

# Speech to Indian Sign Language Translator

Hemang Monga<sup>a,1</sup>, Jatin Bhutani<sup>a</sup>, Muskan Ahuja<sup>a</sup>, Nikita Maid<sup>a</sup> and Himangi Pande<sup>a</sup>

<sup>a</sup> School of Computer Engineering and Technology, MIT World Peace University, Pune, India

**Abstract.** Indian Sign Language is one of the most important and widely used forms of communication for people with speaking and hearing impairments. Many people or communities have attempted to create systems that read the sign language symbols and convert the same to text, but text or audio to sign language is still infrequent. This project mainly focuses on developing a translating system consisting of many modules that take English audio and convert the input to English text, which is further parsed to structure grammar representation on which grammar rules of Indian Sign Language are applied. Stop words are removed from the reordered sentence. Since the Indian Sign Language does not support conjugation in words, stemming and lemmatization will transform the provided word into its root or original word. Then all the individual words are checked in a dictionary holding videos of each word. If the system does not find words in the dictionary, then the most suitable synonym replaces them. The system proposed by us is inventive as the current systems are bound to direct conversion of words into Indian Sign Language on-the-other-hand our system aims to convert the sentences in Indian Sign Language grammar and effectively display it to the user.

**Keywords.** Speech to text, Sign Language Translation, NLP.

## 1. Introduction

Sign language is used worldwide to reduce the communication gap for people with hearing or speech impairment which depends primarily on sign language for everyday communication; currently, we do not have efficient models that convert text to Indian Sign Language. Proper and effective audio-visual support for oral communication is still lacking. There has been significant progress in detecting sign languages in other countries using computers, but very little research has been done to detect ISL. Most of the work done on this topic has focused more on American Sign Language (ASL) or British Sign Language, but for Indian Sign Language, there are hardly any systems developed. The underlying architecture for most of the systems are based on: Direct Translation, Statistical Machine Translation, Transfer-based Architecture The system focuses on Indian Sign Language. For each word, its relevant video will be gathered. For words that are not present in this dictionary, they would be exchanged by other words having similar meanings, considering the duplication of words and their parts of speech. Each language has its own rules and grammar, and therefore translation between two spoken languages is a tedious task. Furthermore, when one language is a spoken language while the other is a sign language, then the translation complexity increases.

<sup>1</sup>Nikita Maid, School of Computer Engineering and Technology, MIT World Peace University, Pune, India  
Email: nikitamaid23@gmail.com.

The users of this system are mainly individuals with hearing impairment, as communication for them in everyday places is very problematic because not everyone understands sign language. Therefore, no conversation can occur between a person who speaks and a person with a hearing impairment. Thus, this system is for people who want to learn sign language or reduce communication gaps.

## **2. Literature Survey**

In India, there is far less research done on sign language and especially Indian Sign Language. Each state in India has its regional language, and therefore India is also known as a multilingual nation. It is also important to note that each state and its language also have its sign language. As India was once a British Empire colony, there are many similarities between the two nations' sign languages.

The ISLRTC (Indian Sign Language Research & Training Center) seeks to balance the principles of program and application of ISL. A widespread practice for deaf people in India is to learn ASL since it is very easily readable and then later learn sign language. 98 percent of 1.1 million deaf people in India are illiterate. Consistent with oralist philosophy, deaf schools try to intervene at an early stage with hearing aids, but this, in particular, does not apply to the community. There have been multiple attempts to provide a new solution to this root cause by researchers worldwide. The techniques for producing gloss in this system are aligned with the Indian sign language grammar. The TSA's data collection method has been carefully analyzed to determine the strategies used in this project.

The primary concern with the transition from Spoken English to Sign Language is the formation of phrases that help translate by giving the correct ASL grammar phrases. This project implements this by using the rule-based technique for building a set of phrases. Each sentence is converted into tokens, and then for each token, we identify the part of speech with which it binds, after which specific actions are taken (Each word is transformed to its root word, and auxiliary verbs are removed). All the words from Spoken English, which are not a part of ASL Gloss, are removed. Then proper nouns (names of countries, cities, people, etc.) are outputted in a different format to display sign language, and other words are processed according to various rules such as verb correction, etc. Each proper noun is stored as alphabets separated by a hyphen, and each alphabet is shown as a sign on display. For Example, Mumbai (name of the city) is stored as M-U-M-B-A-I, and India, in a similar manner, will be stored as I-N-D-I-A, corresponding to the use of conventional sign language. The system is rigid and can only lead to destructive behavior if there is a human error. The human error can be seen while creating an ASL gloss or while storing datum for each sign manually. We can see an increase in efficiency and reduction in time by using a phrase-based algorithm to eliminate repeating words, but it still takes much time in rendering, and the other issue is portability from ASL to ISL [1].

