JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Generating subtitle and sign language of the audio using Artificial Intelligence

[1]Prof.Jaya S S, [2]Vasanthakumar B, [3]Kannan P,[4]Shobika S

[1]Assistant Professor, [2]UG Student, [3]UG Student, [4]UG Student
[1]Department of Computer Science and Engineering,
[1]RVS College of Engineering and Technology, Coimbatore, India

Abstract : The accessibility gap for the deaf and hard-of-hearing communities remains a significant challenge in education and societal communication. Traditional methods of translating spoken language into sign language or subtitles often require human intervention, making them inefficient and costly. This paper proposes a Artificial Intelligence-based system for automated sign language generation and subtitle creation from speech and video audio. The system utilizes speech recognition, natural language processing(NLP), and computer vision to convert spoken words into real-time subtitles and corresponding sign language gestures. This approach enhances accessibility in education and public communication, ensuring equal participation for individuals with hearing impairments. The proposed system integrates automatic speech recognition (ASR),sign language synthesis using neural networks, and real-time subtitle generation. By deep learning models trained on extensive datasets, this system aims to bridge the communication gap for the deaf and improve inclusivity across various domains.

Index Terms - Sign Language Generation, Speech-to-sign language, Machine Learning, Accessibility, Deaf Communication, Education Technology.

I. Introduction

Deaf people and hard of hearing individuals often struggle to understand spoken content in educational, entertainment and professional settings. While subtitles help, they do not convey the expressions and gestures essential to Indian Sign Language (ISL). This project aims to bridge the accessibility gap by developing an AI-powered real-time speech-to-ISL translation system.

The system captures live speech or extracts audio from videos, converts it into real-time subtitles, and then translates the text into ISLavatar . By Natural Language Processing (NLP), and Computer Vision, the system ensures accurate and synchronized ISL gestures, making spoken content more inclusive and accessible.

II. RELATED WORK

TITLE: Advancements in sign language recognition
AUTHORS: Bashaer A.AL Abdullah, Ghada A. Amoudi and Hanan S.Alghamdi
PUBLICATION:IEEE Access,2024

DESCRIPTION:

The development of automated Sign Language Translation Systems (SLTS), focusing on the role of Artificial Intelligence and machine learning in improving sign recognition. By systematically analyzing 58 research papers, including the most cited works up to 2023, the study highlights technological advancements, especially the use of deep learning models like CNNs and RNNs. It also emphasizes the importance of integrating non-manual features to enhance accuracy. The review outlines key achievements, current challenges, and future research directions aimed at improving accessibility and communication for the hearing impaired.

TITLE: A prototype for Mexican sign language recognition and synthesis in support of a primary care physician.

AUTHORS: Candy Obdulia Sosa-Jimenez,Homero Vladimier Rios- Figueroa and ana luisa solis

PUBLICATION:IEEE Access,2022

DESCRIPTION:

A real-time, bidirectional translator system for Mexican Sign Language (MSL) designed to improve communication in primary healthcare settings. Targeting interactions between hearing doctors and deaf patients, the system recognizes and animates MSL signs related to general medical consultations, including fingerspelling for personal information. Using a Microsoft Kinect sensor and Hidden Markov Models (HMMs), it processes sign trajectories and images for accurate real-time recognition. Testing with 22 participants across 82 signs yielded high performance, with an average accuracy of 99% and an F1 score of 88%.

TITLE: Developed a user-independent recognition system based on depth images and PCANet features.
AUTHORS: Aly et al.,
PUBLICATION:IEEE Access,2019

DESCRIPTION:

A user-independent sign language recognition system that utilizes depth images and PCANet features. By focusing on depth data, the system enhances robustness across different users, while PCANet—a deep learning framework based on principal component analysis—enables effective feature extraction for accurate sign recognition.

TITLE: Spanish dictionary—Mexican sign language. (DIELSEME),'' Special Educ. Board, Ministry Educ., Mexico,
AUTHORS:M. T. Calvo-Hernandez, M. de L. Acosta-Huerta, E. D. Maya-Ortega, E. Sanabria-Ramos, and G. A. Zeleni**PUBLICATION:**IEEE Access,2004

DESCRIPTION:

The *Spanish Dictionary of Mexican Sign Language (DIELSEME)*, published by the Special Education Board of the Ministry of Education in Mexico, serves as an essential linguistic and educational resource. It documents and standardizes Mexican Sign Language (MSL), supporting both language preservation and instructional efforts for the deaf community and educators.

