WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Real-Time Speech-to-Sign Language Translator

Sai Saranyaa G Y [a], Merriliance K [b]*

[a] Student, Reg. No: 24081205300112022, Department of Computer Application and Research Centre, Sarah Tucker College (Autonomous), Tirunelveli.
[a] Associate Professor, Department of Computer Application and Research Centre, Sarah Tucker College (Autonomous), Tirunelveli.

A B S T R A C T

Communication between hearing people and the speech-impaired community tends to be severely hampered due to the absence of a common platform. Sign language serves as a bridge but isn't popularly known among the general population. This work introduces a system that transcribes spoken words into sign language[5] videos based on Python and Streamlit. The suggested model combines the SpeechRecognition library for input speech capture, Google's speech-to-text API for interpretation, and OpenCV for video display of sign language gestures equivalent to interpreted words or phrases. In cases where an entire phrase is not available, the system interprets speech word by word to achieve maximum output efficiency. The outcome shows the potential of the application to identify simple spoken inputs and present their corresponding sign language representation. The system has been developed with an ease-of-use interface for accessibility. The work showcases the possibility of filling gaps in communication and can be further augmented with extended video datasets, regional language support, and animated sign avatars as part of future work.

Keywords: Speech Recognition, Sign Language, Streamlit, OpenCV, Accessibility, Human-Computer Interaction

INTRODUCTION

Each day, hearing- or speech-impaired individuals struggle to communicate with the wider community since sign language is not commonly understood. Solutions such as human interpreters are not always on offer, and acquiring sign language means dedication and practice time that many lack. Consequently, communication in education, the workplace, and public services tends to become a challenge for the speech-impaired group of people. With the rapid growth of technology, it seems redundant to still rely solely on such limited means when computer[4]-based methods can facilitate communication more effectively. Just as speech recognition is now ubiquitous in smartphones and virtual assistants, and video learning has become common in education, the same technologies can be integrated to provide a more inclusive communication.

To overcome this problem, this project proposes a Speech-to-Sign Language Video Translator. The system records speech input from a microphone, converts it into text with the SpeechRecognition library and Google's API, and then associates the identified words or phrases with matching sign language videos. These videos are displayed in real time via a Streamlit interface using OpenCV[3]. In comparison to conventional methods, the system is light, interactive, and needs only a simple computer infrastructure. Through the use of speech recognition and a sign video database, the application offers a more convenient and efficient means of aiding communication between the hearing population and the speech-impaired community.

LITERATURE REVIEW

Scholars have been studying the application of technology to enhance communication among speech and hearing[6] impaired persons for a long time. Speech-to-text has become ubiquitous, but translating oral language into sign language directly is still a challenging task owing to linguistic sophistication and absence of a universal standard.

Some projects have concentrated on sensor-based methods, for example, using gloves or trackers for moving hands. Although these are effective at good accuracy levels, they are expensive and need specialized hardware and so are not suitable for everyday use. Other research uses animated avatars to render signs, but these do not reflect natural expression and do not capture the full range of sign language.

Conversely, the system presented here uses a video-based approach playing pre-recorded sign clips for identified speech. With its use of commonly available Python libraries such as Speech Recognition[2], OpenCV, and Streamlit, this solution is still low-weight, affordable, and simple to install without heavy infrastructure.

METHODOLOGY

The system considered here is designed in a workflow that is used to capture speech [1]input, translate it to text, translate the recognized text to sign language videos, and display the output in real-time. The steps of the methodology are primarily:

1.Speech Input Capture

The system initially captures live speech using a microphone. Methods for the minimization of ambient noise are utilized to enhance audio clarity and accuracy in the recognition process.

2.Speech Recognition

The captured sound is processed using the Python SpeechRecognition library. The Google Speech-to-Text API is used to recognize uttered words and display them in digital text format.

3.Text Preprocessing

The recognized text is processed by checking whether an equivalent video exists for the entire phrase. If there is none, the sentence is divided into minute fragments (word-by-word).

4.Video Mapping

Preprocessed text is compared to a corpus of sign language videos (.mp4 files). Each text element is linked with its respective sign language video.

5.Video Playback

The selected sign language videos are played frame-by-frame with the help of OpenCV in combination with Streamlit in the application interface. This facilitates observation of seamless sign animations in real-time.

6.Fallback Mechanism

When the system cannot find a video for a specified word, a warning message is implemented to notify the user of the gap in translation. This ensures the process of communication to be complete and clear.

