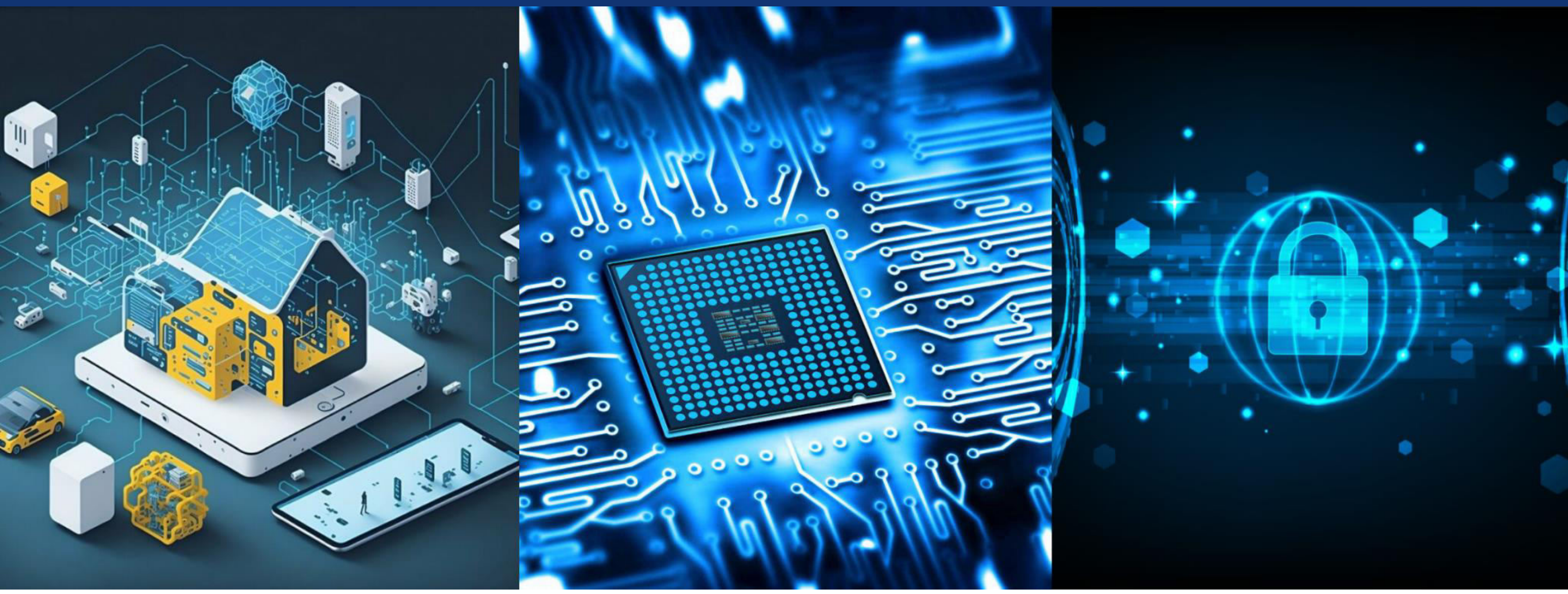
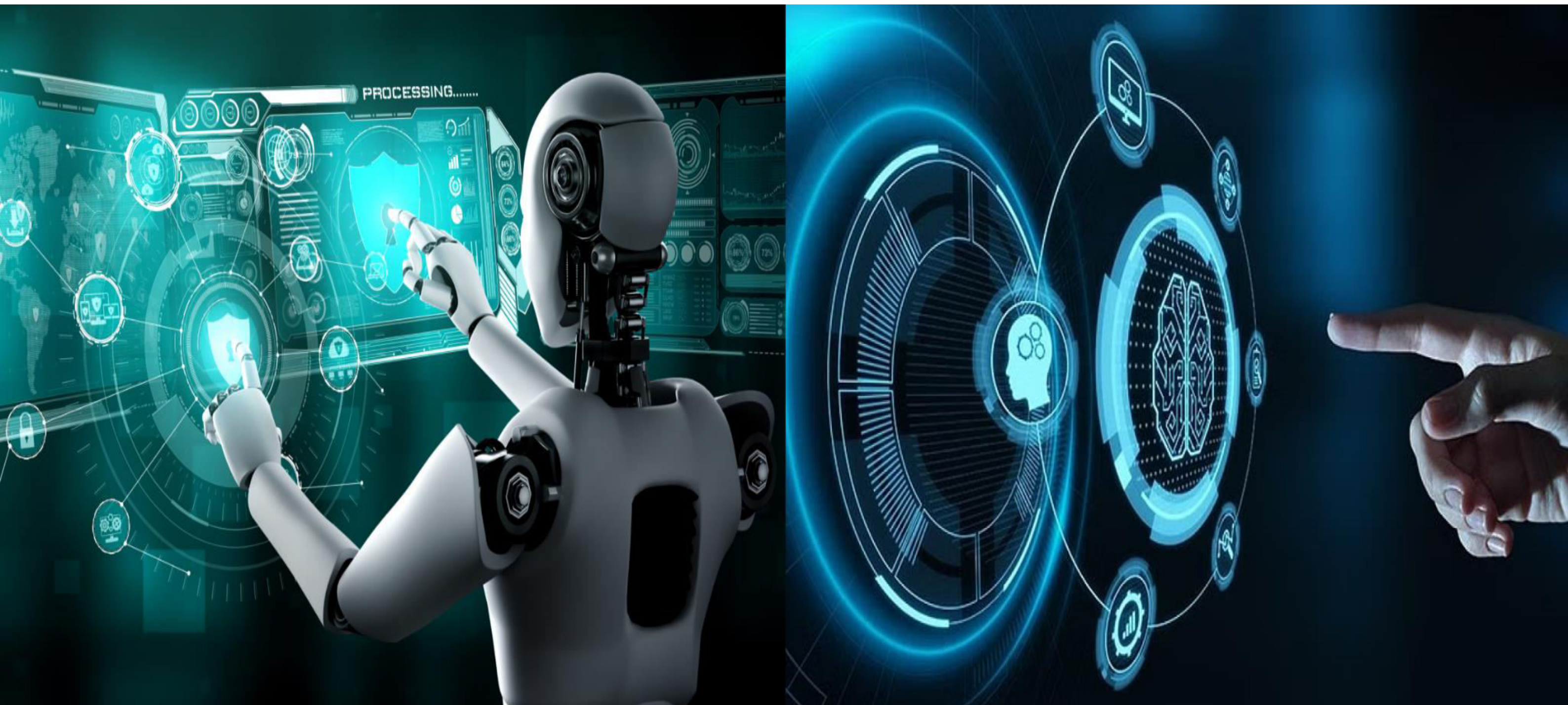
IJIRCCE

