IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 13, Issue 2, April-May 2025
ISSN: 2320 – 8791 (Impact Factor: 2.317)
www.ijreat.org

AI-POWERED SPEECH TO SIGN LANGUAGE CONVERSION SYSTEM FOR HEARING IMPAIRED

K Prakash [1], R Rajesh [2], M Tamil Selvan [3], S Jagatheeswaran [4] and Asso.Prof Dr.S.Nagarajan[5]

**[1]Computer Science and Engineering, Government College of Engineering Srirangam,,
Trichy, Tamil Nadu 620 012, India**
prakashkumaresan2004@gmail.com

**[2]Computer Science and Engineering, Government College of Engineering Srirangam,,
Trichy, Tamil Nadu 620 012, India**
rajeshr45736@gmail.com

**[3]Computer Science and Engineering, Government College of Engineering Srirangam,,
Trichy, Tamil Nadu 620 012, India**
tamilm0236@gmail.com

**[4] Computer Science and Engineering, Government College of Engineering Srirangam,,
Trichy, Tamil Nadu 620 012, India**
jagathees0703@gmail.com

**[5]Computer Science and Engineering, Government College of Engineering Srirangam,,
Trichy, Tamil Nadu 620 012, India**
drsnagarajan_cse@gces.edu.in

Abstract

Hearing and speaking impaired individuals face ongoing challenges to connect with hearing people because of communication gaps during inclusive conversations. This project develops the Audio to Sign Language Converter which functions as a real-time AI web application that changes voice inputs into animated sign language outputs.WebkitSpeechRecognition API enables the software to record voice input that turns into written text in real time. The system distributes text into words before selecting vital keywords. The web application utilizes ISL animations which rely on identified keywords. When the program lacks a direct sign for specific words it follows a procedure of spelling these terms by using each letter in the alphabet.The application employs Django combined with HTML, CSS, JavaScript and showcases a user-friendly interface. The system merges audio to text transformation with NLP capabilities and visual output to create a communication link between real-time speech and digital accessibility for hearing-impaired users.

Keywords – Artificial Intelligence, Speech Recognition, Indian Sign Language, Natural Language Processing, Sign Animation, Real-time Translation, Hearing impaired

1. Introduction

The ability to communicate serves as a vital link that enables people to deliver their ideas and compete feelings while sharing knowledge between one another. People who depend on sign language face significant challenges when they try to engage in daily conversations due to their hearing or speech impairments.

Users now access the Audio to Sign Language Converter as a web-based platform that delivers instant sign language animations through its AI-powered system. WebkitSpeechRecognition API enables the system to convert user vocalization into written text through its operation. The speed and accuracy of real-time speech recognition are enhanced through this method which provides users with better performance.

The text transcription undergoes basic Natural Language Processing (NLP) through tokenization to split each sentence into separate keywords. The system matches these chosen keywords to storage information containing animated signs from the Indian Sign Language (ISL). Whenever the word library lacks an animation for a specific word the system defaults to using ISL alphabet signs for spelling letters one by one thus preserving message content.

The application uses HTML and CSS with JavaScript programming for frontend components which provide a simple user interface that anybody can easily use. The system implements the Django framework in its backend as the primary tool to handle animation management functions alongside application scalability. The system supports convenient modification and expansion of its sign library with its modular design structure.

With the combination of speech recognition and NLP along with multimedia rendering this system establishes communication pathways between people who use speech and those who use non-verbal methods. The tool functions optimally in educational facilities as well as public service counters and healthcare facilities for instant and effective communication with hearing impaired individuals.

2. Existing Work

Speech-to-Sign Language Translation Systems:

The research field focuses on developing assistive technologies to connect hearing people with individuals who have hearing disabilities. Early programs attempted to convert speech and text into sign language through rule-based and templated solutions which struggled in flexible speech translation. A 3D avatar animation system delivered better visual quality through its process but this system required heavy computational power that limited its use with low-performance systems and browsers [1]. A two-way speech-to-sign language model implemented alphabet-based gestures which resulted in poor effectiveness for semantic meaning transmission [2].

An investigation involved developing bilingual system designs which integrate both Indian Sign Language (ISL) and American Sign Language (ASL). Different languages within these linguistically rich systems need independent handling pipelines which makes systems more complex thus reducing scalability [3].

The capabilities of sentence parsing and sign selection rule-based approaches in NLP depend heavily on the quality of tokenization and synonym resolution and they require improved fallback methods for unknown terms in their sign libraries [4].

