IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.151143

Multilingual Speech-to-Sign Language Translator with Avatar

Chandana A C [1], Sandarsh Gowda M .M [2]

Department of MCA, BIT, K.R. Road, V.V. Pura, Bangalore, India[1]

Assistant Professor, Department of MCA, BIT, K.R. Road, V.V. Pura, Bangalore, India[2]

Abstract: Communication barriers between hearing-impaired and hearing individuals remain a significant challenge in everyday interactions. While spoken and written languages are widely supported by modern technologies, sign language communication still lacks accessible and real-time translation solutions. This project presents a **Multilingual Speech-to-Sign Language Translator with Avatar**, designed to bridge this communication gap using artificial intelligence and human–computer interaction techniques.

The proposed system accepts user input in the form of **speech or text**, converts it into a target language using multilingual translation models, and represents the translated content through a **3D animated sign language avatar**. In addition, the system integrates **real-time hand gesture recognition** using computer vision techniques to identify basic sign gestures and map them to corresponding textual meanings. This bidirectional interaction enables both hearing and hearing-impaired users to communicate more naturally.

The system architecture combines **speech recognition, language translation, text-to-speech synthesis, gesture detection, and avatar animation** into a unified web-based platform. By processing inputs locally and rendering sign outputs visually, the system ensures low latency and improved user experience. Experimental evaluation demonstrates accurate speech recognition, smooth avatar animation, and effective translation across multiple languages.

The proposed solution offers an affordable and scalable assistive communication tool that can be deployed in educational institutions, public service centers, and social interaction platforms. By enhancing accessibility and inclusivity, this work contributes toward improving digital communication for the hearing-impaired community while supporting multilingual interaction in real time.

Keywords: Speech-to-Sign Translation, Sign Language Avatar, Gesture Recognition, Multilingual Translation, Assistive Technology, Human-Computer Interaction

I. INTRODUCTION

Communication barriers between hearing-impaired individuals and the general public remain a major social challenge, as sign language is not commonly understood by everyone. Although speech and text translation technologies have advanced significantly, effective solutions for converting spoken or written language into sign language are still limited, especially in real-time scenarios. This project presents a **Multilingual Speech-to-Sign Language Translator with Avatar** that enables users to communicate using speech or text, which is then translated into a selected language and visually represented through an animated sign language avatar. The system also supports real-time hand gesture recognition to convert sign language gestures into text, enabling two-way communication. By integrating speech recognition, language translation, gesture detection, and avatar-based animation into a single web-based platform, the proposed system aims to improve accessibility, inclusivity, and ease of communication for the hearing-impaired community across multilingual environments.

1.1 Project Description

The **Multilingual Speech-to-Sign Language Translator with Avatar** is a web-based intelligent communication system designed to reduce the gap between hearing-impaired individuals and normal users. The system accepts user input in the form of speech, typed text, or hand gestures. Spoken input is converted into text using speech recognition, while typed input is directly processed by the system. The extracted text is translated into multiple languages and simultaneously converted into sign language using an animated avatar. In addition, the system includes a real-time gesture recognition module that captures hand movements through a webcam and converts recognized signs into readable text. By combining multilingual translation, text-to-speech output, and avatar-based sign representation, the project provides an interactive and inclusive platform that supports effective two-way communication.

1.2 Motivation

Communication barriers faced by individuals with hearing and speech impairments remain a major challenge in daily life, education, and professional environments. Most people are unfamiliar with sign language, which limits effective

DOI: 10.17148/IJARCCE.2026.151143

interaction and often leads to social exclusion. Existing communication tools are either language-restricted, lack real-time interaction, or do not provide an intuitive visual representation of sign language. The motivation behind this project is to develop an inclusive system that enables smooth communication between deaf and hearing individuals using modern technologies such as speech recognition, machine translation, and computer vision. By integrating multilingual support and an animated sign language avatar, the system aims to make communication more accessible, natural, and user-friendly. This project is driven by the goal of promoting social inclusion and equal access to communication through technology.

II. RELATED WORK

Paper **[1]** presented a sign language recognition system using image processing techniques to detect hand gestures and convert them into text. Although the system showed reasonable accuracy, it required controlled lighting conditions and lacked real-time performance.

