Emotion Detection Using Speech Analysis

Riya Tyagi[1], Anmol Agarwal[2]

[1, 2]Dept. of Computer Science and Engineering, Graphic Era University, Dehradun, Uttarakhand, India-248002

Abstract— The functionality to handle the audio file consist of human voice and generate speech, is one of the features which make humans different from other living beings. The human speech can be differentiated by some properties. In this paper we detect the basic emotion of human through speech analysis [1]. As we know that today 95% communication is based on vocals which shows different characteristics of human and emotion is one of them which shows us attribute like fear, anxiety, happiness, sadness and angry etc.[3]. Hence, voice and speech analysis causes emotion inside human and very beneficial in different areas of communication like when a victim has been interviewed, he start stammering if he tells lie and in this case we detect victim's lie with the help of emotion analysis. In politics context we use emotion analysis to detect the emotion of minister at different time laps and get the variety of emotion throughout the speech. Our main objective of this paper is to develop such system which is helpful in near future real time systems and improved human-technology interaction.

Keywords— Emotion, rate, frequency, pitch, chunks.

I. INTRODUCTION

Today human can easily communicate and exchange the thoughts through each other in many ways like via signals, speech and networking etc. and speech is most useable method for conversation which is 95% used presently[4]. Human conversation through speech contains so many emotions and this is the reason why emotion detection study takes attentions of all. In this study main focus is on the speech of the human and detects the emotion of his/her speech[5]. Emotion is nothing but the inner feelings of human which he/she actually wants to express to other person through speech like neutral, happiness, fear, anger or frustration etc.

According to study human communication is largely based on what they want to speak[7]. So, it is an important task to detect the emotions using human speech. So, here we are introducing our approach of emotion detection using speech analysis which is user friendly, automatic, efficient with very low service requirement and ability to detect all the emotions throughout whole conversation at different time event with the help of audio chunks. This approach develops in python programming language on pycharm and recorded the speech of human through microphone.

II. OBJECTIVE OF PROJECT

Emotion detection is one of the important and major task which helps in many areas to identify the emotion. So, we made a system which has maximum resemblance with human functioning of emotions. It can perform real time emotion detection through speech.

The main challenge in this project is to develop such system which can detect the emotion in such a manner so that it is time saving, efficient in all test cases and user friendly etc.

We all knew that the probability of making a system as accurate as human is very low so we take it as our main challenge to make such system with high accuracy in those tasks which human can perform easily [6].
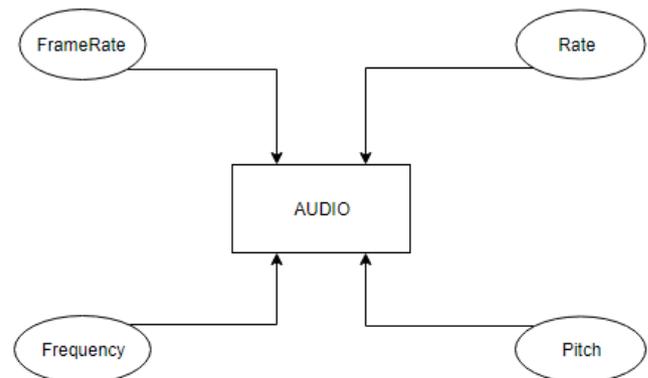
Fig. 1. Attributes of an audio file.

Another secondary objective of the project is to decrease the human effort as much as possible which is very helpful in near future in different areas and we can also record the video of a person and detect the emotion in his/her voice which is a difficult task to do it because of auto frame-rate. Auto frame-rate is different in both cases either we record the audio or record the video and then calculate the emotion. So, it is another objective to solve this problem so that we can add functionality in our system or project. This approach helps us in future like in speech of minister; we can record the whole speech and detect the emotion at different time event which shows all the emotions of minister throughout the speech.

III. CONSTRUCTION

As we detect the emotion using speech or voice of the human, we have to first take the voice as input with the help of microphone and store the voice as .wav file. WAV is a basic extension of audio files. Our system is works in two modes-

1) Record the voice through microphone.
2) Browse the .wav file from the system.

