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

Dr. SenthilPandi[1], Sneha Sajeevan[2], Snekha R [3], Swedha J[4]

[1]Assistant Professor - Department of Computer Science and Engineering, Rajalakshmi Engineering College, Chennai.

[2,3,4]Under Graduate Student - Department of Computer Science and Engineering, Rajalakshmi Engineering College, Chennai.

[1]senthilpandi.s@rajalakshmi.edu.in, [2]220701281@rajalakshmi.edu.in, [3]220701282@rajalakshmi.edu.in [4]220701296@rajalakshmi.edu.in

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.**

Keywords— Speech-to-Sign Language, Real-Time Translation, 3D Animated Avatars, Accessibility, NLP, GPT-4, Whisper, LangGraph, Sign Language Recognition, Assistive Technology

I INTRODUCTION

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.

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.

II LITERATURE SURVEY

[1] This system is based on the latest Microsoft technology of “Kinect v2” which is able to track motion, depth and gestures. This project is under development in Chinese Academy of Sciences for more than two years. The system will record speech on user’s request and pass it to AT&T’s Speech Recognition server. The server will in turn return the text from speech recognition in the form of a string. The system will then parse this string into words and look for their respective animations into the dictionary stored in database. These animations will then be played in sequential manner. This system is reliable enough considering the accuracy figures 84% for sign language to speech and 87% for speech to sign language conversion.

[2] This paper discusses an approach for minimizing the time taken for video processing module. The system architecture consists of three modules - Video Processing • Natural Language Processing • Text to Speech Conversion. The system converts American Sign Language videos into simple English sentences and then into speech and then it matches signs using a video processing module, retrieves the corresponding Sign Writing Image File, and compares it with a dictionary. The matched sign’s details are extracted from an Excel file to fill an SSU frame, which generates a structured English sentence. The system finally converts sentence to speech.

[3] This paper presents a mobile-based sign language translation system for individuals with hearing and speech impairments, specifically using the Sistem Isyarat Bahasa Indonesia (SIBI) on Android devices. The system utilizes Viola-Jones object detection for hand recognition and K-Nearest Neighbors (KNN) for gesture classification The application includes features such as a gesture dictionary, translator, and tutorials for sign language learning. Experimental results indicate that the system performs best when hand gestures are in an upright position and within a 30 cm to 70 cm range for optimal recognition. The research contributes to improving communication between the deaf and normal individuals by providing an accessible, mobile-based solution for real-time sign language interpretation.

[4] This paper proposes a method through which the sign language is conveyed through manual(handshape, hand movement) and non-manual (facial expression, body posture, mouthing) channels. This approach combines deep learning (I3D model) and statistical methods (KL-HMM framework). The proposed system The goal of this paper is to address these limitations by proposing an approach that combines the merits of both deep learning approaches and statistical methods to enable fine-grained assessment of sign language videos.

[5] 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.

[6] 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.

[7] 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] 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, healthcare, and government services.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] 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.

[10] This paper 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] This paper 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] 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] 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.

[14] 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] 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.. This shows strong potential the proposed network has for practical application, in particular, for sign language encapsulation.

[16] 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] 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] This 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] In this paper, 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] 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 communicatio

III PROPOSED SOLUTION

The solution is a strong AI-based real-time speech-to-sign language interpretation system using a fusion of speech recognition, natural language processing (NLP), large language models (LLMs), and 3D avatar animation to facilitate equal communication of the deaf and hard-of-hearing people at public and government events. The stream begins by capturing live conversation directly from microphones embedded inside the event setting, which gets processed into text through high-accuracy speech-to-text engines such as OpenAI Whisper or Google Speech-to-Text and supports multi-accent audio and noisy environments. The resulting text then goes through an NLP normalization and context management unit that tidies up the input, preserves conversation history, and prepares the input for gesture mapping. The normalized input is then submitted to a LangGraph-based orchestrator that interfaces with a variety of language agents, including a GPT-4-based sign language interpreter, to translate the normalized text into structured gesture commands as a function of schema metadata, sign language semantics, and previous knowledge. The gesture commands are then passed through to a gesture generator that transcribes the gesture commands into defined sets of animation from a bank of sign language gestures. The animation engine, programmed using Unity or Three.js, drives a 3D avatar that actually carries out the sign language in real-time in terms of hand signs, face expressions, and body posture for there to be correct and meaningful communication. The animated output is projected on holographic screens or LED screens within the event such that all individuals involved in the event can visualize the oral matter being conveyed by means of their own sign language, for example, ISL, ASL, or BSL. The storage and data aspect comprises sign language data sets, metadata files, and session logs that help enhance the accuracy and responsiveness of the model over time. The system supports multi-turn dialogue with context recall and can therefore be employed during long speeches, announcements, and discussions. It is autonomous, minimizes human interaction, and has fallback modes for the possibility of transcription or translation mistakes. The system can be rolled out and replicated, easily embedding into public event infrastructure and translatable to a variety of domains including education, healthcare, transport, and broadcast media. Globally, the intelligent solution enables accessibility, autonomy, and digital inclusion by providing culturally sensitive real-time sign language via AI-powered avatars so that persons with hearing and speech disabilities receive information at the same time and in the same way as everyone else.

