International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(REVIEW ARTICLE)

Check for updates

Bridging communication gaps: An AI-powered real-time system for sign language, speech and text translation

Naga Sasha Lakshmi Pokanati, Monika Devi Imandi, Yamini Sariki *, Sivaram Sangula and Nagendra Vasamsetti

Department of CSE Aditya College of Engineering and Technology, Surampalem, Andhra Pradesh, India – 533437

International Journal of Science and Research Archive, 2025, 14(03), 1331-1336

Publication history: Received on 13 February 2025; revised on 21 March 2025; accepted on 24 March 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.14.3.0807

Abstract

Communication effectiveness poses an essential challenge to the millions who have hearing or speech impairments in their lives. HandSpeak provides a real-time AI interface through which users can interact easily because it transforms sign language into written messages and spoken words. The Sign-to-Speech module performs its functions through the integration of 3D cameras with Convolutional Neural Networks (CNNs) as well as Long Short-Term Memory (LSTM) networks to detect hand landmarks while tracking hand movements and interpreting sign gesture temporal sequences. Speech input flows through the Speech-to-Sign module to produce text output that gets processed into animated sign language expressions under AI avatar operations. Transformers boost linguistic precision and the integration between Django and Flask provides users with an optimized web-based interface. The application uses SQLite for optimized data storage together with Blender for producing sign animations. The technology serves to create an barrier-free environment for natural communication which enhances inclusivity while providing better accessibility to hearing and speech disabled people in various social and professional environments.

Keywords: Sign Language Translation; Real-Time Communication; Bidirectional; Tracking hand movements; Convolutional Neural Networks (CNN); Long Short-Term Memory (LSTM)

1. Introduction

Although digital communication methods increase accessibility and efficiency most people with hearing and speech impairments experience considerable difficulties while interacting with others. The rich yet expressive communication system known as sign language exists as an inaccessible mode for most people which leads to educational and employment problems alongside social integration difficulties. The conventional methods involving interpreters and text-based platforms function with restrictions that limit their availability and their real-time functionality and their high cost accessibility. New technology solutions remain essential to create inclusive environments which give power to communication disabled individuals to overcome exclusion barriers.

HandSpeak stands as an advanced AI system which delivers real-time communication translation between sign language, speech and text function to connect disconnected language groups. HandSpeak operates through its dual sub-modules Sign-to-Speech which transforms sign language to either audio or written content and Speech-to-Sign that converts speech into computer-generated signed language managed through AI-run virtual characters. The mixture of convolutional neural networks (CNNs) and Long Short-Term Memory (LSTM) networks and 3D animation allows HandSpeak to provide both fast and highly accurate communication. The HandSpeak platform uses Flask alongside Django frameworks to create a simple user-oriented interface accessible by all types of users. HandSpeak serves as a revolutionary tool since it establishes improved communication possibilities for school learning as well as work

* Corresponding author: S Yamini

environments and social spaces while aiming to build an all-inclusive environment between individuals who hear or speak differently

