Ai Powered Accessibility For Enabling Effective Communication For Hearing And Speech Impaired In Virtual Platforms

R.Archana[1], Vikram S [2], Venkadesh R[3], Balamurugan S[4] ,Vihineshwaran M [5]
[1]Assistant professor, Dept of CSE
[2, 3, 4, 5]Dept of CSE
[1, 2, 3, 4, 5] Dhanalakshmi Srinivasan Engineering College(Autonomous) perambalur,India

Abstract- *Sign language is a visual mode of communication that uses hand gestures and movements to convey meaning, serving as an essential communication tool for individuals with hearing or speech impairments. Despite its importance, many virtual platforms lack the ability to recognize and interpret sign language, creating significant barriers to inclusivity in digital communication. As virtual meetings become more integral to professional and personal communication, the need for inclusivity in these spaces has grown. Current meeting platforms often fail to accommodate users who rely on sign language, limiting their ability to engage fully in discussions. This project aims to address this gap by integrating real-time sign language recognition into video calling platforms, ensuring accessibility for all participants. The proposed system employs the Video Calling Vision Transformer (VCViT) to accurately recognize word-level hand gestures. The system captures live video streams from participants, focusing on hand gestures, and translates them into text or speech in real time. By utilizing advanced video processing techniques, gesture segmentation, and the VCViT's ability to model spatial relationships, the system achieves high recognition accuracy, adapting to different signing styles and environmental conditions. This project strives to create inclusive virtual meeting environments, allowing hearing-impaired individuals to actively participate in discussions. Through AI-driven solutions, it ensures seamless communication, fosters equity, and enhances digital collaboration.*

I. INTRODUCTION

Sign language is manual communication commonly used by people who are deaf. Sign language is not universal; people who are deaf from different countries speak different sign languages. The gestures or symbols in sign language are organized in a linguistic way. Each individual gesture is called a sign. Each sign has three distinct parts: the handshape, the position of the hands, and the movement of the hands. American Sign Language (ASL) is the most commonly used sign language in the India.

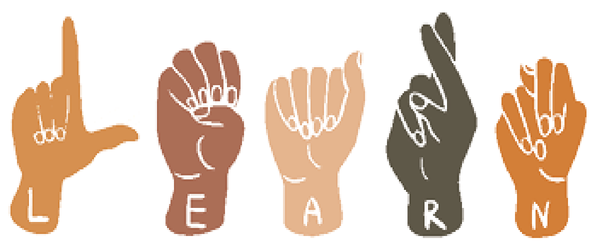


Figure 1.1. Sign Language

A sign language (also signed language) is a language which uses manual communication, body language, and lip patterns instead of sound to convey meaning—simultaneously combining hand shapes, orientation and movement of the hands, arms or body, and facial expressions to fluidly express a speaker's thoughts. Signs often represent complete ideas, not only words. However, in addition to accepted gestures, mime, and hand signs, sign language often includes finger spelling, which involves the use of hand positions to represent the letters of the alphabet. Sign languages have developed in circumstances where groups of people with mutually unintelligible spoken languages found a common base and were able to develop signed forms of communication. Sign languages commonly develop in deaf communities, which include people who are deaf or hard of hearing, friends and families of deaf people, as well as interpreters.

II. EXISTING SYSTEM

Currently, most virtual platforms like Zoom, Microsoft Teams, and Google Meet lack effective support for sign language interpretation.

Assistive Devices

Hearing Aids and Cochlear Implants: Traditional assistive devices include hearing aids and cochlear implants, which aim to improve auditory perception.

Manual Communication Tools

Pen and Paper: Traditional tools such as writing or using pen and paper are often employed for basic communication.

Interpreters

Sign Language Interpreters: In various situations, sign language interpreters are employed to bridge communication gaps between deaf and hearing individuals.

Existing System Algorithm

Support Vector Machines (SVMs)

SVMs have been employed for classification tasks in SLR. They work well with high-dimensional feature vectors extracted from sign language images**.**

Ensemble Learning

Ensemble methods, such as Random Forests and Gradient Boosting, can be applied to combine multiple SLR models to improve overall accuracy and robustness.

