Audio To Sign Language Translation

Neeta Verma, Aditya Jindal, Adit Agarwal, Kritika Jain, Shikha Singh

Inderprastha Engineering College, Sahibabad, Uttar Pradesh
E-mail: Neeta.verma@ipec.org.in, jindaladitya186@gmail.com, adit0612@gmail.com, kritikajain1211@gmail.com, singhshikha2302@gmail.com

Abstract

Each country has its own sign language based on the local language. Unfortunately, many people, even those with normal hearing, often overlook or avoid communicating with individuals who are deaf or hard of hearing. This creates a significant barrier for those who cannot hear, making it even more difficult for them to engage in conversations .Individuals with speech and hearing impairments often face challenges in communication because not everyone knows sign language.

The aim of this research is to create a system that assists those with speech and hearing difficulties by converting spoken language into Indian Sign Language (ISL). Since learning sign language can be a challenge, this study proposes an innovative approach that combines speech recognition and image processing. The development of sign languages has greatly facilitated communication, particularly for the deaf and hard of hearing, ensuring better interaction and fostering versatile communication for all parties involved.

Introduction

Sign language serves as a vital tool worldwide to bridge the communication gap for individuals with hearing or speech impairments, who rely on sign language for daily interactions. However, efficient models for converting text into Indian Sign Language (ISL) are still lacking. There is also a shortage of effective audio-visual support for oral communication in many regions. While significant advancements have been made in detecting sign languages in other countries using computer systems, research on Indian Sign Language remains limited. Most of the research and systems developed so far have concentrated on American Sign Language (ASL) or British Sign Language, with very few focusing on ISL. The existing systems typically rely on models like direct translation, statistical machine translation, or transfer-based architectures.

This system specifically targets Indian Sign Language, focusing on gathering relevant video representations for each word. If a word is not available in the ISL dictionary, alternative words with similar meanings will be used, accounting for word duplication and their respective parts of speech. Each language has its unique set of rules and grammar, making translation between spoken languages a challenging task. When one language is a spoken language and the other is a sign language, the complexity of translation increases significantly. This system is primarily designed for individuals with hearing impairments, as they often struggle to communicate in everyday situations due to the limited understanding of sign language among others. It aims to help reduce communication barriers and assist those interested in learning sign language.

Sign language (SL) is a natural, visual-spatial form of communication that integrates facial expressions, hand shapes, arm positioning, and movements of the upper body to convey meaning in three-dimensional space. It was developed by people in India who are deaf or hard of hearing. Just as there are various spoken languages worldwide, such as Urdu, French, and English, there are also different sign languages used by hearing-impaired individuals across the globe. These sign languages vary based on the specific communities and regions, with each group adapting their own forms of communication.

In India, 6.3% of the population, or 63 million people, have substantial hearing loss, according to the 2011 census. Among them, between 76 and 89 percent of the Indian Deaf are illiterate in spoken or written language.

Communication is challenging for deaf people in settings like banks, trains, and hospitals because of their disabilities. It is necessary to develop a system that will translate text into Indian Sign Language and backward to improve their ability to communicate with the outside world. These systems will raise the community's standard of living. Although sign languages have received less research than spoken languages, there is still much to learn about them. is the idea that individuals or groups of people show up at a location to participate in an event that was previously organized

Literature review

In a 2021 publication, Kajal Jadhav et al. [1] presented a project aimed at assisting individuals with pain or silent speech. Their work focuses on enhancing communication for those who are deaf or mute through the use of technology. In 2021, Prof. Abhishek Mehta et al. [2] proposed a reliable system that can automatically convert spoken words into text and translate written text into animated sign language. The primary aim of Harshita Mishra, Mansi Sharma, Muskan Ali Qureshi, and Shivani Chaudhary[3] is to bridge the communication gap between hearing and deaf individuals, facilitating an easy and efficient way for those using sign language to communicate. Onkar Bidkar et al. (2021)[4] introduced a Sign

Language Recognition system, built with Python, which recognizes hand gestures. The design was inspired by the need to support those who face challenges in linguistic communication, particularly those with hearing loss. Purushottam Sharma et al. (2022)[6] focused on an audio-to-Indian Sign Language translation system to bridge the gap between spoken and sign languages, particularly for individuals with hearing impairments. Vishwas S et al. (2018)[7] emphasize that not everyone can understand sign language when conversing with deaf or speech-impaired individuals. This creates a constant need for effective sign language communication, and without an interpreter, communication becomes difficult. According to Jose, M.J., Priyadharshni, V., Anand, M.S., Kumaresan, A., and Kumar, N.M. (2013)[8], sign language uses visually conveyed signs, such as hand shapes, movements, body orientation, and facial expressions, to communicate meaning. P. Sonawane, K. Shah,
P. Patel, S. Shah, and J. Shah (2013)[9] highlight that hearing and speech impairments impact a significant portion of the Indian population, and Indian Sign Language serves as their primary mode of communication. In the work of Bharti, Ritika, Yadav, Sarthak, Gupta, Sourav, and B, Rajitha [10], the importance of systems that support communication for the deaf community is emphasized, underscoring the necessity of such systems for enhancing interaction. Hemang Monga, S. Darunde, N. Wahie, and
S. Mali [11] note that Indian Sign Language is one of the most widely adopted communication forms for those with hearing and speech impairments. Ankita Harkude et al. (2020)[12] point out that deaf individuals often miss out on experiences that hearing people take for granted, such as casual conversations, video games, seminars, or virtual conferences. In their 2020 study, Renjitha R. et al. [13] discuss the challenges faced by individuals who cannot fully perceive or respond to their surroundings, particularly those with hearing and speech impairments. In their 2021 study, Minh Le et al. [14] proposed a design for a speech-to-sign language converter aimed at assisting individuals who are deaf or hard of hearing. The device is costeffective, energy-efficient, and capable of operating offline.

