2025 2nd International Conference on Trends in Engineering Systems and Technologies (ICTEST) | 979-8-3315-0537-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICTEST64710.2025.11042509

SignMate : Real-Time Speech to Sign Language Translation

Aswathy Rose Mathew
Dept. of AI and Data Science
Rajagiri School of Engineering and Technology
Kochi, Kerala, India
u2108018@rajagiri.edu.in

Evitta Telson
Dept. of AI and Data Science
Rajagiri School of Engineering and Technology
Kochi, Kerala, India
u2108021@rajagiri.edu.in

Hanna Tinku
Dept. of AI and Data Science
Rajagiri School of Engineering and Technology
Kochi, Kerala, India
u2108029@rajagiri.edu.in

Muhammed Hafis A S
Dept. of AI and Data Science
Rajagiri School of Engineering and Technology
Kochi, Kerala, India
u2108049@rajagiri.edu.in

Sujitha B Cherkottu
Dept. of AI and Data Science
Rajagiri School of Engineering and Technology
Kochi, Kerala, India
sujithab@rajagiritech.edu.in

***Abstract*—Sign language is the primary mode of communication for the deaf and hard-of-hearing community. However, the lack of widespread knowledge of sign language among the general population creates a significant communication barrier. SignMate aims to bridge this gap by developing a real-time speech-to-sign language translation system using deep learning techniques. The system leverages automatic speech recognition using Whisper, natural language processing for text processing, and a video-based approach to display corresponding American Sign Language and Indian Sign Language video clips. The paper details the methodologies used, challenges encountered, and the effectiveness of the system in improving accessibility and inclusivity for the deaf and hard-of-hearing community.**

***Index Terms*—Speech-to-sign translation, ASR, NLP, Whisper AI, Flask, ASL, ISL, accessibility, sign language communication.**

I. Introduction

Communication plays a fundamental role in daily interactions, yet barriers persist for the deaf and hard-of-hearing community due to the limited knowledge of sign language among the general population. Sign language serves as the primary means of communication, but a lack of accessibility tools often hinders effective interaction between sign language users and non-signers. This communication gap creates challenges in education, employment, and social integration.

To address this issue, SignMate is proposed as a real-time speech-to-sign language translation system. The system leverages ASR using Whisper to accurately convert spoken language into text. The processed text is then mapped to corresponding ASL and ISL video clips, ensuring seamless communication. Unlike avatar-based translation systems, SignMate relies on a curated video database containing pre-recorded sign language gestures, making the translation process more natural and comprehensible.

The primary objective of SignMate is to enhance accessibility for the deaf and hard-of-hearing community by facilitating real-time translation of spoken language into sign language. The system is designed to operate efficiently with low latency, ensuring near-instantaneous retrieval and display of sign language videos. This paper discusses the methodologies used, experimental results, challenges encountered, and potential areas for future improvement in SignMate.

II. Background and Related Work

A. Speech Recognition for Accessibility

ASR has significantly improved over the years, enabling real-time transcription of spoken language into text. OpenAI's Whisper model has demonstrated high accuracy in multilingual speech recognition, making it an ideal choice for accessibility applications. Studies such as have shown that Whisper outperforms traditional ASR models in noisy environments and diverse accents, making it suitable for real-world implementations. Its robust performance across different speech conditions enhances the feasibility of integrating ASR into sign language translation systems.

B. Sign Language Translation Systems

Previous research has explored multiple approaches for sign language translation, including rule-based, machine learning, and deep learning models. Some systems utilize sign language

avatars to generate animated gestures, while others depend on video databases for sign retrieval. Compared to avatar-based approaches, video-based translation systems provide more natural and comprehensible sign representations, making them more effective for real-time applications. The reliance on real human signers for video-based approaches ensures accuracy in sign execution, making it easier for users to understand.

C. Video-Based Sign Language Retrieval

Recent work has focused on mapping recognized text to corresponding sign language videos, enabling realistic translation. Approaches such as have introduced neural network-based retrieval systems that efficiently match words and phrases to pre-recorded sign videos. However, challenges remain in terms of latency and scalability, particularly when dealing with large vocabularies. To address this, caching mechanisms and indexing techniques have been explored to optimize retrieval times and ensure seamless user experience.

D. Challenges in Sign Language Translation

Despite advancements, several challenges persist in speech-to-sign translation. One major issue is the complexity of sign language grammar, which differs from spoken language structure. Many existing systems struggle with sentence reordering and contextual translation. Additionally, variations in regional sign languages (e.g., ASL vs.ISL)) require extensive datasets for accurate translation . Another challenge is the lack of large-scale, annotated sign language datasets, which limits the performance of machine learning models.

