Volume 5, Issue 5, May 2015 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Emotion Recognition through Speech Using Neural Network

Arti Rawat*
Student, Uttarakhand Technical University
Dehradun, Uttarakhand, India

Pawan Kumar Mishra
Assistant Professor, Uttarakhand Technical University,
Dehradun, Uttarakhand, India

Abstract— Automatic emotion of detection in speech is a latest research area in the field of human machine interaction and speech processing. The aim of this paper is to enable a very natural interaction among human and machine. This dissertation proposes an approach to recognize the user's emotional state by analysing signal of human speech. To achieve the good extraction of the feature from the signal the propose technique uses the high pass filter before the feature extraction process. High pass filter uses to reduce the noise. High pass filter pass only high frequency and attenuates the lower frequency. This paper uses the Neural Network as a classifier to classify the different emotional states such as happy, sad, anger etc from emotional speech database. For the performance of classification use the speech feature such as Mel Frequency cepstrum coefficient (MFCC). The result shows that the Neural Network used as a classifier is a feasible technique for the emotional classification. By using the high pass filter performance should be increase.

Keywords— Emotion Recognition, Feature Extraction, High Pass Filter, Neural Network.

I. INTRODUCTION

Human are emotional beings and it plays an important role behind their thoughts and action. So, it is necessary that emotion processing abilities are absorbed for designing of human environment. The analysis, recognition and synthesis of emotions can design the human environment. In this process the information uses such as audio, visual, written and mental information. Emotion is everywhere in our daily lives.

A novel research topic to be evolved in the Human Computer Interaction field is Automatic Speech Emotion Recognition. Emotion recognition through speech is an important part and current research area in emotion recognition. Accurately recognition of emotion through speech has beneficial for the designing of the human computer intelligent interaction [6]. Due to the ever widening use and demand of computer the need for a more natural communication interface between human and computer had naturally come to the force. To achieve this aim, a computer would have to be responses differently in current situation based on the perception which is perceived by the computer. To make the interaction between human and computer more naturally, it is important for the latter have the ability to respond to the emotions of humans in the same way like human in its position will do. To achieve the target that computer can detect the emotion either by facial expression or by speech. In the field of Human Computer Interaction speech is considered to be as a powerful way for emotion recognition [5]. Basically, Emotion recognition through speech can be explained as a detection of the emotion by feature extraction of voice conveyed by human.

There are some basics emotions given by different researchers as follows:

TABLE I SOME BASIC EMOTIONS RELATED TO DIFFERENT INCLUSION BASIS [1]

Reference	Fundamental Emotion	Inclusion basis
Arnold (1960)	Anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love, sadness, surprise	Relation to action tendencies
Ekman, Friesen, Ellsworth (1982)	Anger, disgust, fear, joy, sadness, surprise	Universal facial expressions
Oatley & Johnson-Laird (1987)	Anger, disgust, anxiety, happiness, sadness	Do not require propositional Content
Plutchik (1980)	Acceptance, anger, anticipation, disgust, joy, fear, sadness	Relation to adaptive biological processes

Emotion can be recognizing through speech by using the acoustical information of the speech. The acoustical information of speech can be divided into two parts. One is the prosodic feature and another one is spectral feature. Prosodic feature depends on the speech elements and their audible nature. The example of the prosodic features is pitch, volume, loudness etc. The spectral features are Mel Frequency Cepstram Coefficient (MFCC) and Linear Predictive Cepstral Coefficient(LPCC)[8].

Emotion recognition from speech can be used in various applications. Some of them are:- psychiatric diagnosis, intelligent toys, lie detection, smart call center, educational software etc.[4].

There are various classifiers used for the emotion detection such as Hidden Markov Model (HMM), k-nearest Neighbor (KNN), Artificial Neural Network (ANN), Gaussian Mixture Model (GMM), Support Vector Model (SVM) etc. [8].

Further in this paper Section 2 shows the literature survey carried out, Section 3 explains the proposed technique in detail. In Section 4, the efficiency of this scheme is tested. Finally, Section 5 involves the conclusion of the paper.

