JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Insta Hearing - Speech to Sign Language using Animations

Varsha Sablani
Dept. Of Computer Engineering
VESIT
Mumbai, India
2018.varsha.sablani@ves.ac.in

Mirudhula Nadar
Dept. Of Computer Engineering
VESIT
Mumbai, India
2018.mirudhula.nadar@ves.ac.in

Surabhi Soni
Dept. Of Computer Engineering
VESIT
Mumbai, India
2018.surabhi.soni@ves.ac.in

Diksha Ramnani
Dept. Of Computer Engineering
VESIT
Mumbai, India
2018.diksha.ramnani@ves.ac.in

Sujata Khandaskar
Asst. Prof., Dept. Of Computer Engineering
VESIT
Mumbai, India
sujata.khandaskar@ves.ac.in

***Abstract*—**Deaf people or people who take birth in deaf families learn sign language as their first language. The system converts audio to sign language. It will first convert the audio to text and then sign language. In this, it takes audio as input, which then with the help of Web Speech API which outputs the text that is matched with the words stored in the database, the displayed text on a screen finally gives the sign code of the given input using the ISL (Indian Sign Language) generator. The results of intervention studies suggest that hearing devices can improve psychosocial and communication outcomes, but behavioural interventions have not shown long-lasting benefits. To help the older ones and their family to manage the impact of hearing impairment for nurses, more research has to be done in many areas where it has been done poorly. This system highlights the important elements of interaction between the disability and community.

***Keywords—Hearing impairment, Speech to Sign, Speech to Text, Animations, Natural language processing, Application, Storytelling, Sign Language, Tokenization, Voice recognition, Sign Language Avatar*.**

I. INTRODUCTION

Sign language is a language used by deaf and dumb people as their mother tongue. It is used by people who cannot speak or have problems in speaking and by normal people to communicate with hearing disabled people. For deaf people having access to sign language is essential and helpful for their social, emotional and linguistic growth. Unlike sound patterns, sign language uses body language and communication to easily convey the person's thoughts.

Sign language is a visual means of communicating through hand signals, gestures, facial expressions, and body language. It is a form of communication for the patients who belong to the Hard-of-Hearing community, but sign language is useful for other groups of people also. The language arises because of the deaf, dumb and hard of hearing people in India. The Indian Sign Language (ISL) is the standard language used by Indians for expressing thoughts and communicating with each other.

In India, approximately 5.07 million people suffer from hearing disabilities. Among them, more than 30% of people are below 20 years of age and about 50% are between 20 years and 60 years of age. The only source of communication possible for them is with the help of sign language. As sign languages don't have proper structure and grammar, there is very little acceptability of the signs outside the small world of these Hard-of-Hearing communities. Research on American Sign Language proved that sign language is a full-fledged language that has its grammar and syntax, and also other linguistic attributes. And thereafter research on Indian Sign Language(ISL) started in 1978 where the conclusion was found that ISL is a complete natural language with its grammar and syntax.

Communication for deaf people in public places like railway stations, bus stands, banks, hospitals etc, is challenging because a hearing person may not understand the sign language used by the deaf person to communicate. Also,

a hearing person cannot convey any message to a deaf person as he/she may not know sign language. To make communication between the deaf and non-deaf communities, language translation is a must.

Sign languages have not been studied extensively as spoken languages, and there is still much left to be learned about them. Audio to sign language translator is a web-based application developed for deaf or hard of hearing people. It translates English words and Audio into Indian Sign Language i.e. Avtaar. The system takes simple English sentences and breaks them down into words takes that as an input and generates ISL-gloss which can be converted into SigML which provides signing instructions to the model and thereby generates an animated representation of ISL to the user.

II. RELATED WORK

1. In this paper Vladyslav Tsap, Nataliya Shakhovska, Ivan Sokolovskyi [authors] have converted audio to text using Adaptive Boosting, which is a technique used as an Ensemble Method in Machine Learning Ensemble is a technique that combines several base models to produce one optimal predictive model. It uses n number of decision trees during the data training period.

The first model is prepared and when the algorithm is implemented errors generated from that model are calculated. The samples that are wrongly classified in the first model are then given as input for the next model. This procedure is continued until the given state is met which means that all data sets are given equal weights and incorrectly classified data sets get higher weights and the higher weights are given more importance in the next model. It will continue to train models till minimum error is obtained. In this way, authors have translated audio files into text.

This paper aims to examine “weak classifiers” and to select the effective number of classifiers and their hyperparameters.

