Real-Time AI Speech-to-Sign Language Translation with Animated Avatars for Inclusive Communication

GE19612 - PROFESSIONAL READINESS FOR INNOVATION, EMPLOYABILITY AND ENTREPRENEURSHIP PROJECT REPORT

Submitted by

SNEKHA R	**(2116220701282)**
SWEDHA J	**(2116220701296)**
SNEHA SAJEEVAN	**(2116220701281)**

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE

ANNA UNIVERSITY, CHENNAI

MAY 2025

RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI

BONAFIDE CERTIFICATE

Certified that this Project titled **“Real-Time AI Speech-to-Sign Language Translation with Animated Avatars for Inclusive Communication”** is the bonafide work of **“SNEKHA R (2116220701282), SWEDHA J (2116220701296), SNEHA SAJEEVAN (2116220701281)”** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE	**SIGNATURE**
Dr. P. Kumar., M.E., Ph.D.,	Dr.SenthilPandi., M.E., Ph.D.,
HEAD OF THE DEPARTMENT	**SUPERVISOR**
Professor	Assistant Professor
Department of Computer Science	Department of Computer Science
and Engineering,	and Engineering,
Rajalakshmi Engineering College,	Rajalakshmi Engineering
Chennai - 602 105.	College, Chennai-602 105.

Submitted to Mini Project Viva-Voce Examination held on ______________

Internal Examiner

External Examiner

ABSTRACT

Efficient communication is still a challenge for the hearing and speech impaired, particularly in public meetings and government functions where live accessibility is important. Current sign language translation technologies are available in the form of sensor-based gloves, vision-based recognition through cameras, and avatar-based interfaces, all providing incomplete solutions. The methods are bound to have constraints such as hardware dependency, high expense, one-way translation, and incompatibility with ongoing and expressive communication. To overcome such difficulties, we suggest an AI-driven real-time speech-to-sign language translation system that utilizes state-of-the-art speech recognition, natural language processing (NLP), and animated 3D avatars to facilitate inclusive communication. The system records live speech through microphones, translates it into text using speech-to-text engines such as Whisper or Google STT, processes the text through NLP and context-based LLMs, and translates it into respective sign gestures through a gesture generator. These indicators are live translated by a 3D avatar onto public screens or holographic projectors. This end-to-end solution eliminates the need for human interpreters, facilitates multi-turn and multi-lingual conversation, and provides more accessibility with cultural and linguistic integrity. Dynamic and scalable, this architecture is designed to empower the hearing-impaired community through equitable access to oral material in real-world, high-impact environments.

ACKNOWLEDGMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavor to put forth this report. Our sincere thanks to our Chairman **Mr. S. MEGANATHAN, B.E, F.I.E.**, our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN**, **Ph.D.,** for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P. KUMAR, M.E., Ph.D.**, Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guides **Dr. SENTHILPANDI.** and **Dr. M. RAKESH KUMAR**, We are very glad to thank our Project Coordinator, **Dr. SENTHILPANDI** Assistant Professor Department of Computer Science and Engineering for his useful tips during our review to build our project.

SNEKHA R 2116220701282
SWEDHA J 2116220701296
SNEHA SAJEEVAN 2116220701281

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

S. No	ABBR	Expansion
1	AI	Artificial Intelligence
2`	API	Application Programming Interface
3	AJAX	Asynchronous JavaScript and XML
4	ASGI	Asynchronous Server Gateway Interface
5	AWT	Abstract Window Toolkit
6	BC	Block Chain
7	CSS	Cascading Style Sheet
8	DFD	Data Flow Diagram
9	DSS	Digital Signature Scheme
10	GB	Gradient Boosting
11	JSON	JavaScript Object Notation
12	ML	Machine Learning
13	RF	Random Forest
14	SQL	Structure Query Language
15	SVM	Support Vector Machine

CHAPTER 1

INTRODUCTION

1.1 GENERAL

Communication is a basic human right, however, millions of individuals with hearing and speech disabilities encounter daily obstacles to access spoken information in public and formal ways, like at government functions, schools, and live events. Solutions such as human interpreters or captioning systems provide partial access but are limited by availability, cost, and expressiveness. Sign language is the predominant language of the deaf community and as such relies very much on facial expressions and body movements in addition to hand movements in order to convey meaning accurately. Given the limitations of the methods available, and to promote digital inclusion, it is time to invest in automated and real-time sign language translation actuated by a systems-based approach.

The project proposes a an AI driven system that translates live speech into sign language using an animated 3D avatar. Using advances in speech recognition, natural language processing, large language models, and real-time animation - the system produces expressive and contextually appropriate translations into sign language. It is human interpreter-independent, language-agnostic, and can be domain-adaptive for mass deployment. The proposed system interfaces with the speech of the speaker making it a unique translation solution that does not simply close the communication gap, but ensures the hearing-impaired community can access information in an equal and independent manner.

1.2 OBJECTIVE

This project intends to develop and implement an AI-based real-time speech-to-sign language translation tool based on animated 3D avatars to facilitate the lives of individuals with hearing and speech impairments, especially during public events, government functions, and live performances. The system aims to eliminate the need for human sign language interpreters by utilizing automation based on modern technologies like speech recognition, natural language processing, and large language models (LLMs). The system interprets speech into accurate and context-based sign language gestures so that deaf and hard-of-hearing individuals gain access to information in a timely and meaningful manner. The inclusion of animated avatars renders the conversation visually expressive and interesting, covering not just the hand signs but also facial and body posture all critical to clear sign language performance. Besides, the system is easy to support multiple sign languages (e.g., ISL, ASL, BSL) and can readily be adapted across various domains and languages with little reconfiguration. Finally, the project aims to close communication divides, promote digital inclusion, and empower individuals with disabilities by enabling them to enjoy equal access to information in both physical and online spaces.

1.3 EXISTING SYSTEM

Since years, various systems have been proposed and implemented for enabling communication between the hearing-impaired community and the general population using sign language recognition and synthesis. They vary from sensor-based wearable systems to vision-based recognition to AI-based translators and avatar-based interpreters. A good example is the "Deaf Talk using 3D Animated Sign Language" system that uses Microsoft's Kinect v2 sensor to record gestures and translate speech into animated signs based on a dictionary of pre-recorded animations

CHAPTER 2
LITERATURE SURVEY

[1] Deaf talk using 3D animated sign language: A sign language interpreter using Microsoft's kinect v2

This system relies on Microsoft's most up to date technology of "Kinect v2" which can track gestures, depth and motion. This project has been under development at Chinese Academy of Sciences for over two years. The system will store speech on user's request and send it to AT&T's Speech Recognition server. The server will then send the text from speech recognition as a string back. The system will then tokenize this string into words and search for their corresponding animations into the dictionary saved in database. These animations will then be played sequentially. This system is trustworthy enough given the accuracy rates 84% for sign language to speech and 87% for speech to sign language conversion

