Recognition of Human Emotion by Speech Processing

R.D. Bodke\(^1\)*, M.P. Satone\(^2\)

\(^1,2\)Dept. of Electronic And Telecommunication, K. K. Wagh Institute of Engineering Education And Research Nashik-422003,Maharashtra, India

*Corresponding Author: roshanibodke1995@gmail.com

Available online at: www.ijcseonline.org

Accepted: 14/Oct/2018, Published: 31/Oct/2018

Abstract— The emotion recognition from speech is used for in human computer interaction. Most of researchers doing research on emotion recognition using speech signal. This project attempts language emotion recognition using speech signal of English language. The emotional speech samples are stored in database and used for Training And Testing. The feature extraction MFCC, PSD and Pitch detection algorithms are used. For classification of different emotions like Angry, Happy/Joy and Normal state SVM classifier is used. The all steps are implemented using MATLAB software. Raspberry pi is used for detection of emotion on hardware This classified emotions can be used for various application areas like medical, security, military etc.

Keywords—MFCC, PSD, SVM etc.

I. INTRODUCTION

The speech is very important factor of human to express the emotions. All emotions like Happy, Angry, Sad are helps to connect the people. The Speech signal is nothing but the electric signal varies with time and frequency both with help of speech the emotion detection for human computer interface[1]. The emotion detection from speech signal is a relatively interesting field of research, This field have various applications. The human computer or human-human interaction systems, emotion recognition systems could provide users with improved services by being adaptive to their emotions. In real worlds, emotion recognition could help simulate more realistic avatar interaction. [2].There are two different process has to be done by speech emotion recognition first one is to train the dataset and second one is to testing the speech emotion. If these two results are matching then is said to correct emotion detection by computer.

In this paper attempts the training and testing of database of emotion of English language samples and get the result for accessing the high accuracy.

II. LITERATURE REVIEW

Speech features are influenced due to emotis like Quality,Spectral energy which related to articulation rate, fundamental frequency\(f_0\),TEO-CB(Teaser Energy Operator) to classify the emotions[3].The speaker recognition is the process the MFCC feature having high accuracy to recognised the speaker with its signal characteristics[4].LSTM based modelling is very useful for childern speech recognition. DNN-HMM modelling required less coefficient for classification and low resource task maneger[5].For male and female speech recognition the PLP,LPC and MFCC feature techniques are benificial to get grate performance [6]. Uttrance, Pitch, Phoneme are properties of speech and they are useful for speech recognition in this training and testing the data base are helps to get exact emotion[7]. Noisy speech training in MFCC-based speech recognition system helps to noise suppression toward robot-assisted autism therapy[8].

III. PROPOSED METHOD

There are three basic steps for speech processing filtering, feature detection and classification.

![Fig. 3.1 Training Dataset](image)

In this project The design system deals with speech emotion recognition and improve the result of MFCC feature and Pitch extraction. This whole method is divided into two parts.
training and testing is done by MATLAB software. In fig 3.1 estimate that each block performs different operations on speech signal. There are main three stages of speech emotion recognition system there are filter signal, feature extraction and classification.

In Database there are three emotion speeches are Angry, Happy and Normal etc. The English language emotions of males and females are used by taking audio clip of different scenes of expression. There are total 90 samples 30 for Angry, 30 for Happy and remaining are Normal. For training used 80 samples and 10 for testing.

A. Training The Dataset
In this method first covert all speech signals into .wav file format so that MATLAB can easily plot the signal in time domain with help of waveread command.

1) Filter
The pre- emphasis filter is used for to remove the noise most of noise present in speech signal so pre- emphasis filter helps to increase the amplitude of frequency bands with respect to other frequency so that it improve the signal to noise ratio(SNR)[9]. The following equation (1) is used for calculating pre-emphasis filtered signal.

\[
y(ns) = x(ns) - ax(ns)
\]

The y(ns) is output signal and x(ns) is input speech signal. a is set to be 0.97.

