International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Two -Way Communication Using Avatars

[1]Ms. Samreen Begum
Assistant Professor, Department of Computer Science and Engineering
Vignan's Institute of Management and Technology for Women, Hyd.
Email: samreen15101997@gmail.com

[2] Puram Reena
UG Student, Department of Computer Science and Engineering
Vignan's Institute of Management and Technology for Women, Hyd.
Email: reenapuram5b1@gmail.com

[3]Boddupally Hema
UG Student, Department of Computer Science and Engineering
Vignan's Institute of Management and Technology for Women, Hyd.
Email: boddupallyhema255@gmail.com

[4]Motar Bhargavi
UG Student, Department of Computer Science and Engineering
Vignan's Institute of Management and Technology for Women, Hyd.
Email: bhargavi26102003@gmail.com

***Abstract*— Deaf and hard-of-hearing individuals primarily rely on sign language, which involves specific hand movements, body language, and facial expressions to communicate. However, inclusivity for this community is still lacking in various sectors, especially in educational institutions like schools and colleges. Students who are hearing-impaired often face difficulties in expressing themselves or understanding lessons in traditional classroom environments, which can lead to feels isolated. In recent years, the development of virtual avatar technology has opened new possibilities for improving communication accessibility. Sign language dictionaries today often incorporate video or animated avatar demonstrations of signs, which are more intuitive and easier to understand than static images. Some also use sign writing notations, allowing gestures to be stored in a textual format that can later be converted into visual signs using avatars. This project aims to generate sign language gestures using animated avatars. These avatars can be implemented in educational tools to help hearing individuals learn and understand sign language, making two-way communication more effective. With this technology, hearing-impaired students can participate more actively in classroom discussions, while their peers and teachers can better comprehend their input. Ultimately, this approach supports the inclusion of deaf and hard-of-hearing individuals in mainstream education and promotes equal learning opportunities.**

Keywords - Mediapipe hand tracking, pyttsx3 speech engine, HamNoSys notation system, SiGML (Sign Gesture Markup Language)

I. INTRODUCTION

Hearing impaired communities often face challenges in expressing themselves and understanding others, which can lead to feelings of exclusion—especially in deaf and dumb communities and colleges where inclusive communication tools are lacking. however, many people without hearing impairments are unfamiliar with sign language, making effective interaction difficult. this gap has limited the development of intelligent systems capable of truly bridging communication between the hearing and non-hearing communities. this project proposes a solution that uses a digital avatar to support two-way communication between hearing-impaired people and normal people. the system allows sign language gestures to be translated into text or speech, and spoken language to be converted back into animated signs through an avatar. by integrating such tools into hearing impaired communities, the goal is to foster better understanding, promote inclusivity, and ensure that hearing-impaired people feel more connected and supported in their learning journey.

II. LITERATURE REVIEW

The Indian Sign Language Research and Training Centre [1], under the Ministry of Social Justice and Empowerment, has made significant strides in formalizing and promoting Indian Sign Language (ISL) as a recognized means of communication for the deaf community in India. Its landmark contribution, the *Indian Sign Language Dictionary*, now in its third edition, includes over 10,000 signs covering daily usage, academic terms, and legal vocabulary. This dictionary, accompanied by video demonstrations, provides an essential linguistic foundation for the development of ISL-based technologies, such as sign language recognition systems and virtual avatar communication. ISLRTC's efforts have contributed to making sign language more accessible and standardized across regions, thereby supporting technological interventions aimed at two-way communication for the deaf and hard-of-hearing communities (ISLRTC, 2021). These initiatives serve as a critical base for research into human-computer interaction and avatar-mediated sign language interpretation.