The attempt to translate through grammar systems produces unsuccessful results during processing of free- form speech alongside unstructured and informal oral input [5]. Certain devices restrict vocabulary size and respond to unknown terms through a system that reveals letters one at a time thus interrupting natural communication patterns [6]. Any image-based attempts to recognize ISL demonstrate various vocabulary limitations in addition to needing manually curated datasets that decrease real-time functionality [7].

The successful deep learning neural translation models demand large annotated datasets for training although these datasets remain hard to find for regional sign languages such as ISL [8]. ISL text converter systems exist but they do not have speech input capabilities and operate only with prewritten textual information [9]. The performance of gesture recognition techniques including Fourier descriptors and their equivalents deteriorates under changing lighting conditions and camera positions [10].

The use of rule-based language mapping in corpus- building approaches shows effective syntactic correctness although they prove unable to handle new or changing language structures efficiently [11]. The implementation of Kinect-based sign language detection requires specific hardware which restricts its operation in portable applications [12]. The static implementation of CNN and SVM uses fixed image frames to complete gesture recognition however these models struggle in real-world applications with dynamic conditions [13].

The precise nature of Multi-handed ISL systems makes them require complex image preprocessing operations that slow real-time performance [14]. Systems that recognize speech by processing MFCC then match against predefined keywords succeed only when speakers stick to set vocabulary and conventional pronunciation guidelines [15].

Research laboratories conduct studies on continuous sign recognition by using deep video-text embedding models yet these methods demand excessive computing resources which makes them unacceptable for lightweight deployments [16].

Tactile gloves for detecting ISL functions as an immersive hardware-based solution yet their high cost together with limited adoptability makes them impractical [17]. The translators designed for Sign Languages such as Argentinian or Ukrainian fail to work with general audiences because they are bound to specific linguistic rules and cultural settings [18][19]. The translation through synonym substitution in text-to-sign systems does not guarantee meaning retention especially in sentences containing terms that have multiple possible meanings or are context-specific. [20]

3. Proposed System

A browser-based web application which runs on AI power exists as the proposed model for the Audio to Sign Language Converter to support instant verbal-nonverbal exchanges mainly targeting deaf and hard-of-hearing users. The system implements speech recognition combined with natural language processing and multimedia playback to produce sign language animations from spoken words. The designed system offers accessibility and lightweight functionality and interactive features to serve educational and social as well as service-based settings.

Speech Recognition Module

The system enters through this module as users communicate through their microphones for voice input. Real-time speech-to-text conversion functions through the WebkitSpeechRecognition API which explains a client-side web technology. The API detects user voice to convert speech into English text so users can bypass using additional software or back-end systems for processing.

Keyword Processing and NLP Module

The converted speech enters an NLP process following text generation. The module performs word tokenization on the sentence while eliminating every common stopword because these frequently used terms carry no substantial meaning during communication.

This system uses an extraction method to obtain important keywords within the input sentence. The system depends on essential keywords since they convey the essential communication meaning the user wanted to express.
Through the NLP process the system increases its efficiency by reducing the number of translation targets for sign language which leads to improved precision and velocity.

Sign Language Mapping and Animation Module

The system operates by matching each extracted word to its matching sign language animation in its predefined asset library which results in the selection of animations for display once keyword matching is successful. When an animation does not exist in the database the system fallbacks to displaying letters of the word through ISL alphabet signs. The system stops at nothing to deliver effective communication of unknown or uncommon words thus ensuring translation continuity.

User Interface and Playback Module

Developer tools HTML CSS and JavaScript create the interface for users to access through a user-friendly interactive platform. The interface contains controls that enable speech recognition activation as well as viewing generated text results and observing sign animations. A built-in video player within the web page shows the sign language animations one after another when the animations have been established. Users can control playback through three interface functions including play start and reset and pause. Django web framework controls backend operations which handle animation flow management for user sessions and enables smooth intermodule communication.

System Flow

When users activate the microphone button through voice activation the text-to-speech process initiates. WebkitSpeechRecognition serves to immediately convert spoken speech into text when the user activates the microphone button. The system uses extracted key terms from processed text to compare with its sign animation database. The screen shows animations which match the processed information. The system converts unrecognizable words into alphabetical letters through sequential letters of the sign language. The system displays real-time sign language animations as a visual sequence following the completion of the processing flow.

