International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 03 | March – 2025 DOI: 10.55041/ISJEM02356
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Real time interface for deaf-hearing communication

Mrs.Jangam Bhargavi*1, Chitikala Sairam*2, Donga Hemanth*3,
Kandula Surya Ganesh*4
*1Assistant Professor, Dept. Of CSE (AI & ML) ACE Engineering College Hyderabad, India.
*2,3,4 Students, Dept. Of CSE(AI&ML) of ACE Engineering College Hyderabad, India.

Abstract

Bridging the communication gap between the deaf and hearing communities using AI is achieved by integrating two key modules: Speech-to-Sign Language Translation and Sign Gesture Detection in Real Time. The first module translates English spoken language into American Sign Language (ASL) animations. It consists of three sub-modules: speech-to-text conversion using the speech_recognition module in Python, English text to ASL gloss translation using an NLP model, and ASL gloss to animated video generation, where DWpose Pose Estimation, and an avatar is used for visual representation. The second module focuses on real-time sign gesture detection, where a dataset is created from the WLASL and MS-ASL datasets. Hand gestures are labeled using LabelImg, and a YOLO-based model is trained for hand pose detection to enable real-time recognition. The system aims to enhance accessibility and interaction between deaf and hearing users through an efficient, automated translation and recognition pipeline.

Keywords: Speech-to-sign translation, real-time sign language recognition, ASL gloss, YOLO hand pose detection, AI for accessibility, deep learning for sign language, gesture recognition, DWpose Pose Estimation, NLP, dataset labeling, real-time gesture recognition.

1 Introduction

Effective communication between the Deaf and hearing communities remains a critical challenge, particularly in real-time interactions. Traditional solutions, such as human interpreters or captioning services, may not always be available or efficient. To address this gap, we propose an AI-driven system that facilitates seamless communication by integrating Speech-to-Sign Language Translation and Real-Time Sign Gesture Detection. This innovative approach leverages deep learning, natural language processing (NLP), and computer vision to provide an automated and accessible solution for both spoken and signed language users. The Speech-to-Sign Language Translation module converts spoken English into American Sign Language (ASL) animations using a three-step process: speech-to-text conversion via Python's speech_recognition module, English-to-ASL gloss translation using an NLP model, and ASL gloss to avatar animation powered by DWpose Pose Estimation. This ensures that spoken content is visually represented in ASL, making conversations more inclusive for Deaf individuals. In parallel, the Sign Gesture Detection module enables real-time recognition of hand gestures, allowing Deaf users to sign naturally and have their gestures recognized and translated. By training a YOLO-based model on a dataset derived from WLASL and MS-ASL, the system can accurately detect and interpret hand movements. Labeled gestures, processed with LabelImg, enhance the model's accuracy in recognizing various ASL signs. Together, these two modules create a bidirectional communication system, significantly improving accessibility in education, workplaces, public services, and virtual interactions.

1.1 The Need for Real-Time Deaf-Hearing Communication

Communication between Deaf and hearing individuals presents significant challenges in everyday interactions, education, workplaces, and public services. Traditional solutions, such as human interpreters and text-based captioning, are often limited by availability, cost, and response time. As a result, there is a growing demand for real-time AI-driven solutions that can bridge the communication gap efficiently. Real-time Deaf-hearing communication systems leverage speech recognition, sign language translation, and gesture recognition to facilitate seamless interactions. These technologies not only enhance accessibility but also empower Deaf individuals to engage in conversations without reliance on human interpreters. By integrating artificial intelligence (AI) and deep learning, this system enables natural and fluid conversations, making communication more inclusive and effective.

1.2 Evolution of AI-Based Sign Language Translation

The development of AI-driven sign language translation has evolved significantly over the past decade. Early systems relied on static databases and rule-based approaches, which were limited in scalability and accuracy. With the advancement of deep learning, computer vision, and natural language processing (NLP), modern systems now offer real-time, context-aware translations that improve communication quality. 1.3.2 Natural Language Processing (NLP) for ASL Gloss Once the speech-to-text conversion is complete, the Natural Language Processing (NLP) module translates the English text into ASL gloss—a structured representation of ASL

syntax. Unlike English, ASL follows a different grammatical structure, making gloss generation a crucial step in accurate sign language translation. Recent breakthroughs in automatic speech recognition (ASR), neural machine translation (NMT), and 3D avatar animation have enabled the creation of highly accurate sign language translation models. Additionally, the use of pose estimation techniques like DWpose and machine learningbased gesture recognition (e.g., YOLO models) has enhanced real-time sign detection capabilities. These advancements have led to more natural and expressive AI-powered sign language solutions.

