I J C T A, 9(16), 2016, pp. 7801-7806

Speech Emotion Recognition based on BPN and SVM Classifier

S. Dhanalakshmi[1], Manideep Kakkireni[2], R. Sathya Narayanan[3] and S. Sreenivasa Reddy[4]

ABSTRACT

We propose a model Speech Emotion recognition on Fourier parameters. Emotional state of voice signal is identified effectively from Fourier parameter features. The proposed features are effective over (MFCC) Mel Frequency Cepstral Co-efficient. The Emotion recognition from voice signal is better when features of MFCC and FP both combined [2] [3].

Index Terms: Fourier parameter model, Features of voice signal, Neural network, Emotion recognition.

1. INTRODUCTION

Speech emotion recognition is defined as determining the state of emotion of a speaker with his or her voice signal. Speech emotion recognition helps in improving the accuracy of identifying the speech of a person [2] [3] [4]. This artificial intelligence is used in human computer interface (HCI) and used widely in man machine interaction, military war-face, health care, surveillance.

The features like MFCC, Linear prediction coefficients (LPC) and (LPCC) Linear prediction cepstral coefficients were used as standard set for feature extractions earlier by many researchers and played significant role in determining the speech emotion. Later on the Fourier parameter features results high accuracy rate compared to other set of features. Here we are implementing the features that are combination of Mel frequency and Fourier parameter features as the recognition of emotion from speech is comparatively accurate.

The intrinsic features are extracted from the input voice signal and also features like zero crossing rate and fundamental energy level are extracted these features vary each other for every individual signal which helps in recognition of emotion. As of each and every feature is compared with threshold values and several voice samples the emotion is identified [4]. NN classifier plays a key role in determining the emotion.

NN classifier organises a network which similars the neurons in brain and neural functioning thus called as neural network. This neural network create a sub space in which all the features comparison is done with in that and simple emotion is given as output.

The following emotions are selected for classification: Happy, Angry, Normal and Sad in this proposed model [2]. We propose a set of features to detect perceptual content of speech signal. These features are classified and examined on speech databases. Both Back propagation network (BPN) and Support Vector Machine (SVM) are implemented.

2. WAVELET ANALYSIS OF SPEECH

Fourier analysis is applied for processing of signal which includes pre-processing, synthesis, Wavelet transformation of signal feature extraction, coding, discrete wavelet decomposition is popular method

[1,2,3,4] SRM University, Kattankulathur-603 203, Kancheepuram-Dist., Tamil Nadu, India, Emails: *Kakkireni_manideep@.srmuniv.edu.in*, *sdhanalakshmi2004@gmail.com*

widely used for feature extraction in signal processing. The signal is decomposed with wavelet transform according to low pass filter and high pass filter [5].

A speech signal x (m) which is divided into l frames and can be represented by a combination on an Fourier Parameter is

$$x(m) = \sum_{k=1}^{M} H_k^l(m)\left(\cos\left(2\prod \frac{f_k^l}{F_s} m\right) + \varphi_k^l\right)$$

F is a sampling frequency of signal x (m) H and φ are the amplitude and phase of the kth harmonic sine component is the index of the frame, and M is the number of speech harmonic components.

In Dwt discrete wavelet transformation the resolution is reduced by one half at each level by subsampling data signal by two [7].

[0.3 0.4 0.2 -0.1 0.2 0.45 0.3 -0.2 -0.5 -0.4....]

c1 c2 c3 c4 c5 c6 c7 c8 c9 c10

c1+c3+c5+c7+c9 c2+c4+c6+c8+c10

C1+1/2 C2 C1-1/2 C2

Low pass High pass

The low pass signal is considered as data signal. The maximum amount of signal is taken into account and the loss of signal is less compared to that of Fourier transform so the harmonic features of the signal loss is less compared to that of Fourier transform signal.

The figure (i**)** show the DWT of a sample speech signal with happy emotion and features are extracted from this signal waveform and further analysed [3].

