

## Recognition of Human Emotion by Speech Processing

R.D. Bodke<sup>1\*</sup>, M.P. Satone<sup>2</sup>

<sup>1,2</sup>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](http://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 fequency( $f_0$ ). 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 manger[5]. For male and female speech recognition the PLP, LPC and MFCC feature techniques are beneficial to get grate performance [6]. Uttrance, Pitch, Phoneme are properities 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.

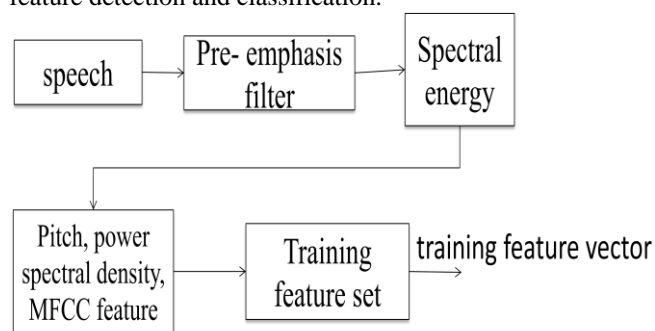


Fig. 3.1 Training Dataset

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 convert all speech signals into .wav file format so that MATLAB can easily plot the signal in time domain with help of wavread 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.

$$y(ns) = x(ns) - ax(ns)$$

(1)

The  $y(ns)$  is an 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

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

$$M(f) = 1125 \ln\left(1 + \frac{f}{700}\right) \quad (2)$$

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.

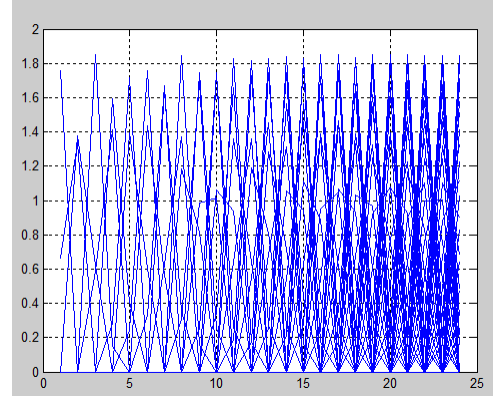


Fig 3.2 MFCC plot for filter 24 in MATLAB

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:

$$h_r(k) = \begin{cases} 0 & k < f(r-1) \\ \frac{k-f(r-1)}{f(r)-f(r-1)} & f(r-1) \leq k \leq f(r) \\ \frac{f(r+1)-k}{f(r+1)-f(r)} & f(r) \leq k \leq f(r+1) \\ 0 & k > f(r+1) \end{cases} \quad (3)$$

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.

$$Z_n = \sum_{m=-\infty}^{\infty} |sgn[x(m)] - sgn[x(m-1)]| w(n-m) \quad (4)$$

$w(n)$  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,

$$P = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T |x(t)|^2 dt \quad (5)$$

The  $x(t)$  is input speech signal. Power Spectral Density in this project is given by figure 3.4.

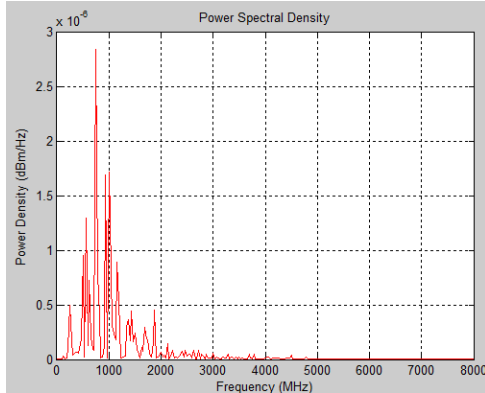


Fig 3.4 PSD of Input Speech

**B Testing**

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.

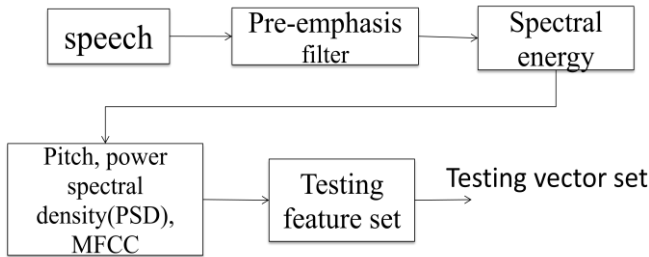


Fig 3.5 Testing

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:

$$c = \sum_i a_i k(s_i, x) + b$$

(6)

Where  $s_i$  support vectors,  $a_i$  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.

