IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 13, Issue 8, August 2024

DOI: 10.17148/IJARCCE.2024.13848

SMART TRANSLATION FOR DEAF PEOPLE

Ananya B A[1], Chaithra U C[2]

Post-Graduation Student, Department of MCA, Vidya Vikas Institute of Engineering and Technology, Mysore, Karnataka[1]

Assistant Professor, Department of MCA, Vidya Vikas Institute of Engineering and Technology, Mysore, Karnataka[2]

Abstract: This project aims to develop an innovative communication system that bridges the gap between deaf and hearing individuals by addressing both verbal and visual communication barriers. The system is implemented in two phases. In the first phase, the focus is on converting audio messages into Indian Sign Language (ISL). Audio input, either live or pre-recorded, is transcribed into text using advanced speech recognition technologies. This text is then mapped to predefined ISL images or GIFs, enabling seamless communication by making spoken language accessible to the deaf community through visual sign representations.

The second phase enhances the system's ability to interpret visual information for the deaf. Images are collected and used to train a Multilayer Perceptron (MLP) model, achieving a 90% accuracy in recognizing and interpreting these images. The model processes the images and converts them into corresponding text or speech outputs, allowing deaf individuals to understand visual cues through textual or spoken descriptions. This dual-phase approach not only facilitates effective communication between deaf and hearing individuals but also enhances the interaction of the deaf community with their environment.

I. INTROUDCTION

Communication is a fundamental aspect of human interaction, enabling the exchange of ideas, emotions, and information. However, for individuals with hearing impairments, traditional forms of communication, particularly those involving auditory signals, present significant challenges. The inability to hear spoken language or respond in kind creates a substantial barrier to effective communication, often leading to social isolation and limited access to information. Bridging this communication gap between deaf and hearing individuals is not only a technical challenge but also a societal imperative, as it directly impacts the inclusivity and quality of life for the deaf community.

Indian Sign Language (ISL) serves as the primary mode of communication for many deaf individuals in India. While ISL is effective within the deaf community, its lack of widespread adoption among hearing individuals creates a communication divide. Current solutions, such as sign language interpreters, while effective, are not always readily available and can be resource-intensive. Automated systems that can translate spoken language into sign language and vice versa offer a promising alternative, leveraging advancements in technology to create scalable, accessible solutions.

In this context, the development of a robust communication system that can convert audio messages into ISL and interpret visual information for the deaf community is of paramount importance. Such a system not only facilitates direct communication between deaf and hearing individuals but also empowers the deaf community by enhancing their ability to interact with their environment.

Problem Statement

A significant challenge faced by individuals with hearing or speech impairments is the absence of effective real-time translation tools that convert spoken language into sign language. Although speech-to-text and text-to-speech applications exist, they often fall short in meeting the specific needs of sign language users.

The growing demand for such technologies spans various fields, including international communication, travel, and commerce, yet current solutions lack the necessary integration of real-time gesture detection, which is crucial for seamless communication. Our approach addresses these deficiencies by creating a comprehensive system that not only translates spoken language into sign language but also recognizes and interprets gestures in real time, thereby facilitating more effective communication for the deaf and hard-of-hearing community.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2024
DOI: 10.17148/IJARCCE.2024.13848

II. LITERATURE SURVEY

[1]In this paper, we propose a marker-free, visual Indian Sign Language recognition system that uses image processing, computer vision, and neural network methodologies to identify hand characteristics in video images captured with a web camera. This method will convert a video of a daily, frequently used full-sentence gesture into text, which will then be converted into audio.

[2] This paper explains how AdaBoost and Haarlike classifiers were used to recognise American Sign Language from live videos with complex backgrounds and translate it into text and speech. The large dataset that was used in the training process contributed significantly to the high accuracy result. To assess the system's recognition rate, we recorded a video with 20 frames per letter in the experiments.

[3] The average accuracy rate of the Android Speech API was 94.06 percent, the average accuracy rate of morphological analysis and ECS stemming algorithm was 95 percent, and the overall accuracy rate was 80.71 percent, according to the test results. These findings suggest that the SIBI voice translator was well-made overall.