II. LITERARTURE SURVEY

In past year, several techniques used for emotion detection through speech. Emotion modeling is the main task of emotion recognition through speech. The various classifier used for the classification of the features of the speech. There are lots of classifiers used such as Gaussian Matrix Model (GMM), Hidden Markov Model (HMM), Support Vector Machine (SVM) etc. B. Yang, M. Luggar [2] proposed a work where the emotion detected by the euphony features. It was based on the music theory. It took the two different pitch intervals. And then found the occurrences which are the reason behind of a consonant or dissonant impression. They can evaluate these harmony features in a more realistic manner. Yashpalsing Chavhan, M. L. Dhore, Pallavi Yesaware [3] proposed the speech features such as, Mel Frequency cepstrum coefficients (MFCC) and Mel Energy Spectrum Dynamic Coefficients (MEDC). They took the utterances of speech of a human voice as a input. And then various features extracted from those utterances. The classifier used for the classify emotions was the support vector machine (SVM). The LIBSVM was used for classification of emotion. Mina Hamidi and Muharram Mansoorizade [7] reported an effort towards automatic recognition of emotional states from continuous Persian speech. Due to the unavailability of appropriate database in the Persian language for emotion recognition, at first, they build a database of emotional speech in Persian. This database consisted of 2400 wave clips modulated with anger, disgust, fear, sadness, happiness and normal emotions. Then they extracted prosodic features, including features related to the pitch, intensity and global characteristics of the speech signal. Finally, they applied neural networks. Igor Bisio, Alessandro Delfino, Fabio Lavagetto, Mario Marchese, Andera Scirrone [9] proposed system which was able to recognize the emotional state of a person. Firstly, they started the registration of audio signals. The system composed of two functional blocks: Gender Recognition (GR) and Emotion Recognition (ER). The Pitch Frequency Estimation method used for the recognition of gender whereas support vector machine (SVM) used for the detection of emotion. Here, they used the two SVMs one for male emotion recognition and other for the female emotion recognition. J. Sirisha Devi, Y. Srinivas and Siva Prasad Nandyala [10] introduced text dependent speaker recognition with an enhancement of detecting the emotion of the speaker prior using the hybrid FFBN and GMM methods. Lingli Yu, Kaijun Zhou, Yishao Huang [11] proposed humans emotional speeches recognition contributes much to created harmonious human machine interaction, also with many potential applications. In proposed system they used three ways to extend binary support vector machines (SVMs) are compared for recognizing emotions from speech by the Chinese and the Berlin speech database. One was standard SVM schemes (one-versus-one), and two other methods are DAG and UDT that could form a binary of decision tree classifiers. Meanwhile, a hierarchical classification technique of feature driven hierarchical SVMs classifiers are designed, whose structure is similar with DAG, it used different feature parameters to drive each layer, and the emotion can be subdivided layer by layer. Finally, analysis of the classification rate of those three extend binary SVMs, DAG performed the best for testing database, and standard SVM was not far behind, the UDT was the poorest due to depend on its upper layer classification accuracy.

III. PROPOSED TECHNIQUE

In this section, the whole process is done in the various steps. These steps generally define the working of proposed model. Here, we will define the propose scheme which is used for recognition of emotion along with the propose algorithm of high pass filter. The aim of designing high pass filter to remove the noise and increase efficiency.

A. Proposed High Pass Filter

A high pass filter is an electronic filter which passes the frequency higher than the cut off frequency. The frequency is lower than the cut off frequency attenuates in the high pass filter. The amount of attenuation depends on the filter design. High pass filter is also known as base cut or low cut filter. The high pass filter is used to remove the unwanted sounds near to the audible range or below.

B. Feature Extraction Using MFCC

Feature extraction is used to extract the feature from the speech signal. Mel Frequency Cepstral Coefficient (MFCC) is the most important and effective method for the feature extraction. Feature extraction aims for data reduction by converting the input signal into a compact set of parameters while preserving spectral and/or temporal characteristics of the speech signal information. The detail of block diagram of feature extraction is as follows:-

1) *Framing:-* It is the process of segmenting the speech sample into no. of frames. The speech samples obtained from the analog to digital conversion. The human voice is of variable length so fixing the size of speech the framing is necessary.