[2] Posterior-Based Analysis of Spatio-Temporal Features for Sign Language Assessment

This paper suggests a technique by which the sign language is transmitted via manual(handshape, hand movement) and non-manual (facial expression, body posture, mouthing) channels. This technique is a combination of deep learning (I3D model) and statistical techniques (KL-HMM framework). The suggested system The aim of this paper is to overcome these shortcomings by suggesting an approach that integrates the strengths of both deep learning techniques and statistical techniques to allow fine-grained evaluation of sign language

[3] An Approach for Minimizing the Time Taken by Video Processing for Translating Sign Language to Simple Sentence in English

The paper explains a method to reduce the time taken for video processing module. There are three modules in the system architecture - Video Processing • Natural Language Processing • Text to Speech Conversion. It converts American Sign Language videos into a plain English sentence and then speech and then matches signs through the use of a video processing module, retrieves a corresponding Sign Writing Image File, and compares to a dictionary. The details of the matched sign are retrieved from an Excel document to populate an SSU frame, which will create a constructed English sentence. The system also converts sentence to speech.

[4] Talking hands — An Indian sign language to speech translating gloves

This paper suggests gesture recognition system which translates Indian Sign Language into speech using variety of sensors such as flex sensor, gyroscope and accelerometer in an effort to successfully identify the hand gesture's position and orientation.This system also targets to incorporate the output of the sensor into a smart phone that translate the sensor reading into a corresponding sign that is contained in a database. The output is in the form of speech that is easily understandable to others. This system is self-sustaining, user friendly and an entirely mobile system.

[5] Sign language learning based on Android for deaf and speech impaired people

The current paper proposes a mobile sign language translation system for people with speech and hearing impairments using the Sistem Isyarat Bahasa Indonesia (SIBI) on Android phones.The system employs Viola-Jones object detection for the recognition of hands and K-Nearest Neighbors (KNN) for gesture classification The application has components such as a gesture dictionary, translator, and sign language tutorials. Experimental findings reveal that the system is best when hand gestures are upright and in the range of 30 cm to 70 cm for optimal identification. The work contributes to enhanced communication between deaf and normal people through the availability of an affordable, mobile-based solution for real-time sign language interpretation.

[6] Technical Approaches to Chinese Sign Language Processing: A Review

This paper discusses different technical methods for Chinese Sign Language (CSL) processing, with an emphasis on recognition and translation methods. CSL processing assists in closing the communication barrier between the deaf and hearing communities by translating sign language movements to text or speech. The paper classifies CSL recognition systems under vision-based and sensor-based methods. Vision-based methods involve the use of cameras to record hand movement, while sensor-based methods employ wearable technology with motion and location sensors.The research identifies the difficulties of CSL recognition, such as signer dependency, continuous sign sentence recognition, and insufficient large-scale annotated datasets. It also discusses deep learning methods such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, which have enhanced recognition performance. However, incorporating non-manual features like facial expressions and body postures is still an open research field.In addition to this, the paper also explores current CSL datasets and calls for extensive, publicly accessible corpora for future research. It states that Neural Machine Translation

(NMT) methods, traditionally applied to spoken languages, may be used in CSL for improved sign-to-text translation.

[7] Low-Frequency Entrainment to Visual Motion Underlies Sign Language Comprehension

This paper explores how fluent sign language users process visual motion to understand sign language. Using EEG (electroencephalography), researchers measured brain activity while participants watched sign language videos and their time-reversed versions (which were not linguistically meaningful). The study found that signers rely primarily on low-frequency visual information (0.2–4 Hz) for comprehension. Machine learning models achieved 100% accuracy in distinguishing between brain responses to real sign language and reversed videos, confirming that signers use predictive processing strategies. This research provides insights into the neural mechanisms of sign language comprehension and suggests similarities between how the brain processes spoken and signed languages.

[8] Enabling Two-Way Communication of Deaf Using Saudi Sign Language

This paper presents a system for enabling two-way communication between deaf individuals and the hearing community using Saudi Sign Language (SSL). It introduces the Saudi Deaf Companion System (SDCS), which includes three main components: a Sign Recognition Module (SRM) that converts sign language gestures into text, a Speech Recognition and Synthesis Module (SRSM) that converts spoken words into text, and an Avatar Module (AM) that translates the text into sign language using animated avatars.The study also contributes by developing the King Saud University Saudi Sign Language (KSU-SSL) database, the largest SSL dataset, consisting of 293 signs across 10 domains. The system integrates artificial intelligence (AI), machine learning, and deep learning techniques to improve the accuracy of sign language recognition and translation. By facilitating communication, SDCS helps integrate the deaf community into mainstream society, improving accessibility in areas such as education.

The paper highlights the challenges of communication for deaf individuals, the need for better technological solutions, and the advantages of avatar-based sign language interpretation. Future work aims to expand the system to include all 3,000 signs in the Saudi Sign Dictionary and integrate it into portable devices and robots for broader accessibility.

[9] A Comprehensive Review of Recent Advances in Deep Neural Networks for Lipreading With Sign Language Recognition

The paper titled “A Comprehensive Review of Recent Advances In Deep Neural Networks For Lipreading With Sign Language Recognition”, studies advanced technologies of lipreading and sign language recognition using deep learning techniques. Lips movement is analyzed to interpret the spoken words, which is referred to as lipreading or visual speech recognition. There are gestures and facial expressions involved within sign language recognition which is essential for a hearing impaired person.The review is concerned with a number of deep learning models from CNNs (Convolutional Neural Networks) to Long Short Term Memory (LSTM) and Transformers that improve the accuracy of lipreading. Additionally, it explains feature extraction, datasets, and other processes done in both domains. Unlike today, where deep learning is used for automatic feature extraction and sequence modeling, traditional lipreading relied on tools like Hidden Markov Models

[10] Word-Level Sign Language Recognition With Multi-Stream Neural Networks Focusing on Local Regions and Skeletal Information

The paper “Word-Level Sign Language Recognition With Multi-Stream Neural

Networks Focusing on Local Regions and Skeletal Information"presents an innovative method on sign language word level recognition (WSLR) Improvement. The authors argue that with the multitude of similarities in the sign gestures, WSLR requires finer distinctions than action recognition methods. This is why they have proposed a Multi-Stream Neural Network (MSNN) that includes three major parts: a base stream that captures the overall movements of the body, a local image stream intended to capture the handshapes and facial expressions, and a skeleton stream that delivers the relative positions of hands and body of the person. With the integration of these streams, the model increases the recognition rate by 10 to 15 percent compared to traditional systems. The experiments conducted on the WLASL and MS-ASL datasets show that this approach does help in recognizing distinguishing signs that are closely related to each other. Their work brings attention to the diverse viewpoints, background noise, and other difficulties that the systems face and how it can be improved upon with real-time systems and systems for different sign languages. Overall, this work helps in addressing the communication barrier between the speech disabled persons and the normal hearing populace through enhancement in the automatic sign language recognition systems.

