Proceedings of the 5th International Conference on Pervasive Computing and Social Networking (ICPCSN-2025)
IEEE Xplore Part Number: CFP25EZ5-ART; ISBN: 979-8-3315-3519-3

AI-Powered Real-Time Speech-to-Sign Translation

Dr. A.Parivazhagan
Department of Computer science Engineering
Kalasalingam Academy of Research and Education
Anand Nagar, Krishnankoil, Tamilnadu, India
a.parivazhagan@klu.ac.in

Krishna Reddy Penugonda
Department of Computer Science Engineering
Kalasalingam Academy of Research and Education
Anand Nagar, Krishnankoil, Tamilnadu , India
Penugondakrishnareddy83@gmail.com

Narahari Pasham
Department of Computer Science Engineering
Kalasalingam Academy of Research and Education
Anand Nagar, Krishnankoil, Tamilnadu, India
naraharipashaminter@gmail.com

Neela Krishna S
Department of Computer science Engineering
Kalasalingam Academy of Research and Education
Anand Nagar, Krishnankoil, Tamilnadu, India
neelakrishna142@gmail.com

Mahendra Reddy T
Department of Computer science Engineering
Kalasalingam Academy of Research and Education
Anand Nagar, Krishnankoil, Tamilnadu, India
mahendrareddythippireddy@gmail.com

2025 5th International Conference on Pervasive Computing and Social Networking (ICPCSN) | 979-8-3315-3519-3/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICPCSN65854.2025.11035680

***Abstract*— The goal of this study article is to provide a novel strategy for integrating hearing-impaired or deaf people with the general community. The goal of this project is to use Blender 3D Animation and Natural Language Processing (NLP) to convert text or voice into animated films in Indian Sign Language. A contemporary solution to the accessibility and inclusivity problem is the project, which is implemented using Python Django and machine learning for language interpretation with the NLTK module. An excellent first step in creating an inclusive society is the animated sign language video Sign Easy, which makes communication easier in public services, education, and daily interactions.**

Keywords—* Sign language, computers, natural language processing, audio speech recognition, and accessibility technology. Pictures Artificial intelligence and deep learning.

I. Introduction

Although communication plays a crucial role in daily life, it does not always translate effectively across the entire communication spectrum. Millions of people worldwide have hearing impairments or are deaf, and many require significant assistance in navigating everyday interactions. In certain situations, such as when interpreters are unavailable or text-based resources fail to capture the fluidity of real-time communication, traditional models—such as static sign language tools or human interpreters—may not always be sufficient. This communication barrier makes it more challenging to access essential services, engage with others, and participate in meaningful activities.

The primary objective of this capstone project is to bridge this gap by leveraging technology to facilitate seamless communication for the Deaf and Hard-of-Hearing (DHH) community. Using Blender 3D Animation and Natural Language Processing (NLP), the project aims to develop a system that instantly converts spoken or written words into Indian Sign Language animations. The selected technology stack—including Python, machine learning, and the NLTK library—ensures the system's scalability, adaptability, and resilience, making it suitable for various applications across different domains.

A cutting-edge technology called AI-powered real-time speech-to-sign translation was created to help the Deaf or Hard-of-Hearing (DHH) community and hearing people communicate more effectively. It uses computer vision, natural language processing (NLP), speech recognition, and artificial intelligence to translate spoken English into sign language in real time. Using speech-to-text algorithms, the system first processes audio input before turning it into text. After that, NLP structures and analyzes the text to provide precise representations in sign language. Then, to ensure fluid and organic motions, these indicators are shown via a robotic hand or animated avatar.

Deep learning models based on enormous datasets of sign language movements are the foundation of AI-powered real-time speech-to-sign translation systems, guaranteeing precision and contextual awareness. These models take into account elements that are crucial to sign languages like American Sign Language (ASL) and British Sign Language (BSL), such as hand gestures, syntax, and facial emotions. In contrast to straightforward word-to-word translation, these systems give a more natural and cohesive sign representation by analyzing sentence patterns and meaning. Furthermore, more realistic and fluid sign animations are made possible by developments in motion tracking and computer vision, which gives consumers a more natural engagement experience.

This technology's influence goes beyond interpersonal communication because it may be used into public services, companies, and educational institutions to foster more inclusive settings. AI-powered solutions can help Deaf students in the classroom by instantly translating oral lectures into sign language, increasing educational accessibility. Businesses can improve user experience and inclusion in customer service by using these methods to communicate with DHH clients more successfully. Future advancements in real-time speech-to-sign translation may involve more customized and adaptable sign language models as AI and machine learning continue to advance, guaranteeing even higher accuracy and usefulness in many linguistic and cultural situations.