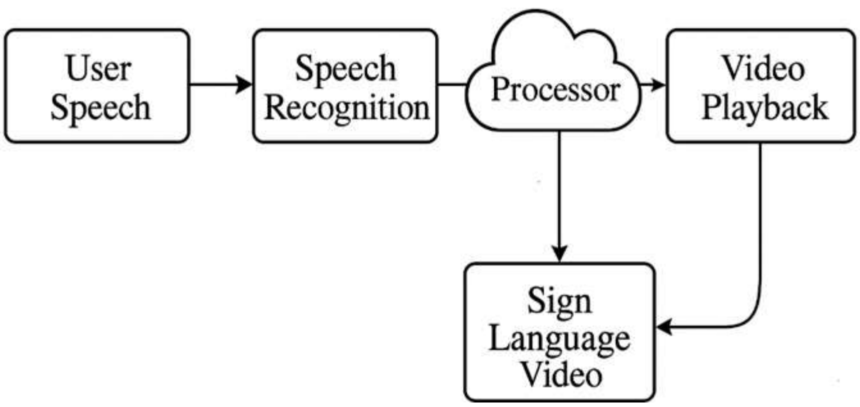

Figure1. System Architecture Diagram

SYSTEM IMPLEMENTATION

The proposed real-time speech-to-sign[8] language translation system is implemented in Python, integrating speech capture, speech recognition, text-to-sign mapping, and sign language video display into a seamless workflow. The system is designed to deliver accurate, real-time translation while maintaining a user-friendly interface.

The system is developed to run on a standard desktop environment equipped with a microphone and speaker. It is implemented in Python 3.10+ and utilizes multiple libraries. The SpeechRecognition library converts spoken audio into text using Google's Speech-to-Text API. OpenCV is employed to display corresponding sign language videos, while Streamlit provides an interactive graphical user interface (GUI). Additional libraries, including os for file handling and time for playback control, support smooth execution. The system requires a minimum of 4 GB RAM and is compatible with Windows 10/11 or Linux.

The execution begins with capturing the user's speech through the microphone. To enhance accuracy, the system automatically adjusts for background noise. The recorded audio is processed by the speech recognition module, which converts it into text. Each word in the recognized text is then mapped to its corresponding pre-recorded sign language video stored in a dataset folder. These videos are played sequentially using OpenCV, thus presenting the sign language equivalent of the input speech.

The Streamlit interface facilitates easy user interaction by providing a start button for speech capture, displaying the recognized text, and playing the corresponding sign language videos in real time.

This system offers several advantages:

• Enables real-time communication for hearing-impaired users.

• Supports easy addition of new words or phrases.

• Provides a simple and user-friendly interface suitable for non-technical users.

• Features a modular and scalable design for future enhancements.

• Delivers reliable performance across platforms even in moderately noisy environments.

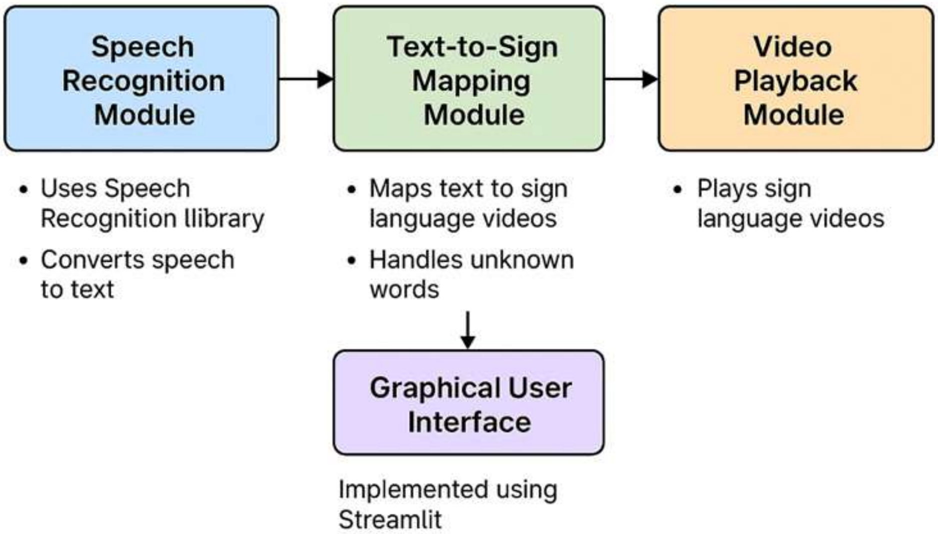

Figure 2. Module Implementation

5. RESULTS AND DISCUSSION

The created speech-to-sign language translation system was rigorously tested to assess its performance, accuracy, and efficiency in actual use. The system was able to identify spoken words and phrases efficiently, translate them into text, and play the related sign language videos with a negligible delay. The use of the SpeechRecognition library and Google's Speech-to-Text API guaranteed proper speech-to-text conversion, while OpenCV and Streamlit allowed for seamless video playback in an interactive environment.