III. EXISTING SYSYEM

•**Limited Domain-Specific Applications of SLTS** – While significant progress has been made in general Sign Language Translation Systems (SLTS), many systems are not tailored to specific real-world use cases such as healthcare or education. The 2022 study by Sosa-Jimenez et al. introduces a rare domain-specific system designed for doctor-patient communication using MSL, highlighting a gap in the broader application of SLTS in critical fields like medical or legal services【1】.

•**User-Dependence and Signer Variability** – Many sign language recognition systems struggle with recognizing signs across different users due to differences in gesture speed, hand shape, or body proportions. The work by Aly et al. (2019) addresses this issue by using depth images and PCANet features to build a user-independent system, but such solutions are still limited and not widely implemented【2】.

•**Lack of Standardized Linguistic Resources** – The absence of standardized sign language datasets for various regions, including Mexico, hampers model training and benchmarking. The DIELSEME dictionary (Calvo-Hernandez et al., 2004) offers foundational support for MSL but similar linguistic resources are missing for many other regional sign languages, creating challenges in language modeling and training accuracy【3】.

•**Inadequate Integration of Non-Manual Features** – Many systems focus only on manual gestures (hands, fingers) and ignore non-manual features like facial expressions or body posture, which are crucial in sign language syntax and semantics. Bashaer et al. (2024) emphasize that deep learning techniques are increasingly integrating these features, but comprehensive solutions are still under research and development【4】.

IV. PROPOSED SYSTEM

Real-Time Speech-to-Indian Sign Language (ISL) Animation Translation
This system enables real-time translation of live speech or video audio into subtitles, which are then converted into Indian Sign Language (ISL) using a 3D animation. It is designed to assist deaf and hard-of-hearing individuals in understanding spoken content more effectively through seamless AI-driven processing and animation.

3.1 Speech-to-Text Processing

- Capture live audio or extract audio from video content.
- Apply AI-based noise filtering to improve audio clarity in diverse environments.
- Convert speech to real-time, word-by-word subtitles using speech recognition engines such as **Google Speech-to-Text** or **Whisper**.

3.2 NLP-Based Sign Language Translation

- Process generated subtitles through **Natural Language Processing (NLP)** models.
- Convert text to grammatically correct ISL structures, accounting for ISL-specific syntax and phrasing.
- Map processed words/phrases to corresponding ISL gestures using a **pre-trained ISL gesture dataset**.

3.3 3D Avatar-Based Sign Language Animation

- Implement **Computer Vision** techniques for facial and hand tracking to support synchronization.
- Use AI to sync gestures with the timing of spoken content.
- Render ISL signs using **Blender-designed 3D animations** and **Three.js**, delivering visually accurate and expressive animations.

3.4 Web-Based Deployment & Real-Time Processing

- Develop a backend using **Flask** and **WebSockets** to handle real-time subtitle translation and gesture generation.
- Deploy an interactive frontend using **Three.js** to display the 3D ISL animation in the browser.
- Ensure low-latency, scalable performance for accessible and inclusive web-based interaction.

V. RESEARCH METHODOLOGY

This research follows a multi-stage methodology integrating speech recognition, natural language processing (NLP), and computer vision techniques to generate real-time subtitles and corresponding sign language animations from audio input. The system is designed to ensure accuracy, synchronization, and accessibility for deaf and hard-of-hearing users. The methodology is divided into the following phases:

4.1 Data Collection and Preprocessing

- **Audio Dataset Preparation**: A curated dataset of speech samples in diverse accents and noise conditions is used to simulate real-world environments.
- **Noise Reduction**: AI-based noise filtering algorithms (e.g., spectral subtraction, deep noise suppression models) are applied to improve audio clarity.
- **Audio Segmentation**: Continuous speech is segmented into smaller time-synced units for precise transcription and gesture synchronization.

4.2 Speech-to-Text Conversion

- **Model Selection**: Pre-trained speech recognition models such as **Google Speech-to-Text API** and **OpenAI's Whisper** are evaluated for real-time transcription capabilities.
- **Real-Time Subtitling**: Audio is transcribed into text using word-by-word or phrase-based streaming recognition. Accuracy and latency are measured for each approach.

4.3 NLP-Based ISL Translation

- **Text Normalization**: The transcribed text is cleaned and formatted for consistent input to the translation module.
- **Grammar Adaptation**: A rule-based or transformer-based NLP model rearranges English grammar into Indian Sign Language (ISL) sentence structures.
- **Gesture Mapping**: Words and phrases are mapped to ISL gestures using a pre-annotated gesture database built from existing ISL corpora and videos.