Ishika Dhall, Shubham Vashisht, and Garima Agarwal [2], the researchers explored an automated system for recognizing hand gestures using a Deep Convolutional Neural Network (DCNN) architecture. Their work addresses the challenge of interpreting dynamic hand gestures accurately—an essential component in enabling more intuitive human-computer interaction, especially for applications like sign language recognition. The proposed model was trained on a labeled dataset of hand gestures and showed a significant improvement in accuracy compared to traditional machine learning approaches. The authors emphasized the importance of preprocessing techniques like background subtraction and contour detection to enhance recognition performance. Their DCNN-based system demonstrated robust classification abilities, making it a promising tool for real-time gesture-based communication systems. This research serves as a strong foundation for building more inclusive technologies that bridge communication gaps for hearing-impaired individuals.

In their 2017 work, Na Zhao and Hongwu Yang [3] introduced a system designed to convert spoken language into gestures using keyword spotting techniques. The approach centers on identifying key words in speech that can be directly mapped to corresponding gestures. This method supports real-time interaction, enhancing communication for users who rely on non-verbal cues, particularly in assistive or human-computer interaction environments. Instead of transcribing entire sentences, their system detects relevant keywords and links them to pre-defined gesture animations, ensuring quicker processing and a more natural response. The authors emphasized the usefulness of their method in situations where full speech recognition may be inefficient or unnecessary, paving the way for simplified gesture-based interfaces driven by voice inputs.

Towards Data-Driven Sign Language Interpreting Virtual Assistant [4] explores the development of virtual assistants capable of interpreting sign language through data-driven methodologies. The authors focus on Kazakh-Russian Sign Language (KRSL) and evaluate four types of signing agents: two data-driven avatars, a motion capture animation avatar, and a human sign interpreter. The study assesses these agents based on user perceptions of naturalness and likeability. Findings indicate that while data-driven avatars perform competently, they do not yet match the human interpreter in terms of naturalness and user preference. This research highlights the potential of data-driven approaches in sign language interpretation and underscores the need for further advancements to enhance the naturalness of virtual signing agents.

Bouzid and Jemni [5] proposed a novel system that leverages avatars to interpret sign language from written notation. Their research focuses on converting HamNoSys—a structured and symbolic notation used to transcribe sign languages—into animated gestures using a 3D virtual avatar. The key objective is to enable accessibility for the deaf community by translating written sign language scripts into visually understandable signs. The system is capable of reading the notation and mapping it to corresponding hand shapes, movements, and facial expressions, all of which are rendered through a virtual human figure. This approach significantly reduces the reliance on human interpreters and supports real-time communication scenarios. One of the strengths of this system lies in its modular design, which makes it adaptable to various sign languages by modifying the input notation and avatar configurations. The work highlights the potential of avatar-based interpretation in bridging communication gaps for the hearing-impaired, while also emphasizing the need for continued refinement to improve the realism and expressiveness of the animations.

III. METHODOLOGY

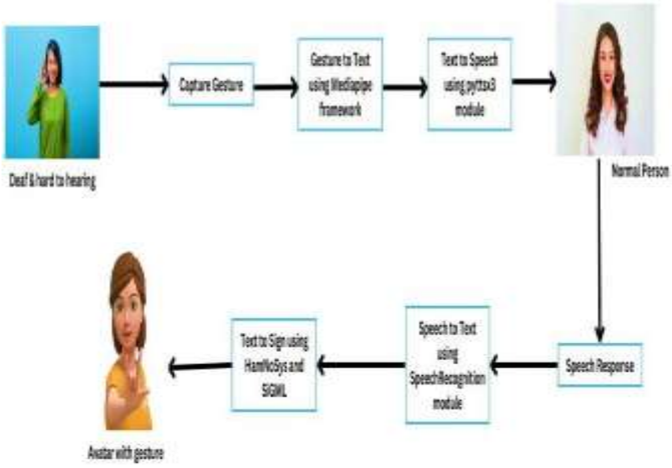

Fig 3: Bidirectional Sign-Speech Communication

Overview:

The diagram outlines a two-way communication framework aimed at bridging the gap between deaf or hard-of-hearing individuals and those who can hear. It starts with the deaf user performing gestures, which are captured by a vision-based input system. These gestures are then processed using the MediaPipe framework to convert them into text, enabling the system to understand the signed input.