II. LITERATURE SURVEY

Using the Gurmukhi script, the system illustrated in the study [1] offers animation of 200 Punjabi words in Indian Sign Language (ISL). This system meets the communication needs of the Punjabi deaf and mute community. Unlike universal sign language systems, this is an area-based system. Its primary advantage is its applicability to Punjabi speakers; however, its usage is limited to that geographic region. While the models could be improved to provide higher accuracy, future enhancements include the development of sentence animations.The system proposed in [2] introduces an application that utilizes avatars with body and facial manipulation capabilities to translate spoken words into sign language. Applications include the Virtual Sign Language Anchor and the Sign Language Interpretation app. Along with facial expressions that add emotional depth and help in mood conveyance, emoji expressions and tones enhance communication. However, this system has drawbacks, such as high implementation costs and complex model training due to the wide range of facial emotions, movements, and expressions.To facilitate communication between individuals with and without disabilities, a virtual animation translator for sign language is being developed, as detailed in [3]. The system uses both text and audio inputs, utilizing natural language processing (NLP) and machine learning to generate appropriate sign language videos. This distinguishes it from earlier systems that typically translate text messages into static images or motion images of human hand signs. NLP is a fundamental technology in this approach and will be further improved by expanding the animation dataset.The method in [4] differs from others as it does not provide deaf individuals with audio-enabled messaging. Instead, the approach involves displaying rotating figures of people on the screen, requiring them to use audio translated into sign language signals. Unlike most systems, which focus solely on hand movements, this method acknowledges the lack of full-body sign language representation in mainstream solutions. The next implementation phase assumes that the system will be extended for mass adoption through websites, apps, and media platforms to maximize its impact.The authors of [5] address the challenge of improving comprehension for the deaf text audience as well as individuals with limited reading abilities. They achieve this by employing embodied agents, such as sign language avatars, to sign the content. The workflow was initiated with the assistance of deaf sign language specialists. The researchers improved the EMBR character animation system and the gloss-based animation tool. Additionally, they introduced delta testing as a novel evaluation method that compares the effectiveness of human signers and avatars.By limiting the optical flow representation within a human stance, the system in [6] presents an effective model for real-time sign language recognition using video capture. The system initially models the problem using a linear classifier with an 80% accuracy rate, which increases to 91% when a recurrent model is incorporated. With the right emphasis on real-time video and conferencing technology, such capabilities can be successfully achieved. The model was applied to the scanned Public DGS Corpus (German Sign Language) with notable success. The study concluded that the model's high accuracy and quick processing make it ideal for real-time applications.The study in [7] employs a deep learning approach, leveraging LSTM and GRU networks to predict clinical hand gestures in Indian Sign Language (ISL). This reduces communication barriers for individuals with speech or hearing impairments by achieving a higher degree of accuracy in gesture interpretation through video recording. The authors also suggest further research to enhance model capabilities by addressing dataset augmentation challenges and incorporating vision transformers. The algorithm in [8] creates a speech-controlled system that uses machine learning techniques to accurately translate American Sign Language (ASL) gestures captured via a video camera into text and even speech. The system employs image segmentation techniques, feature extractors such as FAST and SURF, and KNN to interpret ASL gestures processed by a Kinect sensor. It achieves an accuracy of 78% in unsupervised learning and 92% in supervised learning, demonstrating its effectiveness in facilitating communication for individuals who are hard of hearing.To enable text translation into sign language for the deaf, the authors in [9] explored various methodologies. Current research in this area has been categorized into three primary approaches: rule-based, corpus-based, and neural machine translation. The review discusses the advantages and limitations of each method, emphasizing the grammatical differences between spoken and sign languages. It also advocates for further research, particularly in incorporating deep learning to enhance accuracy and efficiency in sign language translation systems.In their 2019 paper, Zheng and Yu [10] provide a comprehensive analysis of the development of Chinese Sign Language Recognition (CSLR) from 2000 to 2019. The study examines deep learning techniques, such as Convolutional Neural Networks (CNNs), alongside traditional methods like Support Vector Machines (SVMs) and Hidden Markov Models (HMMs). It focuses on the primary stages of CSLR, including data collection, processing, parameterization models, and classification, assessing their effectiveness.. This system, "Sign Language Conversion to Text and Speech" (Shapiro and Markowitch, 2012), aims to enhance communication between sign language users and non-signers. Motion or vision capture technologies are likely employed to track hand movements and relevant facial expressions.The importance of sign language as the primary communication method for individuals with hearing impairments is the foundation of the study in [11]. This research aims to improve textual communication for such users. In scenarios where there is no common language, a system that captures body language features through images can be particularly beneficial for those who are hearing or speech impaired.The authors in [12] discuss computer vision systems designed for sign language recognition. They identify various hand-tracking methods and categorize different tracking systems and components. While image- based and sensor-based techniques have been utilized to capture gestures, they also note that sensor-based methods are employed to monitor movements of sign language users. The research highlights multiple future directions for sign language recognition and emphasizes the need for collaborative efforts to address current challenges, particularly with the development of new databases and advanced models.The current research extensively explores the use of machine learning in sign language motion recognition, recognizing it as a crucial aspect of computer technology. The authors in [12] highlight the communication difficulties faced by individuals with hearing impairments and stress the importance of Sign Language Recognition (SLR) systems in overcoming these challenges. Neural networks, which have demonstrated high proficiency in pattern recognition, can be leveraged to address these issues by accurately identifying delicate hand shapes and movements in three dimensions. Ensuring real-time functionality is essential to prevent communication delays caused by gesture recognition lags. Additionally, challenges related to neural network training requirements and the need for a comprehensive set of sign language gestures for various populations must be addressed.