2. Block diagrams

2.1. Sign language to Speech/Text

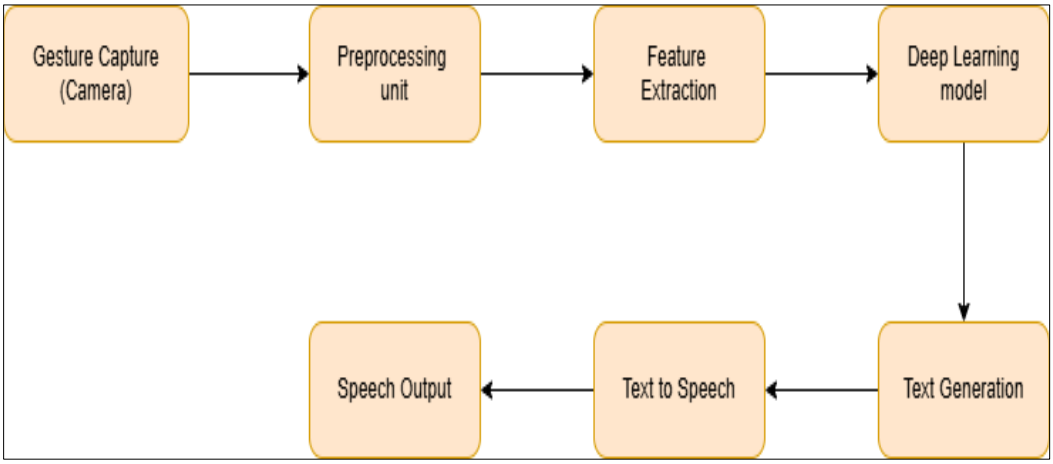

Figure 1 Sign language to Speech/Text conversion workflow

The HandSpeak Sign-to-Speech translation system functions according to the instructions within the diagram. The system starts its process through camera gesture capture which records hand movements together with sign language gestures. Data processing occurs first at the unit before executing the noise reduction operations alongside hand tracking and segmentation operations. The successive operation eliminates fundamental frames while performing analysis. The deep learning model obtains gesture frame data through its input to process spatial contents using CNNs and temporal patterns using LSTMs for gesture classification. User gestures undergo conversion into text until the text-to-speech production transforms them into real-time speech for communicating between users.

2.2. Speech/Text to Sign language

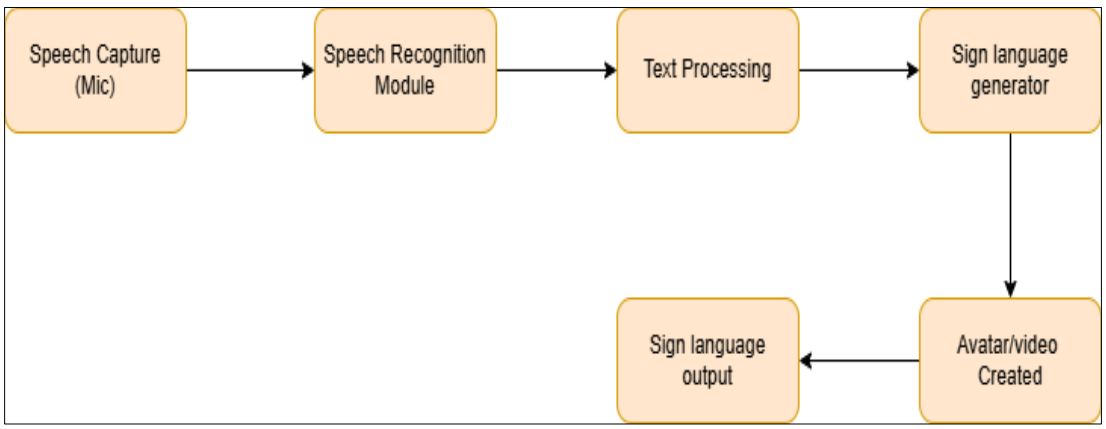

Figure 2 Speech/Text to Sign language conversion workflow

Within the HandSpeak program block diagram the Speech-to-Sign translation sequence takes place. Audio recordings are the first step when the voice capture function utilizes the microphone system. Both speech recognition algorithms and advanced processing algorithms convert audio data into written text through the speech recognition module. Through NLP the text processing module arranges information when it handles text inputs. The sign language generator applies processing results from the previous step to generate sign language gestures. Realistic gestures for animated animations are generated through the combination of 3D avatar/video development tools like Blender. The sign language output featured by real-time visual animations allows users to sustain continuous communication.

3. Methodology

This section outlines the structured approach used to design and develop HandSpeak, an interactive system that facilitates real-time communication for individuals with hearing and speech impairments. The methodology consists of six key components: System Architecture, Data Acquisition & Processing, Model Development, Speech-to-Gesture & Gesture-to-Speech Conversion, 3D Sign Language Animation, and System Integration & Evaluation.

