ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology
(IJARTET) Vol. 5, Special Issue 14, April 2018

SPEECH SIGNAL BASED EMOTION RECOGNITION

SUKUMAR B S[1]
Asst.Professor Dept. of ECE, CBIT Kolar
sukumar.svm@gmail.com

Dr. G N KODANDARAMAIAH[2]
Professor and HOD Dept. of ECE KEC Kuppam
gnk.ramaiah@gmail.com

BALAKRISHNA MASANAM[3]
Research Scholar JNTUA
Balumasanam490@gmail.com

ABSTRACT

This Paper introduces a speech based emotion recognition system. Speech signals are classifies into different emotional states like happy, anger, normal, fear, boredom, disgust. We have created emotional speech signal database for emotion recognition system. For classification of different emotions Support Vector Machine (SVM) is used and for feature extraction MFCC (Mel Frequency Cepstral Coefficients), PLP (Perceptual Linear Prediction) and LPC (Linear Predictive Coding) are used.
Keywords: SVM, MFCC, PLP, LPC

I. INTRODUCTION

Speech is the act of expression or communication of thoughts and feelings by spoken words. speech it not only contains the information but it also contains emotional contents. The work presented here is to make a system which is able to detect human emotion through speech signals. In the past decades most of emotion recognition researchers have focused on using facial expressions and some researchers focused based on speech by extracting features. Here we focused on MFCC, PLP and LPC methods for extracting features and SVM for classification of features for detecting specific emotion.

II. Methodology

The work presented here is to make a system which is able to detect human emotion through speech signals. Many researchers they focused on emotion detection by extracting features only by using MFCC or LPC or PLP. Here we have used combination of all for extracting features and SVC (Support Vector Machine) for classification of features to determine particular emotion as shown in fig.1.

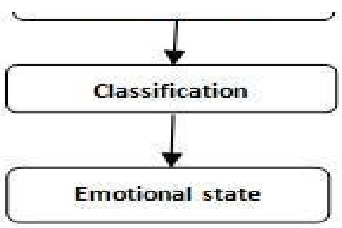

fig.1 Methodology for emotion detection from speech signal

A. Feature Extraction:

The pre-processed Speech signal is used for feature extraction process. We have used three methods for feature extraction.

1) MFCC (Mel-Frequency Cepstral Spectrum)
2) PLP (Perceptual Linear Prediction)
3) LPC (Linear Predictive Coding)

A.1 MFCC (Mel-Frequency Cepstral Spectrum)

The Cepstral coefficient is a set of feature, and this method is considered as one of the robust method for different pattern recognition task related speech. MFC Coefficients are a features widely used in speech and speaker recognition [6]. Here we went through the following steps for extracting features as shown in fig.2.

Initially we frame the signal into short frames (Hamming window), then for each frame calculated the periodogram to estimate of the power of the power spectrum. Then apply the mel filter bank to the power spectra, sum the energy in each filter, take the logarithm of all filter bank energies. Then applied DCT of the log filter bank energies, keep DCT coefficients 2-13 (MFC Coefficients) and discard the rest.

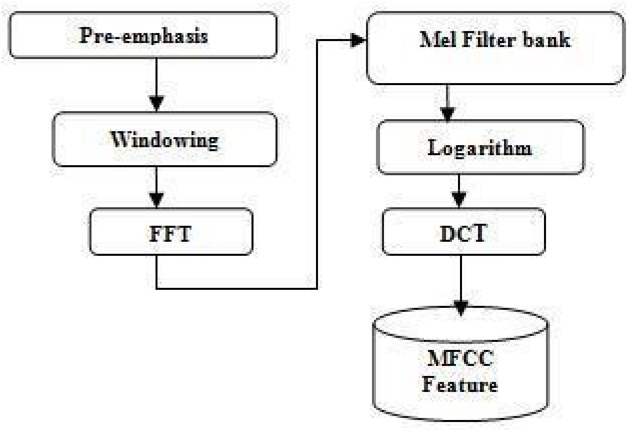

fig.2: Block diagram of MFCC

A.2 PLP (Perceptual Linear Prediction)

The Perceptual Linear Prediction model developed by Hermansky. PLP models the human speech based on the concept of psychophysics of hearing [2]. PLP discards irrelevant information of the speech and thus improves speech (emotion) recognition rate. PLP is identical to LPC except that is spectral characteristics have been transformed to match characteristics of human auditory system. The details of PLP feature extraction steps are as in fig.3. The quantized signal is applied for hamming window and power spectrum is calculated. A frequency warping into the Bark scale is applied.

The convolution is applied between the auditory warped spectrum and the power spectrum of the simulated critical band. The obtained smooth spectrum is down sampled at the intervals of nearly one bark. Frequency warping, smoothing and sampling are integrated into a single filter bank called Bark filter bank. The equalized values are transformed according to the power law of Stevens by raising each to the power of 0.33.

Predictor Coefficient of a signal that has warped spectrum as power spectrum is obtained by applying linear prediction (LP) to the auditory warped line spectrum. The required cepstral coefficients from predictor coefficients are extracted, by a recursion that is equivalent to the logarithm of the model spectrum and followed by inverse Fourier transform.