[4] It is critical to use sign language translation (SLT) to bridge the communication gap between deaf and hearing people. In recent years, there has been a lot of interest in SLT research based on neural translation frameworks. SLT research is still in its early stages, despite its progress. In fact, current systems have difficulty processing long sign sentences, which frequently involve long-distance dependencies and consume a large amount of resources. To address this issue, the authors propose two explainable adaptations to traditional neural SLTmodels using optimised tokenization-related modules.

[5] The goal is to look into different methods for effective communication between Sign and English. A hardware glove with flex sensors is initially implemented, but its accuracy is shown to be low. To improve accuracy even more, a model based on a convolutional neural network was trained on an existing data set. A new diverse data-set was created and the model was improved further because the data-sets were not versatile and the scope was limited. The new model is extremely accurate, and it can correctly predict almost any alphabet.

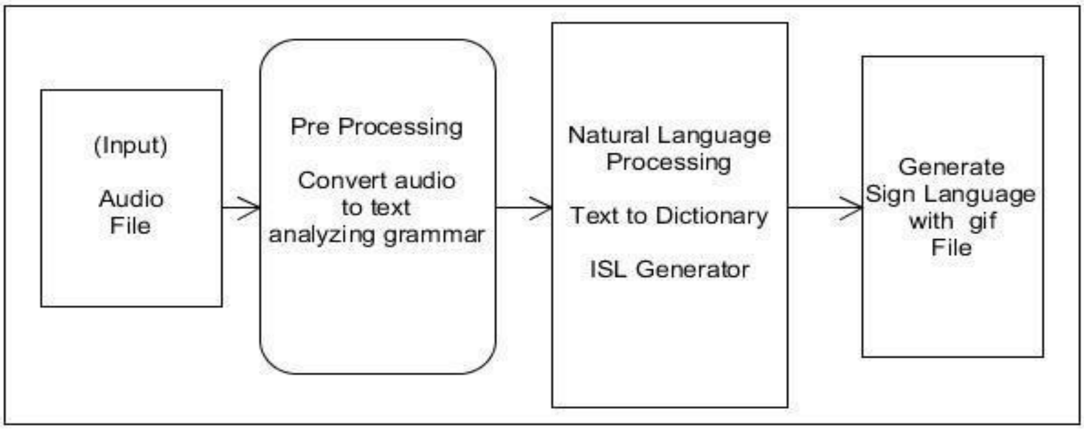
III. METHODOLOGY

Background study & Information Gathering

In today's society, individuals with speech and hearing impairments often face limited and inadequate communication options. Traditional methods primarily rely on manual sign language interpretation, which requires both the speaker and the listener to be proficient in sign language. Unfortunately, the general population usually lacks this level of expertise, leading to significant barriers in communication for those who rely on it. This underscores the need for more accessible and inclusive communication solutions that can bridge the gap and foster smoother interactions for individuals with these impairments.

While there are various applications available that convert text to speech and vice versa, many do not address the specific needs of sign language users. Furthermore, the lack of effective real-time gesture recognition systems continues to pose additional challenges for independent communication among those with speech impairments.

