Speech to Indian Sign Language Translator
Alisha Kulkarni 1*, Archith Vinod Kariyal2, Dhanush V3, Paras Nath Singh4

1,2,3,4 Dept. of CSE, CMRIT, Bangalore-560037, India
*Corresponding author. Email: alis17is@cmrit.ac.in

ABSTRACT
A Sign Language is a non-verbal communication system in which people communicate by visually transferring sign patterns to express their meaning. A large section of the hearing and speech impaired in India uses Indian Sign Language (ISL) as their mode of communication. Recognition of sign language was significant not only from a technical standpoint but also in terms of societal influence. An automatic speech-to-ISL translation system could help the hearing-impaired access more information and services due to the inherent limitations of communicating through written texts. Furthermore, the system can be utilised as an educational tool to learn ISL in addition to improving information access. This paper intends to bridge the gap by providing a system that anybody can use without having to learn ISL.

Keywords: Computational Linguistics, Lemmatizing, Natural Language Processing, Python, Sign Language, Web-scraping.

1. INTRODUCTION

Because it incorporates expressions and gestures, sign language provides greater context and understanding into the topic being communicated. There are around 7139 known living languages in the world, which are divided into 142 language families. Deaf community sign language, which belongs to the Sign Language family, is one of the 142 families of sign languages used by hearing and speech impaired people to communicate. Depending on the region of the world in which the paper is used, this family has around 128 sign languages.

Out of nearly 7.9 billion people on earth, nearly 1.57 billion people suffer from hearing loss. This number accounts for roughly 20.3% of the global population. Close to 6 million people in India use Indian Sign Language for communication making it the 151st most “spoken” language in the world.

Over the years, there has been very little improvement in the access to information for this population. In India, it is still a constant struggle for the deaf-mute people to gain access to education, employment and communications due to the presence of only a small handful of schools with sign language interpreters. Most of these schools are located in big cities, making the access to information even sparser to the rural regions of the country. Therefore, it becomes a major concern to pave the path for these people who would not only minimize the gap in the hearing-impaired population to sign language interpreter ratio, but also make them independent through a platform that ensures self-education and learning of sign language. This paper proposes a novel, easy-to-use and time efficient online platform for deaf-mute people that would serve as an effective mode of communication and learning for them.

Technology is rapidly changing and the way the world operates is improving. This necessitates the breaking of communication barriers for the hearing and speech impaired community. To this day, there are only about 250 certified sign language interpreters in India for a deaf-mute population between 1.8 million and 7 million.

Firstly, sign languages are not international. Many, but not all, countries have unique sign languages. There is no single convention followed for sign language across the globe and varies between regions. BSL (British Sign Language) and ASL (American Sign Language) are popular sign language conventions, used in Europe and America respectively. This proposal intends to target the Indian population.

Secondly, learning to read and write is very difficult for most people with hearing impaired. For those that can read and write, understanding the context of what is being spoken becomes difficult, especially in cases where non-verbal sounds or actions are involved. For these reasons,
most of the times, deaf-mute people resort to using sign language or lip-reading. Sign language is preferred by most because it involves hand movements, lip movements as well as expressions, providing more context and meaning. Lip movements and expressions are critical in sign language as they help in differentiating between similar looking hand signs.

The primary objective of this paper is to convert spoken English into ideal sign language using NLP (Natural Language Processing). This would enhance the communication capabilities of people hard of hearing. In India, there is a huge disparity in the population of the hearing and speech impaired, and the number of sign language interpreters. The system works towards developing a translator to aid a sign language user to convey ideas or easily find comprehensive information from a non-sign language user. The system under consideration provides a means for non-sign language users to communicate with a sign language user without necessitating the learning of ISL. It provides an efficient translator for ISL, a field that has had very little development in recent times. The system would also act like a learning tool for those who wish to learn ISL.