Paper **[2]** proposed a speech-to-text translation framework using deep learning models for multilingual communication. While effective for language translation, the system did not support sign language output or visual representation through avatars.

Paper **[3]** introduced a gesture recognition model based on convolutional neural networks for real-time applications. However, the solution was limited to predefined gestures and supported only a single language.

Compared to these approaches, the proposed system integrates multilingual speech translation, real-time hand gesture recognition, and avatar-based sign language output, making it more inclusive and practical for real-world communication.

III. METHODOLOGY

A. System Environment

The proposed system operates in a web-based client–server environment that supports real-time speech, text, and gesture processing. The frontend is developed using HTML, CSS, and JavaScript to manage user interaction, language selection, and avatar-based sign language visualization. Webcam support and MediaPipe libraries are used for real-time hand gesture detection and tracking. The backend is implemented using Python and Flask to handle speech recognition, multilingual translation, and text-to-speech services. This environment ensures efficient processing, scalability, and smooth communication between spoken languages and sign language representation.

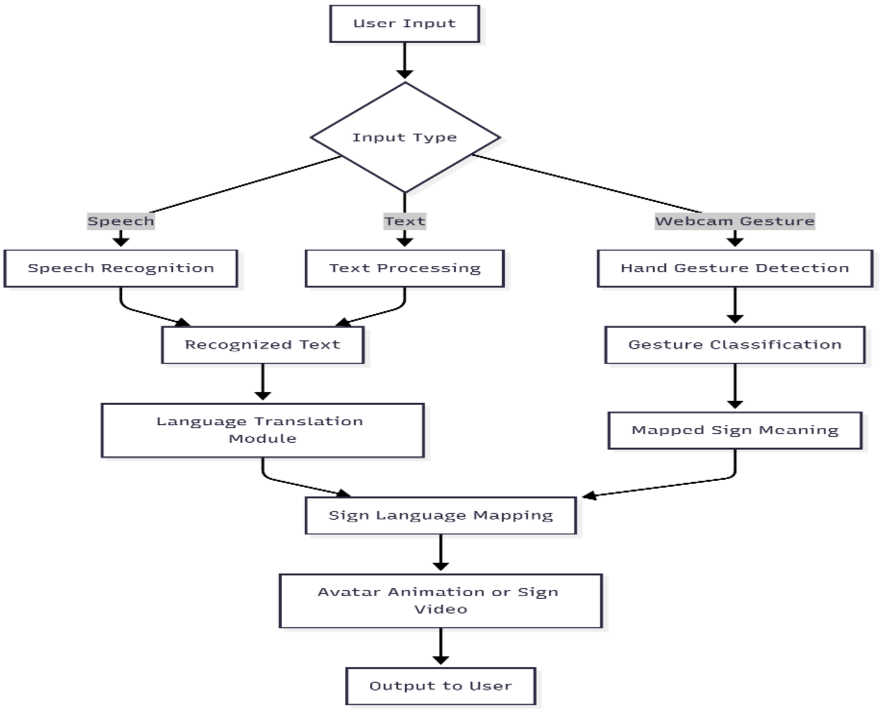


Fig.1.Flowchart of methodology

DOI: 10.17148/IJARCCE.2026.151143

B. System Architecture Description

Client-Side Processing

The client side operates within a web browser and manages all user interactions. Users can provide input in the form of text, speech, or hand gestures using a keyboard, microphone, and webcam. The frontend preprocesses these inputs and sends structured requests to the server. It also handles real-time gesture capture using computer vision techniques and displays outputs such as translated text, synthesized speech, and animated sign language through a virtual avatar.

Server-Side Processing

The server side is responsible for core language and speech processing tasks. It receives requests from the client and performs speech-to-text conversion, multilingual translation, and text-to-speech synthesis. The server also manages sign language mapping logic and sends the processed results back to the client. This separation ensures secure processing, reduced client workload, and smooth real-time communication.