1.3 Key Features of the AI-Based System

The proposed AI-based system consists of four core components, each designed to ensure accurate and efficient real-time Deaf-hearing communication. These components integrate speech recognition, language processing, computer vision, and avatar animation to create a fully automated translation pipeline. Automatic Speech Recognition (ASR) The Automatic Speech Recognition (ASR) module serves as the foundation for the speech-tosign translation process. Using advanced speech recognition algorithms, the system transcribes spoken language into text in real time. Python's speech_recognition module, along with deep learning-based ASR models, ensures high accuracy across different accents, speech patterns, and noise environments. By implementing ambient noise adjustment and adaptive filtering, the ASR module maintains accuracy even in challenging acoustic conditions. This ensures reliable transcription for further processing in the sign language translation pipeline. Natural Language Processing (NLP) for ASL Gloss Once the speech-to-text conversion is complete, the Natural Language Processing (NLP) module translates the English text into ASL gloss—a structured representation of ASL syntax. Unlike English, ASL follows a different grammatical structure, making gloss generation a crucial step in accurate sign language translation. DWpose Pose Estimation and Avatar Animation To create realistic sign language animations, the system employs DWpose for pose estimation. DWpose is a state-of-the-art pose estimation model that accurately tracks hand, arm, and body movements, ensuring precise avatar gestures that match the intended ASL translation. Once the pose is estimated, the system maps these poses onto a 3D avatar, which visually represents the translated ASL gloss. By integrating SMPL-X models, the system generates lifelike and expressive signing motions, enhancing comprehension and engagement for Deaf users. Real-Time Sign Gesture Recognition In addition to speech-to-sign translation, the system incorporates real-time sign gesture recognition, enabling Deaf users to sign directly in front of a camera. The system utilizes a YOLO based hand pose detection model, trained on labeled datasets from WLASL and MS- ASL, to recognize and interpret sign language gestures. DWpose Pose Estimation and Avatar Animation To create realistic sign language animations, the system employs DWpose for pose estimation. DWpose is a state-of-the-art pose estimation model that accurately tracks hand, arm, and body movements, ensuring precise avatar gestures that match the intended ASL translation. Once the pose is estimated, the system maps these poses onto a 3D avatar, which visually represents the translated ASL gloss. By integrating SMPL-X models, the system generates lifelike and expressive signing motions, enhancing comprehension and engagement for Deaf users. To enhance accuracy, the dataset undergoes manual annotation using LabelImg, ensuring highquality training data for sign detection. This module allows for bidirectional communication, where both hearing and Deaf users can interact naturally without requiring an intermediary.

2. Literature Review

2.1 Introduction

The need for a real-time interface for deaf-hearing communication has led to the development of various AI-driven solutions, incorporating speech recognition, natural language processing (NLP), pose estimation, and gesture recognition. Traditional approaches such as human interpreters and text-based captioning have limitations in availability, cost, and efficiency. This section reviews key studies that have contributed to improving real-time sign language translation and gesture recognition.

2.2 Existing Approaches to Real-Time Deaf-Hearing Communication

1.Title : From Audio to Animated Signs

Author: X. Ye, Z. Tang, and S. Manoharan, 2020 This study presents a four-stage pipeline for translating spoken audio into animated ASL gestures. It integrates speech recognition, NLP for ASL gloss conversion, gloss-to-animation mapping, and real-time rendering using a 3D avatar. The system provides accurate ASL representation, ensuring smoother communication. However, the lack of contextual understanding leads to misinterpretations of ambiguous phrases, affecting translation accuracy.

2. Title: Bidirectional Sign Language Translation

Author: Anjali Kanvinde et al., 2019 This project implements a CNN-LSTM model for recognizing sign language gestures from video data. The CNN extracts visual features, while the LSTM captures temporal dependencies in sign sequences. Additionally, a 3D avatar is used to provide visual feedback. Despite its effectiveness in recognizing sequential gestures, the system faces latency issues and struggles with complex hand movements and finger-spelling, limiting real-time performance.

3. Title: Recurrent CNNs for Continuous Sign Language Recognition

Author: R. Cui, H. Liu, and C. Zhang, 2017

This research introduces a Recurrent CNN model for real-time sign recognition using pose estimation techniques. The system tracks hand, arm, and body movements to improve ASL translation accuracy. The use of pose estimation models like DWpose enables precise gesture mapping for 3D avatar-based sign language animation. However, high computational requirements make real-time implementation challenging on low-power devices.