2) *Windowing*:- After framing, the windowing function is performed. For the purpose to minimize the signal discontinuities at the start and end of each frame. Hamming window is a good choice for windowing.
3) *Fast Fourier transform (fft):*- FFT is used ideally for generating the frequency spectrum of each frame. Each sample from time domain to frequency domain converted by the FFT. FFT is used to identify which frequency is present in the particular frame.
4) *Mel scale filter bank :*- The mel scale filter bank specify how much energy exists in particular frame. How much energy exists in the frequency region
5) *Log energy computation:*- After the filterbank energy, take the logarithm of them. It is also motivated by human hearing. A human don't listen high volume on a linear scale. Generally, the need of eight times energy put on it to double the perceived volume. This means that large variations in energy may not sound all that different if the sound is loud to begin with. Here, the purpose is to match the features closely as the human can listen clearly. So this compression operation uses to achieve that target.

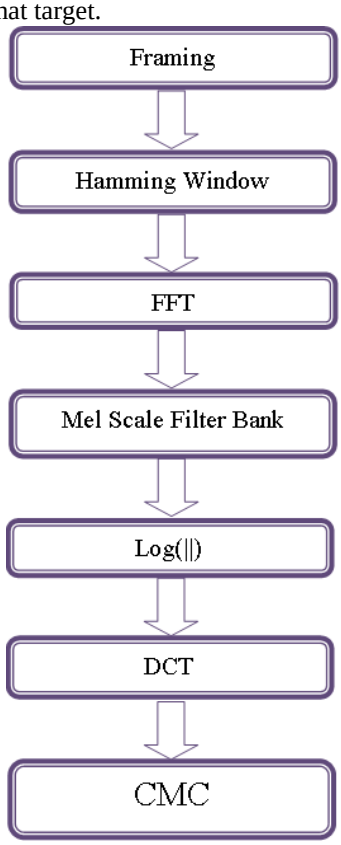

Fig 1:-Feature Extraction using MFCC

6) *DCT:*- The final step is to compute the DCT of the log filterbank energies. The energies of the filterbank correlate with one another. The reason behind this the overlapping of the filterbanks. Because the filterbanks are all overlapping, the filterbank. So, it requires the diagonal covariance matrices which can be used to model the features. That's why here DCT (discrete cosine transformation) uses for decor relate the energy.
7) *Cepstral mean correction:*- The cepstral mean correction was applied to compensate for distortion by subtracting the cepstral mean of a frame from the cepstral coefficients for additional robustness in recognition.

C. Neural Network

The term neural network derives its origin from the human brain which connects the massively large no. of neurons interconnected parallel. So that perceives the different recognition task, different perception in a small amount of time as compare to others. It performs very complex task as compare to today's high computers do not perform. Basically, human brain is made up of massively parallel neurons. It becomes from the neurons which is interconnected among them in a highly complex manner. In this, each neuron is connected to the other neuron. So, it is called highly complex and non linear. Human brain normally contains billions or trillions of neurons.

D. Steps For Emotion Recognition

Steps for Emotion Recognition process are explained below.
Step 1: In the first step, load signal and input the signal which has to be checked. The input and the data sample should be of same time duration and frequency.
Step 2: In this step, the noise has been removed. Here removing of noise only smoothen of the signal. There is no need to manipulate the recorded or input sample data. In the proposed system, high pass filter is used for the removing the noise from the input signal.

Step 3: In the third step of the system the feature of the voice sample has to be extracted. So apply the Mel Cepstral Coefficient to extract the features. This coefficient gives a matrix of 122*64 data.

Step 4: In the fourth step the loaded data and input data is trained with the help of neural network.

Step 5: In this step the Fuzzy theory approach is applied to check the results from the simulated data. Create The Network

P ← Input

T ← Target

Step 6: The network simulation is performed with the help of neural network.

Step 7 : If Output Is Equal To Target then detect the emotion and display the emotion and also play the voice.

