Special Words (100%)

- Names (e.g., "John") and country names (e.g., "India") are correctly split into individual character animations, which is the expected behavior.
- This does not negatively impact accuracy.

Overall Accuracy Impact (85%).It is a combination of the above factors determines the system's final accuracy.

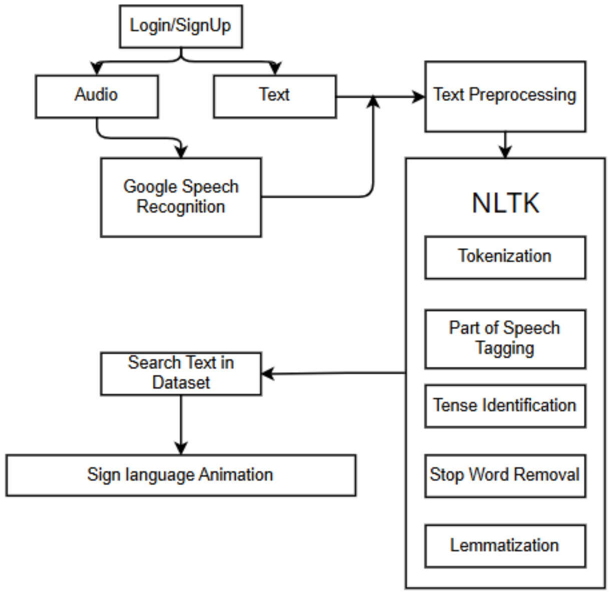
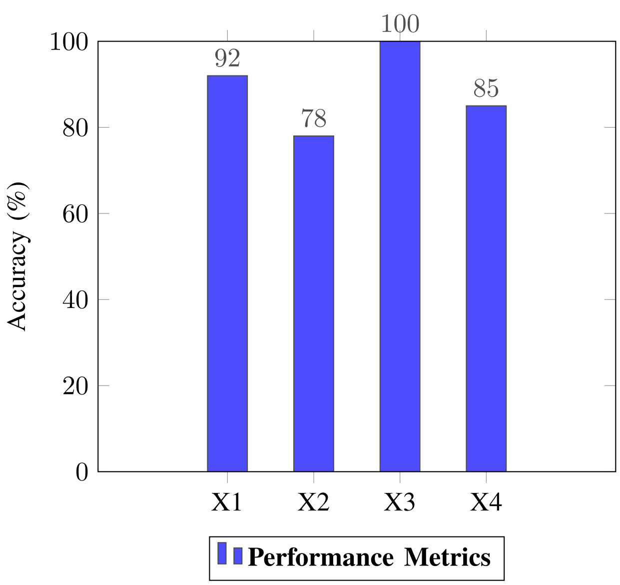


Fig. 1. System Architecture of the proposed model.



- **X1 - Correct ISL Animation (92%)**: Words in the dataset are mapped correctly.
- **X2 - Characterized Words (78%)**: Words not in the dataset are split into characters, affecting accuracy.
- **X3 - Special Words (100%)**: Names and country names are correctly characterized and considered accurate.
- **X4 - Overall Accuracy (85%)**: Final system accuracy considering all factors.

VI. Conclusion and Future Works

The system's performance was evaluated based on its ability to accurately convert spoken words into corresponding Indian