4. Title: How2Sign: A Large-Scale Multimodal Dataset for ASLRecognition

Author: A. Duarte et al., 2021

This study trains a YOLO-based hand detection model on WLASL and MS-ASL datasets to recognize ASL gestures in real-time. The system detects hand landmarks within a single frame, enabling low-latency gesture recognition. While the approach significantly improves processing speed and accuracy, the model lacks facial expression recognition, which is crucial for conveying non-manual ASL components such as eyebrow movements and head tilts.

2.3 Challenges in AI-Based Deaf-Hearing Communication Systems

Despite advancements in real-time speech-to-sign translation and gesture recognition, several challenges persist:

1. Limited Dataset Availability – Most ASL datasets (e.g., WLASL, MS-ASL) have restricted vocabulary, making it difficult to generalize across different users and regional dialects.
2. Grammar Complexity – ASL has a different syntactic structure from English, requiring advanced NLP models to maintain translation accuracy.
3. Computational Constraints – Real-time AI models require high processing power, which limits their deployment on mobile and edge devices.
4. Facial Expression & Context Understanding – Current models focus on hand gestures, but non-manual expressions (e.g., eyebrows, mouth movements) are equally important in ASL.

2.4 Summary

This literature review highlights the evolution of real-time AI-based deaf-hearing communication systems, emphasizing speech-to-sign translation, pose estimation, and gesture recognition. While current approaches enhance accessibility, they still face challenges in real-time processing, grammar accuracy, and non-manual expression recognition. Future research should focus on expanding datasets, improving computational efficiency, and incorporating full-body expression modeling for more natural and accurate ASL representation.

3. Proposed Methodology

3.1 System Overview

The proposed system integrates speech-to-sign translation and real-time gesture recognition to enable seamless communication between deaf and hearing individuals. The methodology consists of three main components: speech processing, sign language translation, and avatar-based animation.

3.2 Speech-to-Sign Language Translation

1. Speech Recognition Module: Converts spoken language into text using deep learning-based ASR models.
2. Natural Language Processing (NLP) Module: Translates English text into ASL gloss using Transformer-based NLP models.
3. Sign Language Animation Module: Uses DWpose pose estimation and a 3D avatar to generate animated ASL signs from gloss text.

3.3 Real-Time Gesture Recognition

1. Dataset Preparation: The system utilizes WLASL and MS-ASL datasets for training a YOLO-based hand gesture recognition model.
2. YOLO-Based Hand Gesture Detection: Identifies hand positions and movements in real-time to recognize ASL gestures.
3. Gesture-to-Text Conversion: Converts recognized gestures into textual output for hearing users to read.

3.4 Integration and Real-Time Communication

1. Bidirectional Communication: The system enables seamless interaction between deaf and hearing users by integrating both speech-to-sign and sign-to-text translations.
2. Real-Time Processing Optimization: Uses GPU acceleration and optimized deep learning models to minimize latency.
3. User Interface Development: Provides a user-friendly interface for real-time communication using AI-driven translation.

3.5 Expected Outcomes

- Improved accessibility for deaf individuals in various domains such as education, workplaces, and public services.
- Enhanced accuracy and speed in speech-to-sign and sign-to-text translation.
- Scalable and efficient AI-based solution for real-time deaf-hearing communication.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 03 | March – 2025 DOI: 10.55041/ISJEM02356
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

4. Architecture

System architecture defines the fundamental structure and design principles of a system, ensuring that all components function together efficiently to achieve a specific goal. It involves organizing hardware, software, data flow, and processing logic in a structured way to support system functionality, scalability, and maintainability. In AI-driven Speech-to-Sign Language Translation and Real-Time Sign Gesture Detection for Video Calls, system architecture plays a critical role in ensuring seamless communication between deaf and hearing users. This architecture is typically designed in a modular format, where each component operates independently but interacts with others to maintain a smooth workflow.

3.1 Overview

The proposed system architecture consists of multiple integrated components designed to facilitate real-time communication between deaf and hearing individuals. It is divided into three main layers: Input Processing, Processing & Translation, and Output Generation.

3.2 Input Processing Layer

- **Speech Recognition Module:** Captures spoken language and converts it into text using deep-learning-based Automatic Speech Recognition (ASR) models.
- **Real-Time Gesture Recognition Module:** Captures hand gestures using video input and processes frames using a YOLO-based detection model.