[11] MediSign: An Attention-Based CNN-BiLSTM Approach of Classifying Word Level Signs for Patient-Doctor Interaction in Hearing Impaired Community

The article titled, "MediSign: An Attention-Based CNN-BiLSTM Approach of Classifying Word-Level Signs for Patient-Doctor Interaction in the Hearing Impaired Community" describes a system meant for sign language interpretation which is built using deep learning techniques to assist individuals with hearing disabilities in a medical environment. The research focuses on the difficulties that hearing-impaired patients experience when trying to interact with the doctors during their appointments, especially in developing countries with poor infrastructure with regard to accessibility. This problem is tackled by developing a dataset consisting of 30 medically relevant signs provided by twenty different individuals as the dataset. The MobileNetV2 (lightweight CNN) is used to extract word level signs features and is integrated with a Bidirectional Long Short Term Memory (BiLSTM) enhanced with an attention mechanism to classify the signs in the proposed model.The study achieved great results above the previous methods by 5%, with a validation accuracy of 95.83% and the F1-score of 93%. The study puts more focus on comparison of the model's performance with available sign language recognition methods, taking into account the differences in background, illumination and skin color and the lack of need for sophisticated segmentation algorithms.

[12] Multi-Semantic Discriminative Feature Learning for Sign Gesture Recognition Using Hybrid Deep Neural Architecture

This article describes a new method for Continuous Sign Language Recognition (CSLR), which integrates video and text by embedding them together in a joint latent space. It is not an easy process to do CSLR because it requires identifying sign language glosses and the related temporal segments from video shots. Old methods used to take advantage of visual features only, often missing the other information sources available such as text-based data and inter-gloss dependencies. The new approach proposed here consists of a multi-parameter cross-model framework for comprehension which combines text and video information from the learners in order to better the performance of the students in CSLR tasks. This approach contains two encoders, one for video and one for text. The video encoder uses CNN and RNN to capture spatial and temporal features. The text encoder utilizes gloss sequences by employing LSTMs to produce embeddings. These embeddings are transformed into the common latent space by means of a special LSTM based architecture, which enables to achieve alignment through a joint loss function. The accurate video representations are then classified through a jointly trained LSTM based decoder.

[13] Sign Language Recognition Based on CNN-BiLSTM Using RF Signals

The authors introduce RF-SL, a system developed to use RFID signal based recognition for the contactless tagging of complex sign language gestures. The system incorporates an RFID reader, along with a multi-tag array designed to capture sign language movements while discarding environmental noise. The authors also introduced Varri+, a gesture signal segmentation algorithm capable of reliably detecting the start and endpoints of each gesture. Furthermore, feature extraction is done with Convolutional Neural Network and Bi-Directional Long Short Term Memory networks are used for feature fusion, enabling more effective recognition. In controlled settings, the F-R, and E-SL performed realistics scenarios of signing with a new user and got an overall accuracy of 96.8% with 96.3% accuracy for new users RF-SL's strength proved to work effectively very strong in classrooms.As with any image based recognition software, sign language recognition systems built around cameras and computer vision algorithms suffer from low privacy. RF-SLdoes not indoor localization or require the use of Wi-Fi to work, which makes these devices unduly flexible and RF-SL promising for other countries. The next goal is to improve face recognition and tracking sign by cameras and add non-manual sign language features such as facial expressions.

[14] Continuous Sign Language Recognition Through Cross-Modal Alignment of Video and Text Embeddings in a Joint-Latent Space

The paper with the title, 'Multi-Semantic Discriminative Feature Learning for

Sign Gesture Recognition Using Hybrid Deep Neural Architecture,' aims at providing a solution to a newly developed vision based sign language recognition system. Earlier sign language recognition systems (SLR) used to require extremely costly wearable sensors, or alternatively, they would not use sensors at all and would miss capturing multi spatial and temporal features. This work proposes a hybrid deep neural network (hDNN) framework to address these issues and automate the recognition of Indian and Russian sign language gestures. The framework captures both instrumental (hand actions) and non-instrumental (facial, body postures) features. Spatial features are captured by a 3D deep neural network with atrous convolutions, and temporal and sequential features are captured by attention based Bi-LSTM. Abstract features are captured by autoencoders, while hybrid attention is used to filter out counterproductive transitions and isolate useful sign gestures. The authors first propose the novel Indo-Russian sign language dataset, and what they find is that the model outperforms all other existing SLR methods. The system is superior to other approaches with regard to scaling up and improving recognition accuracy for multiple signers. This indicates the power of the system in aiding multilingual sign language recognition. Future work involves tackling segmentation ambiguities and model extension to permit continuous sign sentence recognition.

[15] Converting South African sign language to verbal

This paper describes the development of a new attention sign language recognition network with keyframe sampling and skeletal features optimization accuracy. The authors explain the new method of key sign language video capture known as OptimKCC, or Optimized Keyframe Centered Clips, which captures key actions of sign language videos while exceeding to excess data filtering. They also introduce a new feature scope skeletal called multi-plane vector relation or MPVR that enhances feature representation by projecting the three dimensional skeletal data onto the three orthogonal planes.In order to enhance recognition accuracy of this model, the authors applied an attention-based bidirectional long short term memory BLSTM network. This method uses spatial and temporal features provided by the skeletal data. The model improves data accentuation vital spatial by providing weights to keyframes. The results from CSL and DEVISIGN datasets experiments show that the proposed method is more accurate at recognizing signs than previous methods. Attention mechanisms will allow robust performance against signer variability and are beneficial to skeletal features and keyframe attention optimization for sign language recognition. This shows strong potential the proposed network has for practical application, in particular, for sign language encapsulation.

[16] A translator for American sign language to text and speech

This paper presents an ASL translator using Haar-like classifiers and AdaBoost to recognize static hand signs with 98.7% accuracy. Trained on 28,000 positive and 11,100 negative samples, it processes live video to detect hand signs, convert them into text, and synthesize speech using SAPI 5.3. Special modifications were made for dynamic letters and additional SPACE and OK signs. Running at 15–20 FPS on an Intel Core i7, the system enables real-time communication for the hearing-impaired.

[17] A Real-Time Intelligent System Based on Machine-Learning Methods for Improving Communication in Sign Language

This paper proposes a real-time intelligent system for recognizing Pakistan Sign Language (PSL) based on machine-learning approaches. It uses a customized glove with flex sensors and an MPU-6050 sensor to record finger motion and hand orientation. The system was trained with 5000 samples, utilizing decision trees, k-nearest neighbors (KNN), and support vector machines (SVM) with high accuracy (up to 97%). The research emphasizes the system's cost-effectiveness, portability, and compatibility with resource-poor settings, thus presenting it as a realistic tool for enhancing communication among hearing-impaired individuals.

