Voice to Indian Sign Language Translator

Prasanna Shete[1], Pranjali Jadhav[2], Dikshita Jain[2], Pearl Kotak[2]
K. J. Somaiya College of Engineering, University of Mumbai, Vidyavihar - 400077
prasannashete@somaiya.edu, pranjali.j@somaiya.edu, jain.dk@somaiya.edu, pearl.kotak@somaiya.edu

Sign language is a natural way of conversation for disabled people with speech and hearing impairments. It is a non-verbal visual language that is characterized by manual and non-manual signs. Non-manual signs embody facial expressions, head tilting, shoulder raising, mouthing, etc. whereas manual signs include fingerspelling which may be used to signify a word from a spoken language, by spelling out the letters, etc. Indian Sign Language, generally referred to as ISL, is one of the popular sign languages used by people with speech and hearing disabilities to communicate. However, communication with vocal people is not possible as they are not well versed in sign language. This paper proposes a system named Voice to Indian Sign Language Translator that facilitates communication amongst vocal and hearing-impaired people. This is achieved by converting voice to text, which is further translated into corresponding sign language using Natural Language Processing. Experimental results show that the sign language translation provides a fairly good accuracy with commonly spoken English sentences.

CCS CONCEPTS • Human centered Computing• Computing methodologies

Additional Keywords and Phrases: Indian Sign Language(ISL); NLP(Natural Language Processing); speech-to-text; speech-to-sign language; text-to-sign language;

1 INTRODUCTION

Out of 136 residing language families, Sign language is utilized by deaf and hearing-impaired humans to convey their message. Out of the entire human population on earth, nearly 0.1% people are deaf and have hearing difficulties. [1] In crowded and daily use places like shops, malls, schools, bus stands, banks, hospitals, railway stations, etc., it becomes difficult for the general public to communicate with a deaf person because he/she may not know sign language and thus won't be able to efficiently send their message to a deaf or hard of hearing person. Thus, sign language translation holds immense importance because it enhances easier communication between general people and the deaf community.

There is a standard thought that deaf individuals will browse texts since their vision isn't hindered however this doesn't relate to reality. They often use Sign Language as their primary mode of communication. A sign language is a language on its own with its own set of grammar rules and an organized sentence structure; people use different sign languages in different parts of the world. India has its sign language by the name Indian Sign Language (ISL).[2] These differences, and therefore the incontrovertible fact that sign speakers learn their country's written communication as a second language, are the main reason deaf people cannot browse written texts fluently. For example, the phrase "I am eating pizza" in English can be "I pizza eat". In sign language, for a person whose main language is sign language, all other words in the sentence are considered noise making it difficult to understand. [3] To overcome this, a system is needed for converting speech to sign language. In this paper, an efficient method is proposed that focuses on constructing a model that translates the voice to ISL and thus tries to enhance the communication systems of people having hearing difficulties. The proposed model provides a way for the deaf community to master their problem of communication with the public, hence providing a certain level of equal treatment in society.

2 LITERATURE REVIEW

To facilitate communication with hearing-impaired people a lot of research is on-going, and many systems have been developed to convert speech into sign language. An online platform for deaf people has been designed both as a web application and an Android application that effectively means communication and learning.[4] The model has a systematic four-stage functioning: Acquisition of speech using PyAudio, Conversion of speech to text using Google Speech-to-Text API(uses text tokenization and concepts of NLP for text processing), matching of text with visual sign word library(video dataset of sign language) from hand speak, merging of matched videos according to the sequence of processed text and display to the deaf person.

Speech to sign language translation system with new features for increasing adaptability has been proposed in [5]. The system consists of a speech recognizer written in Java and executed using the Eclipse IDE called Sphinx 4.0. The input is taken from the microphone of the system as a sentence and is converted in the form of text thereby displaying a video of the spoken sentence in sign language. Stored videos for the corresponding sentence are displayed to the user. This system concentrates on 4-5 Basic English sentences.