2. In this paper Lalit Goyal[author] has performed the task of converting English text to animation.

Parsing is performed on English text to get the phrase structure and grammar representation of the sentence. Reordering is required as English uses Subject Verb Object order whereas Indian Sign Language(ISL) uses the Subject Object Verb format. Unwanted words are removed as ISL uses words that have meaning other words like the article and helping verbs are removed.

Lemmatization is performed on those words for getting the root word of each word. Each word of the sentence is replaced with its equivalent HamNoSys((Hamburg Notation System)) which consist of 3D sign can be expressed in written form) and the HamNoSys string is translated to SiGML (Signing gesture markup language) code using SiGML rules. The generated SiGML code is then used to perform the required synthetic animation with the help of the SiGML animation tool.

3. The authors of this paper proposed a system of storytelling apps for children. This app accepts a storyline text as input and converts it to sign language and audio output. NLP is used for tokenization and to remove punctuation and stop words before comparing the words to the ISL dictionary to retrieve the acquired sign language gifs. They used the Android Text To Speech API, which employs the HashMap algorithm, for producing audio output.

4. In this paper Khushdeep Kaur, Parteek Kumar has designed a system for HamNoSys to SiGML Conversion.

eSIGNEditor is an existing tool that has a database of words for American and British Sign Language. Authors have developed a system for the Indian Sign language which consists of 210 HamNoSys Symbol and their SiGML tags respectively. A comparison is made between the entered HamNoSys symbol and existing symbols in the database. When a match is found the respective SiGML tag is stored in new SiGML which is then used to give input to a JA SiGML URL app and an avatar player is generated for the tag. In this paper, a test has been performed for 250 words which are frequently used like names of colour, birds or shapes.

5. In this paper Ankita Harkude, Sarika Namade, Shefali Patil, Anita Morey have performed audio to sign language generation.

For this, they have taken input as audio file with the help of the PyAudio module. Then they have used google API to search the recordings and present that text on the screen. With the help of dependency parser relationships between words are established.

Natural Language Processing is used to understand the order of sentences and perform reordering accordingly and is passed to the Indian Sign Language generator. Later on, the ISL generator displays the hand gesture which matches the text.

6. In this paper Alexandru Eugen Ichim, Sofien Bouaziz, Mark Pauly have implemented 3D avatars from handheld videos.

To implement the same authors have started with a two-dimensional representation of the video. Then acquisition is applied to rebuild the blendshapes, albedo texture, detail maps. Afterwards, at the animation level, the recreated rig can be operated by a temporal order of blendshape coefficients.

III. PROPOSED SYSTEM

In this project, we proposed a system that takes audio from users and converts it to text using Javascript Web Speech API. The text is then converted into an ISL sentence by removing the punctuation marks, stop words and rearrangement of the sentence according to the ISL grammar. This ISL sentence will then be illustrated by the avatars. Typically, all existing systems have pre-recorded videos or GIFs of humans performing the sign language. Instead, we'll be using animated avatars that will display sign language.

Avatars are "digitally created humanoids" or "virtual bodies," according to Wikipedia. They take input in the form of SigML or XML text and generate animation based on it. In a sequence, animation frame definitions are fed into the avatar. They describe the avatar's static pose. These sequences also include the avatar's timestamp, which indicates when the avatar will be in that pose. When these avatars are placed in a series of poses, rendering software generates signing animation that corresponds to specific frame definitions

IV. METHODOLOGY EMPLOYED

1. Front-end using HTML, CSS, JavaScript.

We have designed a website for our Insta Hearing project with the help of HTML, CSS, JavaScript. HTML is used to design the structure of our website CSS is added to provide style and colours whereas javascript helps to make the website interactive and make the elements used in Html accept and process the input provided with the help of the Document Object Model.

Users can record their speech/audio with the help of a microphone on our website (for this user will have to give Microphone access) then the following words are translated to text and displayed on the screen. Then on the right hand side, an avatar is present which will translate the displayed text into Indian Sign Language.
In short, an animated video is played which shows the signs for the given input audio.

2. Speech recognition using JavaScript Web Speech API.

Javascript has various functions for events when the onStart function is triggered it displays that webpage is listening to the incoming sound,onEnd event will stop receiving incoming sound, onError is invoked when an error occurs during transmission of voice when some object is recognized on Result function is callbacks to display the defined object.

Javascript uses various properties for speech recognition some of which are interim, language. User has to select a language that they are going to detect. Interim results are not final when an object is detected it can also be added. start() method is used to start the process of speech recognition then the onstart() events gathers all the recognized objects into an array. The first [0] displays the resultSet at the last position

3. Text Preprocessing using Natural Language Toolkit (NLTK).

NLTK consists of the algorithms such as tokenizing, part-of-speech tagging, stemming, sentiment analysis, topic segmentation, and named entity recognition. In tokenization, big quantities of text are divided into smaller parts called tokens. These tokens are very useful for finding such patterns. After tokenization, the normalization of words is done through stemming. In this, a set of words in a sentence are converted into a sequence to shorten its lookup. The words which have the same meaning but varies with the context or sentence are normalized. After normalization text is stated as pre-processed text which is error-free.