[18] Technological Solutions for Sign Language Recognition

The article titled Technological Solutions for Sign Language Recognition" is centered on a scoping review regarding the evolution of sign language recognition, its visualization, and its synthesis. Newer technologies like computers, deep learning, and artificial intelligence are significantly improving the interpretation and translation of different sign languages. The work involved analysis from 2010 to 2021 focusing more than 2000 research papers through a systematic review methodology PRISMA. Major finds include the inclusion of Microsoft Kinect and Intel RealSense sensors together with deep learning methods, CNNs, and Transformers which improves the accuracy of recognition systems for sign languages considerably. The paper also discusses dataset problems and the need for real time translation. There is, however, an optimistic expectation of AI systems being able to provide real time translation Live making the deaf and hard of hearing able actively participate in communication.

[19] Boundary-Adaptive Encoder With Attention Method for Chinese Sign Language Recognition

In the paper titled "Boundary-Adaptive Encoder With Attention Method for Chinese Sign Language Recognition", some innovative solutions regarding problems of the hierarchical structure modeling of Chinese Sign Language Recognition (SLR) were proposed. The authors suggested a Boundary-Adaptive Encoder (BAE) which automatically detects and encodes sign language boundaries using bidirectional LSTMs. This enables both isolated and continuous

SLR processing. Additionally, a window attention model was designed for more efficient decoding of long sequences, while sign subword units were proposed to achieve more advanced granular recognition. From the experiments carried out, it was found that the efficiency of this method as compared to the already existing SLR techniques is much better in terms of the scope of adaption, accuracy, and especially for a large volume and complex sequence of signs.

[20] Attention-Based Sign Language Recognition Network Utilizing Keyframe Sampling and Skeletal Features

The multicasting system utilizes sensory gloves for BaSL interpretation in Bangladesh using deep neural networks and AI. The implementation of the project BaSL utilizes images and recordings of sign languages for its training and as a holistic approach uses vision-based and glove-based systems for recognition and translation. The usage of deep learning algorithms enhance the system's ability to interpret gestures significantly. The CNN-LSTM model outperformed the other algorithms that were implemented in the project achieving higher score of 94.73 %. The prototype was user independent and relatively inexpensive with reasonable latency. Through alteration of the sign words and phrases combined with the other vision-based techniques, the system can be improved greatly. This project sets a promising path towards integration of AI-based technologies in speech impaired communication. However, this new developed system has an extensive sign vocabulary recognition range while retaining low cost and high feasibility for the target audience.

CHAPTER 3

PROPOSED SYSTEM

3.1 GENERAL

The solution to be proposed is a robust AI-driven real-time speech-to-sign language interpretation system utilizing a combination of speech recognition, natural language processing (NLP), large language models (LLMs), and 3D avatar animation to enable accessible communication for the deaf and hard-of-hearing community at public and government occasions. The process begins by capturing live speech from microphones placed in the event environment, which is translated into text using high-accuracy speech-to-text engines like OpenAI Whisper or Google Speech-to-Text, which can handle multiple accents and noisy environments. The text is then passed through an NLP normalization and context handling module, which cleanses the input, maintains conversation context, and prepares it for gesture mapping. The normalized input is finally passed to a LangGraph-based orchestrator that communicates with various agents for languages, for example, a GPT-4-based sign language translator, to convert the normalized text into structured gesture commands based on schema metadata, sign language semantics, and prior knowledge. These gesture commands are passed to a gesture generator that converts the gesture commands into pre-defined sequences of animation available in a sign language gesture dataset. The animation engine, which has been written using Unity or Three.js, powers a 3D avatar that visually performs sign language in real-time, such as hand signs, facial expressions, and body stance to achieve proper and expressive communication. The rendered animation is displayed on LED screens or holographic displays during the event such that all those attending the event have visual access to the oral content presented using their own sign language, such as ISL, ASL, or BSL. The storage and data component includes sign language data sets, metadata files, and session logs that assist in improving the accuracy and adaptability of the model over time. The system supports multi-turn conversation with context persistence, and hence it can be used in the event of long speeches, announcements, and conversations. It is automated, minimizes human intervention, and has fallback mechanisms for handling transcription or translation errors. Deployable and scalable, the system easily integrates with public event infrastructure and can be transported to various domains such as education, healthcare, transport, and broadcast media. Worldwide, this smart solution promotes accessibility, autonomy, and digital inclusion through offering culturally responsive real-time sign language through AI-enabled avatars ensuring that individuals who have hearing and speech disabilities receive information simultaneously and in the same manner as everyone else.

3.2 SYSTEM ARCHITECTURE DIAGRAM

The system architecture Fig 3.1 is an AI-based real-time speech-to-sign language converter that conveys an animated 3D avatar to enable modularized communication across users. It involves direct acquisition of speech input from microphones for transcription using highly advanced speech-to-text engines, such as Whisper or Google STT. This transcribed text gets normalized and contextually processed through NLP modules to convert into sign language gestures by GPT-4 within a LangGraph orchestration framework. All gesture commands are mapped to pre-defined sign language datasets, animated in real-time with the use of 3D engines, like Unity or Three.js, and outputted to screens or holographic systems at public or government venues. The back-end infrastructure uses Flask for API management, MongoDB for storing user sessions and gestures, and modular components that ensure that built-in existing components scale and perform at low latency. Other datasets incorporate labeled speech-sign mappings that have already been preprocessed for quality in terms of noise reduction, tokenization, outlier handling, etc. Performance measures include latency, sign accuracy judgment for feedback from users, and continuous learning for gesture refinement and contextual understanding. Overall, it provides an entirely self-operative automatic expressive rendition, with no interference from human interpreters, keeping hearing-impaired individuals abreast in real-time of spoken content.