The voice received as dataset in a recorded form which will contain different characteristics like rate, frame-rate, frequency and pitch etc[10]. and perform different operation to calculate or detect the emotion like separate the original voice and white voice or background voice because our project work on original voice. After storing the file, we open the audio file in rb mode and calculate some characteristics using python modules. We calculate the pitch of the audio which gives us the overall emotion of the human speech by comparing the pitch value on different emotions. For calculating the emotion

at different time events, we divide the audio file in equal size of chunks at equal time- period.
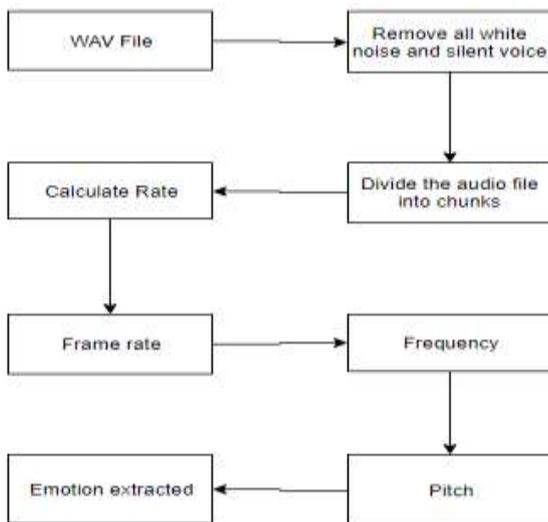
Fig. 2. Control flow diagram of our system.

For detecting the emotion from the video, we convert it into a wav file with the help of command prompt of the system because it is not possible to perform operations on the sound of the video which consist so many other sounds or vibrations. So, we extract the audio from the video file and save it as wav audio file. After conversion of video to audio through command prompt we divide resultant audio file into equal size of chunks for calculating the emotion at specific time event and remove all the other white voices from the audio to get the original voice of the person. We also remove background human voices with the help of MFCC voice pitch and frequency spectrum [8]. It increase the accuracy and reliability of our system or project.

IV. WORKING

Working of our project is very simple and easy to understand to anyone. Our project is work on python programming language[11]. We use some python modules to handle the speech of human. In this project we take the human voice as input of our project which we can record with the help of microphone and save the recorded speech as .wav file or browse the .wav file from the system.

w= pyaudio.Pyaudio()

After getting the voice as .wav file we remove all the white voice or background voice from the original voice.

Resultant voice= Original voice – white voice

And save the resultant voice as .wav file. Now, we open the file in rb mode and perform operations on file as following-

wave.open('filename.wav','rb')

Now, calculate the rate of audio file.

r= w.getframerate()
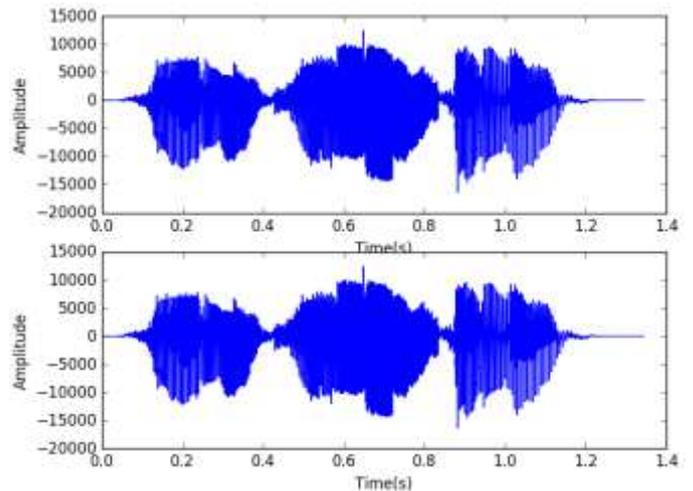
Fig. 3. Graphical representations of rate and time.

Now, divide the audio file into equal size of chunks for different time event.

C_name=”chunk{0}.wav”,format(i)

From the rate we calculate the frequency at different time of the audio.
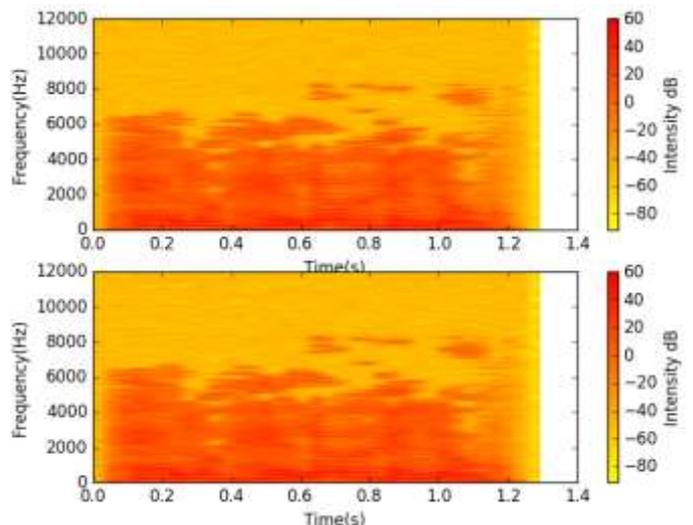
Fig. 4. Graphical representations of frequency and time.