C. Multimodal Translation and Sign Generation Mechanism

The system integrates multiple input modalities including speech, text, and hand gestures to support inclusive communication. Spoken or typed input is translated into the selected target language, while gesture inputs are recognized and interpreted using hand landmark detection. The translated output is then converted into sign language using an animated avatar, enabling clear visual representation. This adaptive mechanism improves accessibility and ensures effective communication across language and hearing barriers

D. Implementation Flow

1. Initialise the web-based client interface and backend server.
2. Accept user input through text, speech, or webcam-based gestures.
3. Preprocess input data on the client side.
4. Send processed input to the server for translation and speech processing.
5. Perform speech-to-text, language translation, and text-to-speech conversion on the server.
6. Map translated content to corresponding sign language representations.
7. Display translated text, audio output, and avatar-based sign animation on the client.
8. Repeat the process for continuous real-time interaction.

E. Hardware and Software Requirements

Hardware Requirements:

- Standard desktop or laptop system
- Minimum 8 GB RAM
- Webcam and microphone for gesture and speech input

Software Requirements:

- Operating System: Windows / Linux / macOS
- Frontend: HTML, CSS, JavaScript
- Backend: Python 3.8 or above with Flask framework
- Libraries: MediaPipe for gesture recognition, speech and translation APIs, text-to-speech modules
- Web Browser: Modern browser with webcam support (Chrome, Edge, or Firefox)

IV. SIMULATION AND EVALUATION FRAMEWORK

This section explains the system design, execution flow, and evaluation approach used for the proposed Multilingual Speech-to-Sign Language Translator with Avatar. The framework integrates speech processing, multilingual translation, gesture recognition, and avatar-based sign language visualization to enable accessible and real-time communication. The system is implemented using a web-based frontend and a Python-based backend to ensure responsiveness and scalability.

A. System Architecture and Workflow

The proposed architecture supports multimodal communication while ensuring real-time performance. It follows a client–server model with integrated gesture recognition and avatar-based sign visualization. The main components are described below:

DOI: 10.17148/IJARCCE.2026.151143

[1] Client Interface Module

The client interface is a web-based platform that enables users to interact with the system using multiple input methods such as text entry, voice input, and hand gestures through a keyboard, microphone, and webcam. It presents the translated output in text form, provides audio playback, and visually displays sign language using an animated virtual avatar.

[2] Processing and Translation Server

This server module handles all core processing tasks, including speech-to-text conversion, multilingual language translation, and text-to-speech generation. It receives requests from the client interface, processes the data efficiently, and sends back translated text along with pronunciation information and audio output.

[3] Gesture Recognition and Avatar Module

The gesture recognition module captures hand movements through the webcam and interprets them using computer vision-based techniques. Identified gestures are matched with appropriate sign language representations, which are then displayed as animations through a virtual avatar to enable clear visual communication.

B. Simulation Setup

The simulation setup is developed to assess the performance of the system under realistic operating conditions. It evaluates the system’s behavior with various input methods and multiple language scenarios to ensure reliability and effectiveness in real-world applications.

Input Configuration

The system is tested using different forms of input, including text entered through the keyboard, spoken voice input, and real-time hand gesture input. This testing helps measure the system’s adaptability, accuracy, and response across multiple interaction modes.

Language and Translation Testing

Multiple target languages are chosen to analyze the accuracy and efficiency of multilingual translation and pronunciation output. This evaluation ensures consistent performance across diverse languages and linguistic conditions.