Proposed Methodology



IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

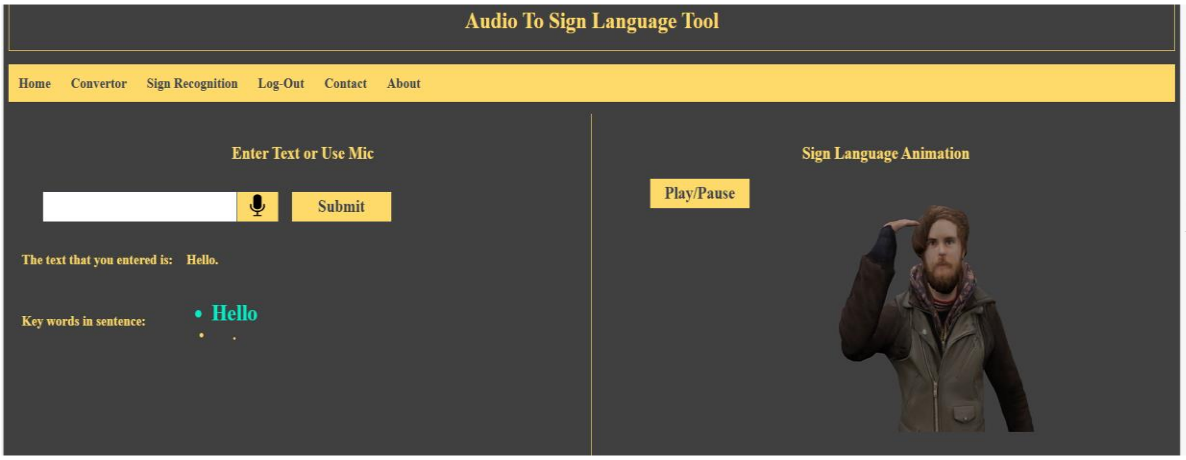
Impact Factor 8.102 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 13, Issue 8, August 2024

DOI: 10.17148/IJARCCE.2024.13848

In the first phase, the focus is on converting audio messages into Indian Sign Language (ISL). The process begins with the acquisition of audio input through a microphone for real-time conversations or by processing pre-recorded audio files. Advanced speech recognition technologies are employed to transcribe the audio input into text accurately. This transcription forms the basis for the next step, where the text is mapped to a predefined set of ISL images or GIFs. These visual elements are carefully curated to represent the vocabulary and grammar of ISL accurately. The mapped ISL images or GIFs are then displayed, enabling deaf individuals to receive and understand spoken messages through visual sign language representations. This phase ensures that the system can effectively translate auditory information into a visual format comprehensible to the deaf community.

The second phase addresses the challenge of interpreting visual information for deaf individuals by converting images into text or speech. This phase involves the collection of various images that need to be interpreted. We utilized the MediaPipe framework for this purpose, which provides robust tools for image processing and pose estimation. The collected images are processed and analyzed using MediaPipe to extract relevant features. These features are then used to train a Multilayer Perceptron (MLP) model. The MLP model is trained on a dataset of labeled images to recognize and classify them accurately. Through rigorous training and validation, the MLP model achieved an accuracy of 90%. Once trained, the model can process new images and convert them into corresponding text or speech outputs. This capability allows deaf individuals to understand visual cues and imagery through textual or spoken descriptions, enhancing their ability to interact with their environment.

The development of our communication system for the deaf involves the implementation of several advanced algorithms in both phases: Audio to Sign Language Conversion and Image to Text/Speech Conversion. This section details the algorithms used, their mathematical foundations, and the step-by-step processes involved.



Phase 1: Audio to Sign Language Conversion

1. Audio Input Acquisition:
The first step is capturing the audio input using a microphone. For pre-recorded audio files, the system processes the files using standard audio processing libraries. The captured audio signal is represented as a time-domain waveform. Let x(t)x(t)x(t) denote the audio signal as a function of time ttt.

2. Speech-to-Text Conversion:
The audio signal x(t)x(t)x(t) is converted to text using Automatic Speech Recognition (ASR) technologies. We use the Hidden Markov Model (HMM) and Deep Neural Network (DNN) hybrid model, which is prevalent in modern ASR systems.

The HMM models the temporal variability of the speech signal, while the DNN models the spectral variability. The audio signal is divided into small frames of 25 ms with a 10 ms overlap. For each frame, feature vectors such as Mel-Frequency Cepstral Coefficients (MFCCs) are extracted. The feature vector at time ttt is represented as ft\mathbf{f}_tft.