3.1. System Architecture

The proposed system follows a layered modular architecture to ensure scalability, accuracy, and efficiency. It consists of the following components:

- **Input Module** – Captures hand gestures through a vision-based system and collects speech input via an audio processing module.
- **Processing Unit** – Implements deep learning-based recognition models, NLP-based text processing, and translation algorithms to facilitate real-time gesture and speech interpretation.
- **Output Mechanism** – Converts text into sign animations using 3D avatars and generates spoken output from recognized gestures through speech synthesis techniques.
- **Backend & Database** – The system backend, developed using Flask and Django, handles user authentication, database operations, and data storage using SQLite. RESTful APIs are utilized for seamless communication between the frontend and backend.

3.2. Data Acquisition & Processing

The system development began with extensive data collection to ensure high-accuracy recognition and translation.

- **Gesture Data Collection** – Captured through a 3D camera setup, recording video samples of various sign language gestures alongside their corresponding text annotations.
- **Speech Data Collection** – Audio recordings were obtained and stored with transcribed text labels to facilitate speech-to-text conversion.
- **Preprocessing Methods** – Applied techniques such as grayscale conversion, adaptive thresholding, Gaussian blur, and feature extraction (HOG, SIFT, MFCC) to enhance gesture and speech recognition accuracy.

3.3. Model Development

The system employs deep learning models to achieve high recognition accuracy for gestures and speech.

- **Gesture Recognition** – Utilizes Convolutional Neural Networks (CNNs) for extracting spatial gesture patterns and Long Short-Term Memory (LSTM) networks for modeling the temporal sequence of sign movements.
- **Speech-to-Text Processing** – Implemented using Transformer-based deep learning models, which learn from transcribed datasets to convert spoken input into structured text.
- **Text-to-Gesture Translation** – Employs Natural Language Processing (NLP) techniques, including tokenization, parsing, and sequence mapping, to transform text input into corresponding sign language gestures.

3.4. Speech-to-Gesture & Gesture-to-Speech Conversion

To facilitate bidirectional communication, the system integrates:

- **Gesture-to-Speech Conversion** – Identified sign language gestures are mapped to their corresponding spoken words using NLP-based translation and speech synthesis models.
- **Speech-to-Gesture Conversion** – Recognized text from speech input is processed using language models, which translate the content into sign language animations, allowing non-verbal communication.

3.5. 3D Sign Language Animation

To provide an interactive and visually accurate representation of sign language, the system includes:

- **3D Avatar-Based Animation** – Uses **Blender** to generate realistic sign language gestures corresponding to textual input.

- **Real-Time Rendering** – Converts recognized text into dynamic motion sequences, enabling a smooth user experience for sign language interpretation.

3.6. System Integration & Evaluation

The final phase of development focused on integrating all components and evaluating system performance.

- **Frontend Development** – The user interface was developed using HTML, CSS, and JavaScript, ensuring ease of use and accessibility.
- **Backend & Database Management** – Flask and Django manage application logic and user authentication, while SQLite stores gesture-speech mappings and system data.
- **Testing & Validation** – The system underwent:
    - **Integration Testing** – Verified end-to-end functionality and communication between different system modules.
    - **User Evaluation** – Involved deaf and mute participants, who provided feedback on usability and real-world applicability.
- **Performance Metrics** – The system was assessed based on recognition accuracy, response time, translation precision, and scalability to ensure reliable performance.

4. Objectives of project

- The development of a real-time communication system requires multilingual capabilities turning sign language into speech and text to make communication seamless for people who suffer from hearing difficulties.
- The platform needs to construct an accommodating framework through efficient communication tools designed especially for deaf and mute users who need minimal assistance from human interpreters in their social networking and educational work and professional activities.
- The user interface design combines features to ensure enjoyable interactions between deaf and hearing users when they engage with the system.
- The system supports two essential functionalities that provide bidirectional translation between sign language and spoken or written language and spoken language and animated sign language communication.
- Multiple dataset training techniques and noise filtering and data preparation methods will enable the system to reach high recognition accuracy levels and robust performance.
- The system must have scalable design properties which enable it to fulfil diverse sign language requirements between different users accessing different applications spanning multiple regions.
- A 3D avatar system needs to generate highly realistic sign language animations to make sign language output more understandable and stronger.
- The technology facilitates better communication by uniting hearing and non-hearing people through improved social integration because it reduces gaps in social hearing support.
- The developed innovative solutions enable researchers to conduct new research developments in real-time gesture and speech recognition through artificial intelligence and computer vision and natural language processing.