C. Translation and Sign Processing Workflow

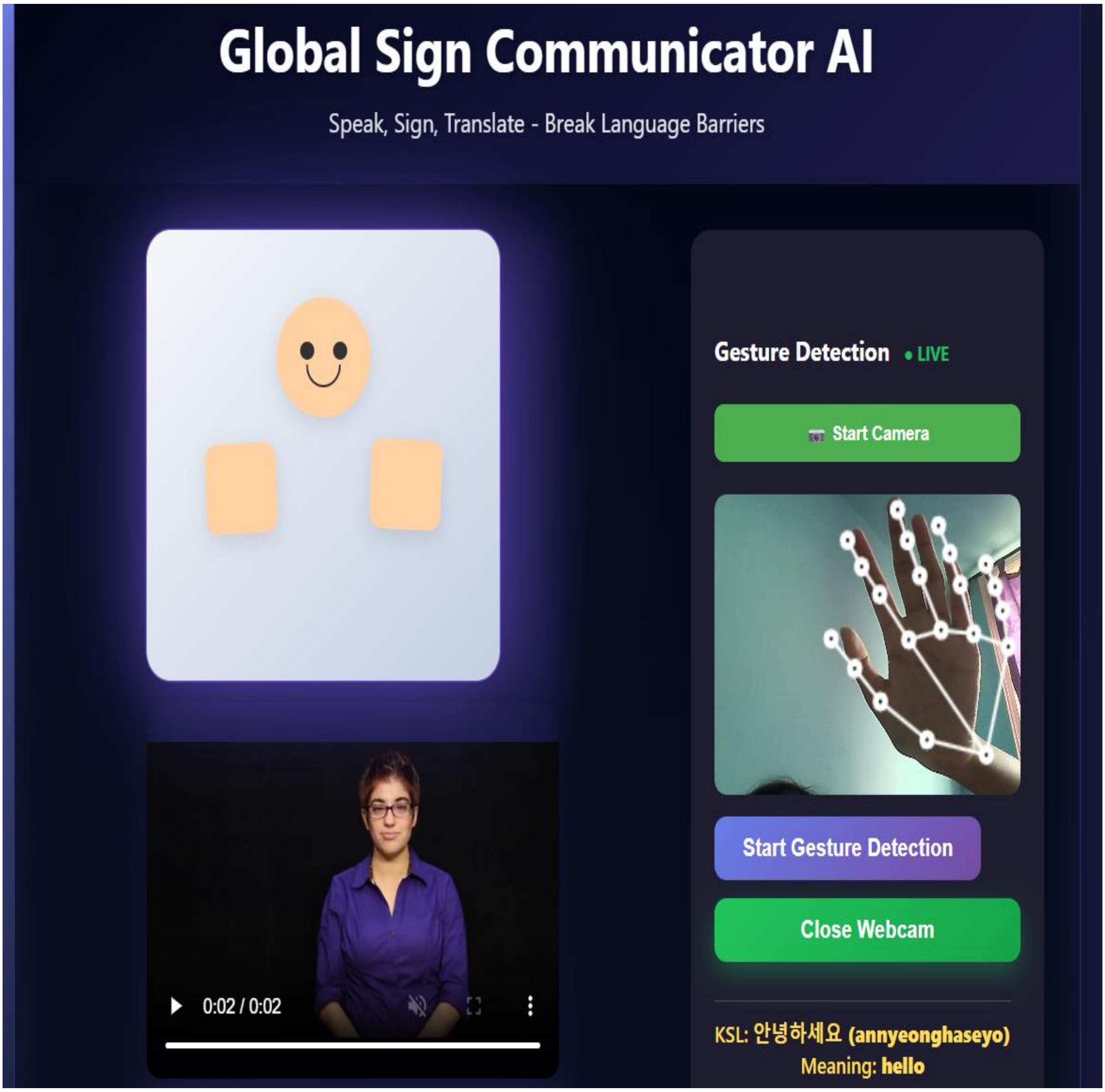
During system operation, user input is initially captured through the client interface and forwarded to the server for further processing. The server handles language translation and speech generation, whereas gesture-related data is processed locally using MediaPipe for efficient recognition. After processing, the results are sent back to the client and presented as translated text, audio output, and animated sign language using a virtual avatar. This continuous cycle supports seamless real-time communication.

a.Results and Observations

System Performance

- The system effectively converted both voice and text inputs into multiple languages while accurately displaying corresponding sign language visuals.
- Real-time hand gesture detection combined with avatar-based animations significantly enhanced user accessibility and engagement.
- The client–server model provided quick response times and maintained stable performance across various input methods.