3.3 Processing & Translation Layer

- **Natural Language Processing (NLP) for ASL Gloss:** Translates English text into ASL gloss using Transformer-based models that maintain grammatical accuracy and context.
- **Pose Estimation & Gesture Recognition:** Uses DWpose to extract skeletal points, ensuring smooth sign language transitions.

3.4 Output Generation Layer

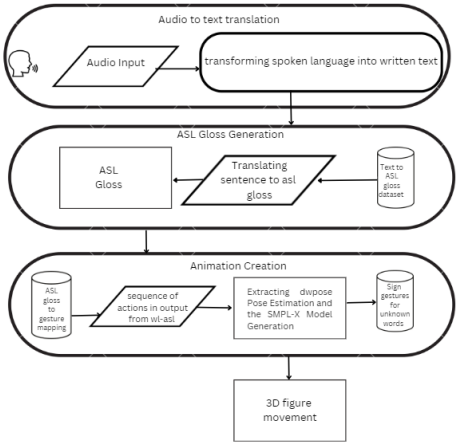
- **3D Avatar Animation:** Converts ASL gloss into animated sign language using a Reallusion-based avatar to enhance comprehension.
- **Text-to-Speech (TTS) for Sign-to-Speech Conversion:** Converts recognized signs into spoken language for hearing users.

3.5 Integration and Real-Time Processing

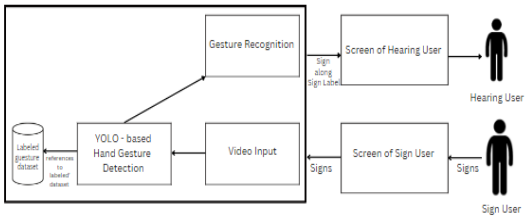
- **Latency Reduction Techniques:** Uses GPU acceleration and model optimization to minimize processing delays.
- **Bidirectional Communication Interface:** Ensures smooth interaction between deaf and hearing users by enabling seamless speech-to-sign and sign-to-text translation.

3.6 Expected Outcomes

- Real-time bidirectional communication between deaf and hearing individuals.
- High accuracy in gesture recognition and ASL gloss translation.
- Scalable architecture that can be adapted for different sign languages and environments.



Architecture 1



Architecture 2

5. Algorithms

Step 1: Speech-to-Sign Language Translation

1. **Speech Input Processing**
   - Capture spoken audio using a microphone.
   - Apply noise reduction and preprocessing techniques.
   - Convert speech to text using an ASR (Automatic Speech Recognition) model.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 03 | March – 2025 DOI: 10.55041/ISJEM02356
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

2. **English Text to ASL Gloss Translation**
    - Process the transcribed English text.
    - Translate the English text into ASL gloss using an NLP model.
    - Restructure the sentence to follow ASL grammar rules.
3. **ASL Gloss to Animated Video Generation**
    - Retrieve gesture data corresponding to the ASL gloss from a dataset.
    - Use DWpose Pose Estimation to extract keypoints for hand and body movements.
    - Apply the extracted keypoints to a 3D avatar using SMPL-X for smooth animation.
    - Generate and render an animated video of the ASL interpretation.

Step 2: Real-Time Sign Gesture Detection

1. **Dataset Preparation**
    - Collect sign language video data from WLASL and MS-ASL datasets.
    - Label hand gestures using annotation tools like LabelImg.
    - Preprocess and augment the dataset to improve model robustness.
2. **Training the YOLO-Based Gesture Recognition Model**
    - Define and configure the YOLO model architecture for detecting hand poses.
    - Train the model using the labeled dataset with bounding boxes.
    - Validate the model's accuracy using test samples.
    - Save the trained model weights for real-time deployment.
3. **Live Gesture Recognition**
    - Capture real-time video input using a webcam.
    - Apply the trained YOLO model to detect hand gestures.
    - Recognize the gesture and match it to a predefined sign language class.
    - Display the detected sign in textual format for hearing users.

6. Data Flow

The system enables **real-time communication** between **deaf and hearing users** by processing speech and sign language simultaneously. It consists of two major data flows that occur **at the same time**:

1. **Speech-to-Sign Animation** (for the deaf user)
2. **Sign-to-Text Conversion** (for the hearing user)

1. Hearing Person Speaks → Sign Language Animation (Speech-to-Sign)

Input:

- The hearing person speaks into a microphone during a video call.

Process:

1. **Speech-to-Text Conversion**
    - Uses **speech recognition** (e.g., speech_recognition module in Python) to convert **spoken words into text**.
2. **English-to-ASL Gloss Translation**
    - Converts **English text** into **ASL gloss** using an **NLP model**
3. **ASL Animation Generation**
    - Uses **DWpose pose estimation** and a **3D avatar (SMPL-X)** to animate the **corresponding ASL gestures**.