5. Output

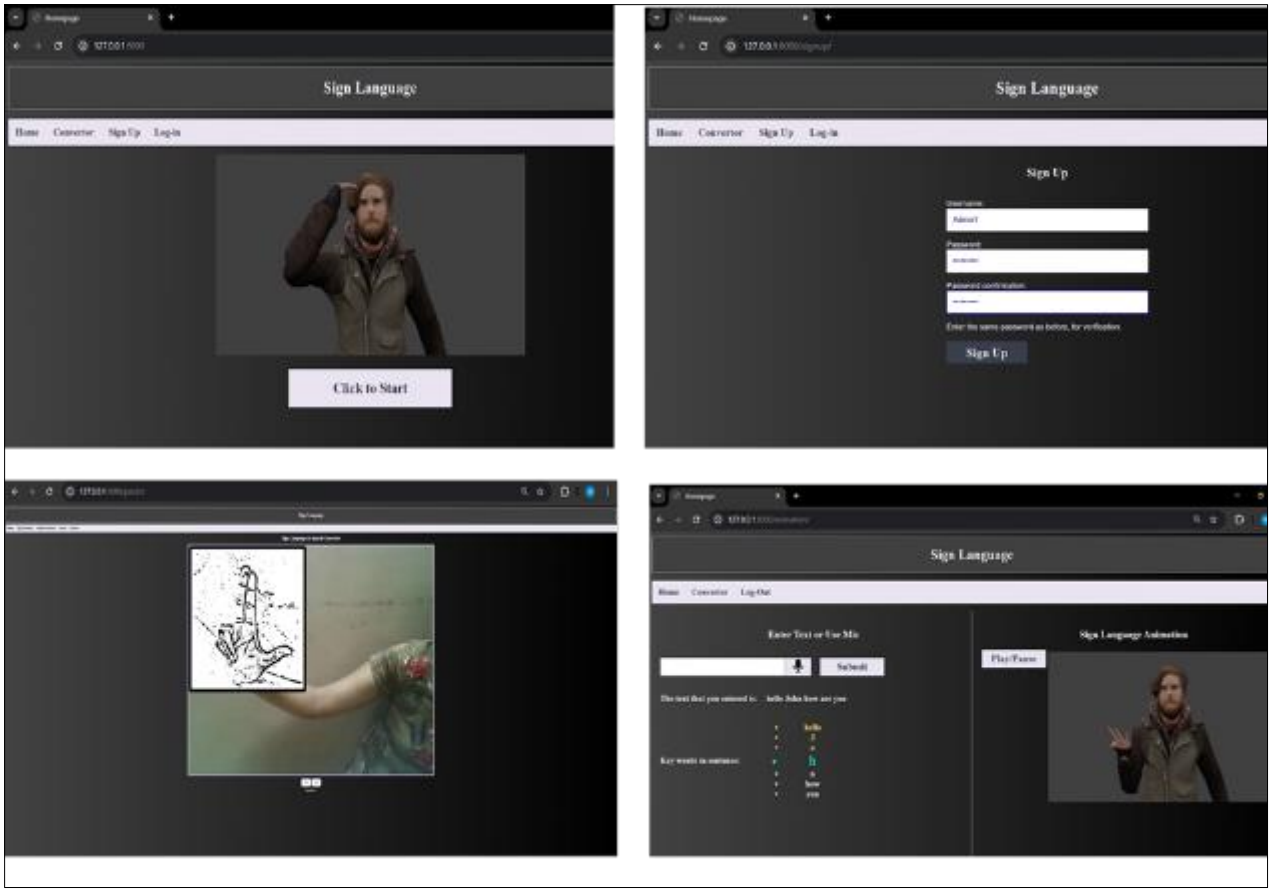

Figure 3 Output Screens

6. Conclusion

A two-way real-time communication system through HandSpeak allowed people with hearing and speech disabilities to interact seamlessly with others. The 3D avatar improvement system made the system more usable by providing users with an easy-to-use interface that was highly interactive. HandSpeak system operates successfully because it continues moving forward with inclusivity across educational facilities and workplaces as well as social environments regardless of its constraints in input quality. Future development phases of the system will focus on language functionality improvement as well as better performance capabilities. With HandSpeak as a communication platform users succeed in exchanging messages which enhances mutual understanding and builds social inclusivity networks.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Patwary Muhammed, Parvin Shahnaj and Akter Subrina, "Significant HOG-Histogram of Oriented Gradient Feature Selection for Human Detection", International Journal of Computer Applications, vol. 132, pp. 20-24, 2015.

[2] Swarnakar, Partha Sarathi, Chakraborty, Ujjal, Swarnakar, Palash, Biswas, Ashim Kumar, and Nandi, Arnab, "Enhancing Communication: A Review on Applications of AI-Based Wearables for the Deaf and Mute", Artificial Intelligence in Healthcare, Elsevier, 2025.

[3] Vivekanand, Maharshi, R G, Shrinivas, Kumar, Piyush, C G, Rahul, and Gowda, Naveen Chandra, "IoT Based Model Bridge Between Deaf and Mute Community with Normal People", Zenodo, 2025.

[4] Saeed Khalid, Tabedzki Marek, Rybnik Mariusz and Adamski Marcin, "K3M: A universal algorithm for image skeletonization and a review of thinning techniques", Applied Mathematics and Computer Science, vol. 20, pp. 317-335, 2010.

