Audio to Sign Language Conversion

Shivangi Saharia[1], Priyanka Kulkarni[2], Amardeep Bhagat[3], Himalaya Prakash[4], Amber Kesharwani[5]

B.E. Final Year Student, Computer Science and Engineering, Government Engineering College, Bilaspur, India[1-5]

Abstract: Audio to video conversion consists of the development of a contributory software for hearing aided people, Audio to video conversion records audio and produces sign language video of that audio. The goal of this application is to create the foundation for a spatial platform for aided people. The application has a user-friendly interface. A folder will consist of videos of alphabets and numeric sets, to convert each line into the sign language word by word. It is a standalone software for personal computers so no user database is maintained and no registration is required, but data of their previously converted audios is maintained. A line on the screen will appear prompting the user to give the input audio which has to be converted, the user can either say it or play the audio, you can also use earphones to give the audio input. This application records the audio maintaining a minimum threshold of audio using python packages like pyaudio and wave. The application will process the audio using the python library called speech recognition which will convert audio into text. Using the requests module of python we extract the video files from an online platform handspeak.com, moviepy.editor is being used for editing the video for giving the desired output. The python module shutil is used to copy the extracted video file into a folder called cache where the user can access the converted audio to video.
"Audio to video conversion" is developed using python and it's advanced libraries. A simple atom editor is used for coding and importing the required modules.

Keywords: Interface, Standalone, Threshold, pyaudio, Shutil.

I. INTRODUCTION

Sign language is a language that uses a visual way of conveying meaning. Sign language uses various hand, arm, facial or body gestures for conveying the meaning rather than using the sound. Sign languages are natural languages with their own grammar, gestures and patterns. Many kinds of sign language are available around the globe. they differ from one another from one geographic region to the next. American Sign Language (ASL), British Sign Language (BSL), French Sign Language (FSL), etc. are amongst some of them. Sign language is considered to be one type of nonverbal communication. For the community of deaf people, sign language has become an easy way to communicate with them. Whenever vocal communication is not possible sign language fills the gap.

The basic purpose of this project is to improve communication with the people having hearing difficulties or disabilities. Our project takes the audio inputs using the pyaudio module of python then this audio is converted into text using google speech API and then this text is converted into sign language video using the python platform. the project intends to provide ease of communication with people having a hearing disability.

Different modules that are been used are:

PYAUDIO: portaudio binding can be done using python module i.e pyaudio. With Pyaudio , python is been used across various platforms to play and record audio. It is cross-platform audio I/O library

MOVIEPY: Video editing in python can be done using the moviepy module of python. The basic operations include concatenation, cuts, insertions, etc. most of the common formats can be read through this module.

SPEECH RECOGNITION: It is a library used for speech recognition, with user support of various engines APIs like

- Google Speech Recognition
- IBM Speech to Text
- Google Cloud Speech API
- Microsoft Bing Voice Recognition
- Houndify API
- SNOWBOY Hotword Detection

GOOGLE SPEECH RECOGNITION API: speech recognition application program interface (API) converts spoken text or speech into written text

II. LITERATURE REVIEW

From an overview of the existing systems in this technology from the researches being happening on the same topic as ours, we can say that this type of technology can also help deaf and dumb people as well as learning the sign language to

communicate with ease with such people. In [1] paper, as per Pooja Balu Sonwane and Anita Nikalje's research, they used raspberry pi and python to take a textual input in Hindi language and convert it into the video, explaining the text in sign language. They are using a database for matching those texts. The use of databases is giving them the liberty to use the system for not only converting the text into sign language but also sign language to text using gesture recognition from python libraries. Our system aims to convert the speech or audio to text and then that text to a video of sign language directly, it is a real-time system and uses python libraries to fetch sign language videos from google and thus uses English which can be understood mostly on a universal level.

In [2] paper, as per Amit Kumar shinde on his research of converting Marathi sign language to text and Marathi text to sign language, he mentions how important this field of research is and how much it is growing state and language-wise. It makes the lives of not only deaf and dumb people easier but also the lives of people dealing with such people become much easier. This type of system can not only help you understand them in real-time but also help you learn it. Just like Stephen Hawking's talking computer, which converted the texts he inputted into voice and this helped him in carrying on normal communication and sharing his wisdom with the world longer than his disease allowed him to. Imagine how many deaf or dumb people can be facilitated with this technology, not only in the sector of communication but also in sectors of learning such as schools and colleges. They can convert the voice coming from their teachers into sign language gestures in real-time which will make their education environment a lot easier.

In some research done from [4] geeks for geeks, we have seen that the audio to video conversion can be more easily implemented using NLP(Natural language processing) and pyaudio is a very important module for taking the audio input.

In [3] paper by Shweta Doura, Dr. M.M. Sharmab, they conducted a report study of recognition of Indian sign language using Sugeno type fuzzy neural networks can be useful for this community and the people developing applications for them. The deaf people are not able to use acoustic means for communication and thus have to depend on the sign language for their communication and this gives rise to another barrier in their communication and that is understanding of sign language. Not everyone understands it or can explain it, this is where projects like these come into play. Their project of converting the Indian sign language to text and vise-versa gives the opportunity of expanding it to other languages as well.

III.   ARCHITECTURE AND WORKING

A. Technological Stack: The various technological stacks required for this project are:

1) **Device**: The device on which the program is to be executed needs to have a working microphone, decent memory specification and processor, working internet.