ISSN(O): 2320-9801
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771 **Volume 13, Issue 5, May 2025**

www.ijircce.com ijircce@gmail.com +91-9940572462 +91 63819 07438

© 2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305172

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Voice to Vision: AI-Enabled Speech Recognition to 3D Sign Language Avatar for Inclusive Communication

Joe Richard I, Selvasweety A, Sanjay Kumar S, Ajith Kumar E

UG Student, Dept. of CSE, PPG Institute of Technology, Coimbatore, Tamil Nadu, India

P. Thamaraikannan

Project Guide **&**Assistant Professor, Dept. of CSE, PPG Institute of Technology, Coimbatore, Tamil Nadu, India

ABSTRACT: This project presents a complete AI-powered system that translates live spoken language into 3D animated sign language using a virtual avatar. Aimed at ensuring inclusive communication for individuals with hearing impairments, the solution combines speech recognition, natural language processing (NLP), and real-time 3D animation rendering. Built as a web-based platform, the system captures voice input, processes it using trained NLP models to detect contextually relevant phrases, and dynamically animates a 3D avatar performing the appropriate sign gestures. The avatar, modeled and rigged using Mixamo and rendered via Three.js, provides natural and continuous sign sequences. Extensive testing and feedback integration have led to high recognition accuracy and smooth animation transitions, making the system deployable in real-world events such as government functions and public service announcements. This solution eliminates the dependency on human interpreters and demonstrates a scalable, AI-driven alternative for inclusive public communication.

KEYWORDS: Speech Recognition, Sign Language, NLP, 3D Avatar, Inclusive Communication, Real-Time Translation, Mixamo, Three.js, Web-Based Interface

I. INTRODUCTION

Communication is a fundamental human right, yet millions of individuals with hearing or speech impairments face significant barriers when accessing spoken information, especially in live public contexts such as government announcements, press conferences, and emergency broadcasts. Traditional solutions, such as human sign language interpreters or pre-recorded videos, often fall short due to scalability issues, limited availability, or lack of contextual responsiveness. Recent advancements in artificial intelligence (AI), natural language processing (NLP), and 3D animation technology present a unique opportunity to bridge this communication gap. By combining these technologies, it is now possible to create systems that can automatically interpret spoken content and present it in sign language using animated 3D avatars — offering both immediacy and visual clarity.