DOI: 10.17148/IJARCCE.2026.151143



Global Sign Communicator AI

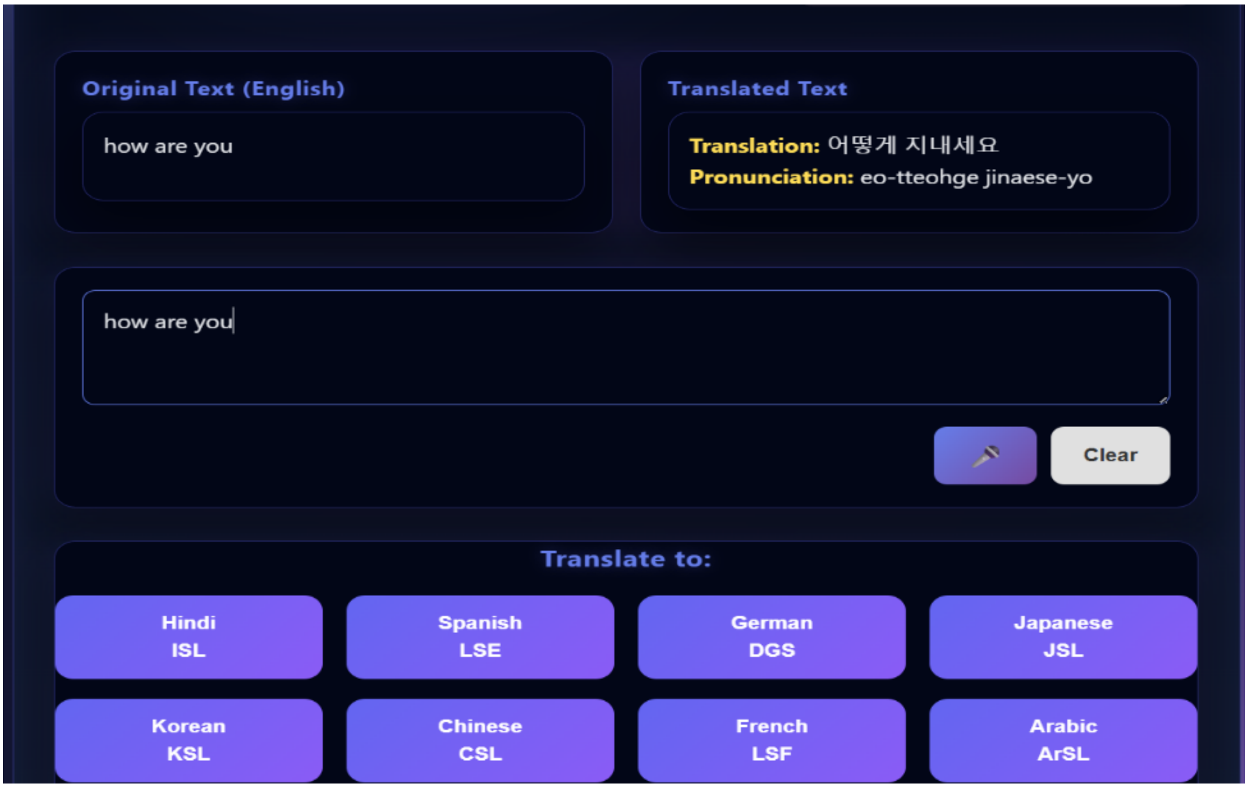
Speak, Sign, Translate - Break Language Barriers

KSL: 안녕하세요 (annyeonghaseyo)
Meaning: **hello**

Fig. 2. Sample output showing live Gesture Detection and Avatar

Figure 2 illustrates the live output of the Multilingual Speech-to-Sign Language Translator with Avatar during real-time gesture recognition. The system captures hand movements through the webcam and applies computer vision techniques to identify sign language gestures using hand landmark detection. Once a gesture is recognized, the corresponding sign meaning is displayed along with its text interpretation, while the animated avatar visually represents the same sign for better understanding. This integrated output demonstrates the system's ability to combine live gesture detection, visual avatar-based sign representation, and semantic interpretation, enabling effective communication support for hearing-impaired users.

