English Speech to Indian Sign Language Translator using Natural Language Processing

1st Dr. Prasanna Shete
dept. name of organization (of Aff.)
K.J Somaiya College Of Engineering (of Aff.)
Mumbai, India
prasannashete@somaiya.edu

2nd Pranjali Jadhav
dept. name of organization (of Aff.)
K.J Somaiya College of Engineering (of Aff.)
Mumbai, India
pranjali.j@somaiya.edu

3rd Dikshita Jain
dept. name of organization (of Aff.)
K.J Somaiya College Of Engineering (of Aff.)
Mumbai, India
jain.dk@somaiya.edu

4th Pearl Kotak
dept. name of organization (of Aff.)
K.J Somaiya College Of Engineering (of Aff.)
Mumbai, India
pearl.kotak@somaiya.edu

***Abstract*—Sign language is a natural way of communication for challenged people with speaking and hearing disabilities.This is a non-verbal visual language that is characterized by manual and non-manual signs. Non-manual signs consist of various facial expressions, head tilting, shoulder raising, mouthing, etc. whereas manual signs consist of a manual alphabet (fingerspelling) which may be used to borrow 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 dumb and deaf people to communicate with the general public. We propose a system called English Voice to Indian Sign Language Converter consisting of two parts -**

1. Speech to text translation
2. Text to ISL translation

Part one is responsible for converting voice to English text. Thus, the text is then passed to the parsing module, which tags each word with its corresponding parts of speech by implementing the CRF algorithm on the Universal Dependencies English EWT corpus. The parsed sentence is then reordered as per the grammar rules defined in the Indian Sign Language (as the grammar of English language and Indian sign language is different). 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 from the Indian Sign Language video dictionary, thereby complying to the part two of the proposed system.

***Index Terms*—Indian Sign Language (ISL), Corpus, CRF (Conditional Random Field) Algorithm, Lemmatization, stop-words eliminator, Universal Dependencies EWT corpus, NLP (Natural Language Processing)**

I. Introduction

Sign language is one of the 136 living language families, which is used by deaf and hard of hearing people convey their message. Out of nearly 7 billion people on earth, almost 72 million are deaf and have hearing difficulties. In common places like railway stations, bus stands, banks, hospitals, etc., it is very difficult for the general public to communicate with a deaf person because a vocal person may not understand sign language and thus won't be able to convey any message to a deaf or hard of hearing person. Thus, sign language translation holds great importance as it focuses on facilitating better communication between the deaf community and the general public.

There is a common perception that even though deaf people use sign language as their main language, they can read texts fluently in their countries' spoken language because their vision is not hindered. Such perception does not correspond to reality. A sign language is a language on its own with its own set of grammar rules and specific 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). These differences and the fact that signers learn their countries' written language as a second language are the main reason why deaf people can't fluently read written texts. Example:- the sentence "I am going home" in English can be "I home go" in sign language; for a person whose main language is sign language, all other words in the sentence are seen as noise that makes comprehension harder. 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 converts the English Voice to ISL that enhances the communication capabilities of people with hearing disabilities. The proposed model provides advantages to the deaf and hard-to-hearing community to overcome the difficulty they face in life thereby hoping that with a better understanding of sign language, deaf or dumb people will find themselves equipotential in society.

Facts about Indian Sign Language:

- NOT the same all over the world.
- NOT just gestures and pantomime, but do have their grammar.
- Have a much smaller dictionary than the other spoken natural languages.
- Finger-spelling for the unknown words.
- Most of the sign languages put the adjective after the noun e.g. Car Red.
- Never use am/is/are/was/were/ (linking verbs).
- Never use word-endings/suffixes.
- Always sign in the Present Tense.
- Do not use articles. (a, an, some, the).
- Do not use I, but use me.
- WH-questions are at the END e.g. "You go where?" Have no gerunds. (-ing).
- Use non-manual expressions as well e.g. use of eyebrows, eyelids, facial expressions, head, and shoulders movement.
- NOT been invented by hearing people.