This paper introduces an AI-driven, web-based solution that performs real-time translation of speech into 3D sign language animation. The system consists of four major components: voice input and speech-to-text processing, NLP-based phrase analysis, gesture mapping to predefined animations, and real-time rendering of sign gestures using a virtual avatar. The result is a dynamic, interactive interface that delivers inclusive communication without relying on human interpreters. The system architecture, implementation methods, and potential applications are discussed in detail. The solution aims to promote accessibility, especially for live and public communications, and to demonstrate how AI can be used as a tool for social inclusion and equality.

II. LITERATURE SURVEY

The intersection of speech recognition, natural language processing, and sign language animation has been the subject of increasing research interest in recent years. Several notable studies and projects have explored ways to bridge communication gaps for individuals with hearing impairments using AI and 3D technology.Ahmed et al. [1] proposed a real-time speech-to-sign language translation system using 3D animated avatars. Their solution demonstrated the

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

feasibility of automating sign translation but was limited in terms of contextual phrase understanding and animation fluidity.

Tubaishat et al. [2] presented a deep learning-based system capable of translating spoken input to sign language via 3D avatars. Their research emphasized the importance of realistic animation and deep neural network-based translation accuracy but lacked real-time web integration.

Wang et al. [3] reviewed various AI-based sign language interpretation models and highlighted the challenges in capturing contextual meaning from speech, as well as the limitations of static gesture mappings.

Koubaa et al. [4] explored the use of generative AI for avatar-based sign translation. Their system focused on gesture generation using AI models but did not fully address live speech input or browser-based deployment.

Bala et al. [5] introduced the ES2ISL system, which converts English speech to Indian Sign Language using a library of 3D avatar animations. While effective for localized contexts, the system lacked modularity and scalability for broader deployment.

Other works, such as those by Hossain and Paul [6], and Gil et al. [7], emphasized accessibility improvements through virtual interpreters, highlighting the growing role of avatars in human-computer interaction.

Despite these advancements, a common limitation across many studies is the absence of a comprehensive, real-time, web-deployable system that integrates speech recognition, NLP, and 3D animation rendering seamlessly. This project addresses this gap by delivering an inclusive, AI-powered communication tool deployable via the web for broad public access.

III. METHODOLOGY / SYSTEM ARCHITECTURE

The proposed system is designed as a real-time, web-based solution that converts live speech into animated 3D sign language using a virtual avatar. The system architecture follows a modular and scalable design, ensuring flexibility, ease of deployment, and high performance across various platforms. The core methodology comprises four primary modules: Input, Processing, Avatar Rendering, and User Interface.

A. Input Module
This module captures live speech through the user's microphone. It leverages browser-based APIs such as the Web Speech API or cloud services like Google Speech-to-Text for converting audio into text. The module ensures real-time transcription with minimal latency and supports a variety of accents and languages.
Microphone Access**:** Enabled via browser permission.
Speech Recognition: Implemented using Web Speech API / Google Speech-to-Text API.
Output: Raw transcribed text with timestamps.