Gesture Recognition By Learning From Poses (GRBP)

GRBP is a method that focuses on learning spatial relations between body joint poses for sign language recognition. It utilizes pose features extracted from skeletal data.

III. LITERATURESURVEY

Sign Language Recognition and Translation

- **KinTrans**: Utilizes machine learning to translate sign language into text and speech, aiming to facilitate real-time communication between deaf and hearing individuals.
- **Google's Sign Language Software**: Developed to interpret sign language into text or speech, enhancing communication accessibility.

Real-Time Speech-to-Text Transcription

- **Ava**: Provides real-time transcription services, enabling deaf and hard-of-hearing users to participate in conversations effectively.
- **AvaScribe**: Offers customizable real-time transcriptions, integrating with various devices to assist individuals with hearing impairments.

Augmentative and Alternative Communication (AAC) Devices

- **Voiceitt**: Employs AI to recognize and interpret speech patterns of individuals with speech impairments, facilitating clearer communication.
- **Deep Learning-Based Assistive Technologies**

Audio-Visual Speech Recognition (AVSR): Combines audio and visual cues, such as lip movements, to enhance speech recognition accuracy for hearing-impaired individuals.

- Deep Learning Approaches for Continuous SignLanguage Recognition: A Comprehensive Review- Deep learning models—such as CNNs, RNNs, 3D-CNNs, and Transformers—are used to extract spatial-temporal features from sign language videos. These features are processed using sequence modeling techniques like CTC loss or encoder-decoder frameworks to recognize and translate continuous sign sequences into text.
- Deep Learning-Based Standard Sign Language Discrimination - The system captures sign language gestures using image or video input. Convolutional Neural Networks (CNNs) are employed to extract spatial features from each frame. These features are then passed through fully connected layers or sequence models like LSTMs to learn temporal patterns, if applicable. The final classification layer assigns the gesture to a predefined standard sign label.
- Sign4all: A Low-Cost Application for Deaf People Communication- The Sign4All application captures sign language gestures through a mobile or webcam interface. Using a lightweight CNN model, it extracts and analyzes visual features from real-time video frames. These features are then classified into corresponding text or speech using a pre-trained classifier. The app provides instant translation of sign gestures into readable or spoken language.
- Hand and Pose-Based Feature Selection forZero-Shot Sign Language Recognition- The system extracts hand and body pose features from sign language videos using pose estimation tools like OpenPose or MediaPipe. Key landmarks from hands, arms, and upper body are selected to form a compact and meaningful representation. These pose features are then encoded and mapped to semantic embeddings of sign labels (e.g., glosses or textual descriptions).
- Advancements in SLR:A Comprehensive Review and Future Prospects- The study categorizes approaches based on model typesas CNNs, RNNs, 3D-CNNs, and Transformers—and their applications in both isolated and

continuous SLR. Research papers are evaluated based on accuracy, efficiency, and real-time performance.

IV. PROPOSEDSYSTEM

The proposed system integrates advanced AI technologies, to create a seamless, real-time communication experience for hearing and speech-impaired individuals in virtual meetings.

Deaf Companion System

The proposed system aims to enhance communication for individuals with hearing and mute disabilities through innovative technologies.