Proposed a model that uses the Hidden Markov Model to convert speech to English text language. The generated text generates Indian Sign Language Gloss using pre-processing and Wordnet, an American Linguistic Database. Pre-processing methods like punctuation removal, selective stop word removal, tokenization, and stems are used. The similarity measurement is done using Leacock Chodorow similarity. This system has a 68% accuracy however faces an issue when word similarity in Wordnet is more remarkable with another word that does not suit the meaning of the statement or

finds a noun for which no other synonym can be found in the ISL dictionary. It has no provision to remove repetition in the ISL gloss, which leads to redundancy [2].

They are considering the most commonly used NLTK libraries and creating a robust and impactful application. They are first converting audio input into text using Google text to speech API. The proposed system performs tokenization and uses rules to convert English to ISL gloss followed by lemmatization and stemming. Then the resultant output is passed to Hamburg Sign Language Notation System, a database retrieval tool. Further, it is passed to the SIGML representation tool to represent the sign language in the form of avatars. Introducing a communication system for people with disabilities for effective communication. Converts speech (English) into three-dimensional avatar animation displays Hindi (Indian) language symbols instead of GIFs, images, or videos for effective memory management [3].

IBM Model 1: The order of words in the target text may not be a correct sentence in syntax and semantics. In order to make a correct sentence from these frames, the arrangement of lexemes is required. This is done using IBM Model -1. For example, "Passengers are" is converted to "passenger," "requested to" is converted to "please," etc.

IBM Model 2: In the IBM Model-2, the design possibilities of glosses depend on their positions as the words in the sentence reside somewhere according to the syntax. Nevertheless, instead of following the target grammar, we will consider the conditional probability of the word in a sentence concerning other words.

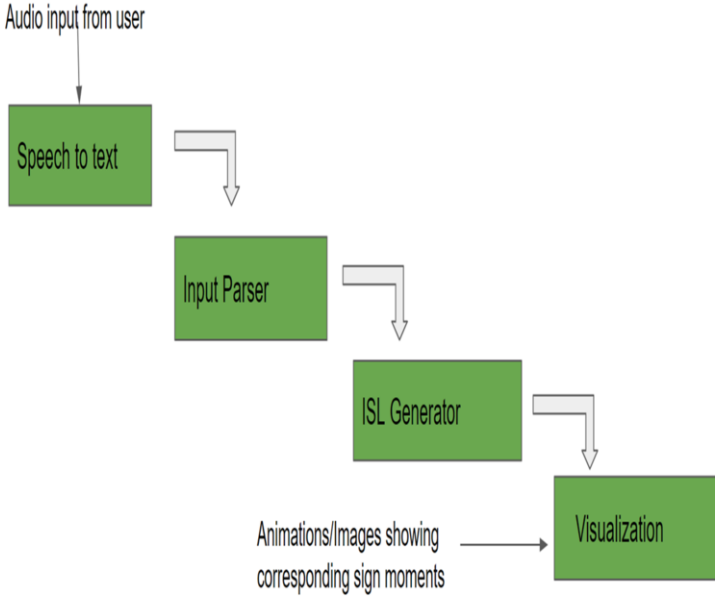
IBM Model 3: The inconsistent name of the targeted sentence will be associated with the invalid source name, e.g., Sometimes, when words are changed using the above models, the whole sentence may not have a meaning, so model 3 helps to arrange the words to make a logical sentence. In the first step, they participate in a translation model where the words of the introductory sentences are translated into corresponding words of the desired language, and the order of these words is converted into a valid target sentence. The rules and grammar of both languages are ignored. The work can also be improved by introducing a sentence-based translation machine as word-based translation can be done successfully in sentences with a few words. The work can be improved by using a sentence-based machine translation process. This research presents an architectural framework of Pakistan sign language recognition tool developed to convert given PSL (Pakistan Sign Language) into voice and text. Using the given tool, sign language can be converted into text using the CNN model and TensorFlow. The system accepts input in the form of an image from the camera and translates the symbol into text and voice. The currently developed translation tool can only translate a few Pakistani sign languages into text and speech because it is an example, and there is a need to build an app or portable device to overcome the barrier of communication while using this user-friendly translation tool. In developing the portable version of the tool, the system on chip devices (SoC) is considered.