Methodology

The methodologies used in this research makes the use of speech recognition, Natural Language Processing(NLP), Automatic Speech Recognition(ASR) and Sign Lnaguage generation along with evaluation metrics.

Speech recognition, also referred to as Automatic Speech Recognition (ASR), is the process of converting spoken language into a text format. This is achieved through algorithms that are implemented in computer programs.

Having made a series of development starting from the 70s and 80s, we have far reached to wider technologies that can easily convert speech to text.

The standard approach to large vocabulary continuous speech recognition (MA Anusuya, SK Katti) is to assume a simple probabilistic model of speech production whereby a specified word sequence, W, produces an acoustic observation sequence Y, with probability P(W,Y). The goal is then to decode the word string, based on the acoustic observation sequence, so that the decoded string has the maximum a posteriori (MAP) probability.

P(W/A)= arg maxwW P(W/A)................ (1)

Using Bayes rule, equation (1) can be written as P(W/A)=P(A/W)P(W) (2)

Since P(A) is independent of W, the MAP decoding rule of equation(1) is W=argmaxw P(A/W)P(W) (3)

The first term in equation (3) P(A/W), is generally called the acoustic model, as it estimates the probability of a sequence of acoustic observations, conditioned on the word string. Hence P(A/W) is computed. For large vocabulary speech recognition systems, it is necessary to build statistical models for sub word speech units, build up word models from these sub word speech unit models (using a lexicon to describe the composition of words), and then postulate word sequences and evaluate the acoustic model probabilities via standard concatenation methods. The second term in equation
(3) P(W), is called the language model.

In 1950, Alan Turing published a paper outlining a test to determine if a machine could be considered "thinking." He proposed that if a machine could engage in a conversation through a teleprinter, and its responses were indistinguishable from those of a human, it could be regarded as capable of thought. A few years later, in 1952, the Hodgkin-Huxley model demonstrated how the brain utilizes neurons to create an electrical network. These developments contributed to the foundation of artificial intelligence (AI), natural language processing (NLP), and the advancement of computing technology.

Empirical results

Table 1: Evaluation metrics involved in Natural Language processing

Metric	Definition	Pros	Cons	Best suited for
Accuracy	the quality or state of being correct or precise.	simple and easy	Not sufficient enough	Classification
Precision	Precision refers to the proportion of correctly predicted positive instances	Useful for measuring the model's ability to identify true positives.	May be sensitive to class imbalance, favoring models that predict the majority class.	Classification tasks
Recall	recall measures the proportion of true positive instances identified	Useful for measuring the model's ability to capture all relevant positive instances.	May be sensitive to class imbalance, favoring models that predict all instances as positive.	Classification tasks
F1 score	The F1 score is a commonly used evaluation metric that combines precision and recall into a single value	Provides a single metric that considers both precision and recall.	sensitive to class imbalance, can be influenced by the relative weights of precision and recall.	Classification tasks
Perplexity	It measures how well a language model predicts a given sequence of words or a corpus.	simple and easy to interpret, can compare models of different sizes.	Doesn't directly measure the quality of the generated text, sensitive to rare words and ngrams.	Language models

Bleu Score

The acronym BLEU refers to a "Bilingual Evaluation Understudy," and it's a statistic for measuring the accuracy of machine translations compared to human translators. IBM's version of BLEU NLP is a popular tool for analyzing data and gauging the quality of machine translations.

The following is the formula for calculating the code BLEU score:

- BLEU = brevity_penalty * exp(sum(w_n * log(p_n))) where:

1. The brevity penalty is explained already.
2. w_n is the weight applied to the n-gram accuracy score. Weights are often set to 1/n, where n refers to the number of n-gram sizes utilized
3. p_n represents the precision rating for the n-gram size.

Word Error Rate

The transcription generated by a system can generate incorrect or misunderstood words, especially if the audio quality is poor. In short, WER is calculated by dividing the number of errors (counting the number of insertions, deletions, and substitutions) between reference and hypothesis transcriptions by the total numbers of words in the reference transcription.