With this paper, the intention is to create a bridge of communication for the hearing and speech impaired community of India by converting English speech to Indian Sign Language using NLP (Natural Language Processing) techniques. The proposed system is an online platform that takes speech as an input and displays a stream of videos depicting the corresponding sign language as the output. The system would not only improve information access to this community, but also act like an educational tool for anyone trying to learn ISL.

2. RELATED WORKS

Distinct scholars have utilised a variety of ways to recognise different sign languages or hand gestures from different parts of the world. Some researchers used static hand motions, while others used video and real-time techniques. In this paper more research papers have been considered which deals with Indian Sign Language.

Dasgupta et al. [1] takes a sentence written in the English language as input and performs syntactic analysis and generates output as ISL structure.

TESSA [2] is a Speech-to-British Sign Language translation technology that attempts to let a deaf person communicate with a post office clerk. A formulaic grammar approach is used by the system. The translation is done using a phrase lookup database and a collection of predetermined phrases. However, because there are just a few sentences to use as templates, the conversation between the participants is limited. This makes TESSA a highly domain specific system.

The SignSynth project [3], developed by Grieve-Smith in 1998 and 1999, uses ASCII-Stokoe model to represent signs. It displays an animated output that is generated through conversion of ASCII-Stokoe to VRML (Virtual Reality Modelling Language). Any source language or communication should be translated to syntactically and semantically valid target language in the architecture by using appropriate transfer grammar rules satisfying both source and target.

The ViSiCAST translator proposed by Bangham et al. [4] is an example of a transfer architecture system. It translates English to BSL by using HPSG (Head-driven Phrase Structure Grammar) and a Prolog based freeware system ALE (Attribute Logic Engine). The phonology of the system is represented in HamNoSys.

The ASL (American Sign Language) Workbench, developed in 2001, is a text-to-ASL system [5] and [19] that uses LFG (Lexical Functional Grammar) to represent English f-structure in ASL. It is one of the most sophisticated machine translation systems and uses a phonological model based on Movement-Hold principle of ASL phonology. The TEAM project [6] is a text-to-ASL translation system that represents the source text in ASL using STAG (Synchronous Tree Adjoining Grammar). To recognize the correct word-sign pair, a multilingual lexicon is kept. The language module generates a written ASL gloss notation as a result of its output. The output of the synthesis module of the system generates an animated human model [20].

An example-based English-to-Dutch Sign Language machine translation system was proposed in 2005 by Morrissey [7]. It was built through a statistical approach using the IBM and the Hidden Markov Models for training the data. The system was able to evaluate small data sets only due to lack of well-annotated corpora.

For the railway reservation domain, INGIT, a Hindi-to-ISL machine translation system, has been developed by Kar et al. [8]. Through HamNoSys, the system takes input from a reservation clerk, converts it to ISL, and presents the result as animated ISL-gloss strings. Unlike TESSA which uses a purely formulaic approach, INGIT is based on a hybrid formulaic grammar approach. One point of appreciation is that Hindi and ISL have the same word order so, it did not require differentiation of structure.

In the Automated Speech to Sign Language Conversion using Google API and NLP approach by Bharati et al. [9] speech is first acquired as input. It is then converted to text using a Google API. The free text obtained contains infected parts that are removed using NLP in the next step. For every word or character, a matching operation is performed with the video database of the sign language. The matched videos are then retrieved from the database and are concatenated to form
a single video that depicts the complete string in sign language. The video is the final output of the system.

The Lancet [10] is a good and big study from 1990 to 2019 for Hearing loss prevalence and disability.

Text-to-ISL translation using LFG (Lexical Functional Grammar) by Goyal & Goyal [11] to represent the syntax in the prototype of text-to-sign language. Statistical machine translation models were not suitable for this system because there is no standard source of ISL corpus. [12], [13], [15], [16], and [18] have talked about Indian Sign Language. Authors of [17] have proposed Markov model for Hand-Movement Trajectory.