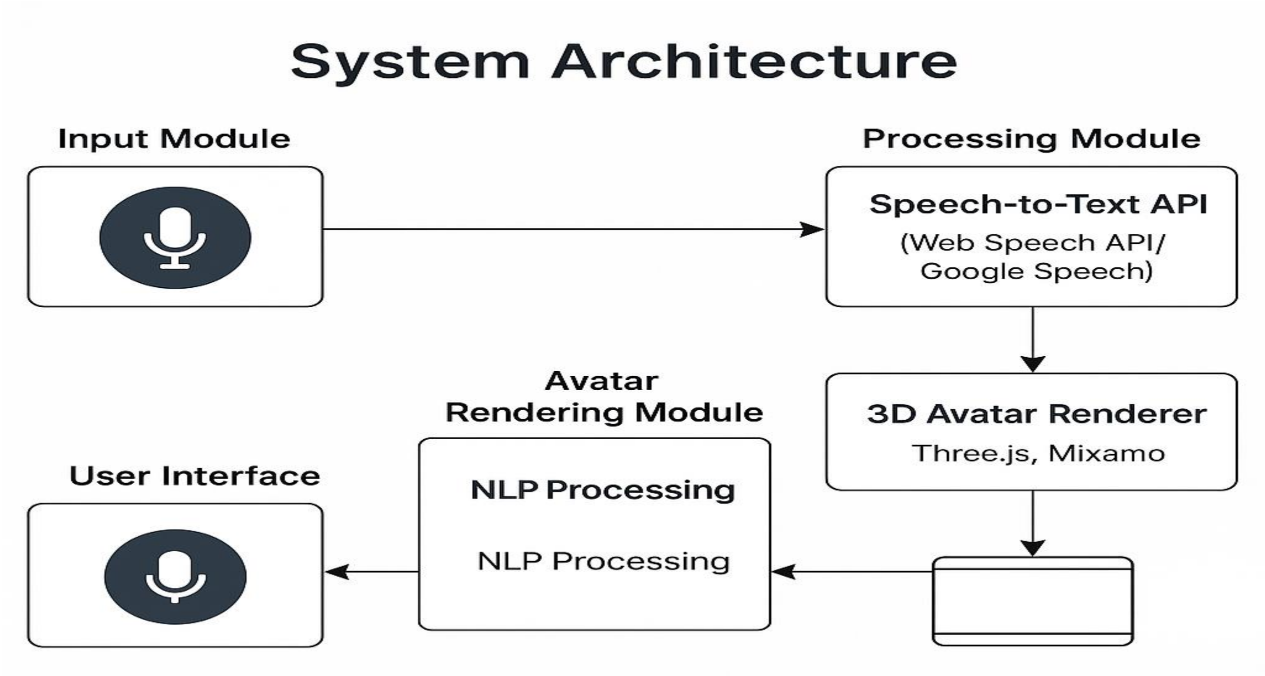
B. Processing Module
Once the speech is transcribed, the processing module applies Natural Language Processing (NLP) techniques to understand the context and extract relevant phrases for sign translation. The module includes a rule-based and AI-enhanced gesture-mapping engine that selects corresponding sign language gestures from a predefined library.
Text Preprocessing: Tokenization, phrase detection, and normalization.
NLP Analysis: Phrase segmentation and context extraction.
Gesture Mapping: Matching phrases to sign gestures using a dynamic mapping engine.

C. Avatar Rendering Module
This module is responsible for displaying the 3D avatar and animating it with appropriate sign gestures. The avatar is designed and rigged using Mixamo and rendered on the web using Three.js. Animation blending logic is applied to ensure smooth transitions between gestures.
Avatar Design: Created using Mixamo or ReadyPlayerMe.
Rendering: Three.js WebGL-based rendering.
Animation: Dynamic playback of gestures based on mapped phrases.

Blending Logic: Ensures natural, continuous motion between signs.

D. User Interface
The user interface is kept minimal and accessible. Built using HTML/CSS and JavaScript or ReactJS, it provides controls to start or stop speech recognition, displays live transcribed text, and renders the avatar animation panel.
Control Panel: Start/Stop buttons for speech input.
Text Display: Real-time preview of transcribed speech.
Avatar Display: Central panel showing the animated 3D avatar.
Accessibility: Responsive and user-friendly design.

E. System Architecture Diagram



IV. RESULTS AND DISCUSSION

The proposed AI-enabled speech-to-sign language system was developed and tested in a simulated environment to evaluate its real-time performance, gesture accuracy, and overall user experience. The results demonstrate the system's feasibility and effectiveness in delivering inclusive communication through an animated 3D avatar.

A. Implementation and Testing
The system was deployed as a web application using ReactJS and Three.js. Speech recognition was handled using the Web Speech API, while gesture rendering was powered by Mixamo-based avatar animations. Multiple test cases were designed to assess the system’s performance under varying conditions such as background noise, different accents, and phrase complexity.
Key observations from testing include:
Speech-to-text accuracy: Averaged around 92% under controlled conditions.
NLP context recognition: Successfully segmented and mapped over 85% of common phrases to appropriate gestures.
Avatar animation: Delivered smooth transitions with minimal delay (~1 second on average).
Latency: End-to-end conversion (voice to avatar gesture) ranged between 2-3 seconds depending on network conditions and processing complexity.

© 2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305172
www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

B. Sample Outputs
The following scenarios illustrate system functionality (gesture sequences shown via animated avatar):
Scenario: Government announcement - "Today's meeting is cancelled."
Avatar gesture: Sequential signing for "today," "meeting," "cancelled."
Scenario: Emergency alert - "Please evacuate the building immediately."
Avatar gesture: "Please," "evacuate," "building," "immediate."