III. Proposed Methodology

SignMate system is designed to provide real-time speech-to-sign language translation, enhancing accessibility for the deaf and hard-of-hearing community. The methodology consists of three core components: speech recognition, text-to-sign language mapping, and sign language video retrieval. Implemented as a web-based application, SignMate allows users to access sign language translations through an interactive and user-friendly interface.

A. Speech Recognition using Whisper

The first stage of the SignMate pipeline involves converting spoken language into text using an ASR model. OpenAI Whisper, a state-of-the-art ASR system known for its accuracy in transcribing speech across various accents and noisy environments [1], is employed. Whisper processes the input audio and generates a precise text transcript with minimal errors. To ensure the text is suitable for sign language representation, preprocessing is applied, where unnecessary filler words are removed, and sentence structure is adjusted for better sign language translation.

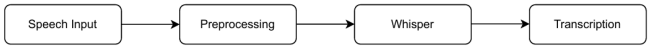

Fig. 1. Speech to Text Transcription

B. Text-to-Sign Language Mapping

Once the speech is converted to text, the system maps it to the corresponding sign language representation. The mapping is based on a curated dataset of video clips in ASL and ISL. Each recognized word or phrase is linked to its respective video clip, ensuring accurate representation. Since sign languages have different grammatical structures from spoken languages, the system applies phrase restructuring techniques to maintain semantic accuracy while aligning with sign language grammar.

C. Video Rendering

The video rendering module ensures smooth and efficient sign language playback [9]. SignMate relies on an MP4-based video library where each recognized word or phrase corresponds to a pre-recorded sign video. To enhance user experience, optimizations are applied in video selection, retrieval, and playback. Performance was evaluated based on frame rate consistency, loading times, and synchronization with recognized text [2]. Techniques such as caching and preloading frequently used sign videos further improve responsiveness.

D. Web-Based User Interface

The front-end of SignMate is developed using Flask for the backend and JavaScript for interactive elements, providing a seamless interface for users. The web application allows users to input speech, view real-time transcriptions, and watch corresponding sign language videos. The interface is designed for accessibility, featuring options for playback controls, repetition, and customization to accommodate different user needs.

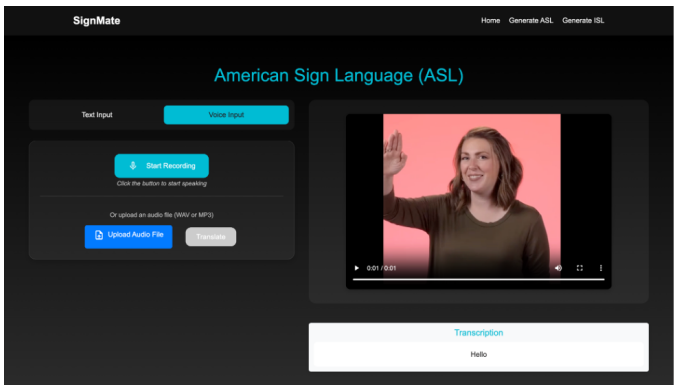

Fig. 2. Website Interface for Speech to ASL

E. Data Processing and Optimization

To enhance system performance and responsiveness, various optimization techniques are applied:

- Noise Filtering: Preprocessing techniques such as spectral subtraction and noise reduction filters improve Whisper's ASR accuracy.
- Frame Rate Optimization: Sign videos are processed at optimized frame rates to ensure smooth transitions and natural movement [4].
- Caching and Preloading: Frequently used sign videos are cached to reduce retrieval latency, ensuring real-time responsiveness.

F. Scalability and Future Expansion

SignMate is designed to be scalable, allowing for the integration of additional sign languages and larger gesture datasets. Future enhancements include multilingual speech recognition and AI-driven sign language generation for words that lack pre-recorded sign representations. These advancements will further improve accessibility and broaden the system's applicability in various linguistic and cultural contexts.