II. Related work

Many systems have been developed to convert Speech into Sign language such as in [1] wherein an online platform for deaf people has been designed both as a web application and an Android application which serves as an effective means of communication and learning for them. 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.

In [2] Shekainah Paulson and Mrs. B. Thilagavathi in their paper discusses a speech to sign language translation system with new features for increasing adaptability. 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 are then displayed for each sentence.

[3] The Audio to Sign Language Translation for Deaf People system introduced by Ankita Harkude, Sarika Namade, Shefali Patil, Anita Morey explores an application that accepts input in the 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 that is taken as input through a microphone uses the PyAudio package and converted to text using Google Speech API. The pre-processing of the text is done using NLP (Natural Language Processing) followed by Dictionary-based machine translation.

[4] Mrs.K.REKHA and Dr.B.LATHA in their paper titled "MOBILE TRANSLATION SYSTEM FROM SPEECH LANGUAGE TO HAND MOTION LANGUAGE" proposed a model in which voice language through the mobile devices has been taken as an input on the client side; it has been translated into text message stored in a cloud database, then text message has been translated into sign language. Two algorithms will be used for the recognition of signs, one for static signs where recognition is performed based on the position and another algorithm is the curve matching algorithm where the trajectory of the motion is analyzed. The software consists of an engine that translates speech to text and then animates a 3D avatar with the equivalent sign language.

III. Data Description

The dataset consists of videos of people representing various English words in Indian Sign Language. The videos can be obtained from the sources: [5] Dictionary of ISL by FDMSE, Coimbatore, and [6] ISLRTC, New Delhi.

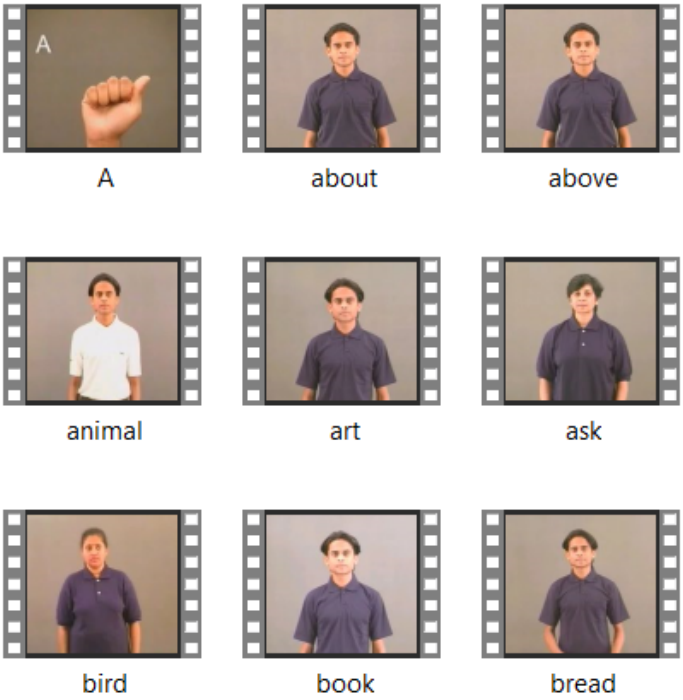

Fig. 1. ISL Dictionary

Every video clip will be labeled and stored in the dictionary as pairs of Key, Value where,

Key: word

Value: video (in .mp4 format)

IV. Implementation Details

The Proposed System consists of two main parts:

PART A: Deals with converting English Voice to text with WebSpeech API and generated text is provided as an input to the following part.

PART B: Deals with Converting English text to the video output of Indian Sign Language Translation.