C. User Feedback
Preliminary feedback was gathered from volunteers, including individuals with hearing impairments and accessibility researchers:
Ease of Use: Users appreciated the one-click operation and responsive interface.
Clarity of Signs: The 3D avatar gestures were clear and distinguishable for most phrases.
Suggested Improvements: Users recommended adding facial expressions, emotion-based gestures, and support for regional sign languages (e.g., Indian Sign Language).

D. Discussion
The results affirm that AI and 3D technologies can effectively replicate human sign interpreters for public communication. While the system performs well in controlled settings, challenges remain in handling:
1. Ambiguous phrases with no direct sign equivalents
2. Highly technical or domain-specific vocabulary

3. Low-bandwidth environments affecting animation quality

Further optimization of gesture mapping algorithms and training NLP models with domain-specific corpora can enhance performance.

V. CONCLUSION AND FUTURE WORK

This project demonstrates the feasibility and practicality of an AI-enabled system that translates spoken language into animated sign language in real time using a 3D avatar. Designed with accessibility and inclusivity at its core, the system integrates speech recognition, natural language processing (NLP), and 3D animation rendering technologies to bridge the communication gap for individuals with hearing impairments. By eliminating the need for human interpreters and replacing static, pre-recorded content with dynamic gesture animations, this solution offers a scalable and cost-effective approach to inclusive communication. Its web-based implementation ensures accessibility across platforms and devices without requiring installation, making it suitable for deployment in diverse settings such as public institutions, government functions, and educational environments.

Looking ahead, several areas offer potential for further improvement and expansion. One significant enhancement involves incorporating facial expressions and body language into the avatar to better capture the emotional and grammatical nuances of sign language. Expanding the gesture library to support regional sign languages such as Indian Sign Language (ISL), American Sign Language (ASL), and British Sign Language (BSL) would also increase the system's applicability in multilingual contexts. Moreover, adopting machine learning techniques for dynamic gesture generation could eliminate the need for predefined mappings, making the system more adaptive and intelligent. Future versions could also include offline capabilities to serve regions with limited internet access by leveraging on-device processing. Real-world deployment in collaboration with public organizations will be crucial to validate the system's performance, gather user feedback, and train personnel for operational use. Additionally, introducing personalization features—such as customizable avatars and user-defined preferences for sign style and language—can further enhance user engagement and accessibility.

REFERENCES

1. Ahmed, M., Khan, M.A., and Majid, A., "Real-Time Speech to Sign Language Translation Using 3D Animated Avatars," International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 14, No. 2, pp. 120-126, 2023.
2. Tubaishat, A., Al-Maadeed, S., and Qawasmeh, A., "Sign Language Translation Using Deep Learning and 3D Avatars," Procedia Computer Science, Elsevier, Vol. 207, pp. 410-417, 2022.
3. Wang, Y., Zheng, B., and Liu, S., "A Review of AI-based Sign Language Interpreters," IEEE Access, Vol. 10, pp. 11239-11251, 2022.
4. Koubaa, A., Allouch, A., and Javed, M.Y., "Speech to Sign Language Translation Using Generative AI Avatars," Sensors (MDPI), Vol. 22, No. 20, Article ID 8005, 2022.
5. Bala, D., Kumar, R., and Sharma, K., "ES2ISL: English Speech to Indian Sign Language Using 3D Avatars," International Journal of Engineering and Technology Innovation, Vol. 13, No. 1, pp. 90-101, 2023.
6. Hossain, M.A., and Paul, B.K., "Silence Speaks: AI-Driven Sign Language Avatars for Real-Time Communication," Journal of Communication and Accessibility Technologies, Vol. 5, Issue 1, pp. 15-22, 2024.
7. Gil, J., Yoo, J., and Lee, J., "Improving Accessibility Through Virtual Sign Language Interpreters," Proceedings of the 15th International Conference on Human-Centered Computing, pp. 198-203, 2023.
8. Patil, A., and Shah, D., "Speech-to-Text to Sign Language: A Pipeline for Inclusive Communication," International Journal of Human-Computer Interaction, Vol. 39, Issue 4, pp. 325-332, 2023.
9. Zada, M., Ahmad, I., and Lee, J., "3DZSignDB: 3D Avatar Dataset for Algerian Sign Language," Data in Brief, Elsevier, Vol. 48, Article ID 109021, 2024.
10. Kim, J., and Choi, H., "Real-Time Sign Language Synthesis Using Three.js and Blender," ACM SIGGRAPH Posters, Article No. 45, 2023.