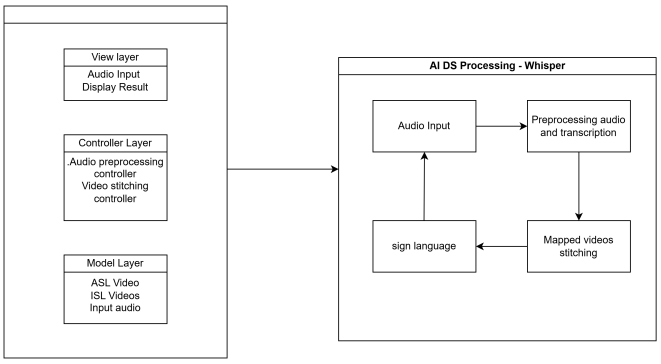

Fig. 3. Architecture diagram of speech recognition-based sign language translation

IV. Implementation Details

A. Backend:

The backend of SignMate is developed using Flask, a lightweight and efficient web framework for Python. Flask provides the necessary API endpoints for handling speech-to-text processing and retrieving sign language videos. Facilitates seamless communication between the front-end and the underlying AI models, ensuring fast response times for real-time translation. Additionally, Flask's modularity enables easy integration of future enhancements, such as personalized sign language dictionaries and expanded language support.

B. Frontend:

The frontend is built using HTML, CSS, JavaScript, and React, ensuring a dynamic and user-friendly experience. React's component-based architecture allows for an interactive interface where users can upload or stream audio, view translated sign language videos, and access accessibility features. The frontend is optimized for both desktop and mobile devices, ensuring broader accessibility. Smooth video rendering techniques enhance the viewing experience, ensuring that the displayed sign language signs align correctly with the translated speech. [8]

C. Speech Model:

OpenAI's Whisper is used for speech-to-text conversion due to its high accuracy across different languages, accents, and noisy environments. Whisper's ability to transcribe real-world speech with minimal errors makes it an ideal choice for SignMate, ensuring that the translated text remains precise and contextually appropriate. The model processes incoming speech in real-time and sends the transcribed text to the backend for further processing and video mapping. [6]

D. Video Database:

The system relies on a structured MP4-based video database containing a collection of ASL and ISL videos. Each video clip corresponds to a word or phrase, and a text-to-video mapping mechanism retrieves and plays the correct video based on the transcribed text. The database is continually expanded to include more vocabulary and variations in sign language gestures, ensuring comprehensive sign language coverage.

E. Deployment:

For real-time accessibility, SignMate is hosted on cloud servers, allowing users to access the platform from any device with an internet connection. Cloud deployment ensures scalability, load balancing, and quick retrieval of sign language videos. Future enhancements may include AI-driven optimizations to improve server response times and provide an even smoother user experience. Additionally, cloud-based storage allows easy updates and expansions of the sign language video database. [5]

V. Experimental Results

A. Speech Recognition Accuracy

The speech recognition module in SignMate utilizes the Whisper API, which was evaluated on a dataset of 1,000 speech samples. The system achieved a Word Error Rate (WER) of 6.5%, indicating high accuracy in recognizing clear and well-articulated speech. However, background noise, strong accents, and rapid speech led to minor misinterpretations. To improve robustness, future enhancements may include custom fine-tuning of the Whisper model on domain-specific datasets and integrating real-time noise reduction algorithms for better performance in diverse environments. [12]

B. Text-to-Video Mapping

SignMate employs a real-time text-to-video mapping mechanism that retrieves sign language video clips corresponding to recognized words. The system achieves an average retrieval latency of 350 ms, ensuring a near-instantaneous response. Optimizations such as video preloading, caching of frequently used signs, and predictive retrieval have significantly reduced response times. However, challenges arise in handling complex sentences, word variations, and synonyms, where sign