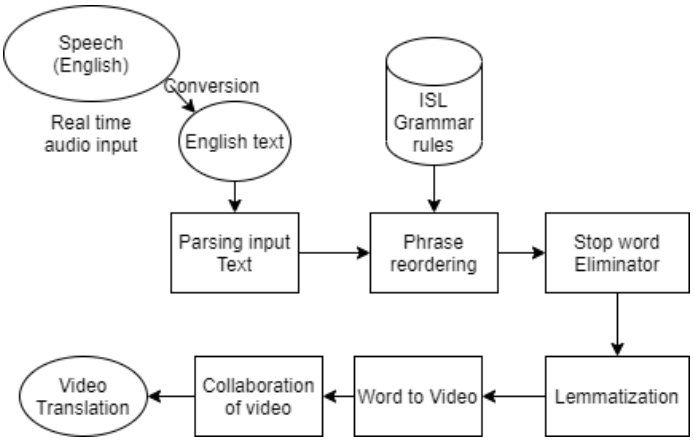

Fig. 2. Proposed System Architecture

PART A: The conversion of English speech/voice to text is done using WebSpeech API. The Web Speech API provides two different interfaces— speech recognition, and speech synthesis. 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 (or list of results) as a text string is returned as output. The Web Speech API has a main controller interface for this — SpeechRecognition — plus several closely related interfaces for representing grammar, results, etc. However, the Web Speech API speech recognition interface gives restricted support to Chrome for Desktop and Android. Hence, Chrome Version 33 uses the prefixed interface - webkitSpeechRecognition. For the earlier versions, the interface to be used is - SpeechRecognition

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

PART B: This part is then deduced into the following modules -

1) PoS (Parts of Speech) Tagger
2) Sentence Reordering
3) Stop Words Elimination
4) Lemmatization
5) Video Conversion Module

1. PoS (Parts of Speech) Tagger: For rule-based conversion of one language to another language, the grammatical structure of the source language is required so that the words of the source sentence can be reordered as per the grammar rules of the target language. So, to understand the structure of standard English grammar and Indian sign language, parsing is needed. We will be using the Universal Dependency EWT Tagset provided by NLTK 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 I
Universal Dependency EWT tagset

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	

There are different techniques for POS Tagging: Lexical Based Methods, Rule-Based Method, Probabilistic Methods(HMM and CRF), Deep Learning Methods. For the training of the dataset in the proposed system, CRF Algorithm (Conditional Random Field Algorithm) is used. CRF is better than other algorithms for training the tagger because the input is a set of features (real numbers) derived from the input sequence using feature functions, the weights associated with the features (that are learned), and the previous label and the task is to predict the current label. The weights of different feature functions will be determined such that the likelihood of the labels in the training data will be maximized. In CRF, a set of feature functions are defined to extract features for each word in a sentence. Some examples of feature functions are: is the first letter of the word capitalized, what is the suffix and prefix of the word, what is the previous word, is it the first or the last word of the sentence, is it a number, etc. These features are called **state features**. In CRF, we also pass the label of the previous word and the label of the current word to learn the weights. CRF will try to determine the weights of different feature functions that will maximize the likelihood of the labels in the training data. The feature function dependent on the label of the previous word is called as **transition feature**.

2. Sentence Reordering: This module 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**. For conversion of English sentences to a sentence as per ISL grammar rules, it is required to shift all the verb patterns

after the corresponding occurrence of the noun.

3. Stop Words Elimination: This module involves removing unwanted words that do not play a role in ISL. It involves linking verbs, determiners, coordinating conjunctions, and so on. The parts of speech that are not part of the ISL sentence are detected and eliminated. The various part of speech which does not form the part of ISL sentence are 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 like Wow, Alas, Yay, Hurray, Congratulations, etc). NLTK(Natural Language Toolkit) in python has a list of stop words stored in 16 different languages. It can be found out in the nltk data directory

4. Lemmatization: Lemmatization finds the core or the root word. This module is used for satisfying the rule of present tense and no gerunds in SL. Any variation 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 (ate, brought, chatting) are back to present tense (to eat, to bring, to chat). We have used two approaches for lemmatization: **Corpus-based and Rule-based.**