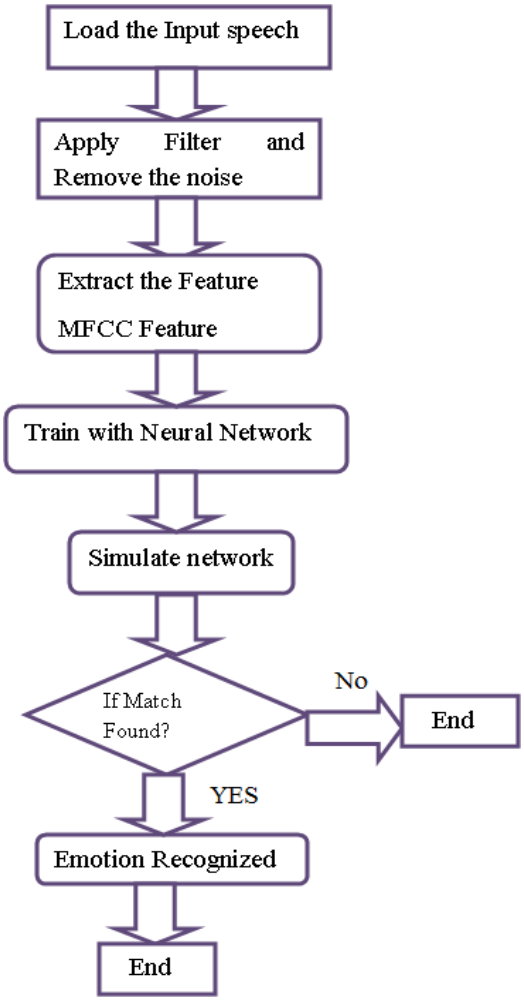

Fig 2:- Block Diagram of Emotion recognition through speech

.

Step 8: If output is not equal to the target then it will be end.

Step 9: End

E. Algorithm For High Pass Filter

1. Begin
   / All frequency values are in Hz.
2. Set Fs ← 32000
   / Fs denotes sampling frequency. The frequency at which the signal to be sampled called sampling frequency
3. Set N ← 32000
   / N denotes Order
4. Set Fc ← 10800
   / Fc denotes Cutoff frequency
   / Sampling Flag,'scale'to n*ormalize the filter so that the magnitude response of the filter at the center frequency of the passband is 0 dB. */*

5. Set Flag ← 'Scale'
   / Create the window vector for the design algorithm.
6. Set win ← barthannwin (N+1)
   / calculate the coefficients using the FIR1 function.
7. Compute b = fir1(N, Fc/(Fs/2), 'high', win, flag);
   / 'high' is for a highpass filter with cutoff frequency Fc/(Fs/2).
   / returns a discrete-time, direct-form finite impulse response (FIR) filter, Hd.
8. Compute Hd = dfilt.dffir(b);
9. End

IV. EXPERIMENTAL ANALYSIS RESULT

The experimental results shown on the test samples. The various samples has been tested for the recognition of emotion. There are 10 samples of 5 different emotions of 5 person recorded. Some of the samples uses for the training and other uses for the testing.

Firstly, when the test the sample the waveform of tested sample generated. The waveform of one of the tested sample is shown in fig 3

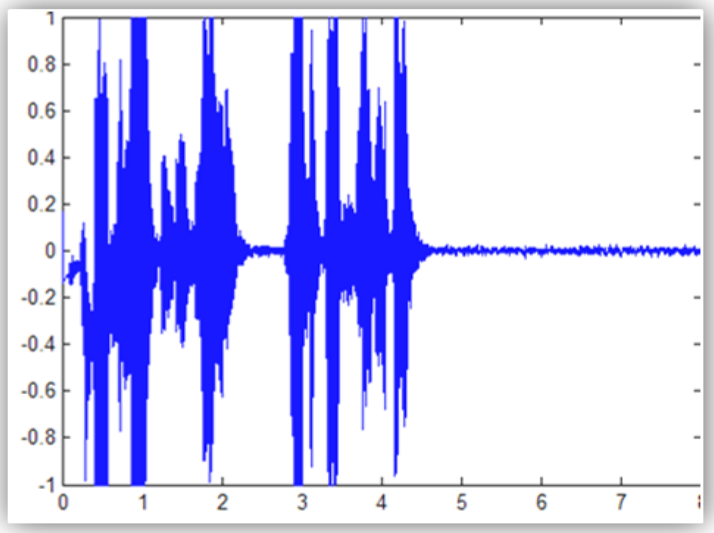

Fig 3:- Waveform of recorded sample

After that the high pass filter applied. The waveform of the tested sample after high pass filter is shown in Fig 2. High pass filter is designed for the reason of reducing the noise from the signal. When the signal is noisy then the feature can not be detect accurate. For the accurately detection of feature the filter is required. Here the high pass filter designed for this purpose. High pass filter is a filter which passes the high frequency and attenuates the low frequency.

