SPEECH/TEXT TO SIGN LANGUAGE CONVERTOR USING NLP

M. Kavya Mounika Research Scholar, Department of Computer science, Vignan's institute of information technology(A),Visakhapatnam
B. Hema Latha Research Scholar, Department of Computer science, Vignan's institute of information technology(A),Visakhapatnam
M. Sai Mounika Research Scholar, Department of Computer science, Vignan's institute of information technology(A),Visakhapatnam
J. Sumanth Research Scholar, Department of Computer science, Vignan's institute of information technology(A),Visakhapatnam (Affiliated to JNTUK Kakinada and approved by AICTE)
Dr. T V Madhusudhana Rao Professor, Department of Computer Science Vignan's institute of information technology(A),Visakhapatnam (Affiliated to JNTUK Kakinada and approved by AICTE)

Abstract
Communication plays an important role for any individual during interactions and people generally communicate using different languages. But challenged people with hearing and speaking disabilities may not be able to communicate like normal people. Such people communicate using sign language. There have been many resources accessible to make an interpretation of or to recognize gestures and convert them to text format, but there is a shortage of systems that are developed to convert these text messages to signs or gestures. This is mainly due to lack of resources and collection related to sign language. The main aim of our project is to create a system which will be able to translate the English text by using Indian grammar rules through a parsing module. The Indian language doesn't support word inflections, so we eliminate the stop words from the re-ordered text. Then, the words are compared with those in the dictionary and match to the videos representing these words. If the word was not found in the dictionary, it will be replaced by its corresponding synonym
Keywords Sign language, Natural Language Processing, JavaScript web speech API, Speech Recognition

1 Introduction

Ordinary people have difficulty communicating with a person with a hearing or speech impairment. To assist in a conversation between a normal and hearing impaired person, we use speech to have the translator sign. This makes interaction less complex and faster for everyday humans to present their ideas to hearing impaired humans. Translator converts text content into sign language using natural language processing algorithms. Accordingly, this system is used to overcome the challenges faced by normal human beings to communicate with hearing impaired humans and it is an ear for the hearing impaired. The model has systematic 4-step operations. Capture audio with PyAudio, convert audio to text with JavyaScript web Speech to Text API (using text tokenization and NLP concepts for text processing), and convert text to visual character word library (video dataset signing language), merge matching videos according to the order of processed text and display to hearing impaired.hardware and software.

2 LITERATURE REVIEW

In the article "Sign Language to Voice Translation ", K Gunasekaran modeled a system consisting of four modules: a sensor unit, a processing unit, a voice storage unit, and a wireless communication unit. This system is implemented by, which integrates the flow sensor and APR9600 with PIC16F877A. Users wear gloves with a flexible sensor that responds to hand movements. It provides the input to the microcontroller with the appropriate circuit response. Therefore, the microcontroller plays the audio recorded using the APR9600.

Oi Mean Foong introduced in the article "Voice-to-Sign Language Translation System for the Hearing Impaired in Malaysia," this project used template-based

recognition to convert voice to sign language. The V2S system was first trained using a voice pattern based on a common spectral parameter set. The database is used to store the spectral parameters as a template. Speech recognition is performed by comparing the input parameter set to the template stored in the database at and finally displaying the sign language in video,

P Vijayalakshmi gave a lecture entitled "Sign Language to Speech Conversion". In this task, we will develop a gesture recognition system that uses a flex sensor. The system recognizes the English alphabet and some words. The system was trained with 8 gestures and the experiment was run 10 times with an average recognition rate of 87.5%.

3 METHODOLOGY

3.1 OVERVIEW

This approach starts by capturing the input voice and converts it to text using the JavaScript web speech API**.** The resulting text is free text, so it consists of infected parts that will be removed in the next stage using the concept of natural language processing. Then, for each word / character in the processed text,use the visual sign word library to perform a mapping operation on the to successfully retrieve the matching video. These videos are then linked to form a single video that renders all the sign language text on the final display, as shown in in Figure .