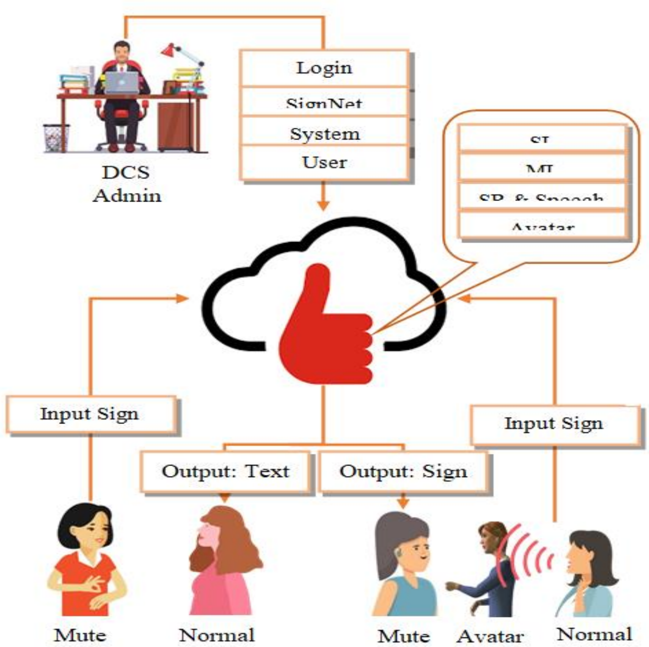
Two-Way Communication

Implementation of a comprehensive two-way communication system, fostering seamless interaction between deaf individuals and the broader community in virtual meeting platforms.

Signnet Model Architecture

Development of the SignNet Model, combining Convolutional Neural Networks (CNN) and Temporal Convolutional Networks (TCN) for robust sign language recognition.

System Architecture



V. MODULES

1.Real-*Time Speech-to-*Text*& Captioning*

- **AppTek.ai**: Provides real-time transcription and speaker identification for meetings and media, facilitating inclusive communication for the deaf and hard-of-hearing community.
- **Ava**: An AI-powered app offering real-time transcription with speaker identification, supporting multiple languages and group conversations.
- **AWS AugmentAbility**: Utilizes Amazon Transcribe and Polly to deliver live transcription and text-to-speech services, including real-time translation across over 75 languages.

2. *Speech-to-*Text *for Non-Standard Speech*

- **Voiceitt**: Offers speech recognition for individuals with speech impairments, converting non-standard speech into understandable language.
- **Echo Labs' CASPER**: An AI-based captioning system providing high-accuracy transcriptions, particularly beneficial in educational settings.

3. *Text-to-Speech & Voice Synthesis*

- **Apple's Live Speech & Personal Voice**: Features that assist non-verbal users by offering pre-recorded response options and text-to-speech during calls.
- **ElevenLabs**: Provides customizable voices, allowing users to design and customize new voices that match their styles and personalities.
- ***4.*** *Multimodal Communication Support*
- **Almawave**: Integrates AI to offer real-time subtitles, sign language interpreters, and content summaries, enhancing accessibility for diverse audiences.

VI. IMPLEMENTATION

1. *Real-Time* Speech-*to-Text (STT) and Captioning*

- **Agora Real-Time STT**: Provides live transcription and closed captions for virtual meetings, lectures, and live streams. It supports real-time translation into multiple languages and can be integrated with AI models like GPT for enhanced interaction.
- **Rev AI Streaming API**: Offers live captions for webcasts, webinars, and broadcasts, ensuring accessibility for diverse audiences.

- **Clevercast Live Captions**: Utilizes AI for automatic speech-to-text conversion with optional real-time human correction, enhancing caption accuracy and readability.

2. *Speech* Recognition *for Non-Standard Speech*

- **Vosk API**: An open-source speech recognition toolkit that supports multiple languages and is suitable for recognizing non-standard speech patterns.
- **Microsoft Speech API (SAPI)**: Provides speech recognition and synthesis capabilities, allowing developers to integrate speech functionalities into applications.

3. *Voice* Synthesis *and Customization*

- **Voice.ai**: Enables real-time voice transformation, allowing users to customize their voice output, which can be particularly beneficial for individuals with speech impairments.

4. Multilingual *and* Multimodal *Communication*

- **CaptionAI**: A system that provides real-time multilingual captioning and broadcasts audio and translated text to personal devices, enhancing accessibility for global audiences.