### **3. System Analysis Proposed Architecture**

Algorithmic Design:

1. Pyaudio is used to obtain user speech and convert it to text.
2. The sentence is restructured using stop words removal and lemmatization.

- 3. Final ISL gloss is generated by converting text in context free grammar.
- 4. YouTube IFrame API is used to display the output in the web application.



**Figure 1.** Low-Level Design

Speech to text phase: Input to the module can be given by an inbuilt microphone, an external source on any Personal Digital Assistant (PDA). Audio given by the user will be recognized using PyAudio, which is a Python binding for PyAudio in Figure 1. We can record and play audio on various platforms including, Microsoft Windows, Apple Mac, and Linux. The module uses Google Cloud Speech for audio to text. The output is given as an English text string. The module will be developed in python script

**Table 1.** POS Tags used for Stopwords removal

POS Tag	POS Tag full-form
CC	Coordinating Conjunction
NNS	Noun, Plural
RP	Particles
NNPS	Proper Noun, plural
SYM	Symbols
DT	Determiners

**Input Parser:** The input section is made into tokens of sentences. Using Machine Learning and Natural Language Processing Tools, each sentence is marked. The issuance of this section is a list of tokens in each line of text with appropriate punctuation marks in Table 1.

**ISL Generator:** This module aims to convert the input text with the grammar of English into text with the grammar of ISL. After the tokenization, we will first POS tag defined as the process where each word is matched with its part of speech. Also commonly known as POS tagging. Part of Speech includes adjectives, conjunctions, verbs, adverbs, nouns, pronouns, and sub-categories. Once the POS tagging is completed, we can go ahead and eliminate all the stop words. Stop words in a simple language are words which does not add much meaning to the sentence. Since they do not add much meaning, they can be removed, and the sentence's meaning will be unchanged.

Furthermore, after removing them, we will use lemmatization to eliminate the inflectional ending and return the root form of the word called the lemma. If the token saw is to be lemmatized, it will return either saw or see depending on whether the token used a noun or verb in the sentence.

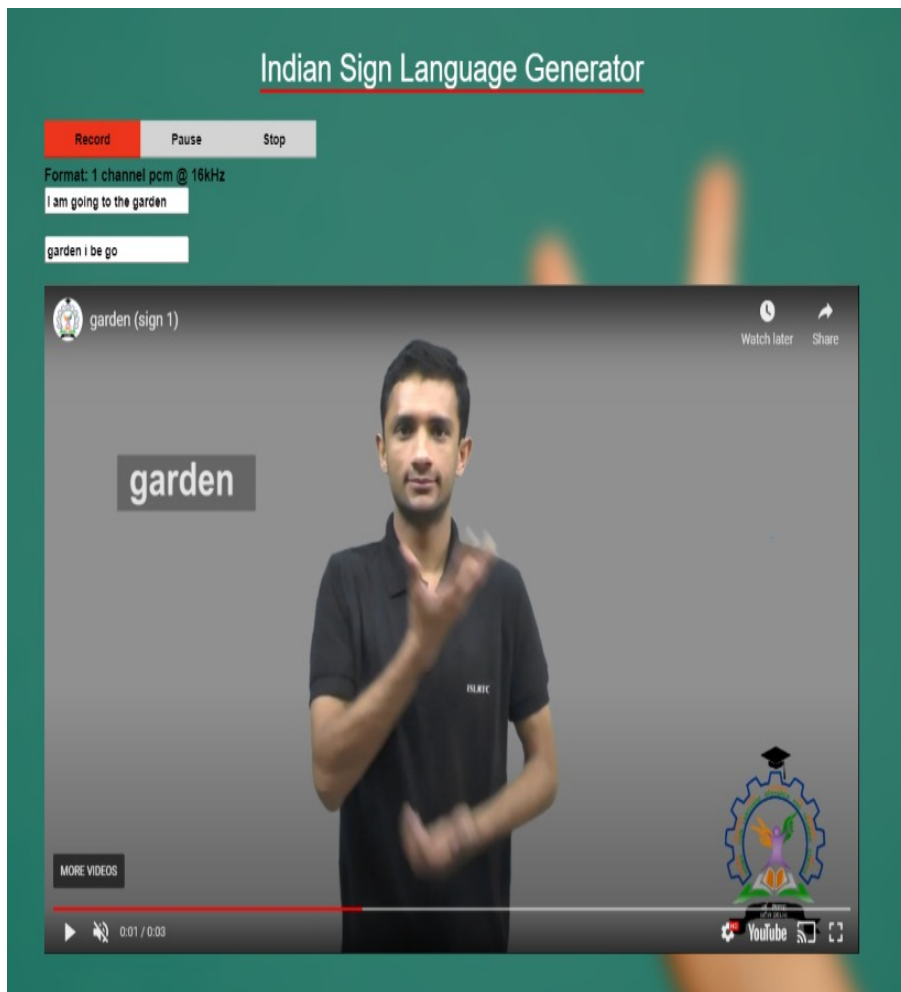
**Visualization:** Input to this module is the ISL text string. In this module, generators map each token of text to a database to fetch the link. For every word present in the database, there is a corresponding file present, and for the words that are not there, we will find their synonym, and if the synonym is also not there, we will display each alphabet individually.