4. 3D animated avatar using Blender 3D tool.

Blender is an open-source 3D computer graphics software toolkit used to create animations, visual effects, 3D printed models, and more. Using this Blender software, we created our 3D model to illustrate sign languages and create GIFs of these words and store them in a database.

V. BLOCK DIAGRAM

Initially, two options are available to the user. Input is taken in voice or text format. If the input is taken in speech format, natural language processing is applied that recognizes speech using the Javascript Web Speech API. Acoustic models represent the relationship between an audio signal and the phonemes that make up speech, while the language model analyzes the human language model to predict words. Using these models, speech is reduced to sentences and then to words. The words are then matched against the ISL dictionary to generate the output. The text input procedure consists of removing the punctuation and reducing the sentence to words to generate the expected output. The output produced will then be animated using Avatar.

VI. MODULAR DIAGRAM

Figure 2 illustrates the modular behaviour of our Audio to Sign Conversion Model.

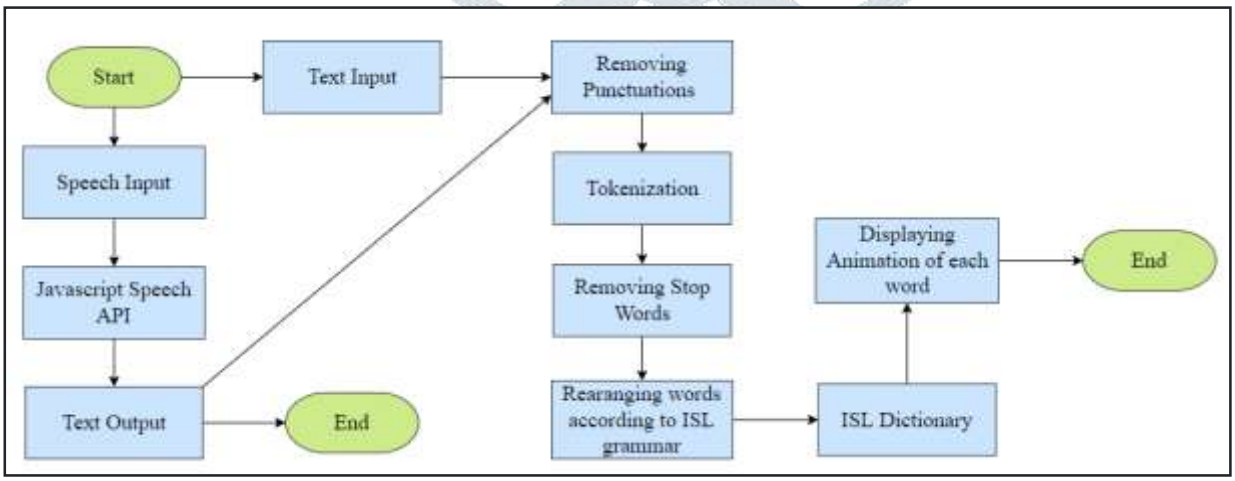

Fig 1. Block Diagram

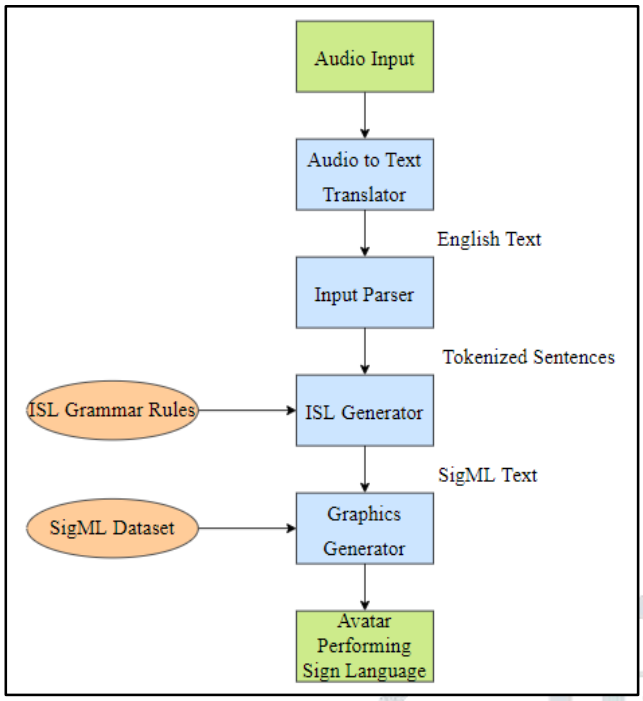

Fig 2. Modular Diagram

Fig 4. Taking Audio Input

VII. RESULTS

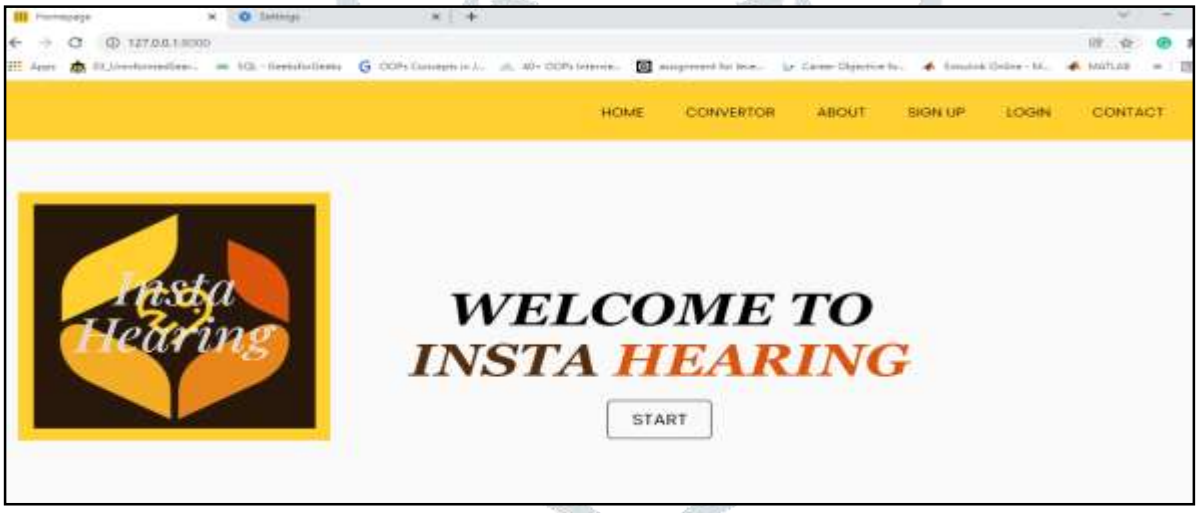

Fig 3. Homepage

Fig 5. Illustration of Sign Language

VIII. FUTURE SCOPE

1. The system includes basic hand movements. It can further be extended to generate all the possible hand movements and non-manual expressions thereby developing a system to automatically convert text or image to its SiGML form for generating animation of any given word.
