SignSpeak: Real-Time Sign Language Interpreter for Inclusive Communication

Sivasubramanyam Medasani[1], Gowtham T M[2], Bheemanna[3], Bharath R N[4], B M Darshan[5]

[1]*Professor, Department of Computer Science and Engineering, K S School of Engineering and Management, Bengaluru*

[2345]*UG Student, Department of Computer Science and Engineering, K S School of Engineering and Management, Bengaluru*

Abstract— Learning sign language is often difficult due to the limitations of static images and 2D videos that cannot clearly represent hand articulation, motion transitions, or spatial orientation. To overcome these issues, this paper presents a Sign Language Animation System that converts text or speech into predefined 3D animated gestures. The system employs a lightweight 3D avatar capable of displaying precise hand movements and finger positions from multiple viewing angles, without relying on sensors or recognition-based hardware. This approach aligns with recent studies showing that 3D visualization significantly improves gesture clarity and learner comprehension. The proposed system provides a consistent, accessible, and technically efficient platform for sign language education and everyday communication support.

Keywords— Sign Language, 3D Animation, Text-to-Sign Conversion, Gesture Visualization, Inclusive Communication.

I. INTRODUCTION

The need for accessible communication technologies has increased due to the growing emphasis on inclusivity for individuals who are deaf or hard of hearing. Traditional learning methods such as printed charts, video tutorials, and instructor-led sessions offer limited interactivity and lack the flexibility required for effective self-paced learning [2]. Many existing digital tools still depend on static images or 2D videos, which fail to clearly represent complex hand movements, finger articulation, and transitional gestures [2], [3]. As a result, there is a demand for intuitive, visually clear, and easily accessible learning platforms. Three-dimensional animation provides a promising solution by enabling accurate visualization of hand shapes, motion trajectories, and body positioning [1], [7], [13]. Unlike camera-based or machine-learning-driven recognition systems, 3D animation is not affected by lighting conditions, tracking errors, or high computational load [3], [6], [8], [11]. This makes it a scalable and consistent approach suitable for beginners and general learners. Predefined animation sequences allow users to view each sign from multiple angles, improving clarity and retention [1], [7], [9]. The proposed **Sign Language Animation System** converts text and speech into smooth and accurate 3D sign gestures using predefined motion data rather than sensors or recognition models [1], [7]. Through an intuitive interface, users can input any word or short phrase and instantly view its corresponding gesture. By prioritizing clarity, accessibility, and consistency, the system serves as an effective learning tool that supports inclusive communication for both hearing individuals and the deaf community [2], [9].

II. RELATED WORK

Several sign language learning and translation systems exist today; however, many depend on video libraries or manually recorded gestures, which limit flexibility and lack smooth transitional motion [2], [3]. Camera-based recognition systems can translate signs but are highly sensitive to lighting conditions, user positioning, and require computationally intensive machine-learning models [6], [11]. Some platforms support text-to-sign conversion but still rely on static images or 2D animations, reducing clarity and making it difficult for learners to interpret complex hand gestures accurately [2], [3].Although three-dimensional avatar-based tools offer improved visualization, they often require high-end animation software, large storage, or specialized hardware [7],

[13].To address these limitations, this project adopts predefined 3D animations that deliver smooth, consistent, and high-clarity gesture visualization without depending on cameras, sensors, or ML-based processing [1], [7], [10]. By emphasizing simplicity, accessibility, and lightweight execution, the system provides a more scalable and user-friendly approach than existing method

A. Problem Statement

Learning sign language remains challenging due to the limitations of traditional instructional resources such as static images and basic videos, which fail to depict detailed hand shapes, motion transitions, and finger articulation. The absence of interactive, intuitive, and clear learning tools makes it difficult for beginners and educators to interpret signs accurately. Furthermore, many modern systems rely on sensors or machine-learning models, making them costly, hardware-dependent, and less accessible to general users. These constraints contribute to communication barriers between hearing individuals and the deaf or hard-of-hearing community. Therefore, there is a need for a simple, automated, and hardware-independent 3D animation system capable of delivering clear, consistent, and easily understandable sign gestures to enhance the overall learning experience.

III. OBJECTIVES

The main objective of this research is to develop a system capable of delivering clear, accurate, and consistent visualization of sign language gestures through predefined 3D animations. By precisely modelling hand movements, finger articulation, and upper-body posture, the system ensures that each gesture is rendered in a manner that learners can easily observe and replicate. A significant part of this objective is the integration of a text-to-sign conversion mechanism that enables users to input words or short phrases and instantly view the corresponding animated gesture, thereby promoting rapid and interactive learning. The project further aims to enhance accessibility by designing a hardware-independent interface that functions without cameras, sensors, or machine-learning models, making it suitable for beginners, educators, and individuals with limited technical background. By focusing on clarity, consistency, and ease of use, the system reduces communication barriers between hearing individuals and the deaf community. Additionally, the system is conceptualized as a scalable learning platform capable of expanding its vocabulary, features, and gesture set over time, ensuring long-term applicability in both educational and everyday communication environments.