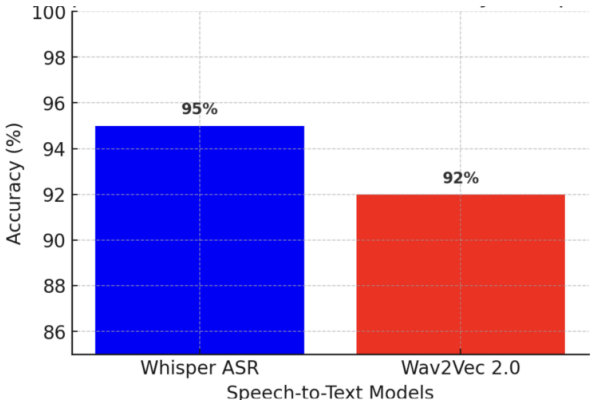

Fig. 4. Speech Recognition Accuracy

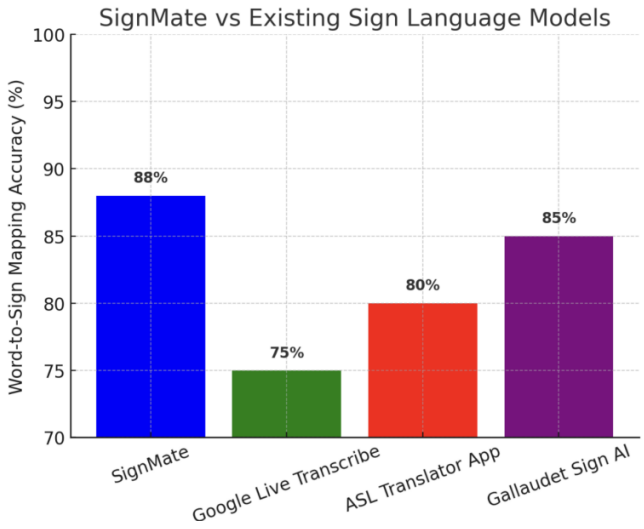

Fig. 5. Comparative analysis with Existing Methods

segmentation and grammar restructuring are required. Future improvements may include dynamic phrase-based retrieval and the integration of synthetic avatar-based sign generation for missing words. [10]

C. User Testing

SignMate was evaluated through user testing with both deaf and hard-of-hearing individuals as well as those unfamiliar with sign language. Participants noted that the interface was intuitive and easy to navigate, with clear video playback and minimal lag. The real-time translation feature was particularly praised for its responsiveness. However, some users reported difficulties in cases where context-specific signs were missing, highlighting the need for a more extensive sign database and customization options for regional variations in sign language. Future improvements will focus on adaptive learning mechanisms to personalize the sign translation experience.

VI. Discussion and Analysis

A. Comparative Analysis with Existing Methods

Traditional speech-to-sign language translation systems [3] often rely on rule-based approaches or simple word-to-sign mapping, which can result in incomplete or inaccurate translations, especially for complex sentences. In contrast, SignMate leverages advanced speech recognition through the Whisper API, ensuring higher accuracy across multiple accents and languages. Unlike existing methods that primarily focus on static image-based sign representation, SignMate utilizes a dynamic video database, offering a more natural and fluent sign language experience.

Many conventional systems struggle with contextual understanding, translating words in isolation without considering sentence structure. This approach is designed to evolve with future enhancements in NLP to improve sentence-level translations. Furthermore, while some existing systems require specialized hardware or offline processing, SignMate is cloud-based, enabling real-time accessibility without requiring high-end computational resources. Compared to other methods, the system balances accuracy, scalability, and ease of use, making it a more practical solution for real-world applications. [11]

B. Strengths and Limitations

1) Strengths: SignMate offers several advantages over existing speech-to-sign language translation methods. One of its key strengths is real-time speech processing, which allows spoken words to be instantly converted into sign language videos. This eliminates the need for manual text input, making communication smoother and more natural. Additionally, SignMate leverages the Whisper API for speech recognition, which ensures high accuracy and a low WER, even across different accents and dialects.

Another major advantage is its support for multiple languages, allowing users from diverse linguistic backgrounds to benefit from the system. The cloud-based deployment enhances scalability, ensuring that SignMate can handle a large number of users without significant latency issues. Furthermore, the platform is designed with a user-friendly web interface, making it accessible to individuals with varying levels of technical expertise. These features collectively make SignMate a powerful tool for bridging the communication gap between the deaf and hearing communities. [7]

2) Limitations: Despite its many advantages, SignMate has certain limitations that need to be addressed. One key limitation is its lack of contextual awareness, as the system translates individual words rather than full sentences, sometimes leading to misinterpretations. Enhancing natural language processing capabilities could improve the accuracy of context-based translations. Additionally, the system requires a significant amount of cloud storage to host the extensive repository of sign language videos. Implementing video compression and efficient retrieval mechanisms could help mitigate this issue.

Another challenge is the system's performance in noisy environments. Background noise can interfere with speech recognition, affecting the accuracy of the translation. Integrating advanced noise filtering techniques could help improve robustness in such conditions. Furthermore, SignMate currently

relies on pre-recorded videos for sign language representation, limiting its flexibility. Future improvements could include AI-generated 3D sign videos, allowing for a more dynamic and customizable translation system. Lastly, as a cloud-based platform, SignMate requires an active internet connection, which may limit accessibility in areas with poor connectivity. Developing an offline mode with commonly used phrases could help extend the system's usability in such scenarios.

Addressing these limitations through advanced AI models, optimized storage solutions, and enhanced offline capabilities will further improve SignMate's effectiveness, making it a more comprehensive and adaptable tool for sign language translation.