3. Text to Sign Language Mapping:
Once the text T={t1,t2,…,tn}T = \{t_1, t_2, \ldots, t_n\}T={t1,t2,…,tn} is obtained from the ASR system, the next step is to map this text to the corresponding Indian Sign Language (ISL) representations. This involves a predefined dictionary that maps each word to its corresponding ISL image or GIF.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

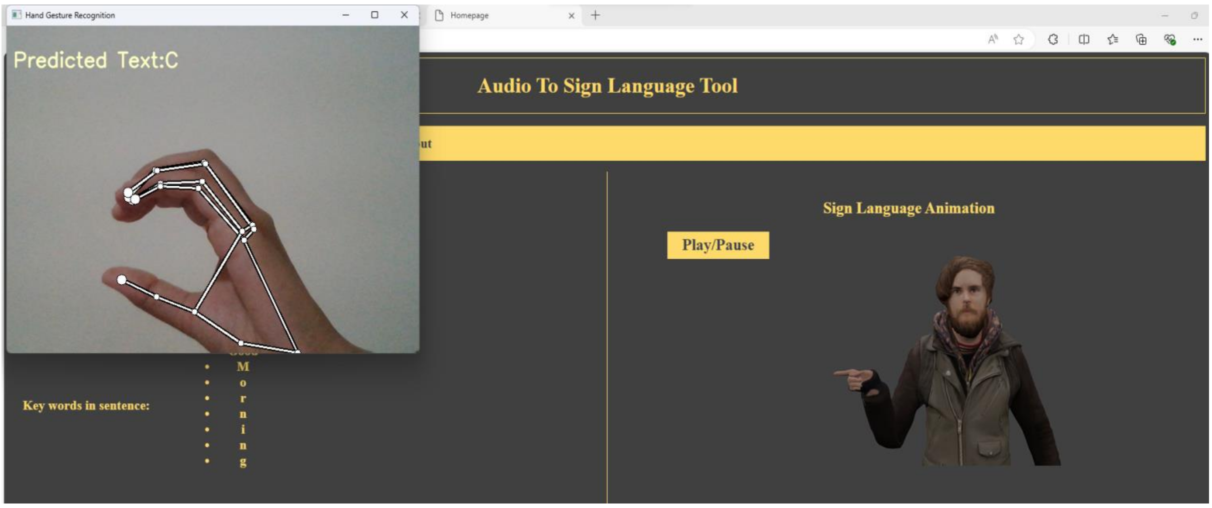
Impact Factor 8.102 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 13, Issue 8, August 2024

DOI: 10.17148/IJARCCE.2024.13848

Let DDD be the dictionary where each word tit_iti maps to an ISL representation d(ti)d(t_i)d(ti). The sequence of ISL images or GIFs is given by:

D(T)={d(t1),d(t2),…,d(tn)}D(T) = \{d(t_1), d(t_2), \ldots, d(t_n)\}D(T)={d(t1),d(t2),…,d(tn)}

These images or GIFs are then displayed in sequence to convey the spoken message in sign language.



Phase 2: Image to Text/Speech Conversion

1. Image Collection and Preprocessing:
In this phase, the system collects images that need to be interpreted. Each image is processed using the MediaPipe framework to extract key features. MediaPipe provides tools for pose estimation, face detection, and hand tracking, which are essential for interpreting sign language gestures.
Let III denote the image, and F(I)\mathbf{F}(I)F(I) represent the feature extraction process using MediaPipe, which produces a feature vector v\mathbf{v}v:

v=F(I)\mathbf{v} = \mathbf{F}(I)v=F(I)

2. Multilayer Perceptron (MLP) Model:
The extracted feature vectors are used to train a Multilayer Perceptron (MLP) model. An MLP is a feedforward artificial neural network that consists of an input layer, one or more hidden layers, and an output layer. The MLP is trained using supervised learning with labeled data.
The feature vector v\mathbf{v}v is the input to the MLP. Let the input layer be represented by x=v\mathbf{x} = \mathbf{v}x=v. The MLP consists of LLL layers with weights W(l)\mathbf{W}^{(l)}W(l) and biases b(l)\mathbf{b}^{(l)}b(l) for layer lll.

3. Training and Validation:
The MLP model is trained using a dataset of labeled images. The training objective is to minimize the cross-entropy loss between the predicted probability distribution y\mathbf{y}y and the true labels. The cross-entropy loss is given by:
L=−∑i=1Cyilog⁡(y^i)\mathcal{L} = - \sum_{i=1}^{C} y_i \log(\hat{y}_i)L=−∑i=1Cyilog(y^i)
Where CCC is the number of classes, yiy_iyi is the true label, and y^i\hat{y}_iy^i is the predicted probability for class iii.
The model parameters W\mathbf{W}W and bmathbf{b}b are optimized using gradient descent and backpropagation algorithms.