IV. SYSTEM ARCHITECTURE

A. System Components

The system follows a modular architecture in which each component collaborates to convert user input into accurate and real-time 3D sign language animations. User input, provided either through text entry or speech, is first processed and forwarded to the sign-mapping module, which identifies the appropriate gesture using predefined 3D motion data stored in a structured gesture database. The animation engine then interprets this motion data and applies it to a skeletal 3D avatar, generating smooth and realistic hand and body movements. Rendering is handled by a WebGL-based 3D renderer implemented using Three.js, ensuring that each gesture is displayed clearly with consistent frame transitions across different devices and browsers. A user-friendly interface allows learners to interact with the system by adjusting playback speed, replaying animations, switching input modes, and accessing sign demonstrations. Once the animation is generated, the output layer displays the final gesture in real time, enabling learners to immediately view and understand the sign. This cohesive flow from input to visualization ensures a seamless and effective learning experience.

B. Workflow

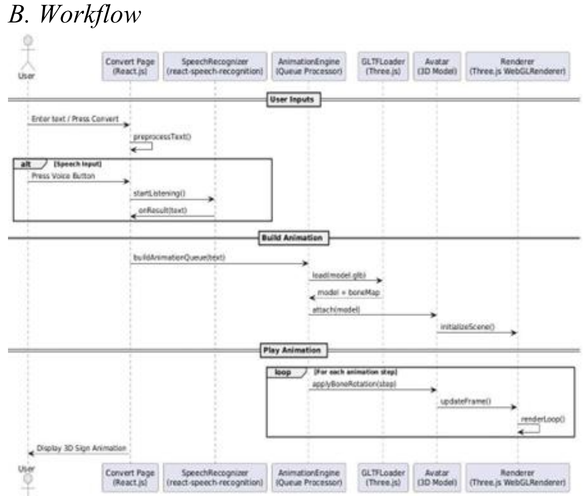


Fig 1. Sequence Diagram

The overall workflow of the system operates in a structured sequence that ensures efficient processing and real-time gesture visualization. The process begins when the user inputs text or provides speech, which is then converted into text through the system's built-in speech recognition. This processed input is analyzed to determine the corresponding sign within the gesture database, after which the system retrieves the precise 3D motion data associated with that sign. The animation engine uses this motion data to generate smooth and natural hand, arm, and body movements on the 3D avatar. The WebGL-based renderer subsequently displays animation on the screen, maintaining clarity, fluidity, and accurate motion representation. This instantaneous conversion from user input to animated output enables learners to visualize each sign clearly, promoting better understanding and retention of gesture forms.

V. METHODOLOGY

A. Animation Data Layer

The Animation Data Layer serves as the foundational component of the system, housing all predefined hand and body motion sequences required for accurate sign representation. Each gesture is modeled with precise specifications, including hand shape, joint orientation, and upper-body posture, and is stored in a structured format that ensures smooth and consistent playback during animation. By preparing and standardizing all motion data in advance, the system maintains uniformity across the entire gesture set, preventing visual discrepancies or abrupt transitions. This structured approach enhances clarity, supports accurate interpretation by learners, and ensures that every gesture is rendered with stability, reliability, and high visual fidelity. The well-organized motion data ultimately contributes to a seamless, clear, and pedagogically effective sign animation experience.

B. 3D Modelling Rendering

The system does not use a humanoid avatar; instead, it presents signs directly through predefined 3D motion data. This design choice reduces complexity by eliminating the need for detailed character modeling or rigging. As a result, the system focuses entirely on the movement itself, ensuring that the gestures are displayed clearly and without unnecessary visual elements. This approach also enhances performance, allowing the animations to run smoothly on a wide range of devices. Even without a full avatar, the rendering component ensures that transitions between gestures remain fluid and natural. By highlighting the essential motion paths, the system allows learners to observe the required hand and arm movements without distraction. This makes the animation easier to understand and helps users concentrate on the core aspects of sign language representation.