Once the text is generated from the gesture, it is passed to a text-to-speech (TTS) module, which produces a spoken version of the message. This allows a hearing individual to receive and understand the communication in an audible form, ensuring inclusivity and smoother interaction between users of different communication needs.

For the response, the hearing person speaks, and their voice input is processed using a speech recognition module that converts the audio into text. This text is then translated into sign language using a combination of HamNoSys notation and SiGML (Sign Gesture Markup Language). These tools ensure accurate representation of sign language gestures in a machine-readable format.

Finally, the translated sign language is visualized through a digital avatar that performs the appropriate gestures. This avatar helps convey the hearing person's message back to the deaf user in a format they can understand. Overall, the system enables real-time, interactive communication between users who rely on different modes of expression—sign language and spoken language—by integrating gesture recognition, speech processing, and avatar-based animation.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

IV. RESULTS AND ANALYSIS

4.1 GESTURE TO SPEECH CONVERSION:

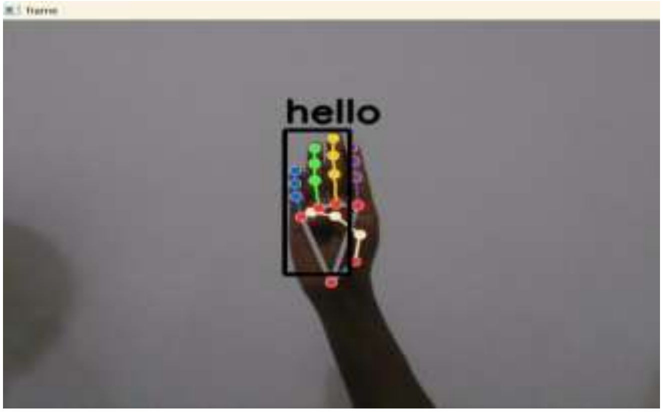

Fig 4.1.1 Recognition of "Hello" Gesture

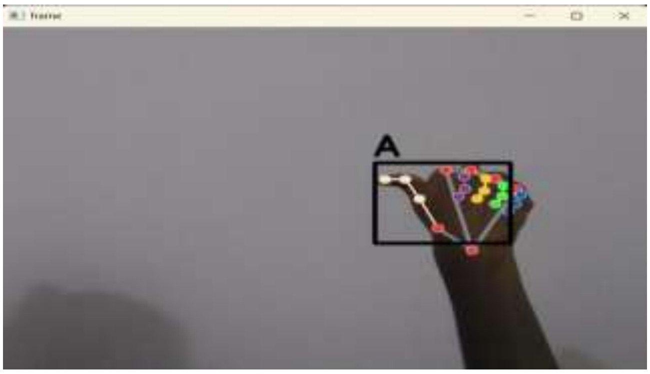

Fig 4.1.2 Recognition of "A" Gesture

Table 1: Accuracy of Gesture Recognition Model

Metric	Value
Classes Recognized	10 (e.g., "ok", "hello", "A", etc.)
Validation Accuracy	~90–95% *(estimated)*
Real-Time FPS	~12–15 FPS on CPU
Delay in Response	< 1 second (TTS latency)
Avg. Prediction Time	~50–100ms

Table 1 presents the evaluation results of the gesture recognition model, which was assessed using key performance metrics to determine its accuracy, reliability, and effectiveness in real-time applications. It was trained to identify 10 distinct hand gestures, including signs like "ok," "hello," and "A." In terms of real-time performance, the system maintained a frame processing rate of around 12 to 15 frames per second, which is sufficient for smooth gesture tracking using a standard webcam. The average response time from gesture input to model prediction was between 50 to 100 milliseconds, and gesture detection to audio feedback was less than one second.

