IJCAT International Journal of Computing and Technology, Volume 1, Issue 2, March 2014
ISSN : 2348 - 6090
www.IJCAT.org

Design of Human Emotion Recognition System from Speech using Particle Swarm Optimization

[1] **Anjali Shelke** , [2] **Sonali Joshi**

[1] PG Scholar,Department of Electronics Engineering, G. H. Raisoni College Of Engineering,
Nagpur, India 440016

[2]Assistant Professor, Department of Electronics Engineering, G. H. Raisoni College Of Engineering,
Nagpur, India 440016

Abstract - Emotional speech recognition is an area of great interest for human-computer interaction. The system architecture contains feature extraction & Selection, emotional speech database, classifier Algorithm, which is used as a emotion classifier and recognizer. The system must be able to recognize the user's emotion and perform the actions accordingly. Many of emotion recognition systems have been constructed by different researchers for recognition of human emotions in spoken utterances. This paper proposed the technique used for recognition of human emotion from speech using different methods of feature extraction for the emotion recognition. For experimental study, the database for the speech emotion recognition system is the emotional speech samples and the features extracted from these speech samples are the energy, pitch, Mel frequency cepstrum coefficient (MFCC).. The Algorithm is used to differentiate emotions such as anger, happiness, sadness, fear, Surprise, normal state, etc. This paper proposed design of human emotion recognition system from speech using Particle Swarm Optimization.

***Keywords* - Introduction, feature extracted, Proposed Work.**

1. Introduction

The term bio-inspired has been introduced to demonstrate the strong relation between a particular system and algorithm, which has been proposed to solve a specific problem, and a biological system, which follows a similar procedure or has similar capabilities. Bio-inspired computing represents a class of algorithms focusing on efficient computing, e.g. for optimization processes and pattern recognition. The bio-inspired algorithms are very easy to understand and simple to implement. These algorithms especially dominate the classical optimization methods. It is often closely related to the field of swarm intelligence (SI), artificial immune system (AIS) and cellular signaling pathways.

Speech emotion recognition is one of the latest challenges in speech processing. Besides human facial expressions speech has proven as one of the most promising modalities for the automatic recognition of human emotions. Affective computing is concerned with emotions and machines. By giving machines the ability to recognize different emotions and make them able to behave in accordance with the emotional conditions of the user, human computer interfaces will become more efficient.

Speech is a complex signal which contains information about the message, speaker, language and emotions. Emotions are produced in the speech from the nervous system consciously, or unconsciously. Emotional speech recognition is a system which basically identifies the emotional as well as physical state of human being from his or her voice. Emotion recognition is gaining attention due to the widespread applications into various domains: Intelligent Tutoring System, Lie Detection, emotion recognition in Call Center, Sorting of Voice Mail, and Diagnostic Tool by Speech Therapists.

2. Speech Emotion Recognition System

Speech emotion recognition aims to automatically identify the emotional state of a human being from voice. Fig.1 indicates the speech emotion system components.

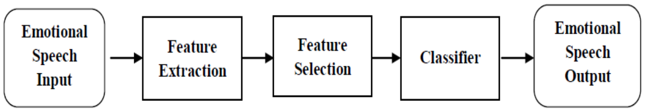


Fig1: Block diagram of Speech emotion recognition system

Speech emotion recognition system contains four main modules: emotional speech input, feature extraction, feature selection, classification, and emotion output. A human cannot classify easily natural emotions, so it is difficult to expect that machines can offer a higher correct classification. A typical emotions set contains 300 emotional states which are decomposed into six primary emotions like happiness, sadness, surprise, fear, anger, neutral. Success of speech emotion recognition depends on naturalness of database.

There are two databases which are available publically: Danish Emotional Corpus (DSE) [4] and Berlin Emotional Database (EMO-DB) [2, 12]. From these two databases Danish Emotional Corpus (DSE) is used. With respect to authenticity, there seems to be three types of databases used in the SER research: type one is acted emotional speech with human labeling. This database is obtained by asking an actor to speak with a predefined emotion. Type two is authentic emotional speech with human labeling. These databases are coming from real-life systems (for example call-centers) [3] and type three is elicited emotional speech with self report instead of labeling. [10]

3. Emotional Speech Database

In this research study we use the Danish Emotional Speech (DES) Database as a Emotional Speech Database. This includes expressions by four actors familiar with radio theatre, two male and two female. In this database speech is expressed in five emotional states: anger, happiness, neutral, sadness and surprise.

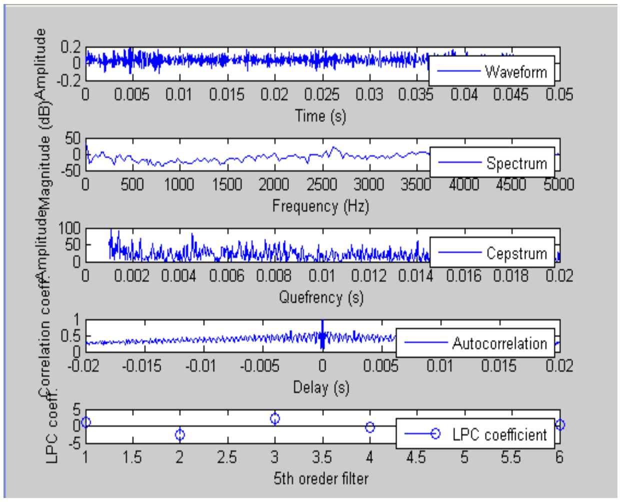
.