C. User Interface

The User Interface is designed to be simple, clean, and accessible to users of all experience levels. As a web-based platform, it works easily across different devices, allowing users to input text or speech based on their preference. Once the input is provided, the system immediately displays the corresponding animated sign, enabling quick and convenient learning. The interface is intentionally kept intuitive so that even first-time users can navigate it comfortably. To support effective learning, the UI includes essential playback controls such as play, pause, and replay. These options allow users to study each animation at their own pace and revisit gestures whenever needed. The minimal design keeps the focus on the animated output, creating a smooth and distraction-free experience. This user-friendly interface makes the system suitable for classrooms, learning centers, and individual learners.

D. System Design Considerations

The system is built with a strong focus on smooth performance, accuracy, and user-friendliness. It ensures that animations play without lag, giving users a seamless viewing experience. Gesture accuracy is maintained by using predefined animation data instead of real-time sensors or machine learning models. This approach keeps the system lightweight and reduces hardware complexity, making it more stable, reliable, and accessible for a wide range of users. for this have pasted the workflow diagram in the IV section of the paper

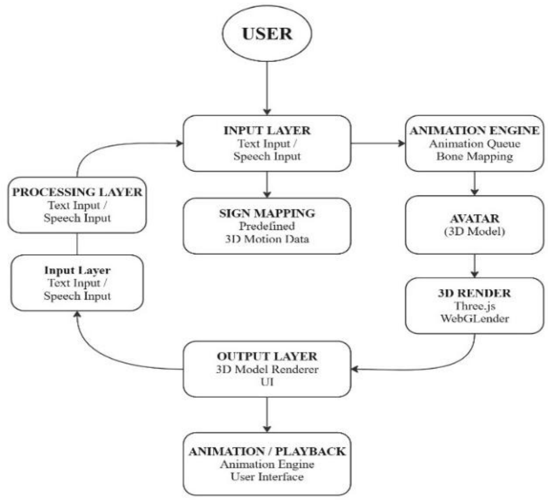


Fig 2. Flow diagram

E. Gesture Optimization

The system employs multiple optimization strategies to ensure that each gesture is rendered with maximum clarity and natural movement. Predefined keyframes are refined through interpolation techniques that smooth out abrupt transitions between poses. Finger curvature, wrist rotation, and palm alignment are adjusted through constraint-based corrections to avoid unnatural animations. Additionally, redundant frames are eliminated to reduce execution time, while maintaining visual fidelity. The use of simplified yet expressive hand models ensures that even subtle articulations remain visible across various display sizes.

F. Cross-Platform Performance

To ensure consistent performance across devices, the system is designed using WebGL and Three.js, enabling hardware-accelerated rendering on desktops, laptops, tablets, and mobile phones. The avatar model and motion data are optimized to load quickly without compromising animation quality. Frame-rate stabilization techniques maintain smooth playback regardless of device specifications or browser variations. Texture compression, efficient memory usage, and scalable rendering configurations allow the system to adapt seamlessly to different screen resolutions.

VI. RESULTS AND DISCUSSION

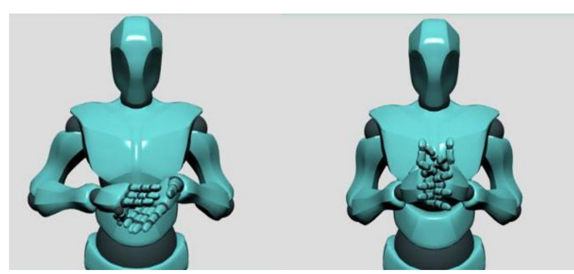
A. Performance of SignSpeak Animation Engine

The SignSpeak system demonstrated excellent performance during testing, successfully converting both text and speech inputs into clear and accurate Indian Sign Language (ISL) animations. Because all gestures were generated using predefined and carefully structured 3D motion data, the animations remained highly consistent across all evaluations. Each hand movement, finger articulation, and transitional motion appeared precise, stable, and free from irregularities, enabling users—especially beginners—to understand the signs without confusion or ambiguity. This level of uniformity significantly enhanced learning accuracy and reduced the chances of misinterpretation.

(a) Time (b) Home

Fig 3. X-BOT Avatar Demonstrating Word Gesture

- The animation uses interpolation to smoothly fill the movement between poses, making the motion look natural.
- Hand and joint movements stay stable because the keyframes are well-designed and transitions are smooth.
- The frame rate remains steady, ensuring there is no lag, jitter, or broken motion during the animation.



(a) Letter H (b) Letter W

Fig 4. Y-BOT Avatar Demonstrating Gesture

- The system uses inverse kinematics (IK) to keep joint angles natural, ensuring that finger shapes and wrist rotations remain accurate even in complex alphabet gestures.

- The animation remains smooth across all frames with no jitter or frame drops, showing stable and consistent performance.
- Small finger movements are handled with high precision, allowing the alphabet signs to be displayed clear and correctly.