#### **4. Result and analysis**

For a given English audio, the system aims at generating its equivalent sign language depiction. Our system generates these outputs in the following ways:

**Video generation** - Output from the ISL conversion phase of the input sentence is passed to the video generation phase, wherein each of the words in the sentence is looked up in the database for its corresponding video file. Then these files are all concatenated to produce a more structured, informative, and easy-to-understand visual depiction of Indian Sign Language.

For example, we provide an audio input "I am going to the garden" the input parser converted this audio into text and tokenized this sentence as "I", "am", "going", "to", "the", "garden", then ISL generator converted this tokenized sentence into ISL Gloss "garden I be go" then video output for the ISL gloss is played by extracting the video of the words from the database and playing it using Youtube IFrame API.



**Figure 2.** Result output

In comparison to existing approaches, we are using a YouTube player in our project for visualization, which makes us feel like we are communicating with a person instead of an avatar or animation. Figure 2. Our output gives results with the help of ISL Gloss sentence, which increases its understanding and reduces sentence complexity for generating signs.

## 5. Conclusion

Here, we have attempted to create a model that will allow people with disabilities to express themselves distinctly, which will help them blend with the rest of the world without any difficulty. Our proposed model will successfully convert the given input audio into an animation. Many improvements along this route can be made as and when the ISL Dictionary grows. The words in the ISL are small, so many improvements can be made by adding new words to their dictionary to increase their breadth. In addition, text-and-speech integration can be done on a project to enable

better communication techniques that will allow users to convert Text into Indian Sign Language by hand-input text.

## References

- [1] F. Shaikh, S. Darunde, N. Wahie, and S. Mali, "Sign Language Translation System for Railway Station Announcements," 2019 IEEE Bombay Section Signature Conference (IBSSC), 2019, pp. 1-6, DOI: 10.1109/IBSSC47189.2019.8973041.
- [2] K. Saija, S. Sangeetha and V. Shah, "WordNet Based Sign Language Machine Translation: from English Voice to ISL Gloss," 2019 IEEE 16th India Council International Conference (INDICON), 2019, pp. 1-4, DOI: 10.1109/INDICON47234.2019.9029074.
- [3] B. D. Patel, H. B. Patel, M. A. Khanvilkar, N. R. Patel and T. Akilan, "ES2ISL: An Advancement in Speech to Sign Language Translation using 3D Avatar Animator," 2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), 2020, pp. 1-5, DOI: 10.1109/CCECE47787.2020.9255783.
- [4] Mishra Gouri Sankar, Ashok Kumar Sahoo, and Kiran Kumar Ravulakollu, "Word-based statistical machine translation from English text to Indian sign language." *ARPJN Journal of Engineering and Applied Sciences* 12.2 (2017)".
- [5] I. Abbas A, Sarfraz S. Developing a Prototype to Translate Text and Speech to Pakistan Sign Language With Bilingual Subtitles: A Framework. *Journal of Educational Technology Systems*. 2018;47(2):248-266. doi:10.1177/0047239518794168
- [6] Advaith Sridhar, Rohith Gandhi Ganesan, Pratyush Kumar, and Mitesh Khapra. 2020. INCLUDE: A Large Scale Dataset for Indian Sign Language Recognition. In *Proceedings of the 28th ACM International Conference on Multimedia*. Association for Computing Machinery, New York, NY, USA, 1366–1375. DOI:<https://doi.org/10.1145/3394171.3413528>