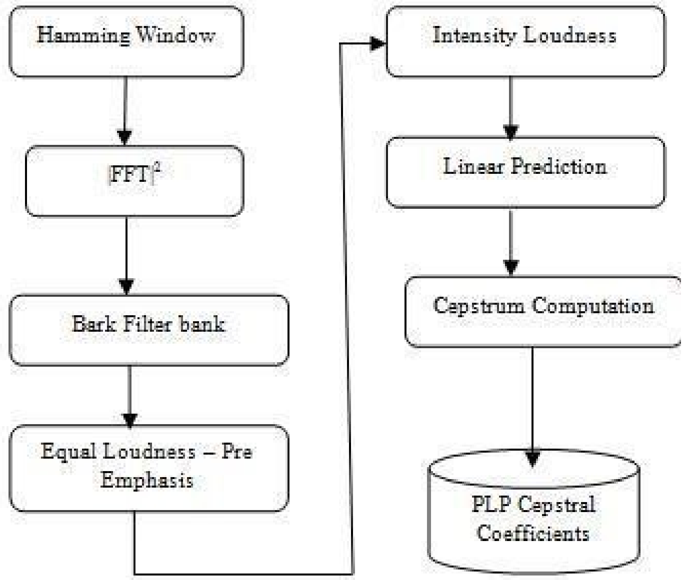

fig.3 Block diagram of PLP

A.3 LPC (Linear Predictive Coding)

A Speech analysis and synthesis with LPC exploit the predictable nature of speech signal. The linear Prediction problem can stated as finding the coefficients which results in the best prediction of the speech sample in terms of the past samples. LPC is mainly used for formant analysis and it is widely used in medium or low bit rate coder, it calculates power spectrum of the signals. Linear prediction models the human vocal tract as an Infinite Impulse Response (IIR) system that produces the speech signal.

For removing redundancy in signal we pass the speech signal in speech analysis filter. In this process residual error is generated as an output. Now we can transfer this residual error and speech parameters instead of transferring entire signal to generate the original signal. In this technique formants are described by LPC Coefficients. The frequencies at which the resonant peaks occur are called the formant frequencies [2]. Thus, with this method, the locations of the formants in a speech signal are estimated by computing the linear predictive coefficients over a sliding window and finding the peaks in the spectrum of the resulting LP filter. We have excluded 0th coefficient and used next ten LPC Coefficients.

B. Classification

After features extraction, specific emotion state was determined by using classification. For classification we have used Support Vector Machine (SVM) classifier. SVM makes use of convex quadratic optimization that is advantageous in making a globally optimal solution. SVM has demonstrated good performance on several classical problems of pattern recognition. For emotion recognition also we can use SVM classifier for classifying different emotional states. There are two different families of solutions aiming to extend SVM for multiclass problems. The first solution follows the strategy of “one-versus-all”, while the second solution follows the strategy of “one-versus-one”. We selected the second method by using LIBSVM because it is more convenient in practice.

III. Results:

Feature extraction with combination of MFCC, PLP, LPC gives features more accurate and precise for classification. SVM is used for classification has provided good generalization and its performance. In our work we achieve the following accuracy for different emotions as shown in the following table.1. Avg. Accuracy of 91.89% is obtained from the experimental results. For classifying emotion states can be obtained by using support vector machine.

Table.1 Accuracy for different emotions.

Sl.No.	Emotion	Accuracy in Percentage	Accuracy in Percentage from [1]
1	Anger	96.71	93.38
2	Boredom	89.92	87.65
3	Disgust	91.36	83.91
4	**Fear**	**92.43**	**---**
5	Happiness	91.71	61.66
6	Neutral	90.64	69.21
7	Sadness	91.00	91.83
Avg. Accuracy		**91.89**	**81.23**

IV. Conclusion:

In this paper, the focus is on emotion detection by using speech signals. Here features are extracted by combining MFCC, PLP and LPC methods. Individually MFCC gives beneficial results but if we combine PLP and LPC accuracy will be improved. We are using the SVM, to classify the user in different emotions. The system can be used for different applications.

References

[1] Mohan Ghai, Shamit Lal, Shivam Dugga l and Shrey ManikJ. “Emotion Recognition On Speech Signals Using Machine Learning” 978-1-5090-6399-4/17/$31.00_c 2017 IEEE.

[2] Namrata Dave, “Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition” International Journal For Advance Research In Engineering And Technology, Volume 1, Issue VI, July 2013.

[3] Olivier Lartillot, Petri Toiviainen, A MATLAB TOOLBOX FOR MUSICAL FEATURE EXTRACTION FROM AUDIO, Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07), Bordeaux, France, September 10-15, 2007.

[4] Shrutika Lokannavar, prashanth Lahane, Apurva Gangurde, Pooja Chidre, “Emotion Recognition Using EEG Signals” International Journal of Advanced Research in Computer and Communication Engineering Vol.4 Issue 5, May 2015.

[5] Pratik K. Kurzekar, Ratnadeep R. Deshmukh, Vishal B. Waghmare, Pukhraj P. Shrishrimal, “A Comparative Study of Feature Extraction Techniques for Speech Recognition System”, International Journal of Innovative Research in Science,Engineering and Technology, Vol. 3, Issue 12, December 2014.

[6] Santhosh Gaikwad, Bharati Gawali, Pravin Yannawar, Suresh Mehrotra, “ Feature Extraction Using Fusion MFCC For Continuous Marathi Speech Recognition”.

[7] Santhosh Gaikwad, Bharati Gawali, Pravin Yannawar, “A Review on Speech Recognition Technique”, International Journal of Computer Applications (0975-8887), Volume 10-No.3, November 2010.

[8] Ji Zhengbiao, Zhou Feng and Zhu Ming, “An Algorithm Study for Speech Emotion Recognition Based Speech Feature Analysis” International Journal of Multimedia and Ubiquitous Engineering, Vol.10, No.11 (2015), pp.33-42.