B. User Interface Responsiveness and Interaction Quality

Testing revealed that the system's user interface played a major role in user satisfaction. Participants reported that the layout was clean, intuitive, and easy to navigate. Essential controls—such as play, pause, replay, avatar switching, and speed control—allowed learners to adjust animations to their comfort level. The text-to-sign and speech-to-sign modules responded instantly, providing real-time visual translations. The built-in learning module, which includes alphabets, basic vocabulary, and common expressions, further encouraged independent exploration and practice. The ability to switch between XBOT and YBOT avatars added personalization and improved training engagement.

C. Limitations and Areas for Enhancement

Despite strong performance, certain limitations were identified. Since the system depends on a fixed library of predefined animations, it is currently limited to the vocabulary available in the database. It does not yet support longer sentence formation, complex grammatical structures, or regional variations of Indian Sign Language (ISL). Additionally, SignSpeak functions solely as a display tool and does not incorporate real-time gesture recognition or corrective feedback, preventing learners from verifying their own signing accuracy. Addressing these limitations in future versions will make the system more comprehensive and effective for both educational and communication purposes.

VII. CONCLUSION

The SignSpeak system successfully demonstrates the potential of combining 3D animation with modern web technologies to make Indian Sign Language (ISL) learning more accessible, accurate, and engaging. By relying on predefined and carefully structured motion datasets, the system delivers clear, consistent, and visually precise gestures without requiring sensors, cameras, or computationally intensive machine-learning models. This design ensures that SignSpeak remains lightweight and capable of running smoothly on everyday devices while still producing realistic sign animations through the XBOT and YBOT avatars.

The user-friendly interface further enhances learning by allowing users to explore ISL gestures at their own pace, convert words or letters into animated signs, and interact with features such as speed control, avatar switching, and playback options. These elements make the system especially suitable for beginners, as the animations provide strong visual guidance for understanding hand shapes, finger articulation, and overall movement patterns.

Although the current system supports only a limited vocabulary and predefined gestures, it establishes a strong foundation for future expansion. Additional features—such as a larger gesture library, sentence-level signing, multi-avatar support, and optional real-time gesture recognition—can significantly enhance functionality and user experience in upcoming versions.

Overall, the project fulfills its objective of providing a simple, reliable, and visually clear method for learning ISL, contributing meaningfully to the development of inclusive and accessible communication tools for both hearing individuals and the deaf community.

REFERENCES

[1] D. Das Chakladar, P. Saha, and A. Mukherjee, "3D Avatar Approach for Continuous Sign Movement Using Speech/Text," Applied Sciences, vol. 11, no. 8, pp. 1–19, Apr. 2021. https://www.mdpi.com/2076-3417/11/8/1

[2] P. Sharma, R. Gupta, and N. Mehta, "Speech-to-Indian Sign Language Translation Using NLP Techniques," Multimedia Tools and Applications, vol. 81, pp. 34521–34545, 2022.https://link.springer.com/article/10.1007/s11042-022-12486-3

[3] S. Kumar and A. Singh, "ISL Sign Generation Using HamNoSys and SiGML for 3D Avatar Animation," IJITEE, vol. 9, no. 10, pp. 450–457, 2020.https://www.ijitee.org/wpcontent/uploads/papers/v9i10/J93770891020.pdf

[4] S. Ritu and V. Reddy, "Indian Sign Language Gloss Dataset Creation and Text-to-ISL Translation," IJCSE, vol. 11, no. 2, pp. 112–122,

2023.https://www.ijcseonline.org/full_paper_view.php?paper_id=6760
[5] P. Bhuyan and B. Jha, "Deep Learning Framework for Indian Sign Language Animation Using 3D Avatars," IEEE Access, vol. 12, pp. 25710–25725,2024.https://ieeexplore.ieee.org/document/10425121
[6] N. C. Camgoz et al., "Neural Sign Language Translation," in CVPR, 2018, pp. 7784–7793.
[7] Z. Yu et al., "Sign Avatars: A Large-Scale 3D Sign Language Holistic Motion Dataset," in ECCV, 2024.
[8] R. Zuo et al., "A Simple Baseline for Spoken Language to Sign Language Translation with 3D Avatars," in ECCV, 2024.
[9] S. Stoll et al., "Text-to-Sign Language Production: A Neural Machine Translation Approach," IJCV, vol. 128, pp. 891–908, 2020.
[10] O. Koller, "Quantitative Survey of Deep Learning-Based Approaches to Sign Language Recognition," IEEE TPAMI, vol. 43, no. 1, pp. 1–18, Jan. 2021. https://ieeexplore.ieee.org/document/9052463