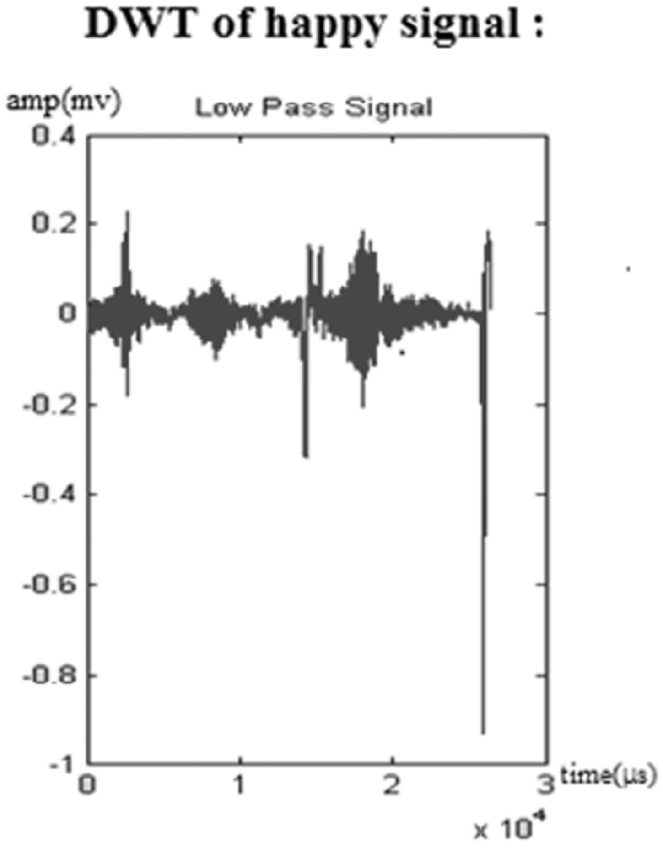

Figure 1: DWT of sample signal

3. FOURIER PARAMETER FEATURES FOR SPEECH EMOTION ANALYSIS

Features are extracted from raw data speech signal, these features are based on Fourier parameters. Each individual speech signal is differed to others, which are differentiated by features.

3.1. Emotion Databases

There are some set of types of emotion database, where we considered three set of databases, German emotional database (EMODB), Chinese emotional database (CASIA), Chinese elderly emotional speech database (EESDB). EMODB has been used by many researchers as a speech emotion recognition standard database signals. CASIA comprises speech in six classes of different emotions with 9600 wave files [5].The EESDB includes seven classes of emotions.

These databases are used for validating the method of FP features extraction and case study to develop the set of new model features.

3.2. Fourier Parameter Features

FP features include harmonics like frequency, amplitude and phase. Every frame is evaluated and the features are generated. It is observed that different classes of emotions amplitude varies and mean of each phase of speech signal varies.

These Fourier parameter features are efficient in differentiating classification, and accuracy. The mean, maximum, minimum, median and standard deviation of the amplitude are calculated primarily [1].

These classify the differences among the emotions for example: average value for sadness and neutral are higher than happy. And the peak values of these emotions are obtained at different individual harmonics for happy and angry emotions, it is obtained at sixth harmonics and for neutral is at fourth harmonics.

4. FEATURES EXTRACTION

Speaker independent emotion recognition is a challenging task in recent field of artificial speech recognition as unknown speakers emotion recognition has generalization better to that of speaker dependent approach of emotions [1] [5] [9].

Both MFCC and FP features are extracted and some other features fundamental frequency (f0) [2] [3], energy and zero crossing rate. MFCC features are first introduced to analyse the speech emotion and successful in determining the result. The MFCC features include mean, maximum, minimum, median and standard deviation. By this process certain speech data signals are filtered by a high-pass filter with a coefficient of about 0.97, which means that tends to a 39-dimensional MFCC features are extracted.

Fourier parameter features include amplitude and its first order difference and second order difference of a signal are computed and this harmonics like mean, median maximum, minimum and standard deviation. Which comprises double standards for speech emotion features which are for helpful in accuracy.

Here from the ‘table (1)’ we infer that, the following features are implemented as a first phase features for a sample signal which are fed as an input to neural network classifier. With the help of these features classification is done and the emotion is determined.

4.1. Normalization of Features

For an emotion recognition system normalization of a signal is an important aspect. It has to eliminate the speaker variability [2]. While including the emotional features effectiveness. There are some normalization techniques as we use one of the technique is Z-score normalization.

Table 1
Features of speech signal

FP parameter	Feature value
Max Signal Level	0.5003
Min Signal Level	-1.0794
Average Signal Level	-2.8817e-05
Peak Level	0.5003
Median Filter Signal Level	0.0106
Standard Deviation	0.0439
Histogram	-3.7941e+06
Entropy Level	2.7243
Zero Crossing Rate	0.0155
Fundamental Energy Level	-1.6521

$$\widehat{H}^s = \frac{H^s - E\left(H^s\right)}{std\left(H^s\right)}$$

Where e (H^s) is its mean value and its standard deviation value std (H^s) respectively.

5. SUPPORT VECTOR MACHINE CLASSIFICATION

Support vector machine is a type of NN classifier. SVM is a new classifier technique which is based on statistical values polynomial functions, radial basis functions and neural networks. It is used in text recognition and also in Facial recognition which performed well than other classifiers. In speech recognition many classifiers like Gaussian Mixture model (GMM), Back Propagation model (BPN) and Artificial Neural networks are considered and studied their experimental performance which results SVM betters other classifiers.