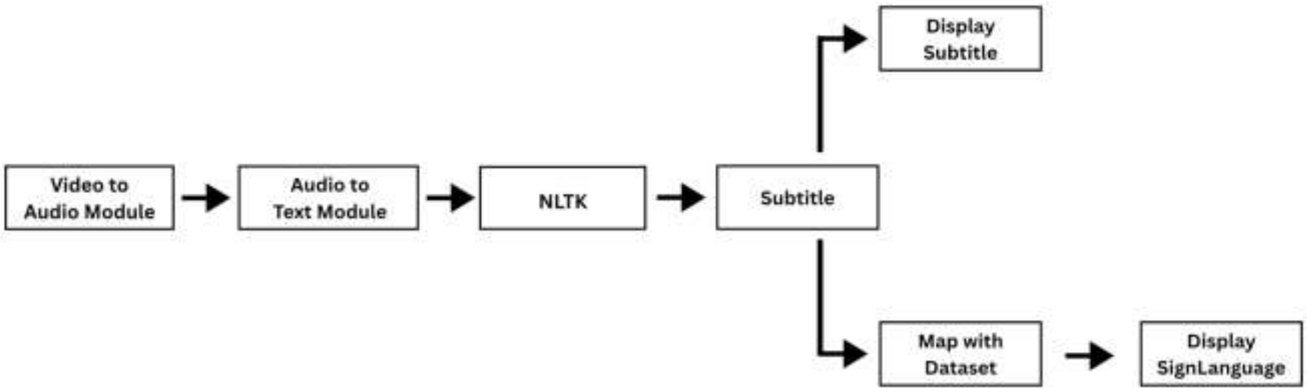
4.4 Sign Language Animation Using 3D Avatars

- **Animation Design**: A 3D animation is modeled using **Blender**, capable of rendering upper body gestures and facial expressions.
- **Animation Engine**: Synchronized with the subtitle timing and gesture transitions.
- **Gesture Generation**: Sign gestures are dynamically rendered based on frame sequences derived from the ISL gesture dataset.

4.5 System Integration and Deployment

- **Backend Framework**: The real-time processing pipeline is implemented using **Django** and **WebSockets**, enabling continuous audio processing and output generation.
- **Frontend Interface**: A responsive web interface displays subtitles and the animated avatar using WebGL-based **Three.js** rendering.
- **Evaluation Metrics**: System performance is evaluated using metrics such as:
    - Subtitle accuracy (WER – Word Error Rate),
    - Gesture correctness (based on expert-labeled ISL validation),
    - Latency (from audio input to animation output),
    - User satisfaction (via surveys with deaf and hard-of-hearing users).