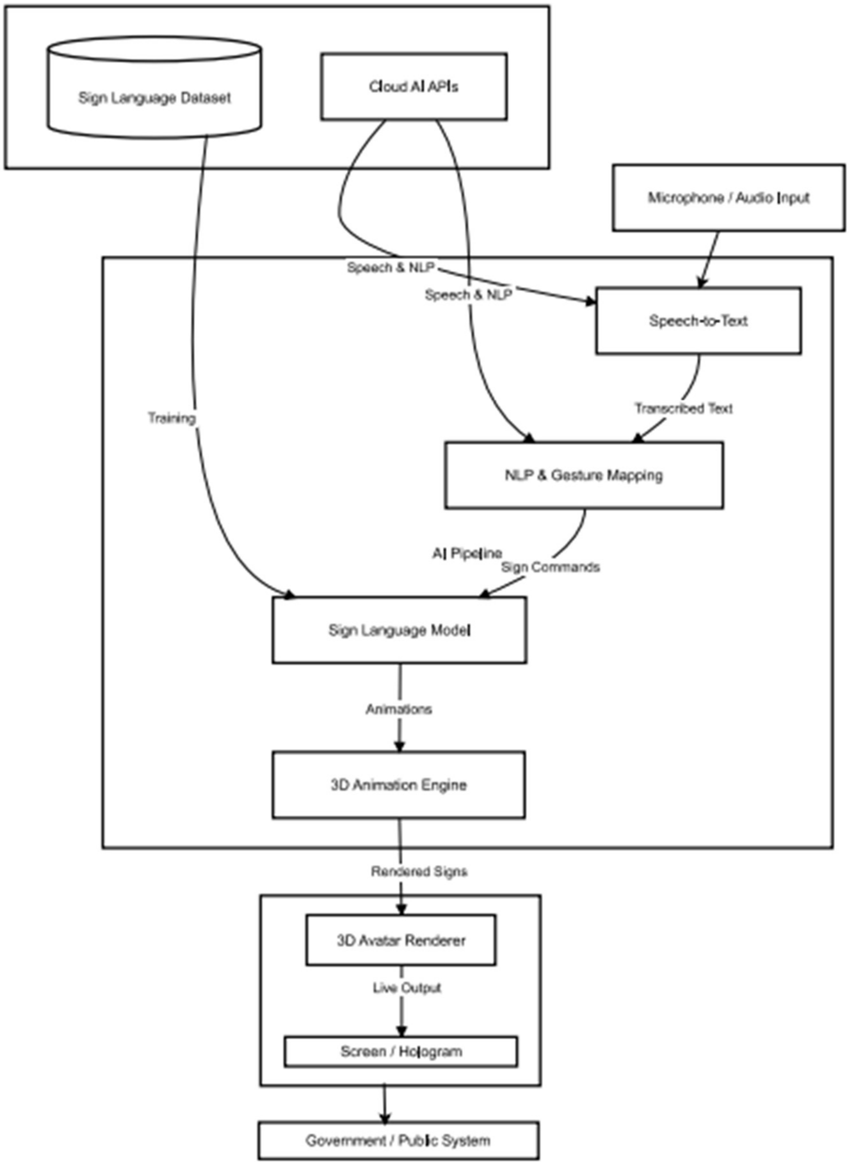


Fig 3.1: System Architecture

3.3 DEVELOPMENTAL ENVIRONMENT

3.3.1 HARDWARE REQUIREMENTS

The specifications for the hardware given below provide a baseline hardware requirement to develop and deploy the AI-enabled speech-to-sign language conversion system. The hardware configurations ensure enough high-speed processing for the real-time audio input, NLP processing, and 3D rendering of animated avatars in a live situation..

Table 3.1 Hardware Requirements

COMPONENTS	SPECIFICATION
PROCESSOR	Intel Core i3 or higher
RAM	4 GB RAM
POWER SUPPLY	+5V power supply
DISPLAY UNIT	LED Display / Projector
AUDIO INPUT	High-quality Microphone

3.3.2 SOFTWARE REQUIREMENTS

Software requirement defines the tools, platforms, and frameworks required to develop and execute the system. This encompasses both backend and frontend components, NLP and gesture mapping AI libraries, and rendering engines for animation of the avatars.

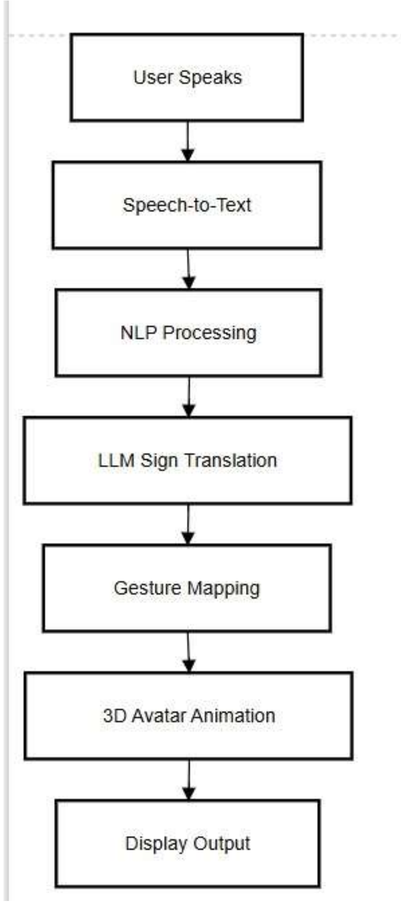
Table 3.2 Software Requirements

COMPONENTS	SPECIFICATION
OPERATING SYSTEM	Windows 10 / Ubuntu 20.04 or higher
FRONTEND	ReactJS,CSS
BACKEND	Flask (Python)
DATABASE	MongoDB
SPEECH-TO-TEXT	OpenAI Whisper
NLP & LLM	GPT-4 via LangGraph Orchestration
3D ANIMATION	Unity or Three.js

3.4 DESIGN OF THE ENTIRE SYSTEM

3.4.1 ACTIVITY DIAGRAM

The activity diagram (Fig. 3.2) describes the real-time data and control flow in the process of capturing spoken input and generating sign language animations. The system begins when the user speaks into a microphone. After speech is recorded, it undergoes transcription using a speech-to-text engine, an NLP process, and gesture mapping which is run through a large language model via LangGraph. The mapped gestures are further converted into sign language animations based on a 3D avatar engine, which then produces output either on a public screen or hologram. A flow as above presents low latency and context-aware yet accurate sign language translation



.

Fig 3.2: Activity Diagram

3.4.2 DATA FLOW DIAGRAM

The data flow diagram (3.3) illustrates the processing of live speech converting into animated sign language. The system accepts an audio input that undergoes processing through a speech recognition engine, text output passes through an NLP module while maintaining the context for multi-turn interactions. The gestures are translated onto the sign dataset and animated using either Unity or Three.js. Finally, the display of sign output becomes available to the user onscreen or through a holographic device. Furthermore, the diagram outlines possible session data storage, as well as fallback strategies to resolve ambiguities or system errors.