4.2 SPEECH-TO-AVATAR CONVERSION:

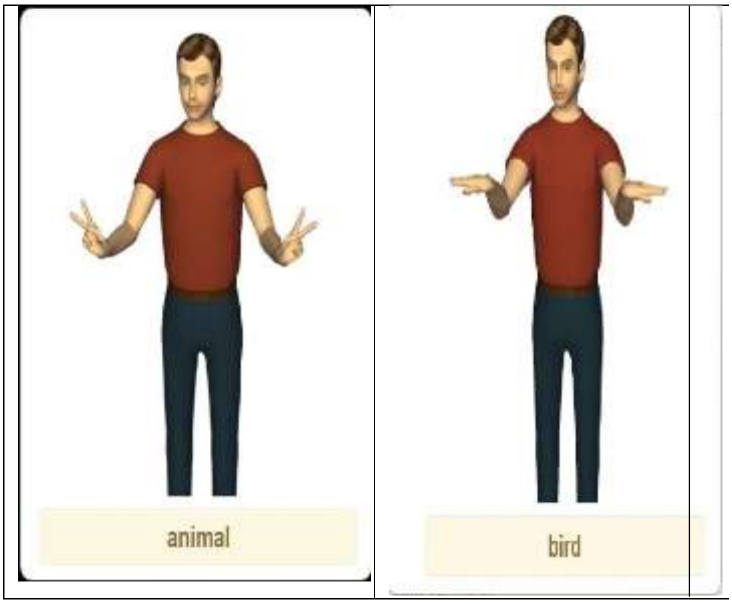

Fig 4.2 Animal and Bird Avatars

Table 2: Speech-to-Avatar System

Component	Observation
Single Word Commands	>95% success rate
Multi-word Input	Successfully split and matched to available videos
Video Merge Time	2–4 seconds for 3–5 videos
Browser Playback	Immediate (opens HTML)

Table 2 presents the evaluation of the Speech-to-Avatar system, which was tested to determine how accurately and efficiently it could convert spoken commands into corresponding avatar gestures. When users spoke single-word commands, the system successfully recognized and matched them with a pre-recorded video in over 95% of cases, demonstrating high accuracy in speech recognition. For longer or multi-word inputs, the system intelligently filtered out known gesture words and sequenced the relevant avatar videos correctly. On average, merging 3 to 5 video clips took about 2 to 4 seconds using MoviePy, and the final output was rendered smoothly. The generated videos were then automatically displayed in a web browser through an HTML file, providing an interactive visual response. Overall, the system proved responsive and accurate in translating voice commands into visual sign language, enabling a more accessible and engaging communication interface.

V. CONCLUSION

An animated avatar acts as an interactive bridge to facilitate communication between individuals with hearing impairments and those without. This system incorporates both gesture-to- speech and speech-to-gesture-conversion mechanisms, specifically tailored to Indian Sign Language (ISL), promoting inclusive and accessible interaction for the hearing-impaired community. For the gesture-to-speech module, sign language gestures made by a user are captured typically through a camera using hand-tracking frameworks and translated into audible speech in real time, allowing hearing individuals to understand the conveyed message effortlessly. Conversely, the speech-to-gesture module interprets spoken language input from a hearing person and converts it into corresponding sign gestures, which are then performed by the avatar. To support this system, a custom Avatar Dataset has been developed.

VI. FUTURE SCOPE:

Sentence-level speech-to-gesture conversion can be enhanced by using Natural Language Processing (NLP) to include more sentences. Avatar Dataset can be extended for more words from ISL. Emotions and Expressions can be added to Avatar for enhancing communication. It can also be extended to various other languages apart from English.

VII. REFERENCES

[1] Indian Sign Language Research and Training Centre, ISLRTC, https://islrtc.nic.in/.

[2] Ishika Dhall, Shubham Vashisht, Garima Agarwal, "Automated Hand Gesture Recognition using Deep Convolutional Neural Network model", https://leeexplore.jece.org/document/9057853

[3] Na Zhao, Hongwu Yang, “Realising speech to gesture conversion by keyword spotting”, https://ieeexplore.ieee.org/document/7918458.

[4] Towards data-driven sign language interpreting Virtual Assistant, https://openreview.net/pdf?id=HpQA6JhTL7x.

[5] Yosra Bouzid, Mohamed Jemni, “An Avatar-based Approach for Automatically Interpreting a Sign Language Notation”, https://ieeexplore.ieee.org/document/660187.