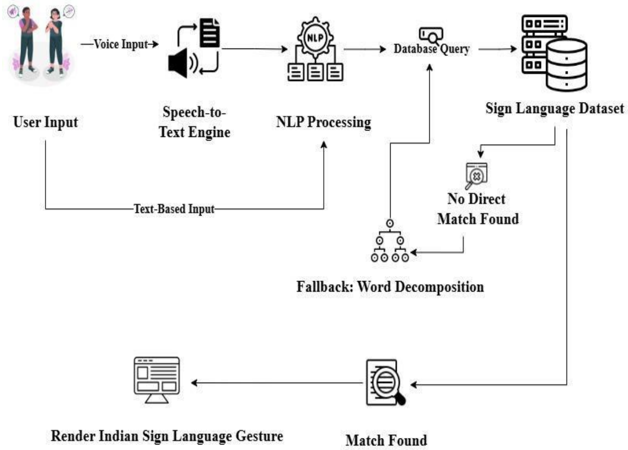

Fig 1. Block diagram of the proposed system

4. System Architecture

The system architecture design of Audio to Sign Language Converter makes it possible for verbal speech users to communicate effortlessly with sign language users. Users shift to an audio input mode through microphone use at the start of processing. The basic input to the system comes from its users' spoken voice data. Real-time conversion of speech occurs through the implementation of WebkitSpeechRecognition API. Digital interpretation of verbal language depends on this transformation since it makes verbal messages ready for subsequent system processing.

The text transcription of speech allows the system to move forward with basic Natural Language Processing (NLP) techniques. The sentence processing begins with tokenization that separates words then proceeds to extract key terms which require translation. The system separates unimportant words from the text so that only the main message components can be analyzed.

The system obtains keywords and looks for corresponding ISL animation entries in a pre-defined Indian Sign Language (ISL) animation database. The system will retrieve ISL animations from its database when it identifies a direct word match. After database retrieval the system will show the video or animated GIF in queue order. The system implements a backup plan to display requested information when database lookup fails to produce results. The system separates words into single characters before it searches for possible ISL representations from a sign database based on alphabet letters. The system operates as a backup to provide visual representation of unknown words through sequential letter delivery from the alphabet- based sign database.

The system delivers the retrieved animations whether they stem from word-based or character-based elements to the display module. Programmable display controls the animation display so viewers understand the broadcasted message in the correct order. The interface demonstrates messages through displays of both complete words or sequential letters to offer viewers straightforward accessible translations.

The system manages time-based speech recognition and NLP analysis to query databases and play animations for real-time sign language presentation to users. The system was created to be both usable and reliable and comprehensive so deaf or hard-of-hearing people can effectively communicate through the integration of verbal speech and visual gestures.

5. Implementation

All essential components within the Audio to Sign Language Converter work together to convert speech into visual sign language outputs in real time. The system integrates web technologies together with Natural Language Processing methods and animation handling protocols through an approach directed toward usability modulation and accessibility across the entire implementation.

Users can interact with the application frontend through HTML and CSS framework together with JavaScript which enables responsive and simple interface. A basic user interface presents system interaction buttons to activate Start, Stop and reset speech recognition functions. The browser microphone enables speech-to-text conversion via WebkitSpeechRecognition API which delivers precise and rapid reactions within the supported browser environment including Google Chrome.
After text generation completes its process JavaScript-based tokenization logic begins to work on the text. The processing separates the sentence into words before eliminating unimportant stop-words such as “is” or “a” or “the” to retrieve essential phrases. The cleaned and filtered text from the English message gets employed to search for matching sign language animations in the Indian Sign Language (ISL) database.

The system retrieves stored sign animations from its database when a word matches database entries which are usually displayed as either video or GIF formats for the user. When the system encounters a word without an exact listing in its database it will split the word into its character components.

It then searches for ISL signs corresponding to each character, effectively spelling out the word using letter-based sign representations.

Animation playback module manages the proper sequence of displayed matched signs throughout the presentation. The module arranges animation sequences to maintain proper timing between movements thus creating outputs that viewers easily understand. The HTML <video> and <img> tags are applied to embed the animation content according to the asset type which is either video or GIF.

The animation platform uses Django as its backend component to supervise the animation assets and manage user interactions with the database. Additionally it operates through a Python-based web framework platform. The database maintains an index which connects ordinary words alongside characters to their corresponding ISL animations.

A modular system configuration with Speech Recognition and Text Processing and Search & Matching followed by Display Module operates to guarantee modular design. The modular system design enables both better maintenance and independent unit improvements such as swapping basic tokenizers with advanced NLP tools like nltk or spaCy.