1) ***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:–
(A) Universal Dependency Treebank Corpus
(B) British National Corpus

```
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

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 are 35078.***

2) ***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

5. Video Conversion: This is the final module wherein the sentence which is transformed through the above 4 modules is taken as input. 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 English voice input using the python library MoviePy. The videos are fetched corresponding to the word obtained from the lemmatizer. The concatenated video is then formed and displayed to the user.

Results and Discussion

The dynamic tool for converting English Voice to ISL translation video is demonstrated by entering a voice input - “Museum is beautiful”.

Figure 5 represents entering the English Voice input using the microphone button.

Fig. 5.

Figure 6 represents the processing of the English voice input through these modules - PoS Tagging, Sentence Reordering,

Stops Word Elimination and Lemmatization using NLP (Natural Language Processing).

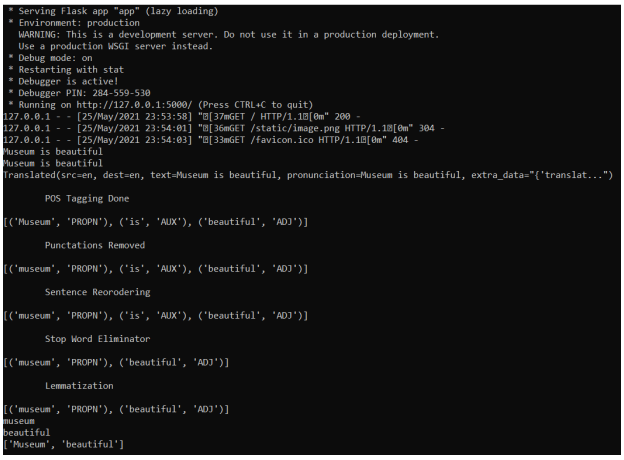

Fig. 6.

Figure 7 represents the final module which involves the output as ISL translation video.

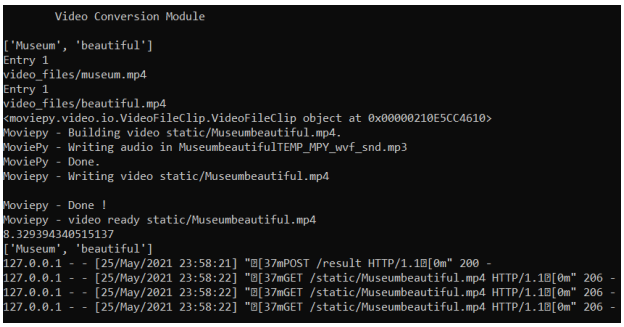

Fig. 7.

Figure 8 is the frontend representing the video in Indian Sign Language for the corresponding English Voice Input.

Fig. 8.

Conclusion and Future Work

The English Voice to ISL Converter is very useful to improve the communication between hearing-impaired individuals and vocal people. The novelty of the proposed system is that involves Indian Sign Language and its rules and not much progress has been done in this field. The proposed model successfully converts the voice input sentence into a single video giving a model a much realistic and lively appeal.

The future scope involves the ISL dictionary expansion by the creation of more videos corresponding to the words and their respective PoS tags may help in more accurate results. The proposed system revolves around English Voice Input which may reduce the liberty of languages thereby restricting it to one language. The ISL videos of words by the same interpreters would help in increasing the usability and reliability of the proposed system.

References

[1] 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

[2] 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

[3] 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

[4] Mrs.K.REKHA, Dr.B.LATHA. MOBILE TRANSLATION SYSTEM FROM SPEECH LANGUAGE TO HAND MOTION LANGUAGE. 2014 International Conference on Intelligent Computing Applications

[5] A Dictionary on Indian Sign Language (ISL) signs by FDMSE, Coimbatore

[6] A dictionary of Indian Sign Language Research and Training Center (ISLRTC), Government of India