It is efficient in solving two class classification. SVM uses hyper-linear plane for separating classifier (figure 2). This creates a nonlinear transformation of input space into high dimensional feature space.Seperating planes are optimal that results maximal margin classifier to that of training data set [4].

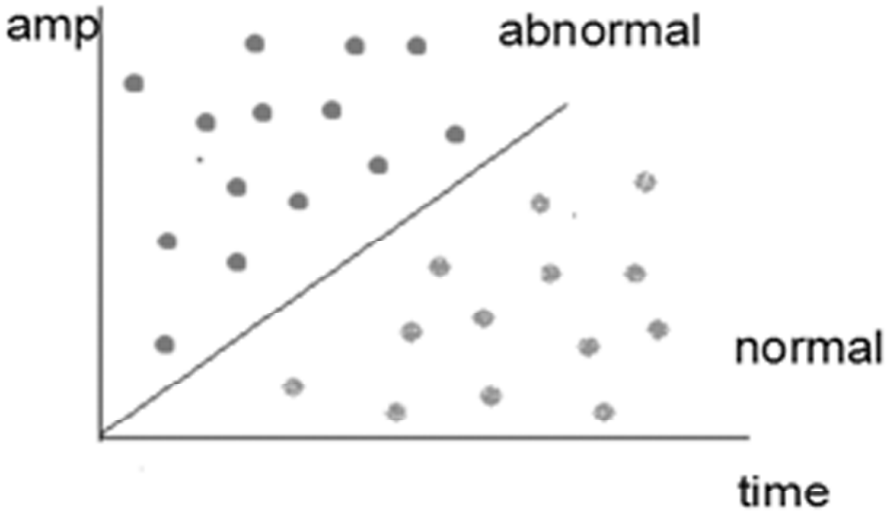

Figure 2: SVM classification

5.1. Experimental results

We first experimented with 10 FP features for emotion recognition. The rate of accuracy and result increased with increase in features and parameters of first order and second order differences. The experimental result also show that FP features are far better than phase features.

We developed the emotion recognition by PNN (probabilistic neural network) classifier using Back Propagation Network (BPN) algorithm classifier [1]. The emotion recognition is experimented and result was not as satisfaction as that of SVM. The recognition rate is high for angry 82.7 percent and 54 percent for sad and 69.7 percent for normal signal and 72.0 percent for happy emotion.

Where the results are improved as the application of combined.

The output of the sample speech signal is demonstrated as an example in figure (3) and is compiled in MATLAB

6. FUNCTIONING OF THIS MODEL

This block diagram (Figure 4) explain the functioning of our proposal model. The speech signal given as input noise reduction is done and wavelet transformation of signal converts into low pass signal. Features