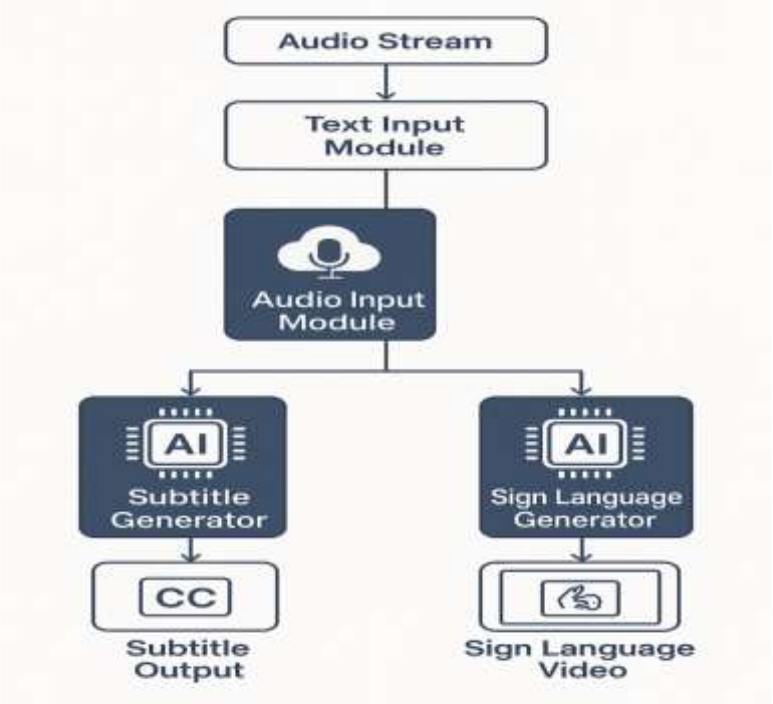
VI. BLOCK DIAGRAM



This block diagram illustrates a system designed to convert spoken audio into both subtitles and sign language for accessibility. The process begins with capturing **audio**, which is then processed by an **audio-to-text module** that transcribes the spoken words into written text. The resulting text is then analyzed using **Natural Language Processing (NLP)** to improve understanding and contextual accuracy. The processed text is passed to a **text-to-subtitle conversion** module, where it is formatted into subtitle-friendly content. This output branches into two parallel processes: one sends the subtitles directly to a **subtitle display**, while the other maps the subtitle content to a corresponding **sign language dataset**. This mapping facilitates the generation of sign language, which is then shown via a **sign display**, enhancing accessibility for individuals who are deaf or hard of hearing.

GENERATING SUBTITLE AND SIGN LANGUAGE

The process of generating subtitles and sign language from audio using Artificial Intelligence involves multiple interconnected stages. Initially, the audio input is processed through Speech Recognition systems, which convert spoken words into written text in real-time. Advanced AI models trained on large datasets ensure high accuracy in transcribing different accents, languages, and speaking speeds. Once the audio is transcribed, Natural Language Processing (NLP) techniques are used to enhance the readability of the text, including punctuation, segmentation, and contextual understanding. The clean subtitle text is then synchronized with the audio timeline to ensure accurate timing. For sign language generation, the subtitle text is passed to a Sign Language Generation Module, which uses AI-driven animation engines. These modules convert the text into sign language gestures by mapping words and phrases to a set of predefined sign language symbols. The final output consists of synchronized subtitles displayed on screen and a virtual 3D animation delivering the message in sign language, making content more inclusive for the deaf and hard-of-hearing community.



CONCLUSION

The proposed AI-powered real-time speech-to-Indian Sign Language (ISL) translation system enhances accessibility by converting live speech or video audio into synchronized ISL animation using Automatic Speech Recognition (ASR),Natural Language Processing (NLP)) and Computer Vision. The system effectively captures and processes speech using ASR model ensures grammatically correct ISL translation using NLP and renders real-time 3D sign language animation using computer vision and Three.js animation. Flask and WebSocket's for real-time deployment .

I. FUTURE WORK

- **Multilingual Support**: Expanding the system to handle multiple languages and dialects will improve its applicability across different regions and cultures.
- **Emotional Expression in Sign Language**: Integrating emotional tone detection from the audio to reflect corresponding facial expressions and body language in the sign language avatar.
- **Personalized Animations**: Developing customizable avatars that can adapt to user preferences, including different signing styles (e.g., SEE vs. ASL).
- **Edge Computing Deployment**: Optimizing the model for deployment on edge devices like smartphones and AR glasses to enable offline usage with low latency.
- **Robustness to Noisy Environments**: Enhancing the model's resilience to background noise to ensure consistent performance in various real-world settings.
- **3D Sign Language Animation**: Introducing full 3D animations for sign language output to create a more realistic and engaging user experience.
- **User Feedback Loop**: Incorporating real-time user feedback to continuously refine and improve subtitle and sign language generation accuracy.

II. REFERNCE

[1] Candy Obdulia Sosa-Jimenez , Homero Vladimier Rios- Figueroa and ana luisa solis , Mexican sign language recognition and synthesis, National council of Mexico(CONACYT), 2022.

[2] Bashaer A.AL Abdullah , Ghada A.Amoudi and Hanan S.Alghamdi,Advancements in sign language recognition ,Department of information system,king abdullaziz university,2024.

[3] S.-O. Caballero-Morales and F. Trujillo-Romero, ‘‘3D modeling of the Mexican sign language for a speech-to-sign language system,’’ Comput. Sistemas, vol. 17, no. 4, pp. 593–608, 2013.

[4] M. T. Calvo-Hernandez, Mexican Sign Language Dictionary. Mexico City, Mexico: Ministry of Education, 2010

[5] M. T. Calvo-Hernandez, M. de L. Acosta-Huerta, E. D. Maya-Ortega, E. Sanabria-Ramos, and G. A. Zeleni, ‘‘Spanish dictionary—Mexican sign language. (DIELSEME),’’ Special Educ. Board, Ministry Educ., Mexico, (in Spanish), 2004.