2) MFCC Feature
The first step in speech recognition system is to extract features means to identify the basic component property of the audio signal that are good for identifying the linguistic content. Sounds generated by Humans are filter by vocal tract including teeth, tongue, cheeks etc. The vocal Throat microphone captured the vocal fold vibration to capture the exact changes in word [10]. By taking the 16 KHz sampling frequency to sampled the speech signal. So framing the signal using hamming window with 25ms frames and each frame having the 0.025*16000=400 samples.66% overlapping is allowed that means 264 samples are overlapped. So that first frame is starts with 0 to 400 next one 401 to 800 and so on .After this plot frequency plot of each sub-frame using FFT.

a) Applying the Mel scale filterbank
Incorporate this scale makes features match closely with humans hear. The formula for convert from frequency to Mel scale is given by equation (2):

\[
M(f) = 1125 \ln(1 + \frac{f}{700})
\]

Where the f is sampling frequency.

To Compute the Mel-spaced filter bank. This is a set of 24 triangular filters that are apply to power spectral signal FFT.

In figure 3.2 filter bank gives an output in form of 24 vectors of length 256. Each vector is largely zeros. Take the each log of all 24 energies from equation (2). For now to create proposed filter bank for this project the first filter bank will start at the first filter coefficient reach its max value the second filter coefficient is return to zero at the 3rd filter coefficient. The second filter bank will start at the 2nd, reach its max at the 3rd, then be zero at the 4th etc is shown in figure 3.2. A formula for calculating these is as follows:

\[
h_y(k) = \begin{cases} 
0 & k < f(r - 1) \\
\frac{k - f(r - 1)}{f(r - 1) - f(r - 1)} & f(r - 1) \leq k \leq f(r) \\
\frac{f(r) - k}{f(r + 1) - f(r)} & f(r) \leq k \leq f(r + 1) \\
0 & k > f(r + 1)
\end{cases}
\]

Where \( r = 24 \) number of filter and \( f() \) is list of \( m + 1 \) mel-spaced frequencies. This plot is shown in figure 3.2.After this apply the Discrete Cosine Transform (DCT) of all 24 energies which gives 24 spectral coefficients. These coefficients are given to classifier.

3) Pitch Detection
Pitch is nothing but the fundamental frequency of speech signal. The Zero-crossings rate method is used for to divided voice and unvoiced signal from speech [11][12].Equation (4) is used for calculating fundamental frequency.

\[
Z_n = \sum_{m=-\infty}^{\infty} |\text{sgn}[x(m)] - \text{sgn}[x(m - 1)]| w(n - m)
\]

w(n) is windowing function here used Hamming window.

4) Power Spectral Density
Power spectral density is used for calculating the power of speech signal as function of frequency. By simply calculating by given formula,

\[
P = \lim_{T \to \infty} \frac{1}{T} \int_0^T |x(t)|^2 dt
\]

The x(t) is input speech signal. Power Spectral Density in this project is given by figure 3.4.
Testing is similarly done as Training process but for testing different samples are taken there are additional 10 samples used for testing. Figure 3.5 shows flow of testing the speech emotion signal.

Both Training vector set and Testing vector set are given to SVM classifier.

C. SVM Classifier

The classifier is basically classify the data into different classes with compared each of weight with one to one class or one to others class[13]. In MATLAB svmclassify function is used for classification of emotion. The svmclassify function uses results from svmtrain to classify vectors \( x \) according to the following equation:

\[
    c = \sum_l a_l k(s_l, x) + b
\]

Where \( s_l \) support vectors, \( a_l \) are the weights, \( b \) is bias and \( k \) is kernel function. If \( c \geq 0 \) the \( x \) is classify as member of first group otherwise classify as other group.

The feature extraction the MFCC is give more accurate result for estimating emotion. The Feature PSD and Pitch are also helps to improve performance and accuracy of system. Our methods are reasonably accurate at recognizing emotions of all speakers. The SVM classifier is helps to classify emotion with various sizes of coefficients sets.

V. CONCLUSION

The feature extraction the MFCC is give more accurate result for estimating emotion. The Feature PSD and Pitch are also helps to improve performance and accuracy of system. Our methods are reasonably accurate at recognizing emotions of all speakers. The SVM classifier is helps to classify emotion with various sizes of coefficients sets.

REFERENCES