For detecting the emotion we have to calculate pitch of the audio and for calculating the pitch of the audio we use pitch and frequency relationship formula.[2]

Pitch= 69+12*ln2(frequency/440)

After calculating the pitch we use following table to detect the emotion-

Emotions	Pitch Range
Aggression	-110-100
Happiness	195-200
Graceful	-105-100
High	95-105
Despair	130-140
Chilled	120-130
Annoyed	150-160
Solemn	110-120

For converting the video into audio wav file we use command prompt command of ffmpeg-

```
cm="ffmpeg -i VID-20171218-WA0008.mp4 n4.mp3"
```

this command convert mp4 which is an extension of video file into mp3 file which is an extension of audio file. After conversion we convert mp3 file into wav file for performing the operations on wav file and shows the output on the screen.

V. CHARACTERASTICS

- Our project has less complexity.
- It is user friendly.
- It is time-saving and run at real time.
- It is efficient in every test case.
- It detects the emotion at different time events.
- It is approximately 75% accurate.
- It can record and also browse the file from the system.
- It can also use MFCC feature to detect the emotion [9].
- It excludes all the white noise from the audio. So, that we get the emotion of the original voice.
- It calculates different attributes of an audio.
- It detects 7 different type of emotions.

VI. CONCLUSION

Our project is giving the result with approx. 75% accuracy and with a reliable speed in real time. It helps in many areas of detection. Because this is a very difficult task to detect the exact emotion of human while speaking or in real time. So, we develop such system which makes this task possible in reliable approach and give the result in accurate time. Our system give the result at different time event. So, our system is very beneficial, user friendly and reliable in near future technology of real time analysis systems.

REFERENCES

[1] S. Casale, A. Russo, G. Scebba, and S. Serrano, "Speech emotion classification using machine learning algorithms," *IEEE International Conference on Semantic Computing*, 2008.

[2] Rohit S. D and Pallavi S. Deshpande, "Neutral to emotional speech conversion by pitch contour modification for Marathi," *International Journal of Engineering Research & Technology (IJERT)*, vol. 3 issue 5, pp. 2228-2231, 2014

[3] R. Banse and K. R. Scherer, "Acoustic profiles in vocal emotion expression," *Journal of Personality and Social Psychology*, vol. 70, pp. 614-636, 1996.

[4] T. Bänziger and K. R. Scherer, "The role of intonation in emotional expression," *Speech Communication*, vol. 46, pp. 252-267, 2005.

[5] F. Yu, E. Chang, Y. Xu, and H. Shum, "Emotion detection from speech to enrich multimedia content," *Lecture Notes in Computer Science*, vol. 2195, pp. 550-557, 2001.

[6] V. A. Petrushin, "Emotional recognition in speech signal: experimental study, development, and application," *ICSLP-2000*, vol. 2, pp. 222-225, 2000

[7] M. E. Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognition*, vol. 44, pp. 572–587, 2011.

[8] P. Y. Oudeyer, "The production and recognition of emotions in speech: features and algorithms," *International Journal of Human Computer Studies*, vol. 59, pp. 157–183, 2003.

[9] T. Iliou and C. N. Anagnostopoulos, "Statistical evaluation of speech features for emotion recognition," in *Fourth International Conference on Digital Telecommunications*, Colmar, France, pp. 121–126, July 2009.

[10] Y. Wang and L. Guan, "An investigation of speech-based human emotion recognition," In *IEEE 6th Workshop on Multimedia Signal Processing*, New York: IEEE Press, pp. 15–18, 2004.

[11] Y. Zhou, Y. Sun, L. Yang, and Y. Yan, "Applying articulatory features to speech emotion recognition," in *International conference on research challenges in computer science, ICRCCS*, pp. 73–76, 28–29 Dec. 2009.