DOI: 10.17148/IJARCCE.2026.151143



Original Text (English)

how are you

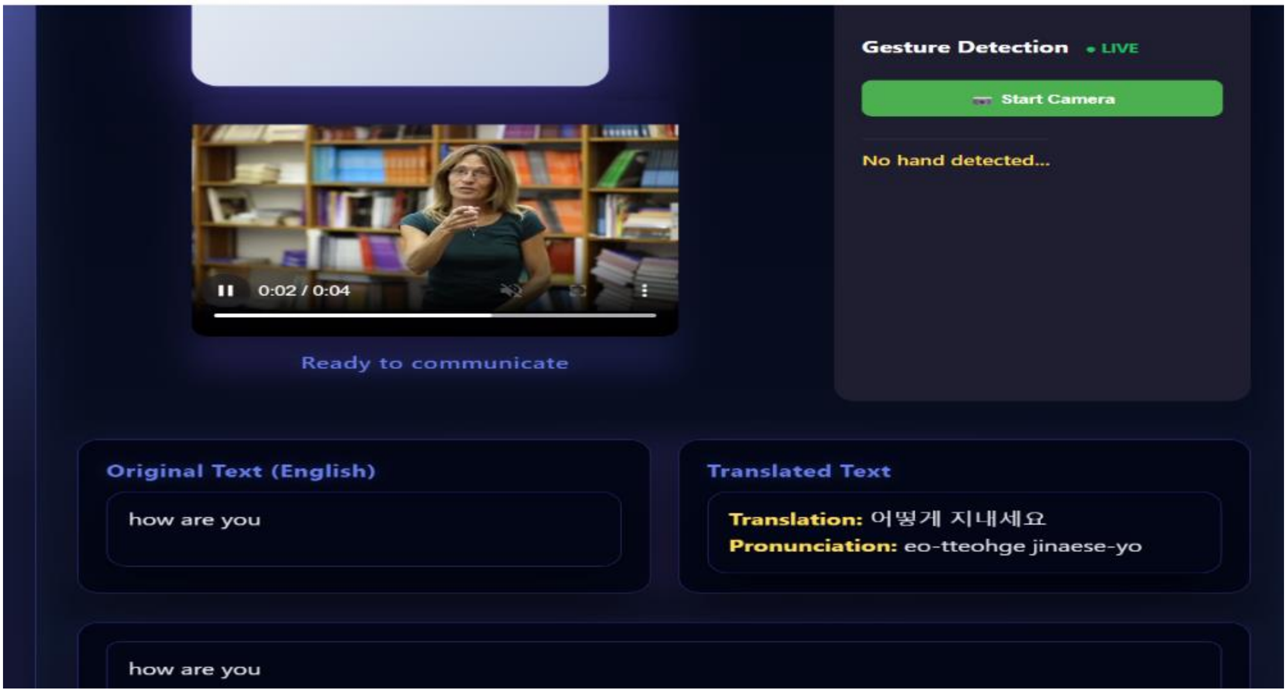
Translated Text

Translation: 어떻게 지내세요
Pronunciation: eo-tteohge jinaese-yo

how are you

Fig. 3. Multilingual Text Translation Interface

This figure shows the text-based translation module where English input is converted into a selected target sign language with corresponding translated text and pronunciation. It demonstrates the system's ability to support multilingual translation through an interactive input area and language selection buttons.



Gesture Detection • LIVE

No hand detected...

Original Text (English)

how are you

Translated Text

Translation: 어떻게 지내세요
Pronunciation: eo-tteohge jinaese-yo

how are you

Fig. 4. Real-Time Sign Output and Translation Result

DOI: 10.17148/IJARCCE.2026.151143

This figure illustrates the conversion of English text input into translated text along with a corresponding sign language video. It confirms successful integration of text processing, multilingual translation, and sign video playback in the system.

V. RESULTS AND DISCUSSION

The Multilingual Speech-to-Sign Language Translator with Avatar was successfully implemented and evaluated under different input scenarios, including text input, speech input, and hand gesture recognition. The system demonstrated reliable performance in converting spoken or written language into translated text and corresponding sign language animations using an avatar. Real-time speech recognition worked effectively with minimal delay, and the translated output was displayed clearly along with pronunciation support.

The gesture recognition module accurately detected basic hand signs through the webcam using computer vision techniques. Recognised gestures were correctly mapped to predefined sign representations, enhancing communication for users with hearing or speech impairments. The avatar-based sign visualization provided an intuitive and user-friendly way to understand translated content, making the system accessible even for non-technical users.