IMPLEMENTATION STEPS

1. **Assess Needs**: Identify the specific requirements of your target audience, such as preferred languages, types of impairments, and communication preferences.
2. **Select Appropriate Tools**: Choose AI-powered tools that align with the identified needs. For instance, if real-time captioning is essential, platforms like Agora or Rev AI may be suitable.
3. **Integrate with Virtual Platforms**: Ensure that the selected tools can be seamlessly integrated into existing virtual platforms like Zoom, Microsoft Teams, or custom applications.
4. **Customize Features**: Tailor the functionalities to meet specific requirements, such as enabling multilingual support or adjusting caption display settings.
5. **Test and Optimize**: Conduct thorough testing to ensure accuracy and reliability. Gather feedback from users to make necessary adjustments.
6. **Provide Training and Support**: Offer training to users on how to utilize the accessibility features effectively and provide ongoing support to address any issues.

VII. SOFTWARE TESTING

Software testing is the process of evaluating and verifying that a software application functions correctly, meets specified requirements, and is free of defects. It ensures the reliability, security, and performance of the system. Testing is essential to identify bugs, improve functionality, and enhance user experience before deployment. For this project, testing is to ensure accurate sign language recognition, seamless virtual communication, proper speech-to-text conversion, avatar generation, and integration with virtual platforms like Jitsi.

VIII. CONCLUSION

- This AI-powered system bridges the communication gap between hearing and speech-impaired individuals and the broader community in virtual meetings.
- By integrating sign recognition, speech synthesis, and an avatar module, it enables seamless, real-time communication.
- The project enhances inclusivity, fostering a more accessible digital environment. With continuous advancements in AI, gesture recognition, and AR/VR integration, the system will evolve to provide even more accurate and immersive experiences.
- This innovation marks a significant step toward a more connected and inclusive future for virtual collaboration.

REFERENCES

[1] T. Reddy Gadekallu, G. Srivastava, and M. Liyanage, “Hand gesture recognition based on a Harris hawks optimized convolution neural network,” Computers & Electrical Engineering, vol. 100, Article ID 107836, 2022.
[2] G. T. R. Chiranji Lal Chowdhary and B. D. Parameshachari, Computer Vision and Recognition Systems: Research Innovations and Trends, CRC Press, 2022.
[3] M.M. Riaz and Z. Zhang, "Surface EMG Real-Time Chinese Language Recognition Using Artificial Neural Networks" in Intelligent Life System Modelling Image Processing and Analysis Communications in Computer and Information Science, Springer, vol. 1467, 2021.
[4] G. Halvardsson, J. Peterson, C. Soto-Valero and B. Baudry, "Interpretation of Swedish sign language using convolutional neural networks and transfer learning", SN Computer Science, vol. 2, no. 3, pp. 1-15, 2021.

[5] P. Likhar, N. K. Bhagat and R. G N, "Deep Learning Methods for Indian Sign Language Recognition", 2020 IEEE 10th International Conference on Consumer Electronics (ICCE-Berlin), pp. 1-6, 2020.

[6] F. Li, K. Shirahama, M. A. Nisar, X. Huang and M. Grzegorzek, "Deep Transfer Learning for Time Series Data Based on Sensor Modality Classification", Sensors, vol. 31, no. 20(15), pp. 4271, Jul 2020.

[7] J. J. Bird, A. Ekárt and D. R. Faria, "British sign language recognition via late fusion of computer vision and leap motion with transfer learning to american sign language", Sensors, vol. 20, no. 18, pp. 5151, 2020.

[8] S. Sharma, R. Gupta and A. Kumar, "Trbaggboost: an ensemble-based transfer learning method applied to Indian Sign Language recognition", J Ambient Intell Human Comput Online First, 2020, [online] Available: https://doi.org/10.1007/s12652-020-01979-z.

[9] M. Oudah, A. Al-Naji and J. Chahl, "Hand Gesture Recognition Based on Computer Vision: A Review of Techniques", *J. Imaging*, vol. 6, no. 73, 2020.

[10] Z. M. Shakeel, S. So, P. Lingga and J. P. Jeong, "MAST: Myo Armband Sign-Language Translator for Human Hand Activity Classification", IEEE International Conference on Information and Communication Technology Convergence, pp. 494-499, 2020.