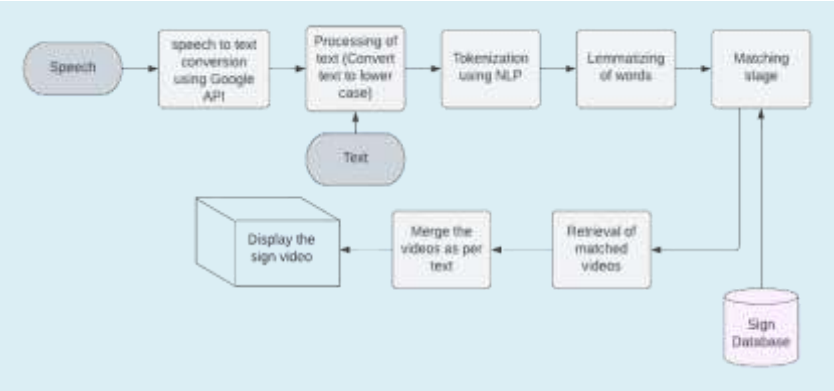

Fig: Flow Chart

3.2 SPEECH RECOGNITION

The microphone is used to record audio or speech and is taken as input to the system. The Packages in Python, PyAudio, On various platforms we use this to record audio. The JavaScript Web Speech Recognizer API is used to convert this recorded audio to text format. This API uses neural network models to convert audio format to text format. We also use JavaScript web Speech Recognizer to translate input audio into text format. In the case of large files of audio, the audio is divided into smaller parts. This break is done based on the occurrence of silence. Parts of these audio files are sent to JavaScript web Speech Recognizer, which converts them into text files.

3.3 SPEECH TO TEXT CONVERSION

Speech recognition is an interdisciplinary field that uses statistical modelling to specify natural language from a computer perspective. Speech recognition involves converting the above words into meaningful text. This involves recording a human voice and using a speech recognition algorithm to map it to exactly the word the speaker said. This process requires training the model to recognize human voice and spoken words using a pre-stored vocabulary of words and their mapped spoken versions. In this way, the computer creates a model that can accurately map the spoken language to the text by converting the spoken language into a meaningful audio signal. Once voice detection is done, the further speaker recognition architecture can be considered to work in three stages:

– Speech Feature Extraction
– Speech to Text Modelling
– Matching method

3.4 PRE-PROCESSING OF TEXT

Fill steps are used to fill in the blanks in the sentence and do not provide much context or structure to the text / sentence. Such complimentary words are also present in the English language and do not greatly affect the context of the sentence. So, the application removes these complementary.

3.5 LEMMATIZING ALGORITHM

Lemmatization is used in natural language processing (NLP) and machine learning. This is also done using the steaming algorithm. In both stemming and lemmatization, we try to reduce the given word to its original word. But we have some differences in choosing the lemmatizing algorithm over the steaming algorithm.

In stemming, some piece of the word is removed at the end of the tail of the word, but they don't actually know the definition of the word in the language to which it belongs. This knowledge is known to the lemmatizing algorithm. The lemmatizing algorithm refers to the dictionary to understand its meaning before reducing it to the root word (lemma) of the word. Although lemmatization takes a lot of time, it is far more accurate than the steaming algorithm results. There is also a computational overhead for lemmatization, however, in machine learning, computational resources are rarely of concern. Therefore these are the reasons for choosing the lemmatization algorithm over the steaming algorithm.

3.6 TOZENIZATION OF TEXT USING NLP

Natural Language Processing, popularly known as NLP has brought a revolution in the computation world by providing a means through which computers can understand the human language. When the Javascript web Speech To Text API operates in the first stage, an output text is generated from the input audio signal. This text undergoes tokenization to separate each word individually which may include various forms of the word as per the grammar rules but the root word remains the same such as growing and grown contain the same root word grow but used differently in English. Therefore, Stemming and Lemmatization come into the picture. Both these techniques are normalization techniques for text or words. python NLTK or Natural Language Toolkit Package. Stemming and lemming help remove the infected part of the text or document in , leaving only the stem and root of the word. Porter Stemmer is working on text stored at this feature level in the NLTK package.

3.7 MATCHING WITH VISUAL SIGN WORD LIBRARY

For each word / character in the processed text retrieved after the second stage of the application, perform a match using the tags in the video's visual character word library that resides in the character database. If a match is found, matching videos are retrieved from the signature database and moved to the desired location.

3.8 MERGING VIDEOS AS PER TEXT AND FINAL DISPLAY

The framework repeats each word in the message / sentence that has been acquired from the past and seeks out communication video successes based on comparable gestures in a nearby framework. Assuming the word is found, the framework displays the result as a video grouping.

OpenCV is an open source library used for many other features related to picture handling, video examination, and PC Vision. It represents an outline of video succession through Outline. In the event that the word is not found in a nearby framework, the framework will search for the word in a gesture-based communication vault called "Indian Sign Language Portal".

At this point, when you use the word "what is your name" in the mouthpiece, the message is completely changed to "what is your name", speeding up the removal of the filled words.

The result that accompanies it shows each video in the grouping as follows -

Fig. 2(a),2(b) Display the output sign languages for the given phrase "Hello world".

Fig-2(a) output sign-Hello

Fig-2(b) output sign-World

4 PERFORMANCE ANALYSIS

The model is tested with 150 audio input the final result obtained with the samples where the text conversion accuracy was better in Online mode where java web speech API was used than in Offline mode. The matching accuracy is better in Online mode due to the vast video dictionary by Hand speak available online and the average total delay in operation is almost same as the online mode searches the entire dictionary and the offline mode has to enter into sub-word searching until a character is left for the word in case the original word is not found.