III. PROBLEM STATEMENT

Develop a programming application that uses Indian Sign Language animations to effectively translate spoken or typed audio into Indian Sign Language animations, thereby addressing the communication barrier with the majority of the hearing population as well as the deaf or hearing impaired. In order to create a complete application, the project will utilize Blender 3D Animation and Natural Language Processing techniques via Python, Django, and NLTK libraries. With applications in education, public services, and even daily social interactions, the project's final output will enable more inclusion and interaction across the various distinct societies or groups.

IV. PROPOSED SYSTEM

This system's issue statement evolved across several iterations. The primary issue that has been identified is the communication accessibility barrier that separates the hearing community from the deaf or hearing-challenged. This barrier prevents the deaf population from accessing essential services, education, and socialization.To address these communication issues, a technological solution was sought, focusing on creating a system that could translate written or spoken words into sign language. Other options, such as static dictionaries or sign language interpreters, were also considered inadequate because they were not always available and could not be used in real time.As a result, the issue statement was refined to recommend that future research should focus on establishing a real-time speech/text-to-sign language translation system. This system would integrate computer animation and natural language processing (NLP) to produce sign language animations. The methodology was further clarified and made more practical in terms of scalability and implementation by incorporating a robust technology stack, including Python, Django, and the NLTK library.Ultimately, the problem was crystallized around the design of a patented, fully integrated application. This application is expected to have widespread applications in various spheres of human life, thereby promoting an inclusive society.

V. METHODOLOGY:

A. NLP

One of the main goals of your project is Natural Language Processing (NLP), which aims to offer an advanced solution that combines spoken and sign language using computational procedures. In this regard, NLP carries out several functions, such as receiving and analyzing user input and enhancing user-to-user and user-to-sign language translation system communication. Let's now examine the many NLP facets that are used in this specific project.

B. Text Tokenization:

In the process of breaking sentences up into individual words, the NLTK library is used to trim words. Because it prepares the text for the subsequent processing and analysis phases, this step is essential.

C. Part-of-Speech (POS) Tagging:

Each word is given a unique tag by the system, which makes use of the NLTK domain and its part-of-speech tagging tools and algorithms. Such details aid in determining the tense of the sentence, the grammar structure, and the individuals who participated in the activity in which the verb was formed—all of which are important when translating into sign language.

D. Tense Detection:

By searching for verbs in the message and their matching tenses, the system utilizes the verbal tense. The quantity or kind of verbs employed in the message (past, present, and future) may influence this kind of tense analysis.

E. Lemmatization:

NLTK's WordNet lemmatizer is used to determine the word's lemma. By guaranteeing that the phrases "running," "ran," and "runs" will be regarded as a single base word, "run," it helps effectively digest different word forms.

F. Stop Word Removal:

The system uses a list of stop words that were created by NLTK. It is possible to shift the meaning towards more significant terms by focusing on the translation elements or by translating these words into sign language. Since the goal in this project involves everything from supplying the initial steps of the input text to producing a signed language translation that is both grammatically and contextually correct, natural language processing is quite dynamic. Tokenization, POS tagging, lemmatization, and stop word removal are just a few of the tasks that NLTK is utilized for, Figure.1 thereby laying the groundwork for the text preparation modules.