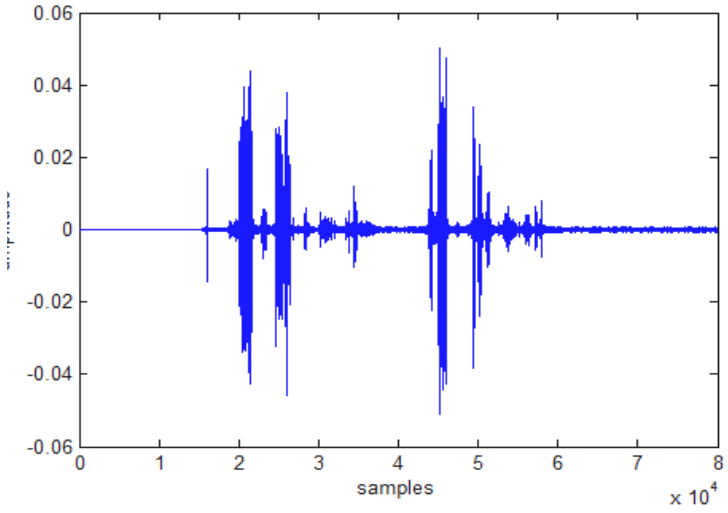

Fig 4:- After applying high pass filter, waveform of a recorded sample

After that the various feature extracted such as MFCC. The waveform of these features is shown in Fig 5.

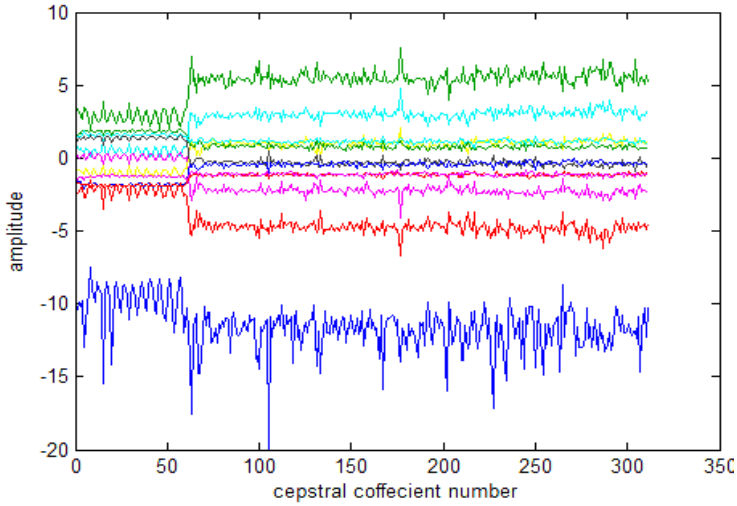

Fig 5:- MFCC feature extraction

These are the following parameters to check the performance:-

Performance (mse): It measures the network performance according to the means square error. To prepare a custom network to be trained with MSE, set net performance to 'mse'.

Epoch: Epochs are directly related to the iterations in which the network is trained.

Time: Time function is related to training time. The training time shows how time is taken by the network to be trained and give the simulated output.