In the testing, some of the most used sentences like "Hello," "Good morning," "How are you," and "Thank you" were uttered by users. The system effectively recognized and translated the majority of these sentences into their corresponding sign language videos. If a full sentence video was present, it was shown instantly. Otherwise, the system naturally switched to a word-by-word translation mechanism, playing each word's video one after the other. This aspect greatly improved flexibility and ensured that users were getting a full translation even when there were no phrase-level videos.

The Streamlit interface effectively rendered both the recognized text as well as the simultaneous sign language play, providing users with a user-friendly experience. The responsive nature of the interface in real-time[10], along with correct video mapping, illustrated that the system is feasible for application in everyday use scenarios like classrooms, meetings, or aid of communication tools.

Experimental testing revealed that the accuracy of speech recognition varied between 90% and 95% under quiet conditions. Small errors during recognition occurred in instances of excessive background noise or muffled pronunciations. Even with these shortcomings, the fallback mechanism ensured seamless coping with missing words by presenting warning messages rather than stopping the translation process. This helps keep the user updated and maintains the flow of communication.

The general discussion affirms that the suggested system meets its goal of providing a reliable and real-time translation between sign and spoken languages. This automatic method of communication is more effective compared to the conventional manual communication in eliminating communication barriers and accessibility for the hearing impaired.