5 RESULTS

The user interface is created using HTML,CSS and JavaScript. The web pages are designed with main heading and buttons on Navigation bar to redirect to different pages such as home page, login/signup and convertor .These pages are shown as screenshots in below figures fig 5.1, fig 5.2 and fig 5.3 respectively

5.1 Home Page

Below figure shows the front cover of proposed system.

Fig: 5.1

5.2 Sign Up Page

The user need to register before making use the services ,below figure shows website of sign up page.

Fig: 5.2

5.3 Input page(Convertor)

Below web page shows the main page which takes input from user and plays the role of convertor.

Fig 5.3

5.4 Output Page

This page shows the test results which is an animated video of respective sentence (input).

Fig 5.4

6 CONCLUSION

This paper helps deaf and dump people to communicate with normal people with an ease. This paper helps layman who doesn't know the sign language. In the first part we used speech processing methods to process the speech, it takes input as and converts into text. In the second part is processed and displayed as video in sign language. This removes the communication barriers.

7 REFERENCES

[1] J.P. Bonet. "Reducci_on de las letras y arte para ense~nar a hablar a los mudos", Coleccion Cl_asicos Pepe. C.E.P.E., 1992.

[2] William C. Stokoe. Sign Language Structure [microform] / William C. Stokoe. Distributed by ERIC Clearinghouse, [Washington, D.C.], 1978.

[3] William C. Stokoe, Dorothy C Casterline, and Carl G Croneberg. "A Dictionary of American Sign Language on Linguistic Principles" Linstok Press, [Silver Spring, Md.], New Edition, 1976.

[4] Bridle, J., Deng, L., Picone, J., Richards, H., Ma, J., Kamm, T., Schuster, M., Pike, S., Reagan, R., "An Investigation of Segmental Hidden Dynamic Models of Speech co-articulation for Automatic Speech Recognition.", Final Report for the 1998 Workshop on Language Engineering, Center for Language and Speech Processing at Johns Hopkins University, pp. 161, 1998.

[5] Ma, J., Deng, L., "Target-directed Mixture Linear Dynamic Models for Spontaneous Speech Recognition", IEEE Transactions on Speech and Audio Processing, Vol. 12, No. 1, January 2004.

[6] Mori R.D, Lam L., Gilloux M., "Learning & Plan Refinement in a Knowledge Based System for Automatic Speech Recognition", IEEE Tra. on Pattern Analysis Machine Int., 9(2):289-305, 1987.

] Pavlovic, V, Sharma, R., &Huang T. (1997). Visual Interpretation of Hand Gestures for Human-Computer Interaction (HCI): A Review. IEEE TOPAMI, VOL. 19, NO. 7.

[7] Vogler, C.M., &Dimitris (2004). Handshapes and Movements: Multiple-Channel American Sign Language Recognition Gesture-Based Communication in Human-Computer Interaction. Ed. A.V.Camurri, Gualtiero. Vol. 2915: Springer Berlin Heidelberg. 247-258.

[8] Gunasekaran, K., &Manikandan, R. (2013). Sign Language to Speech Translation System Using PIC Microcontroller. International Journal of Engineering and Technology (IJET), Vol 5 No 2.

[9] Kalidolda, N., &Sandygulova, A. (2018). Towards Interpreting Robotic System for Finger spelling Recognition in Real-Time. HRI Companion: 2018 ACM/IEEE International Conference on Human-Robot Interaction Companion.

[10] Liang, R., & Ouhyoung, M. (1998). Real-time continuous gesture recognition system for sign language. Proc Third. IEEE International Conf: on Automatic Face and Gesture Recognition, pp. 558-567.

[11] LeCun, Yann, Lon, Bottou, Yoshua, Bengio, & Haffner, Patrick (1998). Gradient-based learning applied to document recognition', Proceedings of the IEEE 86, vol. 11, pages 22782324.

[12] Simonyan, Karen, & Zisserman, Andrew(2015). Very Deep.... Recognition. Conference paper ICLR.

[13] Ankita Harkude, Sarika Namade, Shefali Patil, Anita Morey "Audio to Sign Language Translation for Deaf People" ISSN: 2277-3754, International Journal of Engineering and Innovative Technology (IJEIT) Volume 9, Issue 10, April 2020.

[14] Farahanaaz Shaikh, Shreya Darunde,Nikita Wahie, Swapnil Mali "Sign Language Translation System for Railway Station Announcements", Institute of Electrical and Electronics Engineers (IEEE), IEEE Bombay Section Signature Conference (IBSSC), 2019.

[15] Vajjarapu Lavanya, Akulapravin, And M.S., Madhan Mohan "HandGesture Recognition And Voice Conversion System Using Sign Language Transcription System" ISSN:2230-7109 (Online) | ISSN:2230- 9543 (Print) IJECT Vol. 5, Oct-Dec 2014.