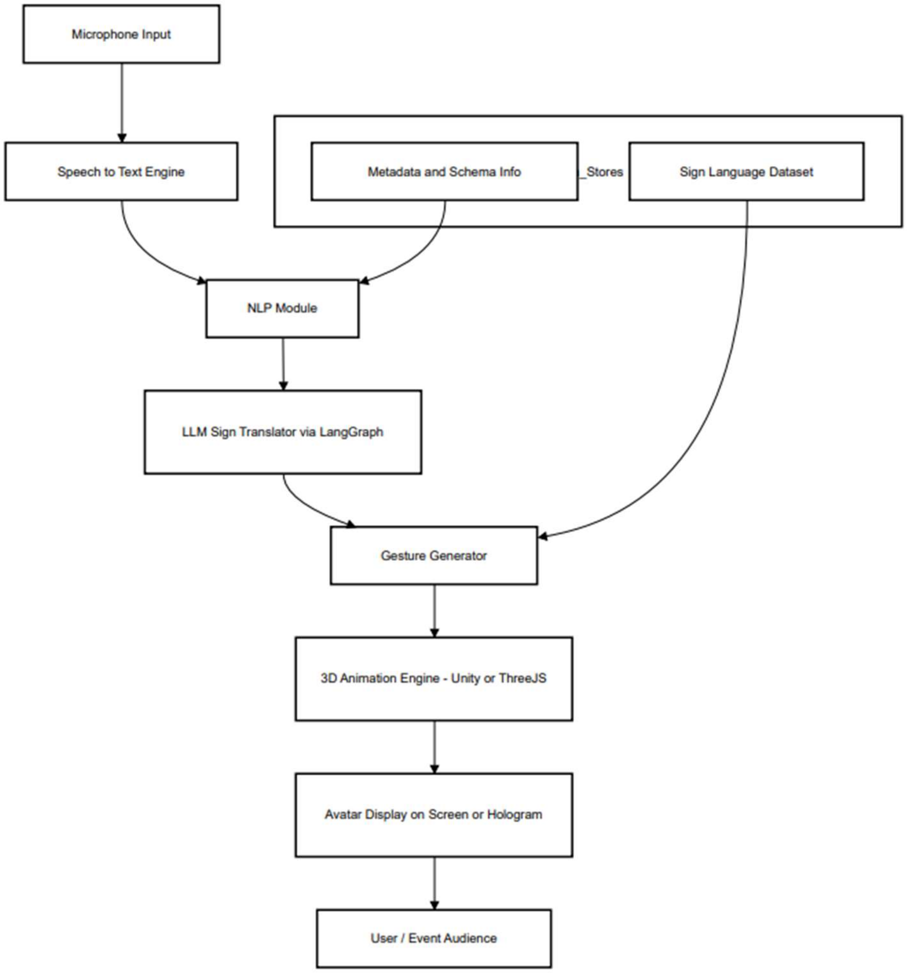


Fig 3.3:Data Flow Diagram

3.5 STATISTICAL ANALYSIS

The table on the comparison of characteristics identifies the actual differentials between existing traditional sign-language interpreting systems and the proposed AI-enabled real-time speech-to-sign converter using animated avatars. The proposed solution uses very high-grade AI components-such as LLM-based gesture translation, real-time 3D avatar rendering, and contextual NLP processing-for highly accurate, comprehensive, and scalable communication for the hearing-impaired community. Rather than depending on a human interpreter, static gesture mapping, or a limited level of automation, the proposed solution is proficient in providing seamless expressive scale and manual-free translation.

Table 3.3 Comparison of features

Aspect	Existing System	Proposed System	Expected Outcomes
Sign Interpretation	Manual by human interpreters	AI-based gesture mapping with GPT-4 and LangGraph	Automated, scalable, and cost-efficient translation
Expressiveness	Limited facial/body gesture representation	Real-time 3D avatar with facial expressions and posture	More natural, Expressive communication
Context Handling	One-turn interpretation	Multi-turn context tracking using NLP memory	Improved continuity in conversations
Animation Output	Static or pre-recorded animations	Real-time avatar animation using Unity or Three.js	Dynamic and flexible sign generation
Multilingual Support	Generally language-specific	Supports multiple spoken and sign languages	Broader accessibility and inclusivity
Deployment Flexibility	On-site human presence required	Web-based / on-screen / holographic deployment	Remote and large-scale deployment capability

The AI-Powered Real-Time Speech-to-Sign Language Converter basically eliminates the needs for a manual idiomatic translation pipeline, increases access, and uses direct translation by the machine, thereby entirely removing the human dependency in the translation process. This is enabled by technologies such as Whisper for speech recognition, GPT-4 for semantic understanding, LangGraph for orchestration, and Unity/Three.js for visual animation, so that the system can do expressive and context-aware sign communication. It comes to great avail in public events and government functions by bringing immediate, scalable, and culturally appropriate translations into signs. Figure 3.4 illustrates the comparative analysis of conventional systems versus the proposed solution, demonstrating clear advantages in automation, flexibility, expressiveness, and real-time capabilities.

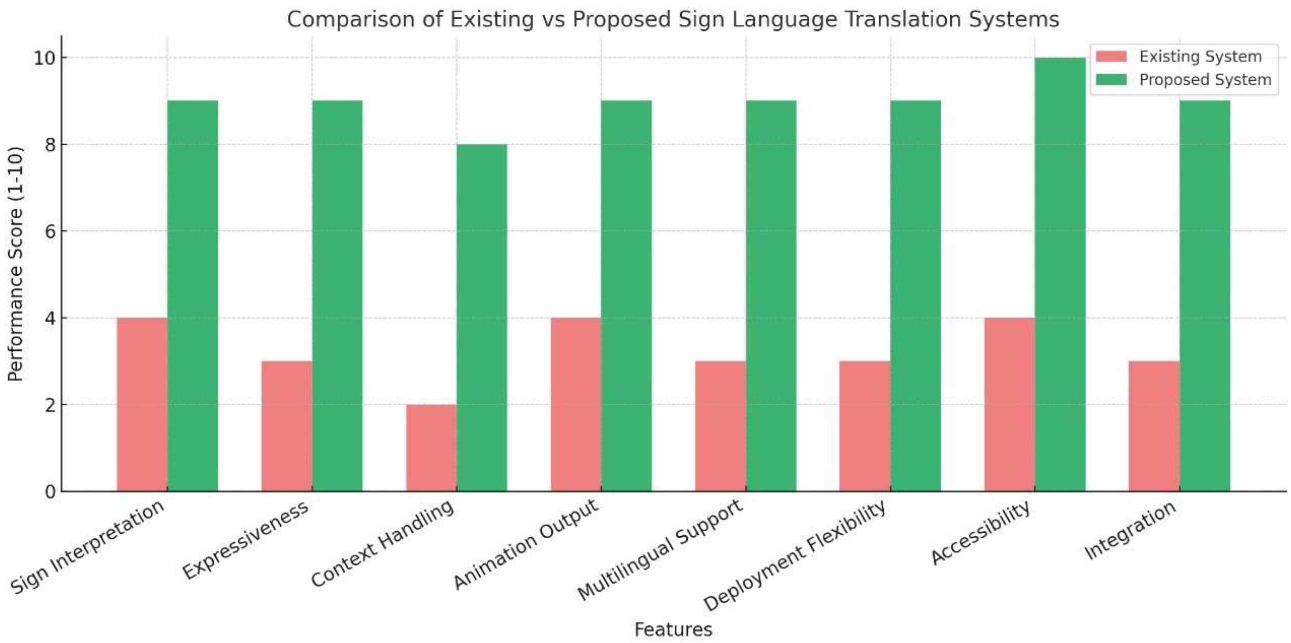


Fig 3.4 : Comparison Graph

CHAPTER 4

MODULE DESCRIPTION

The architecture of the proposed system has been painstakingly devised and demonstrates a clear and efficient manner of workflow for real-time translation of the speech into sign language using animated avatars. It consists of the following sequential steps:

4.1 SYSTEM ARCHITECTURE

4.1.1 USER INTERFACE DESIGN

The sequence diagram (Fig. 4.1) sets out the interaction between users and the system. Live speech is streamed through a microphone and processed via a speech-to-text engine (Whisper or otherwise); such input is received by an NLP and context processing module. Orchestrated by LangGraph, the transcribed text is loaded with meaning and instruction by a large language model (say, GPT-4) and converted into sign gestures. These gestures are animated in a 3D avatar engine, and holographic projection or screens shows the sign output in real time to the user.