2) **Speech Recognition API**: Any of the available speech recognition API's can be used for audio to text conversion. Several speech recognition engines/APIs are:
- Microsoft Bing Speech Recognition
- SNOWBOY Hotword DetectionGoogle
- IBM speech to text
- CMU sphinx
- Google Cloud Speech API
- Google Speech Recognition

3) **Dedicated Libraries:** It includes python/python3 inbuilt libraries like:

- pyaudio : Used for recording and processing audio.
- speech recognition: Used for audio to text conversion.
- moviepy : Used for video processing.
- Requests: Used to link the client with the video source.
- OS : Used for system-related manipulations.
- wave : Used to save the audio file in .wav format.

B. **Workflow:** The workflow of the project can be broadly categorised into three phases namely:

1) **Audio Recording and Processing:** The audio is recorded using the inbuilt microphone(*threshold=6000*) or headphone jack(*threshold=1300*) of the user's device. This process ids carried out using the inbuilt library package 'pyaudio'. The desired specifications of the audio could be defined based on the requirements in terms of format, bit rate and chunk size. A stream is opened to do so using the pyaudio package and byte order is used to save the recorded audio. The recorded audio is then processed by trimming the silence before and after the recorded audio in each chunk of data received. The resulting data is then saved in .wav format using wave package and os library.

2) ***Allow adjusting of the ambient noise***: Since this noise is in the surrounding varies, the program is allowed to adjust its energy threshold according to the external noise level.

3) ***Speech to Text Conversion***: The saved audio file will then be fed to the Speech Recognition API using the speech recognition package. Various APIs could be used to do so such as Google Speech Recognition, IBM Text to Speech, Microsoft Bing Voice Recognition, etc. and a stable internet connection is required to work with these APIs. Google Speech Recognition is user-friendly and easy to use. The speech recognition API then converts the audio to the text which is then used for the further process of converting the text to the audio.

4) ***Text to Video Conversion***: The text file is further processed where each word is separated and converted into video. The sample sign language videos are taken from the website "https://www.handspeek.com" in real-time. Each word is separately converted into a video using the moviepy.editor package's functions. The chunk size of the word could be set according to the memory specifications of the device. The video files of each of the words are finally merged using moviepy's inbuilt library functions.
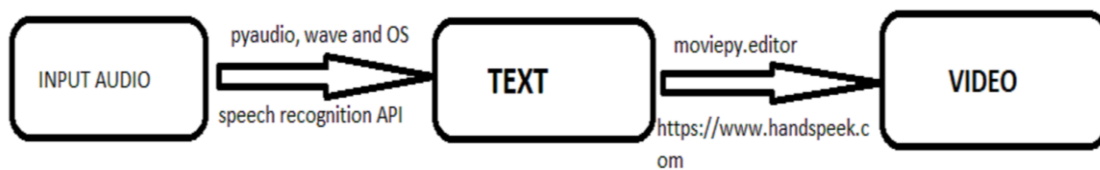

Figure 1 Workflow diagram

Figure 1 shows the block diagram representation of the proposed system. The system takes speech input through the inbuilt microphone or headphone jack of the device, this process is been done by the python library pyaudio. This audio is then processed with the help of wave and os packages and converted into text with the help of speech recognition API of google. The recorded stream is then checked for silence at the start and the end, if found then it is trimmed from every chunk. Further, the stream is adjusted according to the ambient noise level and for the same reason, the energy threshold is adjusted. Speech recognition API converts the audio to text. The text file generated is processed where each word is separated and converted into video. The rest of the video concatenation and further video editing is done by the moviepy.editor package which divides the video into chunks and gives the output.

IV.   CONCLUSION

There are a lot of communication barriers between normal people and people with hearing disabilities. This project aims at devising a useful tool for disabled people so that the communication barrier can be reduced. This project is an attempt to make it easy for deaf people by converting speech into sign language video. The project displays an equivalent sign language symbol for the speech given as input. The proposed system has more possibilities in the future. It could be converted and used for multilingual speech. Even this system can be used in airports, hospitals, courts and several different places for the convenience of disabled people

ACKNOWLEDGMENT

REFERENCES

[1].   Text to sign language conversion by using python and database of images and videos by pooja balu and anita nikalje.
[2].   Amit kumar shinde and ramesh khagalkar "sign language to text and vise versa recognition using computer vision in Marathi".
[3].   Shweta Doura, Dr. M. M. Sharmab "the recognition of alphabets of Indian language by sugeno type fuzzy neural network"
[4].   https://www.geeksforgeeks.org/project-idea-audio-sign-language-translator/

BIOGRAPHIES

Ms. Shivangi Saharia pursuing Bachelor of Engineering [2016-2020] with major in Computer Science and Engineering from Government Engineering College, Bilaspur (C.G) Affiliated to Chhattisgarh Swami Vivekanand Technical University, Bhilai (C.G). Proficient Web designer and Python developer.

Ms. Priyanka Kulkarni pursuing Bachelor of Engineering [2016-2020] with major in Computer Science and Engineering from Government Engineering College, Bilaspur (C.G) Affiliated to Chhattisgarh Swami Vivekanand Technical University, Bhilai (C.G). Proficient Python developer.

Mr. Amardeep Bhagat pursuing Bachelor of Engineering [2016-2020] with major in Computer Science and Engineering from Government Engineering College, Bilaspur (C.G) Affiliated to Chhattisgarh Swami Vivekanand Technical University, Bhilai (C.G). Proficient in C++ and Web development.

Mr. Himalaya Prakash pursuing Bachelor of Engineering [2016-2020] with major in Computer Science and Engineering from Government Engineering College, Bilaspur (C.G) Affiliated to Chhattisgarh Swami Vivekanand Technical University, Bhilai (C.G). Proficient Python developer.

Mr. Amber Kesharwani pursuing Bachelor of Engineering [2016-2020] with major in Computer Science and Engineering from Government Engineering College, Bilaspur (C.G) Affiliated to Chhattisgarh Swami Vivekanand Technical University, Bhilai (C.G). Proficient in C and Java.