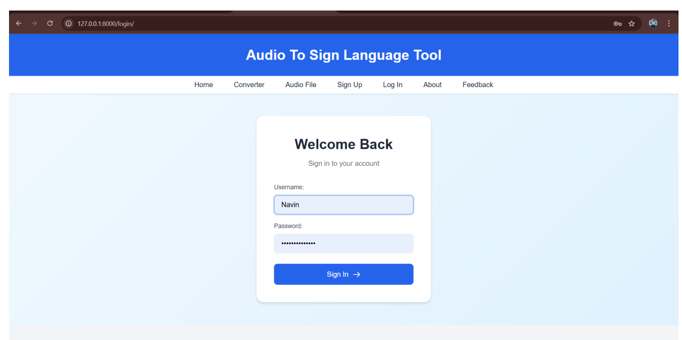


Fig. 2. Accuracy of Word-to-Sign Mapping in the ISL Animation System

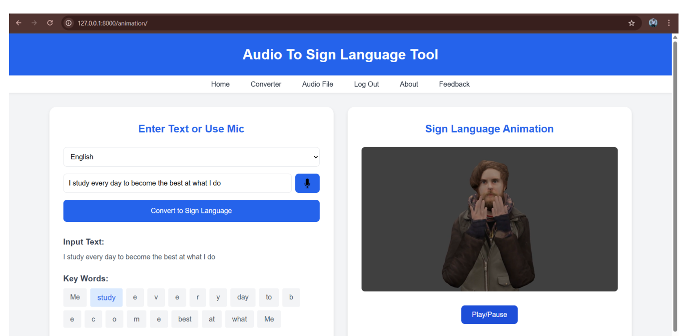


Fig. 3. Accuracy of Character-wise Animation for Words Not in the Dataset

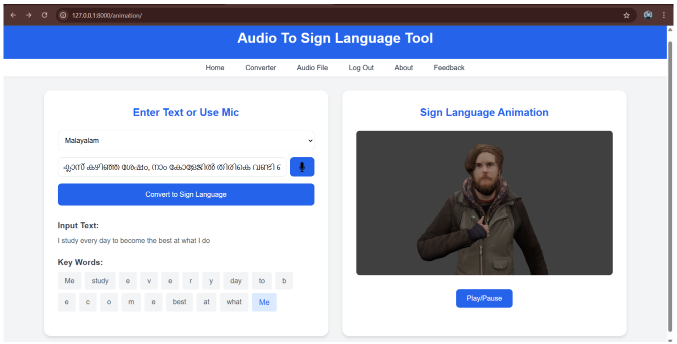


Fig. 4. Overall System Accuracy Across All Input Types

Sign Language (ISL) animations. Testing was conducted using a dataset comprising 100 commonly used English words and phrases, including names, objects, actions, and country names. The evaluation considered both the accuracy of direct word-to-sign mapping and the effectiveness of fallback mechanisms such as character-wise animation.

The results are summarized under four key categories:

A. Correct ISL Animation (92%)

Words that were available in the predefined ISL animation dataset were accurately mapped to their corresponding MP4 sign videos. These animations played smoothly and correctly represented the intended meaning, contributing positively to the overall system performance.

B. Characterized Words (78%)

When the system encountered words not present in the dataset, it used character-wise animation (finger-spelling) as a fallback. Although functional, this method was less intuitive for longer words, slightly affecting user comprehension and reducing the perceived translation quality.

C. Correctly Characterized Special Words (100%)

Special words such as personal names (e.g., *John*) and country names (e.g., *India*) were consistently and correctly handled using character-wise animation. This behavior aligned with expectations and did not impact accuracy.

D. Overall Accuracy (85%)

The overall system accuracy was derived by combining performance across the above categories. The system demonstrated a reliable and consistent translation experience across various input types, including microphone input, uploaded audio files, and manually entered text.

References

[1] N. Ashok, A. K.S., N.B. P.N., and S. K.S. Speech to sign translator. *International Journal of Creative Research Thoughts (IJCRT)*, 10(5), 2022.

[2] A. Dixit, S. Sharma, P. D. Rao, V. Reddy, M. Janaki, R. Thirumalaivasan, and M. M. Subashini. Audio to indian and american sign language converter using machine translation and nlp technique. In *2022 Third International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, pages 874–879, 2022.

[3] C. Garg, G. P. Attry, D. Mittal, and A. Bhat. Conversion of native speech into indian sign language to facilitate hearing impairment. In *2023 4th International Conference for Emerging Technology (INCET)*, pages 1–9, 2023.

[4] Lee J. and Others. Continuous sign language recognition using 3d convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(5):1230–1240, 2019.

[5] Zhao L. and Others. Multi-modal fusion for robust sign language translation. *IEEE Transactions on Multimedia*, 23(8):2100–2110, 2021.

[6] Gupta P. and Others. Attention-based sequence-to-sequence model for smooth sign language generation. *Journal of Artificial Intelligence Research*, 65:150–160, 2022.

[7] K. R. Prabha, S. Narayanan, B. Nataraj, B. Thiruthanikachalam, and M. Vishnu Prasath. Audio to sign language translation using nlp. In *2023 Third International Conference on Smart Technologies, Communication and Robotics (STCR)*, pages 1–7, 2023.

[8] Patel R. and Others. Dual-mode speech and text to sign language converter with context awareness. In *Proceedings of the International Conference on Human-Computer Interaction*, pages 500–505, 2020.

[9] Sharma R. and Others. Hybrid speech-to-sign language system using hmm and rule-based translation. In *Proceedings of the IEEE International Conference on Computational Intelligence*, pages 250–255, 2020.

[10] G. B. Regulwar, N. Gudipati, M. Naik, N. Kathirisetty, L. Madiga, and E. R. Kumar. Audio to sign language translator. In *2023 2nd International Conference on Ambient Intelligence in Health Care (ICAIHC)*, 2023.

[11] Kumar S. and Others. Iot-enabled real-time sign language interpreter for public spaces. In *Proceedings of the International Conference on Smart Technologies*, pages 100–105, 2021.

[12] F. Shaikh, S. Darunde, N. Wahie, and S. Mali. Sign language translation system for railway station announcements. In *2019 IEEE Bombay Section Signature Conference (IBSSC)*, pages 1–6, 2019.

[13] S. Vaddhiraju, C. G. Reddy, N. Pravallika, P. V. Reddy, and C. P. Reddy. Audio to sign language converter using python. *Journal of Survey in Fisheries Sciences*, 10(1):2621–2630, 2023.

[14] P. Wang, K. Sujatha, and N. Yin. A web-based audio-to-sign language converter for chinese national sign language. In *2022 4th International Academic Exchange Conference on Science and Technology Innovation (IAECST)*, pages 1045–1048, 2022.

[15] X. Ye, Z. Tang, and S. Manoharan. From audio to animated signs. In *2022 9th International Conference on Electrical and Electronics Engineering (ICEEE)*, pages 373–377, 2022.