Validation check: Validation check describes at what value

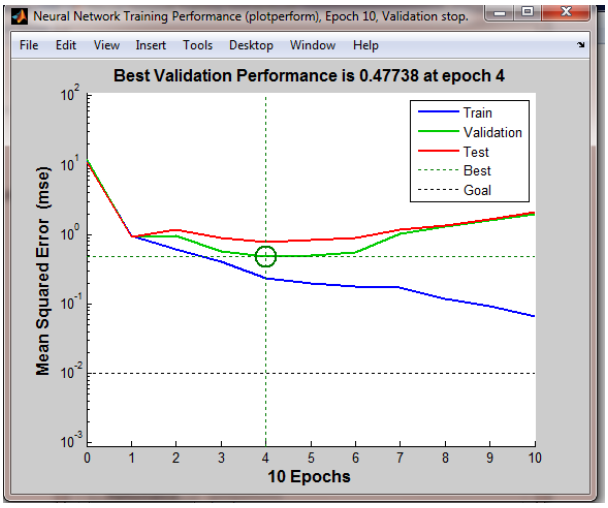

Fig 6:- Performance(Mean square error**)**

TABLE II ACCURACY RATE OF EMOTION RECOGNIZED ON RECORDED SAMPLES

EMOTION	**ACCURACY(%)** **NN with high pass filter**
HAPPY	95 %
SAD	94.16%
NEUTRAL	91.58%
DISGUST	93.42%
ANGER	92.74%

The various test samples recorded and trained them using the neural network. The results of different emotions as shown in Table II.

Because of the high pass filter and neural network classifier this work performs better than previous technique.

V. CONCLUSION& FUTUTRE SCOPE

The proposed scheme presented an approach to recognize the emotion from the human speech. This approach has been implemented by the neural network. This dissertation focuses on the feature extraction method that is useful in the emotion recognition through speech signal. For the purpose of feature extraction Mel Frequency Cepstrum Coefficient (MFCC) used. To achieve the good extraction of the feature high pass filter designed. High pass filter reduces the noise

from the signal and help to extract the better feature rather than other filter. To achieve the better performance the neural network is used for the training. Using high pass filter before the feature extraction and neural network for the classification gives the higher accuracy. The obtained result shows that the proposed system can reliably identify the single emotion from the speech samples. The performance is highly depends on the emotional speech samples. So, it is necessary to take a proper and correct speech samples. The performance is good of the propose technique but it takes a time at the time of execution. So, it can be reduce in the future. Another future enhancement is that it can be applied for the bigger set of emotions i.e. positive or negative and can be implemented by other classification algorithm

REFERENCES

[1] Ortony and T. J. Turner, "What's basic about basic emotions?," Psychological Review, vol. 97, no. 3, pp. 315–331, 1990.

[2] B. Yang, M. Luggar "Emotion Recognition from Speech Signals using new Harmony Features" Elsevier ISSN 0165-1684 Volume 90 Issue 5 May 2010.

[3] Yashpalsing Chavhan, M. L. Dhore, Pallavi Yesaware, " Speech Emotion Recognition Using Support Vector Machine" International Journal of Computer Applications (0975 - 8887) Volume 1 – No. 20 2010

[4] Simina Emerich, Eugen Lupu " Improving Speech Emotion Recognition using Frequency and Time Domain Acoustic features" EURSAIP 2011.

[5] Bhoomika Panda, Debnanda Padhi, Kshamamayee Dash, Prof. Sanghmitra Mohantay "Use of SVM classifier & MFCC in Speech Emotion Recognition System" International Journal of Advance Research in Computer Science and Software Engineering(IJARCSSE) Volume-2, Issue-3, March 2012

[6] Vaishali M. Chavan, V. V. Gohokar "Speech Emotion Recognition by using SVM Classifier" International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-1, Issue-5, June 2012.

[7] Mina Hamidi and Muharram Mansoorizade "Emotion Recognition From Persian Speech With Neural Network" International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.5, September 2012

[8] Akshay S. Utane, Dr. S. L. Nalbalwar "Emotion Recognition through speech using Gaussian Mixture Model and Hidden Markov Model" International Journal of Advance Research in Computer Science and Software Engineering(IJARCSSE) Volume-3, Issue-4, April 2013.

[9] Igor Bisio, Alessandro Delfino, Fabio Lavagetto, Mario Marchese, Andera Scirrone "Gender Driven Speech Recognition Through Speech Signals for Ambient Intelligent Applications" IEEE Volume-1 No.2, December 2013.

[10] J. Sirisha Devi, Y. Srinivas and Siva Prasad Nandyala, "Automatic Speech Emotion and Speaker Recognition based on Hybrid GMM and FFBNN" International Journal on Computational Sciences & Applications (IJCSA) Vol.4, No.1, February 2014

[11] Lingli Yu, Kaijun Zhou, Yishao Huang, "A Comparative Study on Support Vector Machines Classifiers for Emotional Speech Recognition" Immune Computation (IC) Volume2, Number1, March 2014.