Result and Discussion:
The framework is intended to perceive the hand motions made by the dumb people. The proposed framework is basic and the subject isn't needed to wear any gloves or any electromechanical gadget. The speech is eared by the system and is converted into alphabets. And accordingly the hand gesture is made visible for deaf peoples.eg: Riya is passed as a speech to system. The system coverts it to alphabet like R, I, Y, A where system matches the signs in database and provides output.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2024

DOI: 10.17148/IJARCCE.2024.13848

The proposed web application seeks to develop a translating mechanism or automation that includes a parser element that converts the incoming speech data or English text to a phrase structure grammar representation, which is then used by another module that contains Indi Sign language grammatical format. This is accomplished through the means of removing stop-words from the reordered input format. Because Indian sign language does not provide word inflections, stemming and lemmatization are used to turn words into their root form. Following sentence filtration, all words are tested against the words in the database, which is represented as a dictionary comprising video representations of each word. If the words are missing from the database, the algorithm will then look for its related synonym and replace it with that term.In many ways, the proposed system is more innovative and efficient than existing systems, because Existing methods can only convert words directly into Indi sign language, and they were not as efficient as this system, whereas this in the actual world, the system tries to translate these phrases into Indian sign language grammatical order. Because this is a web-based programmed, it is straightforward to access and use. This technology is platform agnostic and more versatile to use, and it transforms phrases to sign language in real time

IV. CONCLUSION

The communication system developed in this project effectively bridges the gap between deaf and hearing individuals by leveraging advanced technologies in speech recognition, image processing, and machine learning. The first phase of the system converts audio messages into Indian Sign Language (ISL), enabling real-time translation of spoken language into visual sign representations. This phase utilizes a robust speech-to-text engine that accurately transcribes audio input and maps the text to predefined ISL images or GIFs. The second phase enhances the system's capability by converting images into text or speech using the MediaPipe framework for feature extraction and a Multilayer Perceptron (MLP) model trained to achieve a 90% accuracy rate. This allows for the interpretation of visual cues and gestures, making visual information accessible to the deaf community through textual or auditory output. Overall, the system demonstrates significant potential in facilitating inclusive communication, enhancing the interaction between deaf and hearing individuals in various scenarios. By addressing both auditory and visual communication barriers, this project contributes to the development of more inclusive and accessible communication technologies for the deaf community, fostering better understanding and cooperation in diverse social contexts.

REFERENCES

[1]. "Machine learning model for sign language interpretation using webcam images," 2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA), 2014, pp. 317-321, doi: 10.1109/CSCITA.2014.6839279.

[2]. A translator for American sign language to text and speech," 2016 IEEE 5th Global Conference on Consumer Electronics, 2016, pp. 1-2, doi: 10.1109/GCCE.2016.7800427.

[3]. Morphological Analysis of Speech Translation into Indonesian Sign Language System (SIBI) on Android Platform," 2019 International Conference on Advanced Computer Science and information Systems (ICACSIS), 2019, pp. 205-210, doi: 10.1109/ICACSIS47736.2019.8980000

[4]. An Improved Sign Language Translation Model with Explainable Adaptations for Processing Long Sign Sentences Computational Intelligence and Neuroscience Volume 2020, Article ID 8816125, 11 pages https://doi.org/10.1155/2020/8816125

[5]. Convolutional Neural Network based Bidirectional Sign Language Translation System," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2020, pp. 769-775, doi: 10.1109/ICSSIT48917.2020.9214272

[6]. Application for Iraqi sign language translation on Android system", International Journal of Electrical and Computer Engineering (IJECE)Vol.10, No.5, October2020, pp. 5227~5234ISSN: 2088-8708, pp5227-5234

[7]. Electronic Glove: A Teaching AID for the Hearing Impaired", International Journal of Electrical and Computer Engineering (IJECE)Vol.8, No.4,