The mentioned state-of-the-art systems discuss the various approaches to developing a speech-to-ISL translator. It is important to address the pitfalls of the systems. The first approach discusses a method to take speech as an input and generate a stream of videos depicting the sign language. However, this system did not convert the English syntax into the ISL syntax through NLP techniques. In ISL, the syntax is one of the most important aspects to promote understanding in the hearing and speech impaired community. The second approach discusses the usage of LFG f-structure to depict the ISL syntax.

With due regard to all research implementation, computer application is missing. In this paper and project, the application is successful in executing textual and video outcomes for Indian Sign Language [21-23].

3. METHODS

In this system, text is taken as the input which may not be as intuitive and convenient as using speech as the input. Figure 1 depicts the major steps involved in the proposed system.

The primary modules of the system perform the operations of speech-to-text conversion, parsing, lemmatizing, filtering stop words, generating the ISL syntax, web scraping and merging the videos.

Speech-to-Text Conversion: This step involves using a built-in or external microphone to receive speech as an input from the user. The system uses the JavaScript library, Speech Recognition, to perform this operation.

Parsing: In this step, the text received from the first step is tokenized. Tokenization is the process of splitting a string into smaller units like words, sub words or characters. These small units are called tokens and they are the building blocks of natural language. The system tokenizes every sentence into its constituting words.

Further, the Stanford Parser is utilized to create a structure of phrases. Based on the grammatical structure, the Stanford Parser tags the parts-of-speech of the text, provides a CFG (Context-Free Grammar) representation of the parse tree and produces a type dependency representation for the sentence. Once the parse tree is generated, a new parse tree is initialized for the corresponding ISL syntax of the sentence. This new parse tree is constructed in accordance with the ISL grammar rules. According to the defined ISL grammar rules, the noun phrases and prepositional phrases are fixed. If a verb phrase is encountered, the system recursively checks the parse tree. This is because verb phrases may contain other verb phrases, noun phrases, prepositional phrases or complete sentences within them.

Stemming and Lemmatizing: In ISL, the root form of every word is used. It does not utilize words with suffixes or gerunds. Stemming reduces a word to its root form, that is, a form that contains no affixes or suffixes. The system uses stemming algorithms from the NLTK (Natural Language Toolkit) library to reduce tokens or words to their root forms.

Lemmatizing in NLP is the process involving the grouping together of inflected forms of a single word. This is done so that all forms of the word can be analyzed as a single root word, or the dictionary form of the word. The system uses the Word Net Lemmatize from the NLTK package. It uses a Word Net corpus along with the part-of-speech of every token to obtain the correct root form, also known as lemma of the word. For example, the words ‘playing’, ‘played’ and ‘plays’ are all reduces to the root word ‘play’ through the stemming and lemmatizing process. The lemmatizing step helps in improving the accuracy of the model.

Filtering Stop words: The ISL syntax is free of stop words. Sentences in ISL are formed using main words only. Linking words, like ‘am’, ‘are’, ‘is’, ‘be’, ‘to’ and ‘for’, and articles, like ‘the’, ‘a’ and ‘an’, are removed.

Generated ISL Syntax: The generated ISL syntax follows the subject-object-verb sentence pattern, unlike

Figure 1 Major steps in the proposed system

Filter stop words
Lemmatizing
Modify Parse Structure
Speech to Text (Speech Recognition library)
Speech Input
Generated ISL Syntax
Scraping the web for respective images and videos
Merge videos and images
Final Result

Text Conversion:
Filtering Stop words:
Stemming and Lemmatizing:
Speech-to-Text Conversion:
Parsing:
English, which follows the subject-verb-object pattern. It does not have any inflections, like gerunds, suffixes or any other forms of the word, like plurals or different tenses. Only the root form of every word is utilized. The generated ISL sentence is free from conjunctions, articles and linking verbs.

**Web Scrapping:** Web scraping is the process of utilizing bots to extract the underlying HTML code from any website, as well as the data held in a database. To obtain the sign language videos depicting every word, the system scrapes a web-based ISL dictionary. To automate this process, Selenium is utilized. Selenium is a portable web framework that provides tools to write functional tests without having to learn test scripting languages.