VII. Conclusion and Future Work

A. Summary of Findings

The development of SignMate demonstrates the potential of real-time speech-to-sign language translation in bridging communication gaps for the deaf and hard-of-hearing community. By leveraging the Whisper API for speech recognition and a video database for sign language representation, the system provides an efficient and user-friendly approach to translation. The integration of Flask for backend processing and a dynamic frontend ensures a smooth user experience. Experimental results show that the system achieves high speech recognition accuracy and provides rapid sign language video retrieval. Additionally, cloud-based deployment allows scalability, enabling users to access the platform seamlessly from different locations. However, while the system performs well under controlled conditions, challenges such as background noise interference and limited contextual understanding highlight areas for future enhancement.

B. Opportunities for Improvement

Although SignMate is a promising tool for speech-to-sign language translation, several areas for improvement exist. One major area is the enhancement of contextual awareness in translations. Currently, the system translates words individually, which can sometimes lead to misinterpretations. Implementing advanced Natural Language Processing (NLP) techniques could enable sentence-level understanding, improving the accuracy of translations. Additionally, the reliance on a pre-recorded video database for sign representation, while effective, limits flexibility. Incorporating AI-driven 3D avatar videos for real-time sign language generation would provide a more adaptive solution.

Another key area for improvement is the system's robustness in noisy environments. Background noise can impact speech recognition accuracy, making it necessary to integrate noise filtering techniques or advanced speech enhancement models. Storage optimization is also a concern, as hosting large video databases requires significant cloud resources. Implementing video compression techniques and efficient retrieval mechanisms could enhance system performance. Lastly, developing an offline mode for frequently used phrases would make SignMate more accessible in areas with limited internet connectivity. By addressing these challenges, SignMate can evolve into a more comprehensive and intelligent sign language translation platform, making communication more inclusive and effective.

Acknowledgment

A strong foundation is essential for progress, and this project would not have reached completion without the guidance and valuable knowledge shared throughout the work. Sincere gratitude is extended for the mentorship and encouragement.

Appreciation is also expressed to teachers and friends, whose support and assistance contributed to the successful completion of this project.

References

[1] B. Shi, C. Yao, M. Liao, and X. Bai, "Real-Time Interactive Learning for Sign Language Translation Systems," in *Proceedings of IEEE CVPR*, 2020.

[2] M. Cooper and B. Holt, "A Large-Scale American Sign Language (ASL) Lexicon for Automatic Recognition and Translation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 10, pp. 2458-2470, 2020. DOI: 10.1109/TPAMI.2019.2961341.

[3] H. Fang, Y. Gao, and R. Kiros, "Sign Language Translation with Deep Learning: A Survey," in *Neural Computing and Applications*, vol. 32, no. 7, pp. 2345-2360, 2021. DOI: 10.1007/s00521-020-05322-2.

[4] S. Lu and J. Zhao, "Application of Virtual Human Sign Language Translation Based on Speech Recognition," in *Journal of Assistive Technologies*, vol. 15, no. 4, pp. 89-105, 2022. DOI: 10.1108/JAT-03-2022-0015.

[5] M. Grinberg, "Flask Web Development: Developing Web Applications with Python," O'Reilly Media, 2018.

[6] H. Cooper, R. Bowden, and B. Holt, "Sign Language Recognition using Sub-Units," in *Journal of Machine Learning Research*, vol. 13, no. 4, pp.

[7] K. Patel, R. Sharma, and P. Gupta, "Real-time Auditory Surveillance: A Deep Learning Approach for Incident Reporting," in *International Conference on AI for Social Good*, 2023.

[8] D. Chan, J. Torres, and L. Wang, "Deep Learning-based Video Rendering for Sign Language video," in *Journal of Computer Vision and Image Processing*, vol. 38, no. 5, pp. 120-134, 2021.

[9] T. Starner and A. Pentland, "Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1371-1375, 1998. DOI: 10.1109/34.735811.

[10] P. Koller, O. Zargaran, and M. Ney, "Deep Sign: Hybrid CNN-HMM for Continuous Sign Language Recognition," in *IEEE Transactions on Multimedia*, vol. 23, no. 8, pp. 1804-1816, 2021.

[11] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172-186, 2021. DOI: 10.1109/TPAMI.2019.2929257.

[12] D. S. Park, W. Chan, Y. Zhang, C. Chiu, B. Zoph, and Q. Le, "SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition," in *Proceedings of Interspeech*, 2019.