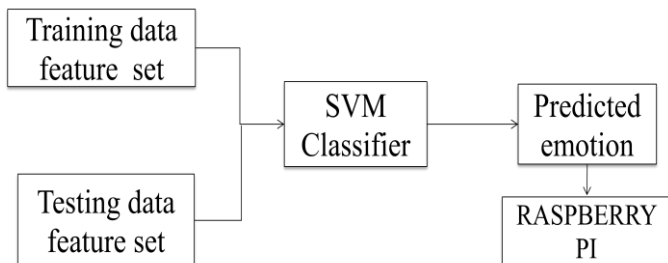


Fig 3.6 Final detected emotion

SVM classifier gives output of predicted emotion. After Training is completed the system is ready to testing the emotion from testing dataset and give the output A for Angry , J for joy and N for normal. If training and testing result is match then concluded that emotion is correct.

**D. Raspberry Pi**

This application development board Raspberry pi is used to Audio listening through speaker which identify speech emotion.

**IV. RESULTS**

**A. Accuracy**

Accuracy is being calculate by,

$$a = \frac{A+J+N}{total\ no\ of\ speech\ signals} \%$$

(6)

Table 3.1 Table for Estimate Result of Emotion output

	Training Speech Emotion Signal			
	Emotions	Angry Person	Joy/Happy Person	Normal Person
Testing Speech Emotion Signal	Angry person	29	0	1
Joy/Happy person	0	30	0	
Normal person	0	2	28	

So,

$$a = \frac{29 + 30 + 28}{90} = 96\%$$

The expecting accuracy is upto 96% is achieved so this system is getting high accuracy and accurate result for exact speech.

**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**

- [1] Moataz El Ayadi, Mohamed S. Kamel, Fakhri Karray, "Survey on speech emotion recognition: Features, classification schemes and databases", Waterloo, Ontario, Canada, July 2011, , Pages 572-587.
- [2] <http://cs229.stanford.edu/proj2007/ShahHewlett%20>
- [3] Surekha Reddy Bandela, T. Kishore Kumar "Stressed Speech Emotion Recognition using feature fusion of Teager Energy Operator and MFCC," IEEE 2017.
- [4] Jeet Kumar, Om Prakash Prabhakar , Navneet Kumar Sahu," Comparative Analysis of Different Feature Extraction and Classifier Techniques for Speaker Identification Systems: A Review", IJRCCE 2014.

- [5] Sreeram Ganji, Rohit Sinha, "Exploring Recurrent Neural Network based Acoustic and Linguistic Modeling for Children's Speech Recognition", IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017.
- [6] Namrata Dave, "Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition," Ieee International Journal For Advance Research In Engineering And Technology, July 2013.
- [7] Dr.V.AjanthaDevi,Ms.V.Suganya," An Analysis on Types of Speech RecognitionandAlgorithms,"IJCST, April2016.
- [8] Pavol Harár1, Radim Burget1 and Malay Kishore Dutta," Speech Emotion Recognition with Deep Learning," 2017 4th International Conference on Signal Processing and Integrated Networks.
- [9] Markus Niermann, Peter Jax, Peter Vary," Joint Near-End Listening Enhancement And Far-End Noise Reduction," 2017 Ieee.
- [10] Amritha Vijayan, Bipil Mary Mathai, Karthik Valsalan, Riyanka Raji Johnson, Lani Rachel Mathew," Throat Microphone Speech Recognition using MFCC," International Conference on Networks & Advances in Computational Technologies, 2017.
- [11] Pooja A, Pravena D, Govind D," Significance of Exploring Pitch only Features for the Recognition of Spontaneous Emotions from Speech Signals,"IEEE 2017.
- [12] D.S.Shete, Prof. S.B. Patil, Prof. S.B. Patil," Zero crossing rate and Energy of the Speech Signal of Devanagari Script," IOSR Journal of VLSI and Signal Processing, Jan 2014.
- [13] Mohan Ghai, Shamit Lal, Shivam Duggal and Shrey Manik," Emotion Recognition On Speech Signals Using Machine Learning,"IEEE2017.