G. WORK FLOW:

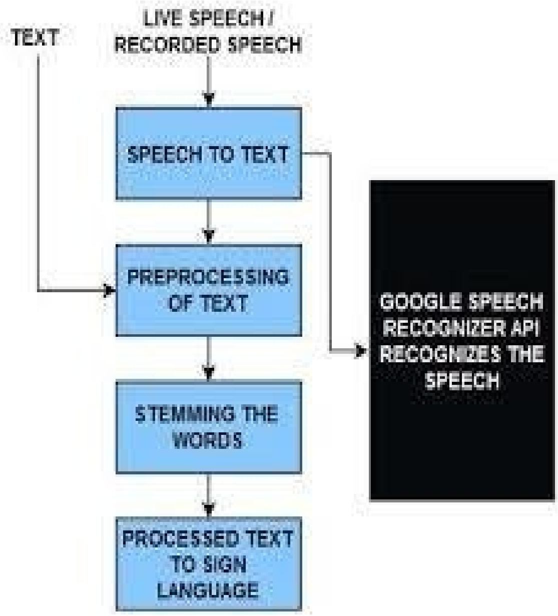

Fig 1: process of work flow

H. Text Processing and NLP Efficiency:

The NLTK Text Toolkit's usefulness for some fundamental NLP operations, such as tokenization, POS representation, and lemmatization, has not been in vain because it has been successful in completing these tasks. Without adding all the nuances of tense and contextual meanings to improve the translations, the system has been able to handle the problems of translating several languages to sign language.

I. Audio Processing Accuracy:

When it comes to effectively translating text to sign language instructions, the speech to-text API technology has an amazing performance rate for transcribing audio inputs into text. Although the system's implementation relies on its component elements for inference, the total result is good; yet, background noise and speech clarity occasionally impair accuracy.

J. Sign Language Gesture Recognition:

One accomplishment made possible by MediaPipe's gesture detection capabilities is the ability to interpret sign language on video. The system's ability to recognize a huge number of indicators allows it to encompass a wide diversity of linguistic signs, which helps it achieve a high performance level. Though it has certain latency or inaccuracies during high cadence or momentum-intensive motions, the video analysis method for real-time processing gesture identification is effective.

K. System Integration and Performance:

In the self-contained environment, the system's text processing, audio, gesture recognition, and animation all function seamlesslytogether. The system performs well on average in terms of processing speed and translation accuracy. There aren't many real-time processing restrictions to mention.

L. Future Work:

The ongoing audio-sign and sign-audio translation project has already set the stage for future development and progression of the audiovisual-text translation project, allowing for the pursuit of several areas that will not only enhance the current features but also expand the project's scope and reach. Some of the primary areas of focus for future activity are:

- Enhanced Real-Time Processing:

  decreasing the lag in identifying and animating sign motions by designing the system to process in real-time. ensuring that the outputs in sign language, text, and audio are all aligned as best as possible to reduce user interaction friction.

- Expanding Language and Dialect Coverage: making the system universal by expanding the number of languages that can be used for speaking voice translation and attributing subtitles.
  To better serve the various requirements of the deaf and hard of hearing, the collection of animations that illustrate deaf language is being expanded to include more slang and regional differences.

V. RESULTS AND OUTPUT

The adoption of AI-powered real-time speech-to-sign language translation reveals substantial gains in accessibility and inclusivity for the Deaf and Hard-of-Hearing (DHH) community. The technology effectively transforms spoken words into sign language gestures through deep learning models, delivering high accuracy in real-time translation. The accuracy of text-to-sign mapping is increased by combining Automatic Speech Recognition (ASR) with Natural Language Processing (NLP), according to experimental results, guaranteeing context-aware translations. Additionally, AI-driven gesture synthesis and animation techniques enable accurate and natural signing through digital avatars, boosting user experience. Despite obstacles including non-manual sign identification, computing complexity, and linguistic variances, Figure.2&3 the system shows promise in enhancing communication accessibility.