**Merging Videos:** Once the sign language videos are obtained, they are merged or concatenated into a single video. This video is essentially a stream of videos that represents the spoken sentence in ISL. The videos are concatenated with the help of MoviePy, a Python module that enables video editing. The final video is displayed as the result on the webpage [24-26].

4. **AUTHORS’ CONTRIBUTIONS**

**Flow of activity:** The first activity performed is that of taking the input from the user as audio. This is done by the speech or voice recognition module of the system. If the voice is not recognized, the user must record his/her voice again or enter the sentence to be translated in the text are provided on the landing page. If the speech is recognized, the speech is converted into its textual form. Further, the parts-of-speech are tagged for the sentence generated through the parsing process. Next, the grammatical structure of the sentence is analyzed and modified to the requirements of the ISL syntax. The next activity performed includes lemmatization and the removal of stop words. This is done because ISL does not possess stop words or inflected forms of words. Once the sentence is ISL syntax is ready, the sign language videos are fetched using a web scraper. The primary reason behind using a web scraper is to make the application as lightweight as possible. If the sign language video exists, the video is scraped. If it does not exist in the sign language dictionary, the word is depicted as a sequence of signs for every character constituting the word.

To perform this operation, the video clips depicting every alphabet in the word is merged. The scraping and/or merging alphabet videos is repeatedly performed until all the words of the sentence are fetched as videos. Once all the words are received, they are concatenated to form the final result video.

**Use-Case Diagram:** A use case diagram is designed to summarize the minute details of the operations performed by the different kinds of users of the system. It helps in visualizing the interactions made between the users and the system. Figure 2 describes the use case diagram of the Speech-to-ISL translator system.

The user performs the actions of giving input in the form of speech or in the form of text. The admin of the system is responsible for maintaining the sign language video database links. The admin also updates and improves the voice models. The system performs the operations of voice or speech recognition, modifying the sentence structure to suit the ISL syntax, generate the sign and produce the final sign language video.

![Use-Case diagram of the proposed system](image)

**Figure 2** Use-Case diagram of the proposed system

**Front-End Implementation:** Figure 3 describes the flow in the user interface of the system.

![User Interface Flow](image)

**Figure 3** User Interface Flow

The proposed system has a fairly simple user interface. The intention was to make it as intuitive as possible in order to enable ease-of-use amongst all people. The home page or landing page has a microphone button. On clicking this button, the built-in or external microphone on the device is activated to record the speech audio. In a situation where a microphone is unavailable, a text box is also present on the homepage where the user can enter the sentence he/she wishes to translate. Once the speech to be translated is recorded, a
loading icon appears. This has been included to assure the user that no lag, delay or failure of the system has occurred. Once the result video has been generated, the site redirects to another webpage. On the final webpage, the sign language video is displayed as the result.[27-30]

**Back-End Implementation:** Figure 4 depicts the sequence of operations in the back-end implementation of the system. First, the speech is given as the input and is converted into text by the JavaScript library, Speech Recognition. Once the text is generated, EJS sends a request to Node.js to begin the video processing. In the request, the generated text is sent as a parameter. On receiving the request, Node.js spawns the Python child process to execute the NLP algorithm to apply ISL syntax transformation rules, and perform video processing.

This child process is turned into a promise so that Node.js can send back a response to EJS after the completion of the Python script execution. Figure 5 describes how promises are handled by EJS.

Once the promise is resolved, a status code 0 is sent back to Node.js. This implies that the execution of the ISL syntax transformation rules and video processing is completed without any errors. If not, an error handling process takes place. Once the ISL video is generated, it is sent to EJS in the response header by Node.js. The ISL video is buffered as a response when requested.[31-34]

5. **RESULTS**

The developed Speech-to-ISL translator is an effective web-based tool to aid communication with the hearing and speech impaired community of India. Figure 7 is a snapshot of the landing page or home page of the web portal. As seen in the image, the microphone button is used to record audio. Once the audio is recorded, the textual version of the recorded sentence is displayed in the text area as shown. The text area can also be used to type in the sentence if a microphone is not available. Once the sentence has appeared in the text area, the ‘Generate’ button can be clicked in order to begin the processing of the final result video. To cancel the recording of the audio, the ‘Cancel’ button can be pressed. The Landing Page is shown in Figure 6.