Output:

- The **3D avatar signs in ASL**, displaying the translated message to the deaf user.

2. Deaf Person Signs → Text for Hearing User (Sign-to-Text)

Input:

- The deaf person performs **sign gestures** in front of the camera.

Process:

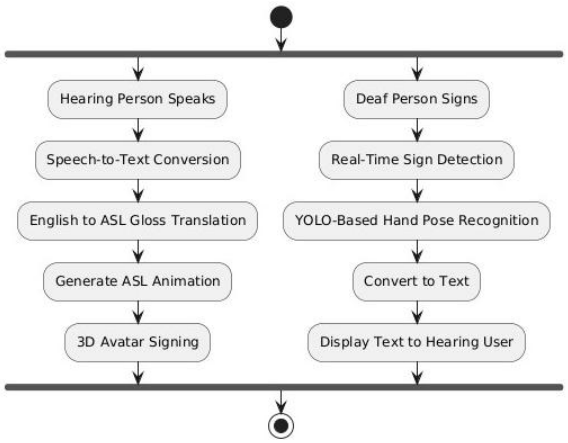
1. **Real-Time Sign Detection**
    - Uses a **live video feed** to capture hand movements and body gestures.
2. **YOLO-Based Hand Pose Recognition**
    - Identifies **hand shapes, motion, and positions** using a **trained YOLO model** for sign recognition.

Output:

- The **translated text** is displayed on the screen for the hearing user.

Simultaneous Processing (Real-Time Communication)

- Both processes happen **at the same time** during the video call.
- The system ensures **low-latency communication** by using **optimized deep learning models** for **real-time** execution.



Data Flow

Conclusion

The **Real-Time Interface for Deaf and Hearing Communication** successfully bridges the communication gap between deaf and hearing individuals by leveraging **AI-driven speech recognition, sign language translation, and gesture detection**. The system operates in **real-time**, enabling smooth and efficient interaction through:

1. **Speech-to-Sign Translation** – Converts spoken language into **ASL animations** using an NLP-based model and a **3D avatar**, allowing deaf users to understand the conversation visually.
2. **Sign-to-Text Translation** – Detects and recognizes **sign gestures** through a **YOLO-based hand pose detection model**, translating them into text for hearing users.

By integrating these two processes simultaneously, the system ensures **seamless, bidirectional communication** in video calls. It eliminates the reliance on human interpreters, making **digital conversations more inclusive and accessible** for the Deaf and Hard-of-Hearing communities.

This work contributes significantly to the **advancement of AI-based sign language processing**, paving the way for more **inclusive communication technologies** in education, workplaces, and daily interactions. Future improvements could involve **enhancing avatar realism**, supporting **multiple sign languages**, and **optimizing real-time performance** for large-scale deployment.

References

[1] O. Koller, S. Zargaran, and H. Ney, "Re-Sign: Realigned end-to-end sequence modeling with deep recurrent CNN-HMMs," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017.

[2] L. Naert, C. Larboulette, and S. Gibet, "A survey on the animation of signing avatars: From sign representation to utterance synthesis," *Comput. Graph.*, vol. 92, pp. 76–98, Nov. 2020.

[3] J. Forster, O. Koller, C. Oberdorfer, Y. Gweth, and H. Ney, "Improving continuous sign language recognition: Speech recognition techniques and system design," *Proc. 4th Workshop Speech Lang. Process. Assist. Technol.*, 2013.

[4] H.-D. Yang, S. Sclaroff, and S.-W. Lee, "Sign language spotting with a threshold model based on conditional random fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 7, pp. 1264–1277, 2009.

[5] O. Koller, J. Forster, and H. Ney, "Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers," *Comput. Vis. Image Underst.*, vol. 141, pp. 108–125, 2015.

[6] R. Cui, H. Liu, and C. Zhang, "Recurrent convolutional neural networks for continuous sign language recognition by staged optimization," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017.

[7] A. Duarte et al., "How2Sign: A large-scale multimodal dataset for continuous American sign language," *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021.

[8] R. Wolfe et al., "The myth of signing avatars," *Proc. 18th Biennial Mach. Transl. Summit*, 2021.

[9] V. N. T. Truong, C. Yang, and Q. Tran, "A translator for American sign language to text and speech," *Proc. IEEE 5th Global Conf. Consum. Electron.*, Kyoto, Japan, 2016.