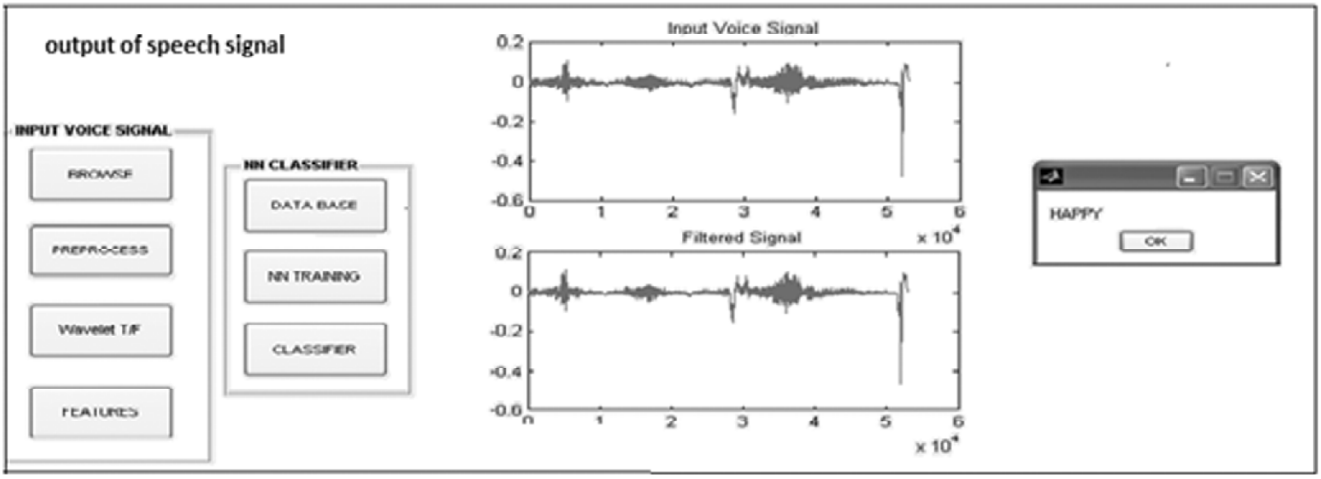

Figure 3: Output of Speech Signal

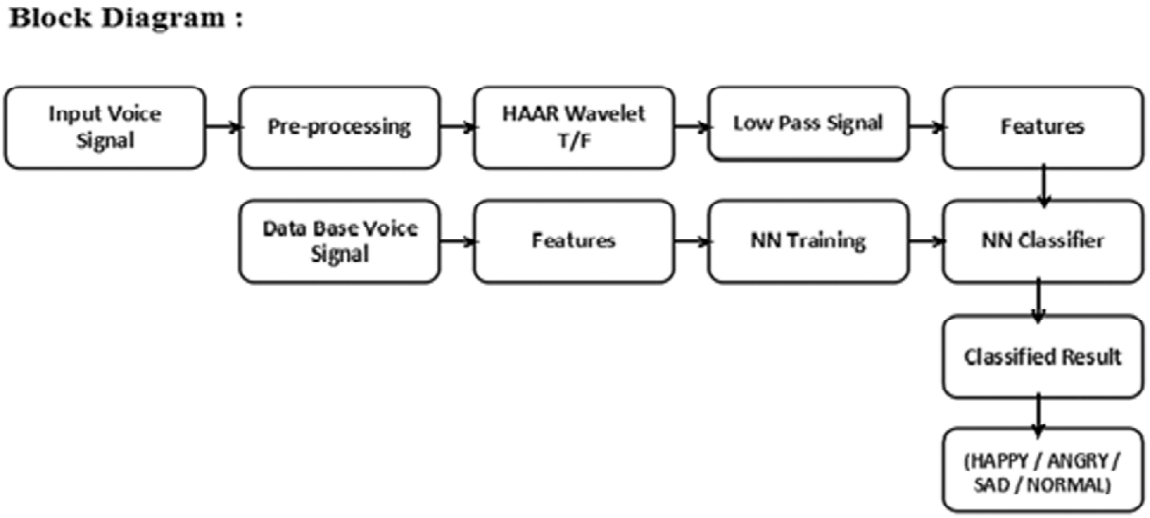

Figure 4: Block diagram

are extracted and further fed to NN classifier as it organize a neural network and classify the signal and result is obtained

7. CONCLUSION

In previous studies, several emotion databases were employed and different set of features were experimented and the results were effective.

We implemented the MFCC features on Fourier parameters and combination of MFCC and FP features [2] [3] on SVM classifier. The recognition of emotion in a speech signal is effective and the accuracy rate is high compared experimentally with BPN algorithm [1].

This established model on FP would be helpful in future developments and speaker-independent speech emotion recognition.

REFERENCES

[1] S. Dhanalakshmi, C. Venkatesh, "Classification of Ultrasound Carotid Artery Images Using Texture Features", International review on computers and software" Vol. 8, No. 4, 2013.

[2] Kunxia Wang, Ning An, Bing Nan Li, Yanyong Zhang members of IEEE, "Speech Emotion Recognition Using Fourier Parameters", IEEE transactions on effective computing, vol. 6, No.1, Jan-Mar 2015.

[3] Yayuan Yujin, Zhao Peihua, Zhou Qun "Research of Speaker Recognition Based on Combination of LPCC and MFCC", 978-1-4244-6585-9/10/©2010 IEEE.

[4] M. E. Ayadi, M. S. Kamel, F. Karray, "Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases", Pattern Recognition 44, pp. 572-587.

[5] Y. Li and Y. Zhao, "Recognizing Emotions in Speech Using Short- Term and Long-Term Features," Proc. Int'l Conf. Spoken Language Processing, pp. 2255-2258, 1998.

[6] B. Vlasenko, B. Schuller, A. Wendemuth, and G. Rigoll, "Combining Frame and Turn-Level Information for Robust Recognition of Emotions within Speech," Proc. Int'l Conf. Spoken Language Processing, pp. 2225-2228, 2007.

[7] Je Hun Jeon, Rui Xia, Yang Liu, "sentence level emotion recognition based on decisions from subsentence segments", ICASSP 2011, 978-1-4577-0539-©2011 IEEE.

[8] M. Kotti and F. Paterno, "Speaker-independent emotion recognition exploiting a psychologically-inspired binary cascade classiûcation schema," Int. J. Speech Technol., vol. 15, pp. 131–150, 2012.

[9] M. Hayat and M. Bennamoun, "An automatic framework for textured 3D video-based facial expression recognition," IEEE Trans. Affective Compute., vol. 5, no. 3, pp. 301–313, Jul.-Sep.2014.