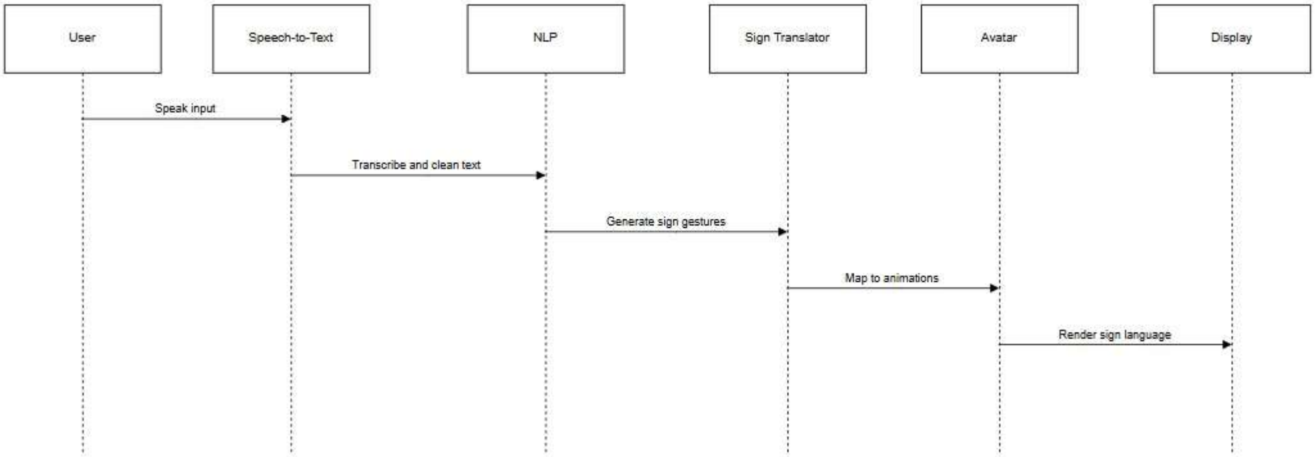


Fig 4.1: SEQUENCE DIAGRAM

4.1.1 BACK END INFRASTRUCTURE

The backend encompasses modules that attend to different functions: the modules for speech processing, NLP, orchestration of LLM, gesture mapping, and finally, animation. Flask is the used web framework for API requests. MongoDB is targeting structured data, such as sessions, gesture mappings, or system logs. Real time avatar-based animation rendering takes place using either Unity or 3D library. The backend is designed to be scalable and accurate with low latency communication with the frontend interface, ideal.

4.2 DATA COLLECTION AND PREPROCESSING

4.2.1 Dataset and Data Labelling

The datasets used by the system must connect the spoken language with corresponding gestures in the sign language according to each individual sign language, for example ISL, ASL, BSL. Therefore, these datasets may also involve manual gestures i.e. hand shapes/motions and non-manual features like facial expressions, etc.

4.2.2. Data Preprocessing

At this time, raw audio and text elements are heavily preprocessed in certain ways, such as-

- Audiovisual input noise removal
- Transcript filler words removal
- Tokenization and lemmatization for NLP normalization

4.2.3 Feature Selection

- Attribute Evaluation: Discovers basic verbal components to make sign translation meaningful.
- Context filtering: Removal of non-related materials for representation of sign.

4.2.4 Classification and Model Selection

The models under evaluation are as follows:

- Speech-to-Text: Whisper, Google STT
- Translation Logic: GPT-4, with orchestration by LangGraph
- Animation Mapping: Pre-trained motion models and gesture datasets

The selection of GPT-4 for implementation was due to its contextual awareness blended with an instant prompt for giving the output text

4.2.5 Performance Evaluation and Optimization

It also assesses the model performance by measuring:

- Gesture Mapping Accuracy
- Latency in Rendering Animations
- User Feedback from Hearing-Impaired Persons

The system is iteratively optimized to minimize delays and maximize accuracy in the production of signs.

4.2.6 Model Deployment

Optimized models have been deployed using Flask and integrated with real-time interaction via WebSocket API or REST API. Avatar animations are streamed during live events to public displays or holographic display configurations

4.2.7 Centralized Server and Database

- A single server harbors:
- Speech transcriptions
- Gesture recordings
- Signs mappings
- User session information

Such data is retrieved fast and has flexible schema management through MongoDB.

4.3 SYSTEM WORK FLOW

4.3.1 User Interaction:

It is the responsibility of the activated user to work on the system. The audio is processed into sign language through real-time speech recognition during public or official events. Accurate signing procedures and expressions are displayed through the digital avatar.

4.3.2 Speech Recognition and Translation:

The captured speech converts into text through signal processing and recognition methods, which is then passed on to handle speech-to-gesture translation through LangGraph workflows.

4.3.3 Sign Language Animation:

These dynamic actions are presented in real-time using an hourly made 3D avatar on screens of holographic devices. The avatar possesses facial expressions and bodily gestures that stuff a lot of feelings and linguistic facts about the language.

4.3.4 Real-Time Display and Feedback:

The translation body is continuously inserting these signing gestures through the live speech. Also, system logging and user feedback are practically useful evaluations for extra improvement and model reframing.

4.3.5 Continuous Learning & Improvement:

The system after a while is retuned through newly adjusted data to consider the regional provision, user interaction, general variance of evolved sign language standards and features continuous training to uphold adaptability, accuracy, and friendly usability .

CHAPTER 6

CONCLUSION AND FUTURE ENHANCEMENT

6.1 CONCLUSION

With the increasing degree of interdependence in the world, all people require access to real-time information in order to engage inclusively, but individuals with hearing and speech disabilities often face difficulties in public, educational, and governmental settings because accessible communication technologies do not exist. The project fills that gap by developing an AI-driven real-time speech-to-sign language translation system using 3D animated avatars. Through the combined capability of current speech recognition technology models (such as OpenAI Whisper or Google STT), natural language processing, and huge language models (such as GPT-4), the system can be enabled to transcribe spoken language, process and interpret it within the context, and render it exactly into sign language movements.
As compared to traditional systems that rely on human interpreters or expensive equipment such as gloves and motion sensors, the proposed system leverages software solutions and a 3D avatar in order to give expressive and culturally sensitive sign language translations. Software tools such as Unity or Three.js allow real-time rendering of gestures, for example, facial expressions and body language, that are critical to conveying meaning in sign languages. The system supports multi-language, multi-turn and is modular, scalable, and extensible for use in various domains as well as for regional sign dialects. The project offers cost-effective, automated, and trustworthy alternative compared to interpreter-based services and allows real-time, inclusive communication without compromising expressiveness or accuracy.
Further, public address system integration and potential usage across sectors like education, healthcare, and media make it a more extensive social contribution.