The Audio to Sign Language Translation for Deaf People introduced in [6] explores an application that accepts input in form of speech, transforms it into text, and then the output is shown in the form of Indian Sign Language images. EasyGui is used to design the front end of the system. Speech input through a microphone uses the PyAudio package and is converted to text using Google Speech API. Text pre-processing is completed using NLP (Natural Language Processing), followed by dictionary-based machine translation.

A mobile system is proposed in [7] that translates the language of speech into the language of hand gestures. Here the language of speech was used by mobile devices as client-side input; translated into text messages stored in a cloud database, then the text message has been translated into sign language. Here algorithms are used for both static and dynamic gesture recognition. The software consists of an engine that translates speech into text and then animates the 3D avatar with the appropriate sign language.

The proposed system differs from the above systems in various ways-

- Most of the previous work has been done on American Sign Language whereas our system focuses on Indian Sign language for commonly spoken simple sentences.
- Real-time videos rather than 3D avatar/images are incorporated in our system for translation, thus enhancing the communication by including facial and hand expressions.
- The proposed system translates each sentence by fetching the videos of individual words rather than translating the entire sentence by using videos of individual letters

3 INDIAN SIGN LANGUAGE AND ISL DATASET DESCRIPTION

ISL is a commonly used language among deaf people in India. The Indian Sign Language involves its dictionary and rules of grammar. It has not been invented by hearing people. Some facts about the sign language of which ISL is a part is listed below: [8]

Facts about Sign Language:

- It isn't similar all over the world.
- It has grammar as well as gestures along with non-manual expressions.
- Compared to other spoken languages, it has a smaller dictionary.
- Unknown words signified using fingerspelling.
- The adjective follows the noun in most sign languages.
- Never use am/is/are/was/were/ (linking verbs).
- Always uses Present Tense without word-endings/suffixes.
- Avoid using articles.
- Use "me" instead of "I".
- Interrogative words are placed at the last e.g. "You want how much?"
- Have no gerunds. (-ing).

The dataset consists of videos of people representing various English words in Indian Sign Language. The videos can be obtained from the sources: Dictionary of ISL by Faculty of Disability Management and Special Education (FDMSE) [9], Coimbatore, and Indian Sign Language Research and Training Center (ISLRTC), New Delhi. [10]

Fig 1. ISL Dictionary

We have labelled every video clip and stored it in the dictionary as pairs of {Key, Value} where Key represents the word and value represents the video (in .mp4 format). Fig. 1 represents the stored dictionary of ISL videos from the above-mentioned sources.

4 PROPOSED SYSTEM

The approach starts with the recognition of input speech which is then converted to English text. The text is then transferred to the parsing module, which tags each word with its corresponding parts of speech. Since the grammatical structure of the English language and the ISL differ, the parsed sentence is rearranged according to the grammar rules defined in the ISL. This is followed by the elimination, lemmatization using the NLP (Natural Language Processing), and then the output is finally represented in the form of a sign language video by concatenating individual videos of the respective words obtained by the ISL video dictionary.

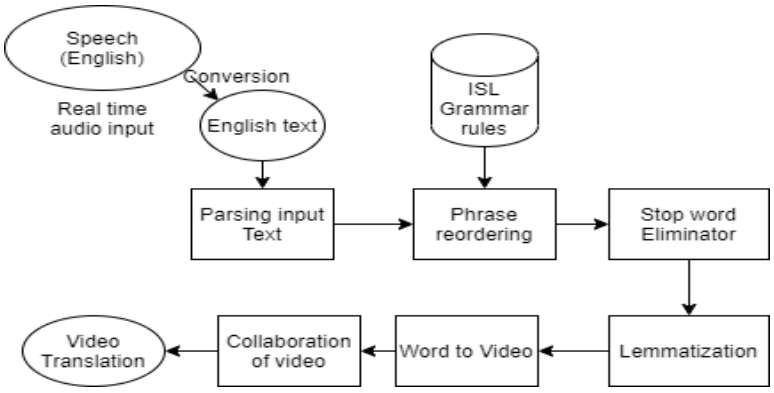

Fig. 2 Proposed System Architecture

The Proposed System consists of two main parts:

1. Voice to English text Conversion
2. English text to Indian Sign Language Video Conversion

Voice to English text Translation:

The conversion of speech/voice to text is done using WebSpeech API [11]. Speech recognition and speech synthesis are two different interfaces provided by Web Speech API. For our system, speech recognition functionality is being used. It receives speech through the microphone followed by scanning through a speech recognition service against a list of grammar. On successful recognition of a word/phrase, a result as a text string is returned as output. SpeechRecognition is the main controller of Web Speech API. However, the Web Speech API speech recognition interface gives restricted support to Chrome for Desktop and Android. It has been supported by Chrome since version 33 with prefixed interfaces. So, we need to use the webkitSpeechRecognition interface. The Web Speech API has an accuracy of 89.4% for native English speakers and 65.7% for non-native English speakers. [12]

Considering the version incompatibilities, the proposed system must have a higher version of Chrome (33 or higher)

English text to Indian Sign Language Translation:

This part is then deduced into the following modules:

I. PoS (Parts of Speech) Tagger:

The grammatical structure of linguistic communication is needed for the rule-based conversion of 1 language to another. So, to know the structure of ordinary English synchronic linguistics and ISL, parsing is needed. We will be using the Universal Dependency English Web Treebank (EWT) Tagset [13] provided by Natural Language ToolKit (NLTK) [14] for parsing the English sentence. This corpus contains 16622 sentences. Each word of the input is tagged with its respective correct part of speech. Table 1 represents the provided tags.

Open-Class Words	Closed-Class Words	Other
ADJ -Adjective	ADP – Adposition	PUNCT - Punctuation
ADV - Adverb	AUX – Auxiliary	SYM - Symbol
INTJ - Interjection	CCONJ - Coordinating Conjunction	X - Other
NOUN – Noun	DET – Determiner	
PROPN - Proper Noun	NUM – Numeral	
VERB – Verb	PART – Particle	
	PRON – Pronoun	
	SCONJ -Subordinating Conjunction	

Table 1. PoS tags by EWT UD Corpus

There are different techniques for POS Tagging: Lexical Based Methods, Rule-Based Methods, and Probabilistic Methods (HMM and CRF), Deep Learning Methods. [15] For the training of the dataset in the proposed system, CRF Algorithm (Conditional Random Field Algorithm) [16] is used. Because CRF is constructed as one exponential model for the joint probability of the entire sequence of labels provided by the observation sequence, it proves advantageous to alternative algorithms for teaching the tagger. Therefore, the weights of various features at different states are determined based on the maximum likelihood of the training data. CRF helps in developing good feature selection and induction algorithms for the training model; which implies instead of specifying beforehand the features of (X, Y) to use, the method starts from

feature-generating rules and evaluating the advantage of generated features automatically on data. [17] The feature generating rules include capitalization of the first letter of the word, suffix and prefix of the word, its previous word, is it the first or the last, etc. The f1 score which shows the accuracy of this system turned out to be 93.2%.

II. Sentence Reordering:

The tagged sentence is then passed to this module which helps to reorder the words in the sentence based on ISL grammar rules. According to the grammar of Standard English, the sentence is in the form of Subject-Verb-Object; however, in the grammar of ISL, the sentence is in the form of Subject-Object-Verb. [18] All the verb patterns are required to follow their respective occurrence of the noun for the conversion of English sentences to an ISL sentence with its grammar rules.

III. Stop Words Elimination:

The Re-ordered sentence is passed to this module which involves removing unwanted words that do not play a role in ISL. It involves linking verbs, determiners, coordinating conjunctions, and so on. The words which are not part of the ISL sentence conforming to its grammar rules are recognized and removed which includes TO, possessive ending, AUX (Auxiliary verbs) like need, must, should, would, foreign words like bona fide, status quo, prima donna, faux pas, CCONJ (coordinating conjunctions like and, or, but, yet, some), DET (determiners like a, an, the), non-root verbs, INTJ (Interjections). Stop words can be found out in the nltk_data directory [19].