Overall, the results indicate that the proposed system achieves effective multimodal communication with good responsiveness and usability. While the system performs well for supported languages and predefined gestures, its accuracy depends on lighting conditions, gesture clarity, and network stability. These observations highlight the system’s practical applicability and also indicate scope for future improvements such as expanding gesture datasets and adding more language support.

VI. CONCLUSION

This project presented the design and implementation of a Multilingual Speech-to-Sign Language Translator with Avatar aimed at improving communication accessibility for individuals with hearing and speech impairments. By integrating speech recognition, multilingual translation, gesture recognition, and avatar-based sign language visualization, the system enables effective interaction across different communication modes.

The developed system successfully converts spoken or written language into translated text and corresponding sign language animations in real time. The use of a web-based interface, combined with computer vision and language processing techniques, ensures ease of use and platform independence. Experimental observations show that the system performs reliably under normal operating conditions and provides meaningful support for inclusive communication.

Although the current implementation supports limited gestures and languages, it establishes a strong foundation for future enhancements. The system can be extended by incorporating advanced deep learning models, expanding sign language datasets, and adding support for additional regional and international languages. Overall, the proposed solution demonstrates practical applicability and contributes toward bridging communication gaps through intelligent and accessible technology.

VI. FUTURE WORK

The Multilingual Speech-to-Sign Language Translator with Avatar can be further enhanced in several directions to improve accuracy, usability, and real-world applicability. Future work may focus on expanding the sign language database to include a wider range of gestures, facial expressions, and regional sign variations, enabling more natural and expressive communication.

Advanced deep learning models can be integrated to improve speech recognition, gesture detection, and translation accuracy, especially in noisy environments or for continuous sign recognition. Support for additional languages and dialects can also be incorporated to make the system globally accessible. Furthermore, real-time emotion recognition and lip-syncing with the avatar can enhance the visual clarity and realism of sign language output.

The system can be extended to mobile and wearable platforms, allowing users to communicate on the go. Integration with assistive devices, educational tools, and public service systems would further increase its impact. With these improvements, the proposed solution has strong potential to evolve into a comprehensive and intelligent communication aid for inclusive and accessible interaction.

DOI: 10.17148/IJARCCE.2026.151143

REFERENCES

[1] T. Starner, J. Weaver, and A. Pentland, “Real-time American Sign Language recognition using desk and wearable computer-based video,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1371–1375, 1998. https://ieeexplore.ieee.org/document/735338

[2] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, “OpenPose: Realtime multi-person 2D pose estimation using part affinity fields,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172–186, 2021. https://ieeexplore.ieee.org/document/8765346

[3] Google, “MediaPipe: Cross-platform machine learning solutions,” MediaPipe, 2023. [Online]. https://mediapipe.dev

[4] M. Lewis, A. Courville, and Y. Bengio, “Deep learning for sign language recognition: A survey,” *ACM Computing Surveys*, vol. 52, no. 4, pp. 1–36, 2019. https://dl.acm.org/doi/10.1145/3363578

[5] A. Vaswani *et al.*, “Attention is all you need,” in *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5998–6008. https://papers.nips.cc/paper/7181-attention-is-all-you-need

[6] F. Chollet, *Deep Learning with Python*, 2nd ed. Manning Publications, 2021. https://www.manning.com/books/deep-learning-with-python

[7] S. Prasad, R. Sharma, and P. Kumar, “Speech-to-sign language translation system using natural language processing,” *International Journal of Computer Applications*, vol. 176, no. 7, pp. 15–21, 2020. https://www.ijcaonline.org/archives/volume176/number7/31027-2020920569

[8] Python Software Foundation, “Python Language Reference, version 3.x,” Python.org, 2024. [Online]. https://www.python.org

[9] Flask Development Team, “Flask: A lightweight WSGI web application framework,” Flask Official Documentation, 2024. [Online]. https://flask.palletsprojects.com

[10] R. Patel and N. Shah, “Avatar-based sign language generation for hearing-impaired communication,” *International Journal of Assistive Technologies*, vol. 8, no. 2, pp. 45–52, 2022. https://www.researchgate.net/search?q=Avatar-based%20sign%20language%20generation%20for%20hearing-impaired%20communication