IV EXISTING SYSTEM

Over time, numerous systems have been proposed to facilitate communication between the hearing-impaired and the general populace using sign language recognition and synthesis. These include sensor-based wearable systems such as flex and gyroscope gloves and vision-based and artificial intelligence-powered translators. For example, "Deaf Talk using 3D Animated Sign Language" utilizes Microsoft's Kinect to translate speech into animated signs, whereas MediSign and lipreading-based systems employ CNN-BiLSTM and attention mechanisms for enhanced accuracy. Furthermore, different deep learning architectures like CNNs, Bi-LSTM, and Transformers have been utilized for both manual (hand) and non-manual (facial, posture) recognition. Yet issues such as dataset limitation, signer dependence, and necessity for good-quality spatio-temporal modeling continue to exist, particularly in subtle scenarios such as hospitals.

Even with technological advances, existing solutions have inherent drawbacks—dependence on costly hardware, inability to support continuous or bidirectional dialogues, and limited expressiveness in avatar systems. Additionally, most modern platforms find it difficult to scale across sign languages and local dialects. Mobile-based and contactless technologies, like Indonesian Sign Language applications and RF-based solutions, hold potential, but a combined, real-time, AI-based solution is still essential. Integrating speech recognition, natural language processing, and dynamic avatars could provide inclusive, scalable communication without the expense or limitation of existing systems

.

V METHODOLOGY

4.1 REQUIREMENTS GATHERING:

Identify the hearing impaired community's needs in general and concerning public or government events. Specify system goals, supported languages, sign languages (e.g., ISL, ASL), and animation capabilities. Review current systems and develop function and technical specifications

4.2 DATA COLLECTION AND PREPROCESSING:

Collect sign language datasets, the dataset should contain gestures, facial expressions, and body positions. Collect speech datasets for various accents and environments. Annotate and structure datasets to train AI models.

4.3 AI MODEL DEVELOPMENT:

Speech-to-Text Engine: Integrate models like OpenAI Whisper or Google STT to give real-time transcription as input. Natural Language Processing (NLP): Normalize polite & letter word transcriptions of spoken text using basic language models. Sign Language translation: use GPT-4 with LangGraph to orchestrate and translate processed text into sign language gestures. Gesture Mapping: map sign language tokens to sequences of motion using a gesture database.

4.4 3D AVATAR ANIMATION:

Model a 3D avatar using Unity or Three.js. Animate 3D avatars based on sequences of gestures, using realistic depiction commonality with ease of use in regards to an avatar, as well as cultural accuracy. Refine animation to give real-time performance with clear visual presentation.

4.5 SYSTEM INTEGRATION:

Integrate input (microphone), back-end AI pipeline, output (display/hologram). Build an interface that acts as a frontend (with (React.js) for monitoring and user interaction in real-time. Evaluate accuracy, speed or latency, and expressiveness of the system. Add fallback and error trails.

4.6 DEPLOYMENT AND EVALUATION:

Deploy the system at live productions or public exhibition events. Monitor the performance, gather the feedback from users and observers and iteratively.

VI SYSTEM OVERVIEW

A. HIGHER-LEVEL ARCHITECTURE

The architecture of the system contains three layers :

1.CLIENT LAYER:
Captures live speech from microphone (web or kiosk interface). Plays back the rendered 3D sign language animation from monitor or hologram.

2. BACKEND LAYER:
Is based on FastAPI to orchestrate the order of operations for the NLP, gesture generation, and rendering. When using LangGraph, same is useful for modularization and performance for tasks requested by AI agents. Real-time speech-to-text and text-to-sign.

3.DATA LAYER:
Contains gesture datasets, sign language mappings, and model parameters. Leverages SQLite or MySQL (for flexibility) and, optionally, stores session data or logs.

B. IMPORTANT COMPONENTS

1. USER INTERFACE
Constructed using React.js (or another modern framework) . Receives voice input and produces animated output. Designed to be inclusive with multiple language and screen reader support.

2. SPEECH – TO – TEXT ENGINE
Translates spoken input to text sing a model like OpenAI Whisper, Google STT, or Vosk. Offers noise filtering, language detection, and timestamp mapping.

3.NLP & TEXT NORMALIZATION:
Cleans the transcribed text for gesture mapping infrastructure. Identifies filler language, determines intent, and follows context through multi-turn scenarios.

4.SIGN LANGUAGE TRANSLATOR MODEL:
Maps the normalized text into formal gesture commands using LLMs (for example, GPT-4) with domain-specific fine-tuning. Works with regional sign languages: Indian Sign Language (ISL), American Sign Language (ASL), and British Sign Language (BSL).