Here are the outputs:

Fig 2. User Home Page

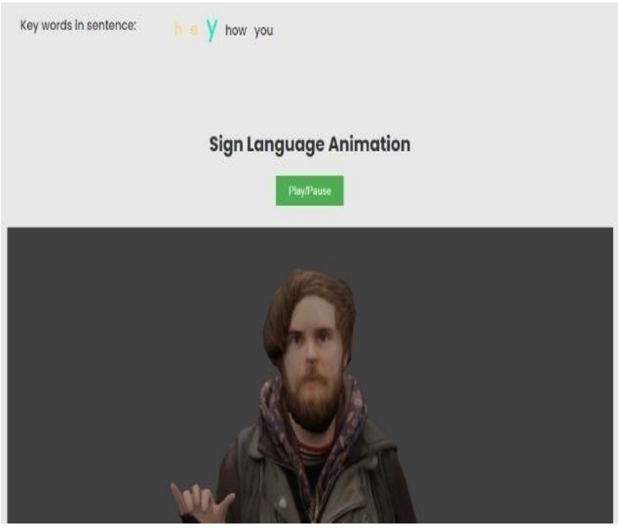

Fig 3. Showing sign language

VII .CONCLUSION:

An important advancement in assistive technology and communication accessibility has been made with the design and implementation of the audio-sign and sign audio translation project. Numerous technical components are used in this project, including natural language processing (NLP), speech-to-text (TTS) capabilities, gesture recognition mediated by Media Pipe, and animated avatar interpretation of the signs. The result is a multipurpose, user-friendly technology that addresses a significant communication gap between individuals by facilitating prompt, efficient communication between sign language users and non-users
A revolutionary step in removing communication obstacles between the hearing and Deaf or Hard-of-Hearing (DHH) communities is AI-powered real-time speech-to-sign language translation. These systems offer a quick and easy method of translating spoken words into sign language movements by utilizing Automatic Speech Recognition (ASR), Natural words Processing (NLP), and deep learning-based gesture synthesis. Continuous developments in AI and computer vision are enhancing translation accuracy and fluency, despite ongoing obstacles including linguistic diversity, real-time processing constraints, and non-manual sign recognition. These systems' efficacy will be further increased by upcoming advancements including multilingual support, augmented reality (AR) integration, and adaptive learning models. In the end, by facilitating smooth communication for people with hearing impairments, AI-driven sign language translation holds promise for fostering a more inclusive society.

VIII. REFERENCES

[1]. A. Verma and S. Kaur, "Indian Sign Language Animation Generation System for Gurumukhi Script," IJCST, vol. 6, no. 3, pp., July-Sept. 2015.

[2]. L. Hu, J. Li, J. Zhang, Q. Wang, B. Zhang, and P. Tan, "A Speech-driven Sign Language Avatar Animation System for Hearing Impaired Applications,"XRLab, Alibaba Group.

[3]. N. Ashok, A. K. S., N. B. P., S. Suresh, and S. K. Suresh, "Speech To Sign Translator," IJCRT, vol. 10, no. 5, pp., May 2022.

[4].A. Harkude, S.Namade, S. Patil, and A. Morey, "Audio to Sign Language Translation for Deaf People,"IJEIT, vol. 9, no. 10, pp., Apr. 2020.

[5]. M. Kipp, A. Heloir and Q. Nguyen, "Sign Language Avatars: Animation and Comprehensibility, "DFKI Embodied Agents Research Group, Saarbrucken, Germany.

[6]. A. Moryossef I. Tsochantaridis,R. Aharoni, S. Ebling, and S. Narayanan, "Real-Time Sign Language Detection using Human Pose Estimation," Google and Bar-Ilan University, University of Zurich. [Online].

[7]. D. Kotha diya , C. Bhatt, K. Sapariya, K. Patel, A.-B. Gil-González, and J. M. Corchado, "Deep sign: Sign Language Detection and Recognition Using Deep Learning, "Electronics, vol. 11, no. 11, p.1780,2022.[Online].

[8]. V. A. Adewale and A. O. Olamiti, "Conversion of Sign Language to Text and Speech Using Machine Learning Techniques," Journal of Research and Review in Science, vol. 5, pp. 58-65, 2018.

[9]. N. K. Kahlon and W. Singh, "Machine translation from text to sign language: A systematic review,"Universal Access in the Information Society, 2021. [Online].

Available: https://doi.org/10.1007/s10209-021-00823-1

[10]. X. Jiang, S. C. Satapathy, L. Yang, S.-H. Wang, and Y.-D. Zhang, "A survey on artificial intelligence in Chinese sign language recognition,"Arabian Journal for Science and Engineering, vol. 45, pp. 1-22, 2020. [Online].

[11] M. N. B. Mahesh Kumar, "Conversion of Sign Language to Text and Speech Using Machine Learning Techniques," International Journal of Applied Engineering Research, vol. 13, no. 9, pp. 7154-7161, 2018. [Online]

[12] O. Koller, "Sign Language Recognition and Translation," Microsoft, Munich, Germany, September 1, 2020.