[5] Kumar Pradeep, Gauba Himaanshu, Roy Partha and Dogra Debi, "A Multimodal Framework for Sensor based Sign Language Recognition", Neurocomputing, vol. 259, 2017.

[6] M. Khan, S. Chakraborty, R. Astya and S. Khepra, "Face Detection and Recognition Using OpenCV", 2019 International Conference on Computing Communication and Intelligent Systems (ICCCIS), pp. 116-119, 2019.

[7] Ur Rehman, Muneeb, Ahmed, Fawad, Khan, Muhammad Attique, Tariq, Usman, Alfouzan, Faisal Abdulaziz, Alzahrani, Nouf M., and Ahmad, Jawad, "Dynamic Hand Gesture Recognition Using 3D-CNN and LSTM Networks", Computers, Materials & Continua, vol. 70, no. 3, pp. 4675-4690, 2022.

[8] Tao, Tangfei, Zhao, Yizhe, Liu, Tianyu, and Zhu, Jieli, "Sign Language Recognition: A Comprehensive Review of Traditional and Deep Learning Approaches, Datasets, and Challenges", IEEE Access, vol. 9, pp. 99300-99318, 2021.

[9] Paul, Snehasish, and Chauhan, Shivali, "Enhancing Accessibility in Special Libraries: A Study on AI-Powered Assistive Technologies for Patrons with Disabilities", arXiv, vol. 2411.06970, 2024.

[10] Sobhan, M., Chowdhury, M.Z., Ahsan, I., Mahmud, H., & Hasan, M.K. (2019). "A Communication Aid System for Deaf and Mute using Vibrotactile and Visual Feedback". In Proceedings of the 2019 International Seminar on Application for Technology of Information and Communication (pp. 15-20). IEEE.

[11] Skaria, S., Al-Hourani, A., & Evans, R.J. (2020). Deep-learning methods for hand-gesture recognition using ultra-wideband radar. IEEE Access, 8, 203580-203590.

[12] Molchanov, P., Yang, X., Gupta, S., Kim, K., Tyree, S., & Kautz, J. (2016). Online detection and classification of dynamic hand gestures with recurrent 3D convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4207-4215).