The system demonstrates functionality for lightweight operation together with browser-compatibility and scalability across different devices with minimal resources. The implementation reaches real-time sign language audio translation through its integration of browser-based APIs coupled with basic natural language processing and multimedia rendering functions which promote digital accessibility for hearing-impaired users.

System	Technology Used	Input/Processing Method	Accuracy (%)	Remarks
Google Live Transcribe	Android Speech API + NLP	Real-time speech-to-text	~85	Provides only text output, no sign language integration
Sign Language Translator (Image-Based)	Image processing + hand gesture recognition	Image input matched with predefined gesture datasets	~80–85	Relies on image quality and hand pose consistency
ISL Gloss Generator (Text to ISL Video)	Text parser + animation renderer	Text-to-ISL gloss translation using pre-recorded sign videos	~87	Limited to known vocabulary and phrase mappings
Voice to Sign Language Mobile App (Basic)	Android Speech API + manual ISL mapping	Speech-to-text with keyword match to ISL image/video	~86	Lacks fallback for unknown words; app-specific support
Proposed Audio to ISL Converter (Web App)	WebkitSpeechRecognition API + NLP + Django + ISL video dataset	Real-time speech input → tokenization → word lookup → fallback to character spelling	89–92	Works in-browser, includes NLP-based parsing, supports unknown words via spelling fallback

Table 1. Comparison of Performance

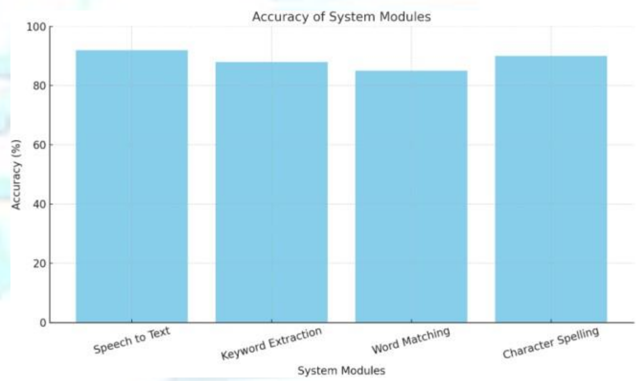

Fig 2. Performance of the proposed system

6. Results and Discussions

Different speech and sign language processing systems are examined through Table 1 which displays their technological approaches combined with processing strategies alongside accuracy results. The Proposed Audio to ISL Converter (Web App) operates at a high level of performance due to its accuracy range between 89–92%. The system succeeds due to its implementation of WebkitSpeechRecognition API alongside NLP and Django along with its unknown word response system based on character-level spelling.

The ISL Gloss Generator and Voice to Sign Language App generated outputs with accuracy levels between 87% and 86%. The systems depend on pre-recorded mappings which reduces their vocabulary flexibility because they lack complex parsing and fallback features. The correctness of the Image-Based Sign Language Translator operated in a range between 80% and 85% because user images and gesture executions were inconsistent.

The performance analysis of system modules in the proposed web application appears in Fig 2. The Speech to Text module reached the highest accuracy level of 92% and was trailed by Character Spelling (90%), Keyword Extraction (88%) and Word Matching (85%). The system demonstrates superior performance throughout its operations because of its multi-layered term-handling strategy for common and out-of-vocabulary terms.

By integrating tabular and graphical evaluations the proposed system demonstrates simultaneous equal and superior performance than current solutions since it provides greater adaptability along with real-time processing and user-friendly web deployment.

7. Conclusion

The Audio to Sign Language Converter takes a major leap to create communication connections between hearing individuals and people who have either hearing difficulties or speech impediments. The current social environment demands advanced inclusive communication tools since verbal interaction functions as an unacknowledged baseline for most societies. The proposed study creates an AI-driven real-time system which changes spoken messages into visual representations of ISL animations to provide speech access for users who cannot communicate verbally.

The system design implements both usability and scalability through its well-thought-out architecture. The WebkitSpeechRecognition API enables the system to work with speech-to-text conversion directly in the browser which eliminates the need for specialized software or additional hardware. Basic Natural Language Processing (NLP) logic processes recognized text while it removes unimportant words to extract key terms for continued processing. The system matches the selected key words against a carefully selected animation database from ISL. When a direct match cannot be found the system will translate the word into individual characters to perform sign language animations at a character level.

The project implements web modules using Microsoft technologies HTML and CSS together with JavaScript and Django. The project's modular structure enables simple system updates including the substitution of the present matching system with deep learning models together with external NLP service incorporation for context assessment enhancement. The current system, although simple in its NLP pipeline, sets a strong foundation for future expansion into more advanced features like emotion recognition, grammar translation, or multilingual support.