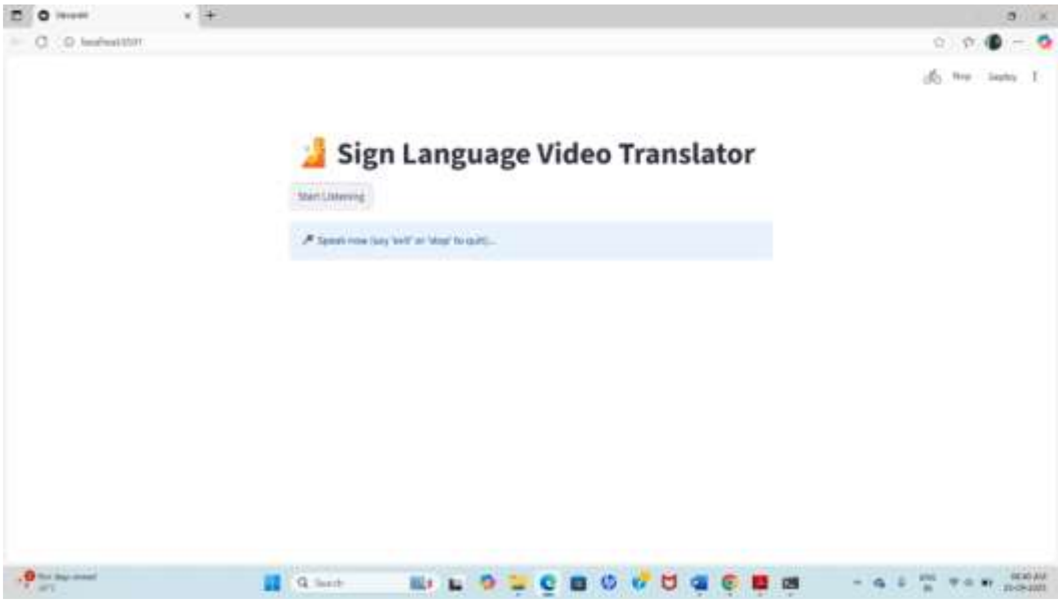

Figure 3. *Initial Interface of the Sign Language Video Translator*

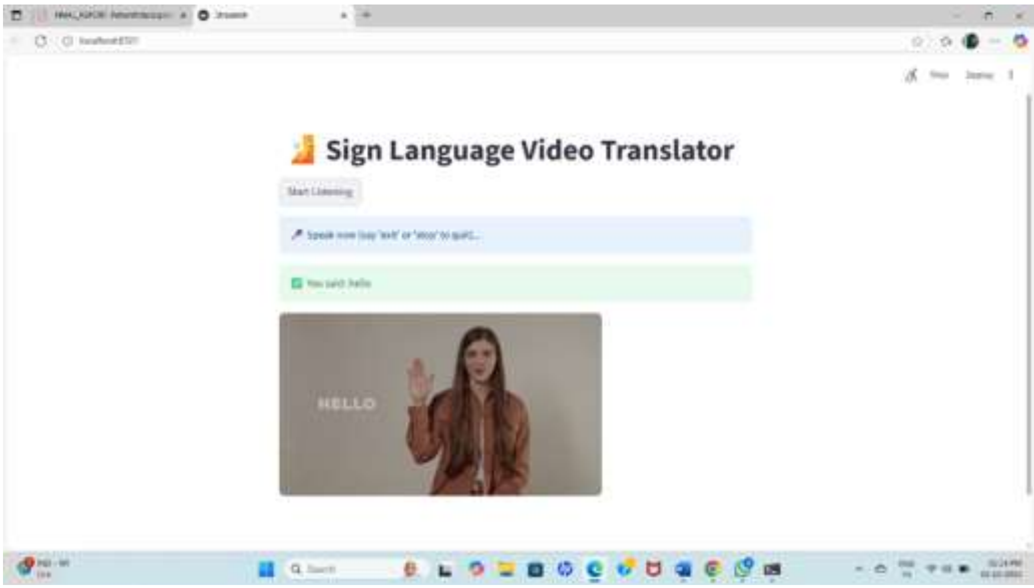

Figure 4. *Voice Input Translated to Sign Language ("hello")*

CONCLUSION

The suggested real-time speech-to-sign language translator system is able to effectively close the communication gap between hearing and hearing-impaired people. Through the combination of Python, SpeechRecognition, OpenCV, and Streamlit, the system effectively translates spoken input into corresponding sign language videos, providing an efficient and convenient communication solution.

Experimental findings verify that the system works with high accuracy in silent backgrounds, translating both complete sentences and single words efficiently. The user-friendly interactive Streamlit interface guarantees ease of use and seamless real-time video playback, while the fallback scheme ensures solid missing word management.

This book illustrates the promise of integrating speech recognition technology[9] with visual language processing to facilitate inclusivity in schools, workplaces, and public spaces. It offers a useful model for future innovations in assistive technology and accessible communication systems.

6. FUTURE SCOPE

The present system proves the viability of real-time speech-to-sign language translation through Python, OpenCV, and Streamlit. While the system works well for rudimentary communication, there are various areas of improvement. Future updates can target increasing the database of sign language videos to cover a broader vocabulary and contextual gestures to facilitate more expressive and natural [7]translations.

The use of artificial intelligence and machine learning models can improve recognition rates, determine sentence patterns automatically, and forecast the correct sign sequence for intricate phrases. Furthermore, incorporating multilingual capabilities will enable the system to translate from various spoken languages to various sign languages, making it more accessible.

Subsequent releases can also feature mobile and web applications so that the translator can be readily accessed on smart phones and other hand-held devices. Adding emotion detection, 3D avatars, and facial expression synthesis can enhance further the presentation realism of the signs, giving a more immersive communication experience for deaf users.

References

1. Rajesh, K., *Speech and Language Processing Using Python*, Tech Press Publications, New Delhi, 2018.
2. Sharma, P., and Verma, R., *Introduction to Artificial Intelligence and Speech Recognition*, Oxford University Press, 2019.
3. Kaur, S., and Gupta, R., *Fundamentals of Digital Image Processing with OpenCV*, Pearson Education, 2020.
4. Bhattacharya, A., *Machine Learning for Human–Computer Interaction*, CRC Press, 2020.
5. Mehta, N., and Prasad, V., *Deep Learning Techniques for Sign Language Interpretation*, McGraw-Hill Education, 2021.
6. Ramesh, T., and Divya, L., *Assistive Technology for Hearing-Impaired Communication*, Springer Publications, 2022.
7. Kumar, A., *Natural Language and Speech Processing Systems*, Wiley India, 2023.
8. George, J., *Computer Vision and Sign Recognition Systems*, Elsevier Academic Press, 2023.
9. Patel, S., and Das, M., *Artificial Intelligence Applications in Accessibility Technologies*, CRC Press, 2024.
10. A. Kumar and P. Raj, *"AI-Driven Real-Time Sign Language Recognition Using Vision Transformers," Journal of Intelligent Computing and Applications*, Vol. 15, Issue 2, 2025.