![Image of the Landing Page](image-url)

**Figure 6 Landing Page**

After the result video is generated, the website redirects to another page. This page has the video embedded into it, making it easy to view for the user.

**Testing:** The three main categories of tests performed on the Speech-to-ISL system were checking of stop word removal, checking of tense conversion and checking inflection handling.

**Stop word Removal:** Figure 7 shows an example for stop word removal. As seen in the figure, the English syntax is converted to the ISL syntax. In addition to this the removal of stop word ‘is’ has taken place.

English Syntax: “My house is in Karnataka”

ISL Syntax: “My house Karnataka in”
Stop Word Removal

**Tense Conversion:** Figure 8 describes an example where tense conversion has taken place. As seen in the figure, the word ‘became’ is transformed into ‘become’. This example proves that the system successfully converts past tense into present tense. This is a critical operation because ISL sentences do not have any words with tenses. In the example above stop word ‘a’ has also been removed.

**English Syntax:** ‘His son became a surgeon’

**ISL Syntax:** ‘His son surgeon become’

Inflection Handling: Figure 9 gives an example of a scenario where inflection handling takes place. The word ‘studying’ is converted to ‘study’ and the stop word ‘is’ gets removed. In ISL, all words are in simple present tense. As seen in the figure, present continuous tense has been converted to simple present tense successfully.

**English Syntax:** ‘He is studying architecture’

**ISL Syntax:** ‘He architecture study’

The Table 1 shows the comparison of different existing techniques in terms of accuracy and sensitivity with proposed Speech to ISL with result from 2016 to 2020.

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proposed Speech-to-ISL</strong></td>
<td>97.86%</td>
<td>94.56%</td>
</tr>
<tr>
<td>Lalit Goel et al. (2016)</td>
<td>94.35%</td>
<td>92.89%</td>
</tr>
<tr>
<td>Krunal Sailza et al. (2019)</td>
<td>94.53%</td>
<td>93.45%</td>
</tr>
<tr>
<td>Sugandhi et al. (2020)</td>
<td>96.67%</td>
<td>94.02%</td>
</tr>
</tbody>
</table>

In machine learning algorithms accuracy is the proportion of true results. Sensitivity is used for reliability of the test from speech to Indian Sign Language. This way it has been implemented successfully and by 3 test cases the proposed methods and implementation give 97.86% accuracy.

6. CONCLUSION

A significant section of the Indian society suffers from hearing and speech impairment. This population uses Indian Sign Language as their primary mode of communication. Due to the difficulty in learning and understanding the meaning and context of written texts, sign language is preferred. Sign language involves the usage of hands, lip movements and expressions in order to communicate words, emotions and sounds. The proposed system provides an efficient method to aid communication between an individual with hearing and speech impairment. It is a field that has had little development over the years particularly in successful implementation in Python programming language. The system would improve access to information for the hearing-impaired population of the country like India. Moreover, the system can also act like an educational tool to learn ISL.

To ensure uniformity in the quality of the video, animated characters enacting the sign language can be incorporated. Since there is a slight variation in the
dialect of ISL depending on the region of India, the system can be modified to suit the needs of the hearing and speech impaired of that region. This can be done by creation of custom corpora by taking the help of schools and other facilities for the deaf-mute. In addition to this, Google Translate APIs can be utilized in order to aid translation from any Indian regional language to ISL. This would improve the communication between an individual not literate in English and an individual who uses ISL. The proposed system can also be used in applications, like video streaming, to dynamically translate the content into ISL. This would help the deaf-mute community gain a better understanding of the context and meaning behind the content being displayed.

REFERENCES