IV. Lemmatization:

Lemmatization finds the core or the root word [20]. The tagged sentence as output from the previous module is passed to this module which is used for satisfying the rule of present tense and no gerunds in SL, thereby any change related to time or quantity is removed. For example, in nouns, plurals (girls, boys) get reduced to their singular form (girl, boy); and in verbs, time/participle variants (playing, slept, eating) are back to present tense (to play, to sleep, to eat).
We have used two approaches for lemmatization: Corpus-based and Rule-based.

A. Corpus-based:

It is built as a dictionary that uses a word and the pos tag as key and the corresponding lemma as the value. The dictionary is built by combining two separate corpora:-Universal Dependency English Web Treebank Corpus [12] and British National Corpus [20].
The word (form), pos tag, and the lemma for each word in the corpus are extracted and appended to the dictionary. The total number of words in this dictionary is 35078.

```
living [ ADJ ] --> living
living [ NOUN ] --> living
living [ VERB ] --> live
guns [ NOUN ] --> gun
wives [ NOUN ] --> wife
thieves [ NOUN ] --> thief
leaves [ VERB ] --> leave
watches [ VERB ] --> watch
leaves [ NOUN ] --> leaves
watches [ NOUN ] --> watches
fairies [ NOUN ] --> fairies
```

Fig. 3 Corpus-based Lemmatization

B. Rule-based:

The corpus-based approach gave pretty good results in many cases, but it does not perform well with plural nouns like leaves, watches, fairies, etc. Hence, a rule-based approach has been used for extracting the lemma of plural nouns. This method uses many rules of grammar related to plural nouns that tell how a word should be modified to extract the lemma.

```
living [ ADJ ] --> living
living [ NOUN ] --> living
living [ VERB ] --> live
guns [ NOUN ] --> gun
wives [ NOUN ] --> wife
thieves [ NOUN ] --> thief
leaves [ VERB ] --> leave
watches [ VERB ] --> watch
leaves [ NOUN ] --> leaf
watches [ NOUN ] --> watch
fairies [ NOUN ] --> fairy
```

Fig. 4 Rule-based Lemmatization

V. Video Conversion:

After this, the lemmatized sentence is given into this module where for each word of the transformed sentence, each video clip is extracted from the collection. If the video clip is unavailable for a particular word, it is represented in the video as a series of letters representing the word. These video clips are then merged to give an output in the video form representing the Indian sign language of the corresponding voice input using the python library MoviePy [21]. The videos are fetched corresponding to the word obtained from the lemmatizer. The concatenated video is then formed and displayed to the user. The two parts of the proposed system are then integrated using Flask as a development framework. As soon as the voice is recognized and interpreted by the API, it is passed through part two wherein the English text is translated into ISL video. The video is then displayed as output to the user.

5 RESULTS

The system is deployed on Heroku[23]. The web link is - https://english-voice-to-isl.herokuapp.com/ . The dynamic tool for converting voice to ISL translation video is demonstrated by entering a voice input - "the Museum is beautiful".

Figure 5 represents entering the voice input by the user on clicking the microphone button using Web Speech API (Speech Recognition Interface)

Fig. 5 User Interface using Web Speech API

Figure 6 shows Frontend representing the video in Indian Sign Language for the corresponding Voice Input.

Fig. 6 Result displaying Translated video

The demonstration of the website on desktop and mobile is uploaded on YouTube link - https://bit.ly/3cZf11B

6 Conclusion And Future Scope

The Voice to ISL Translator is very useful to improve the communication between hearing-impaired individuals and vocal people. The novelty of the proposed system involves Indian Sign Language and its rules since not much progress has been done in this field. The Voice to ISL Translator successfully converted the voice input sentence into a single video giving a model a much realistic and lively appeal for better usage, by the hearing-impaired persons. We tested the system on commonly spoken English sentences which shows an accuracy of 60-70%. However due to the pandemic, we couldn't reach out to the ISL communities for their feedback.

In the future, we plan the expansion of the ISL dictionary by creating more videos corresponding to the words and their respective PoS tags to achieve more accurate results. The ISL videos of words by the same interpreters would help in increasing the usability and reliability of the proposed system. Also, the current implementation is restricted to English Voice input, which reduces the liberty of using various local languages; so, we plan to support voice input in various Indian languages and increase our system's reach and usability.