The math formula for calculating WER is as follows:

$$WER = \frac{S_w + D_w + I_w}{N_w} \quad \text{…………(4)}$$

Where:

- S (Substitution): is the total of words replaced in the hypothesis transcription. For example, if the reference transcription is “cat” and the recognized word is “bat”, then there is one substitution error.
- I (Insertion): is the total of words added in the hypothesis transcription. For example, if the reference transcription is “the book is on the table” and the recognized phrase is “the book is on table”, then there is one insertion error.
- D (Deletions): is the total of words omitted in the hypothesis transcription. For example, if the reference transcription is “the book is on the table” and the recognized phrase is “the ugly book is on the table”, then there is one deletion error.
- N (Total Words): is the total number of words in the reference transcription.

Conclusion

We aim to develop a model that enables individuals with disabilities to communicate more clearly, helping them integrate with the wider world without barriers. Our proposed system will efficiently transform spoken input into an animated form. As the Indian Sign Language (ISL) dictionary expands, there will be opportunities for further enhancements. The current ISL vocabulary is limited, and by incorporating more words into the dictionary, we can improve the model’s effectiveness and increase its coverage.

As per our study, one of the best review paper for “Audio to sign language translation” system is https://www.jetir.org/papers/JETIR2303011.pdf. It highlights the importance of Indian Sign Language (ISL) as a key communication tool for individuals with hearing and speech impairments in India. Since written materials may not always provide an effective or accessible means of communication, ISL offers a more intuitive and contextual approach to conveying words, emotions, and concepts through hand gestures, lip movements, and facial expressions. This system aims to bridge communication gaps by enabling users to convert spoken language into ISL animations, allowing for clearer expression and better integration of hearing- and speech-impaired individuals into society. The proposed model also focuses on improving the functionality of a "Speech to ISL" translator that would interpret spoken language and convert it into visual representations of sign language.

References

[1] Kajal Jadhav, Shubham Gangdhar ,Viraj Ghanekar , SPEECH TO ISL (INDIAN SIGN LANGUAGE) TRANSLATOR, International Research Journal of Engineering and Technology (IRJET), Volume: 08 Issue: 04,Apr 2021.

[2] Abhishek Mehta, Dr. Kamini Solanki, Trupti Rathod, 2021, Automatic Translate Real-Time Voice to Sign Language Conversion for Deaf and Dumb People, INTERNATIONAL JOURNAL OF ENGINEERING RE-
SEARCH TECHNOLOGY (IJERT) ICRADL – 2021 (Volume 09 – Issue 05).

[3] Harshita Mishra1, Mansi Sharma2, Muskan Ali Qureshi3, Shivani Chaud- hary4 1, 2, 3, 4Department of CSE, Krishna Engineering College, GhaziabadISSN: 2321-9653; IC Value:

[4] 45.98; SJ Impact Factor: 7.538 Volume 10 Issue V May 2022.

[5] Onkar Bidkar,, Kunal Chanda, Soma Mitra, “A survey paper on Speech/Audio to Sign Language Translator for Deaf People”, Interna- tional Conference on Information Science (ICIS) 2021.

[6] Ashmi Katariya, “Voice to Indian Sign Language Conversion for Hearing Impaired People” in COEN 296B-Natural Language Processing, 2020.

[7] Sharma, Purushottam Tulsian, Devesh Verma, Chaman Sharma, Prat- ibha

Nancy,Nancy. (2022). Translating Speech to Indian Sign Lan- guage Using Natural Language Processing. Future Internet. 14. 253. 10.3390/fi14090253.

[8] Vishwas, Gowda, H., Chandra, V. (2018). SIGN LANGUAGE TRANS-
LATOR USING MACHINE LEARNING. International Journal of Ap- plied Engineering Research ISSN 09734562 Volume 13, Number 4 (2018) Spl.

[9] Jose, M.J., Priyadharshni, V., Anand, M.S., Kumaresan, A., Kumar,

[10] N.M. (2013). Indian Sign Language (ISL) Translation System For Sign Language Learning. International journal of innovative research and development.

[11] P. Sonawane, K. Shah, P. Patel, S. Shah and J. Shah, ”Speech To Indian Sign Language

[12] (ISL) Translation System,” 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2021, pp. 92-96, doi:10.1109/ICCCIS51004.2021.9397097. Bharti, Ritika and Yadav, Sarthak and Gupta, Sourav and B, Rajitha, Automated Speech to Sign language Conversion using Google API and NLP (June 2019). Proceedings of the International Conference on Advances in Electronics, Electrical Computational Intelligence (ICAEEC) 2019.

[13] Hemang Monga, S. Darunde, N. Wahie, and S. Mali, ”Speech to Indian Sign Language Translator,” 2021 Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0 2021, pp. 1-6, DOI: 10.1109/IBSSC47189.2019.8973041.