The converter serves as an accessible practical deployment method which promotes digital accessibility within society. The device gives deaf and hard-of-hearing people live access to spoken dialogues through automated translation without an interpreter.

Through this project we demonstrate both the technical accessibility and meaningful contribution toward creating inclusive communication systems which use AI with browser-based methods. This assistive tool may achieve wide-scale adoption as an assistive technology throughout India and other territories through additional enhancements to natural language processing capabilities and graphics quality as well as equipment compatibility features.

References

[1] Patel, Bhavinkumar Devendrabhai, et al. "ES2ISL: an advancement in speech to sign language translation using 3D avatar animator." 2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). IEEE, 2020.

[2] Suman, Shashikant, et al. "Two-way Speech to Sign Language Converter application using Python, OpenCV and NLP." 2023 6th International Conference on Contemporary Computing and Informatics (IC3I). Vol. 6. IEEE, 2023.

[3] Dixit, Aditi, et al. "Audio to indian and american sign language converter using machine translation and nlp technique." 2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT). IEEE, 2022.

[4] Neog, Anannya Priyadarshini, Arunabh Kalita, and Ms Nithyakani Pandiyarajan. "Speech/text to indian sign language using natural language processing." Indian Journal of Computer Science and Engineering (IJCSE), 2023.

[5] Kajal Jadhav, Shubham Gangdhar, Viraj Ghanekar: "SPEECH TO ISL (INDIAN SIGN LANGUAGE) TRANSLATOR", International Research Journal of Engineering and Technology, 2021

[6] Nighila Ashok, Aswin K.S., Nikhil Babu P., Sajith Suresh, Sharun K. Suresh. " Speech To Sign Translator." International Journal of Creative Research Thoughts (IJCRT), 2022.

[7] Mariappan, H. Muthu, and V. Gomathi. "Real-time recognition of Indian sign language." 2019 international conference on computational intelligence in data science (ICCIDS). IEEE, 2019.

[8] Camgoz, Necati Cihan, et al. "Neural sign language translation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[9] .Goyal, Lalit, and Vishal Goyal. "Automatic translation of English text to Indian sign language synthetic animations." Proceedings of the 13th International Conference on Natural Language Processing. 2016.

[10] Kharate, Gajanan K., and Archana S. Ghotkar. "Vision based multi-feature hand gesture recognition for Indian sign language manual signs." International Journal on Smart sensing and intelligent systems 9.1 (2016): 124.

[11] Tmar, Zouhour, Achraf Othman, and Mohamed Jemni. "A rule-based approach for building an artificial English-ASL corpus." *2013 international conference on electrical engineering and software applications*. IEEE, 2013.

[12] Çimen, Gökçen. Animation models for interactive AR characters. Diss. ETH Zurich, Gesture-Based Communication with Kinect & Unity3D, 2019.

[13] Katoch, Shagun, Varsha Singh, and Uma Shanker Tiwary. "Indian Sign Language recognition system using SURF with SVM and CNN." Array 14 (2022): 100141.

[14] Dutta, Kusumika Krori, and Anil Kumar GS. "Double handed Indian Sign Language to speech and text." 2015 Third International Conference on Image Information Processing (ICIIP). IEEE, 2015.

[15] Vijaya, J., et al. "An Efficient System for Audio-Based Sign Language Translator Through MFCC Feature Extraction." 2023 International Conference on Sustainable Communication Networks and Application (ICSCNA). IEEE, 2023.

[16] Bull, Hannah, et al. "Aligning subtitles in sign language videos." Proceedings of the IEEE/CVF international conference on computer vision. 2021.

[17] Kunjumon, Jinsu, and Rajesh Kannan Megalingam. "Hand gesture recognition system for translating indian sign language into text and speech." *2019 international conference on smart systems and inventive technology (ICSSIT)*. IEEE, 2019.

[18] Ronchetti, Franco, et al. "LSA64: An Argentinian sign language dataset." arXiv preprint arXiv:2310.17429 (2023).

[19] Olga, Lozynska, Savchuk Valeriia, and Pasichnyk Volodymyr. "The sign translator information system for tourist." 2019 IEEE 14th International Conference on Computer Sciences and Information Technologies (CSIT). Vol. 3. IEEE, 2019.

[20] Shovkovyi, Yevhenii, et al. "Automatic sign language translation system using neural network technologies and 3D animation." (2023).