REFERENCES

[1] World Federation of the Deaf (WFD) - human rights, deaf, deaf people. (2015). Retrieved July 10, 2016, from https://wfdeaf.org/

[2] Zeshan, U., Vasishta, M. N., & Sethna, M. (2005). Implementation of Indian Sign Language in educational settings. Asia Pacific Disability Rehabilitation Journal, 16(1), 16- 40.

[3] Tiago Oliveira, Paula Escudeiro, Nuno Escudeiro, Emanuel Rocha, Fernando Maciel Barbosa. Automatic Sign Language Translation to Improve Communication. 2019 IEEE.

[4] Ritika Bharti, Sarthak Yadav, Sourav Gupta, and Rajitha Bakthula. Automated speech to sign language conversion using Google API and NLP. Proceedings of ICAEEC-2019, IIIT Allahabad India, 31st May - 1st June 2019

[5] Shekainah Paulson and Mrs. B. Thilagavathi. An Adaptable Speech to Sign Language Translation System. 2014 International Journal of Engineering Research & Technology (IJERT) Vol. 3 Issue 3, March – 2014.

[6] Ankita Harkude, Sarika Namade, Shefali Patil, Anita Morey. Audio to Sign Language Translation for Deaf People. International Journal of Engineering

and Innovative Technology (IJEIT) Volume 9, Issue 10, April 2020.

[7] Mrs.K.Rekha, Dr.B.Latha. Mobile Translation System from Speech-Language to Hand Motion Language. 2014 International Conference on Intelligent Computing Applications.

[8] Goyal, L., & Goyal, V. (n.d.). Automatic Translation of English Text to Indian Sign Language http://www.aclweb.org/anthology/W16-6319.

[9] A Dictionary on Indian Sign Language (ISL) signs by FDMSE, Coimbatore. https://indiansignlanguage.org/.

[10] A Dictionary of Indian Sign Language Research and Training Center (ISLRTC), Government of India. http://www.islrtc.nic.in/

[11] Documentation from MDN Web Docs - a Web API named as Web Speech API. https://wicg.github.io/speech-api/

[12] Tim Ashwell and J.R Elam. How accurately can the Google Web Speech API recognize and transcribe Japanese L2 English learners' oral production?

[13] Ann Bies, Justin Mott, Colin Warner, Seth Kulick August 16 2012. Universal Dependencies Contributor. https://catalog.ldc.upenn.edu/LDC2012T13

[14] Steven Bird, Ewan Klein, and Edward Loper. Natural Language Processing with Python. O'Reilly Media, 2009

[15] Ian Goodfellow, Yoshua Bengio, Aaron Courville.Deep Learning (Adaptive Computation and Machine Learning series) Hardcover – 18 November 2016

[16] Charles Sutton, Andrew McCallum. An Introduction to Conditional Random Fields.Volume 4, 2012

[17] John Lafferty, Andrew McCallum, and Fernando Pereira. 2001. Conditional random fields: Prob-abilistic models for segmenting and labeling sequence data. In Proceedings of the Eighteenth International Conference on Machine Learning, pages 282–289.

[18] Gouri Sankar Mishra, Parma Nand, Pooja. English text to Indian Sign Language Machine Translation: A Rule Based Method. International Journal of Innovative Technology and Exploring Engineering (IJITEE). Volume-8, Issue-10S, August 2019- Page 460

[19] Steven Bird, Ewan Klein, and Edward Loper. Natural Language Processing with Python. Chapter 2. O'Reilly Media, 2009

[20] Divya Khyani, Siddhartha B S, Niveditha N M, Divya B M. An Interpretation of Lemmatization and Stemming in Natural Language Processing. Journal of the University of Shanghai for Science and Technology. Volume 22, Issue 10, October - 2020- Page 350

[21] The University of Oxford. British National Corpus http://www.natcorp.ox.ac.uk/

[22] Zulko, MoviePy, released under the MIT license. https://pypi.org/project/moviepy/

[23] A web deployment platform- https://www.heroku.com/