5. 3D ANIMATION ENGINE:
Leverages Unity, Three.js, or Blender animations to render 3D avatars. Avatar hand signs, body posture, and facial expressions are expressed through the gesture data in real-time

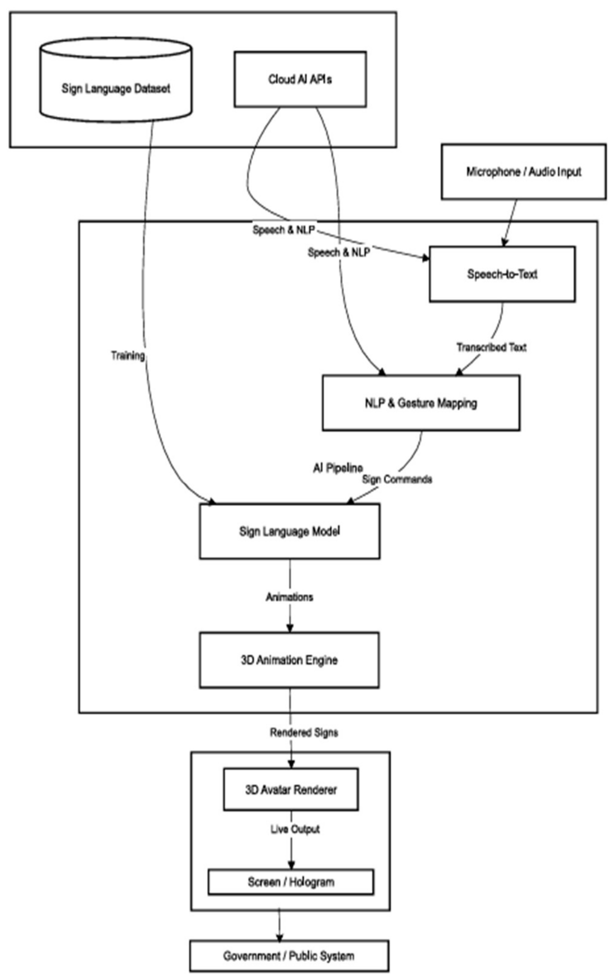
C. INTERACTIVE FEEDBACK LOOP:
Employs LangGraph to store conversations for multi-turn requests. Example:User says "Hello everyone! Welcome to the Independence Day celebration." Then User says "Now moving onto to the awards ceremony!"The avatar resumes the animation flow without a reset and remains in context!.

D. ERROR MANAGEMENT & FEEDBACK OPTIONS
If transcription fails, the fallback options are: Provide subtitles. show standard gestures (i.e., "unable to interpret."). Protocol prevents cultural/signing errors by verifying against a list of sign rules for meaning. Memory logs are kept from model tuning at a later time.

E. ADAPTABILITY AND DEPLOYMENT MODE OF DEPLOYMENT:
Deployable in- Public auditoriums, conference halls news studios, and schools.
Scalable use cases: - Event announcements, government briefings, and classroom lectures. Possible fine-tuning for regional variations with localized sign language datasets.

F. TECHNOLOGY STACK OF AI AVATAR FOR ASL MASS COMMUNICATION:
Frontend - React.js , WebRTC, WebGL / Three.js
Backend API - FastAPI / Node.js
Speech-to-Text - OpenAI Whisper, Google STT, Vosk
NLP & LLMs - GPT-4, LangGraph (LangChain)
3D Rendering - Unity, Blender (Avatar animation)
Data Storage - SQLite, MySQL, Firebase, S3
Visualization Layer - Real-time avatar or holographic scree

VII ARCHITECTURE DIAGRAM



VIII CONCLUSION

As the level of interdependence within the world continues to grow, everyone needs access to information in real time to participate inclusively, yet the hearing and speech disabled are commonly at a disadvantage in public, educational, and government spheres due to a lack of disabling provisions within accessible communications technologies. The project addresses this gap through the creation of an AI-powered real-time speech-to-sign language interpreter using 3D animated avatars. With the combined capability of existing speech-to-text models (e.g., OpenAI Whisper or Google STT), natural language processing, and large language models (e.g., GPT-4), the system can be trained to transcribe spoken language, process and analyze it in context, and reproduce it word for word as sign language movements.

In comparison to the more classical systems using human interpreters or costly hardware like gloves and motion sensors, the system proposed here uses software tools and a 3D avatar in an attempt to provide expressive and culturally aware sign language interpretations. Software such as Unity or Three.js may be used to enable real-time rendering of gestures, including facial expressions and body language, that are important to sign language meaning. It is multi-turn, multi-language, and modular, scalable, and extensible in order to be used in any domain and also for regional sign dialects. The project promotes cost-efficient, automated, and reliable alternative to interpreter-based solutions and supports inclusive real-time communication with no trade-off on expressiveness and correctness.

In addition, public address system integration and possible use in education, health, and the media make it a broader societal contribution. In taking the deaf and hard-of-hearing on par with each other when it comes to access to timely information, the solution addresses not just an independent technological requirement but one which is socially mandated. It sets the stage for future development on AI-enabled accessibility devices in the years to come, and while it makes communication an equality and not a superiority

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 30 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 Reco gnition," 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