2. Also, the context can be taken care of to improve the accuracy while converting the English sentences to ISL sentences.

IX. CONCLUSION

Sign languages, such as BSL and ASL, have a specific grammar that permits rule-based systems and syntax and semantic analysis to be performed to obtain the appropriate translation. In contrast, there is no specific grammatical rule in Indian Sign Language, which makes syntax and semantic analysis difficult since there are no rules to compare the English text with. As a result, an appropriate translation of the English text is not viable. In ISL, negative and interrogative sentences are represented by facial expressions. While the ISL animation for the verb clause is playing, the expressions in the sentence change to denote negation and interrogation. The system has so far not implemented this feature. The handling of directionality and discourse is kept simple.

X. REFERENCES

[1] Keluskar, Akshay, et al. "STORYTELLING APP FOR CHILDREN WITH HEARING IMPAIRMENT."

[2] Ankita Harkude, Sarika Namade, Shefali Patil, Anita Morey "Audio to Sign Language Translation for Deaf People."

[3] Dong, Qianqian, et al. "Consecutive decoding for speech-to-text translation." The Thirty-fifth AAAI Conference on Artificial Intelligence, AAAI. 2021.

[4] Vladyslav Tsap, Nataliya Shakhovska, Ivan Sokolovskyi "The Developing of the System for Automatic Audio to Text Conversion. "

[5] Lalit Goyal, Dr Vishal Goyal" Automatic Translation of English Text to Indian Sign Language Synthetic Animations."

[6] Jingdong Chen, Member, Yiteng (Arden) Huang, Qi Li, Kuldip K. Paliwal, "Recognition of Noisy Speech using Dynamic Spectral Subband Centroids" IEEE SIGNAL PROCESSING LETTERS, Vol. 11, Number 2, February 2004.

[7] Dynamic 3D Avatar Creation from Hand-held Video Input Alexandru Eugen Ichim∗ EPFL Sofien Bouaziz EPFL Mark Pauly