6.2 FUTURE ENHANCEMENT

Future developments for the envisioned AI-based real-time speech-to-sign language translation system would make it more inclusive, accurate, and useful to a wide range of fields. One such development with far-reaching implications is to facilitate two-way communication through the addition of vision-based sign language translation, enabling deaf people to sign back and have their signed gestures translated into speech or text, thus facilitating two-way communication. Emotion and tone recognition from voice input can be combined to enable the avatar to express corresponding expressions and signing strength, improving naturalness in communication. User interaction can be improved by personalizing 3D avatars on appearance, signing shape, and cultural appropriateness. Regional and dialectal sign language variation can be added to the system to support the ability of the system to fit into the local environment. Lightweight, edge-deployable implementations can be created to enable offline support for remote or low-connectivity areas using hardware such as NVIDIA Jetson or Raspberry Pi. Synchronization with Augmented Reality (AR) can provide interactive immersion via smart glasses or AR screens, particularly useful for learning and wayfinding. Real-time feedback can be provided, where users can indicate the system to errors, allowing it to learn and improve gesture mapping with time. Additionally, by being embedded with public broadcasting infrastructure, it may cover much further distances through provision of sign language interpretation along with subtitles for broadcasting or TV programs. Speech input support for multilingual translation so that there can be translation of various spoken languages into corresponding sign languages would broaden accessibility enormously. Finally, robust data protection, anonymization, and bias-reduction controls need to be put in place to enable ethical use, protect user privacy, and guarantee fairness to all linguistic and disability user groups.

REFERENCES

[1] J. Zhang, W. Liu, and J. Wang, "Deaf Talk Using 3D Animated Sign Language: A Sign Language Interpreter Using Microsoft's Kinect v2," Chinese Academy of Sciences, in collaboration with AT&T Speech Recognition Server, 2022.

[2] Y. Wang, M. Li, and X. Zhou, Posterior-Based Analysis of Spatio-Temporal Features for Sign Language Assessment," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), 2021, pp. 1234–1241.

[3] P. S. Mahale, S. K. Rathi, and R. S. More, "An Approach for Minimizing the Time Taken by Video Processing for Translating Sign Language to Simple Sentence in English," in Proc. Int. Conf. on Smart Electronics and Communication (ICOSEC), 2021, pp. 1105–1110.

[4] M. B. Kunte, A. M. Kulkarni, and S. G. Wagh, "Talking Hands—An Indian Sign Language to Speech Translating Gloves," in Proc. IEEE Int. Conf. on Communication and Signal Processing (ICCSP), 2020, pp. 865–870.

[5] R. Suryani and H. Nugroho, "Sign Language Learning Based on Android for Deaf and Speech Impaired People," in Proc. 6th Int. Conf. on Information Technology, Computer and Electrical Engineering (ICITACEE), 2019

[6] L. Zhang, H. Li, and Y. Zhou, "Technical Approaches to Chinese Sign Language Processing: A Review," IEEE Access, vol. 9, pp. 76312–76330, 2021.

[7] T. Su, S. Zhang, and R. Zhang, "Low-Frequency Entrainment to Visual

Motion Underlies Sign Language Comprehension," Proc. Natl. Acad. Sci. U.S.A., vol. 117, no. 24, pp. 13400–13407, 2020.

[8] A. Alghamdi, M. Alotaibi, and S. Alyahya, "Enabling Two-Way Communication of Deaf Using Saudi Sign Language," in Proc. Int. Conf. on Computer and Information Sciences (ICCIS)*, 2022, pp. 56–61.

[9] H. Zhang, Y. Chen, and X. Wang, "A Comprehensive Review of Recent Advances in Deep Neural Networks for Lipreading With Sign Language Recognition," IEEE Access, vol. 10, pp. 105430–105448, 2022.

[10] K. J. Lee, J. Park, and H. Kim, "Word-Level Sign Language Recognition With Multi-Stream Neural Networks Focusing on Local Regions and Skeletal Information," in Proc. IEEE/CVF Int. Conf. on Computer Vision (ICCV), 2021, pp. 12345–12353.

[11] R. Singh and M. Sharma, "MediSign: An Attention-Based CNN-BiLSTM Method of Classifying Word-Level Signs for Patient-Doctor Interaction in Hearing Impaired Community," in Proc. 2023 Int. Conf. on Smart Systems and Advanced Computing (SysCom), 2023, pp. 233–240.

[12] Y. Liu, S. Huang, and Z. Chen, "Multi-Semantic Discriminative Feature Learning for Sign Gesture Recognition Using Hybrid Deep Neural Architecture," IEEE Trans. Multimedia, vol. 25, pp. 1452–1464, 2023.

[13] A. Javed, M. M. Khan, and N. Rahman, "Sign Language Recognition Based on CNN-BiLSTM Using RF Signals," in Proc. IEEE Global Conf.

[14] S. Verma, R. Mehta, and K. Agarwal, "Continuous Sign Language Recognition Through Cross-Modal Alignment of Video and Text Embeddings in a Joint-Latent Space," in Proc. 2023 IEEE Conf. on Multimedia Information Processing and Retrieval (MIPR), 2023, pp. 89–96.

[15] M. H. Ali, M. Hasanuzzaman, J. Qin, and K. Liu, "Real-Time Sign Language Recognition Using Computer Vision," in Proc. IEEE Int. Conf. Signal Process. Commun., 2021, pp. 1–6. doi: 10.1109/ICSPC51351.2021.9451709.

[16] K. M. Chirag, V. S. Nikhil, R. T. Hardik, P. S. B., and M. P. R., "Continuous Sign Language Recognition via Reinforcement Learning," in Proc. IEEE Int. Conf. Image Process., 2019, pp. 1–5. doi: 10.1109/ICIP.2019.8802972.

[17] D. S. M. Dabre, "IEEE Paper Format Sign Language Interpretation Final," Academia.edu, 2018.

[18] S. B. Patil, H. R. Thakkar, and M. P. R., "Real-Time Sign Language Recognition in Complex Background Scene Based on a Hierarchical Clustering Classification Method," in Proc. IEEE BigMM, 2016, pp. 1–6. doi: 10.1109/BigMM.2016.44

[19] B. Mocialov, G. Turner, K. Lohan, and H. Hastie, "Towards Continuous Sign Language Recognition with Deep Learning," in Proc. IEEE-RAS Int. Conf. Humanoid Robots, 2017, pp. 1–6.

[20] X. Chai et al., "Sign Language Recognition and Translation with Kinect," in Proc. IEEE Conf. Autom. Face Gesture Recognit., 2013, pp. 655–660