4. Feature Extracted

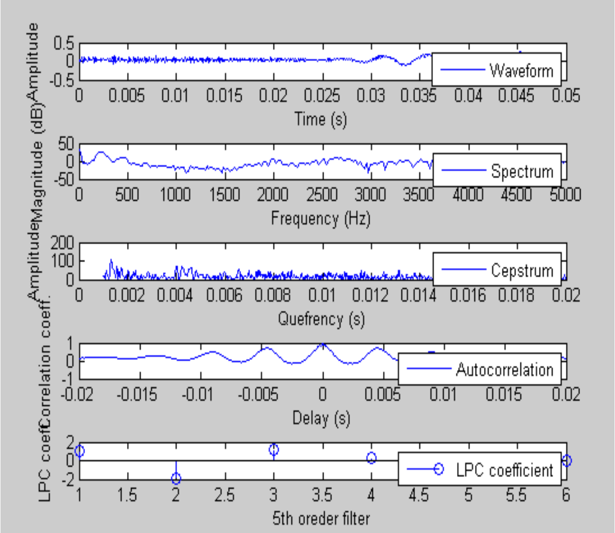
Speech signal composed of large number of parameters which indicates emotion contents of it. Changes in feature parameters indicate changes in the emotions. Therefore the proper choice of feature vectors is one of the most important tasks in emotion recognition. There are several approaches towards automatic recognition of emotion in speech by using different feature vectors. In this research work we use two feature extraction method.
The features extracted here are:

4.1. Statistical Feature

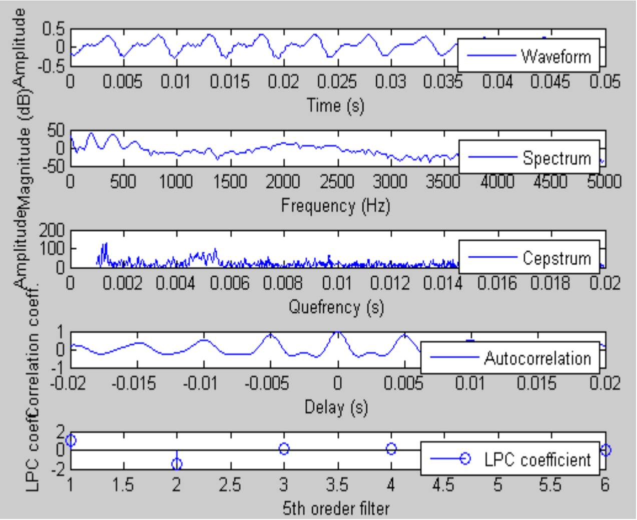
Energy and pitch are two important feature for speech emotion recognition. The Energy is the basic and most important feature in speech signal. We can obtain the statistics of energy in the whole speech sample by calculating the energy, such as mean value, max value, variance, variation range, contour of energy. The value of pitch frequency can be calculated in each speech frame and the statistics of pitch can be obtained in the whole speech sample. These statistical values reflect the global properties of characteristic parameters. For our experiment we calculated some descriptive statistical minimum, maximum, Average, std. deviation, variance. The pitch is estimated based on cepstral analysis.



(a)



(b)



(c)

Fig2:-Result of statistical feature extracted for anger having fundamental frequency for (a) 916.84Hz, (b) 759.036Hz & (c) 746.193Hz.

4.2. Mel Frequency Cepstrum Coefficient (MFCC)

MFCC is a representation of the short-term power spectrum of a sound, based on a liner cosine transform of a log power spectrum on a nonlinear mel scale of frequency.

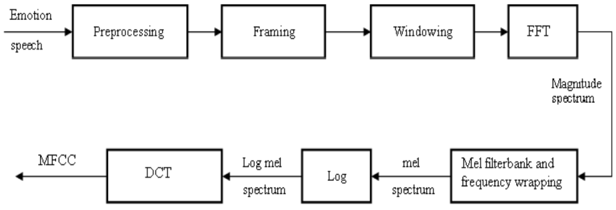


Fig3: Block Diagram of the MFCC feature extraction process

Pre-processing: The continuous time signal (speech) is sampled at sampling frequency. At the first stage in MFCC feature extraction is to boost the amount of energy in the high frequencies. This pre-emphasis is done by using a filter [4].

Framing: it is a process of segmenting the speech samples obtained from analog to digital conversion (ADC), into the small frames with the time length within the range of 20-40 msec. Framing enables the non-stationary speech signal to be segmented into quasi-stationary frames, and enables Fourier Transformation of the speech signal. It is because, speech signal is known to exhibit quasi-stationary behaviour within the short time period of 20-40 msec [4].Windowing: Windowing step is meant to window each individual frame, in order to minimize the signal discontinuities at the beginning and the end of each frame [4].

FFT: Fast Fourier Transform (FFT) algorithm is widely used for evaluating the frequency spectrum of speech. FFT converts each frame of N samples from the time domain into the frequency domain [4].

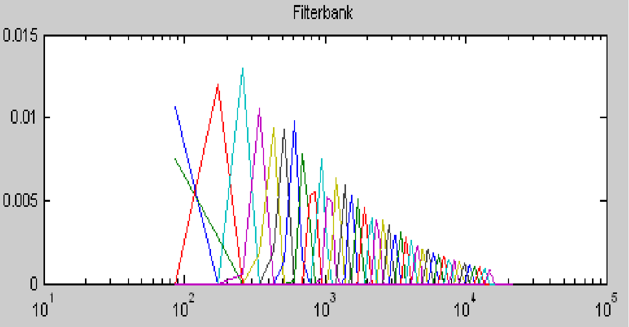
Mel Filter bank and Frequency wrapping: The mel filter bank consists of overlapping triangular filters with the cut-off frequencies determined by the centre frequencies of the two adjacent filters. The filters have linearly spaced centre frequencies and fixed bandwidth on the mel scale [4].

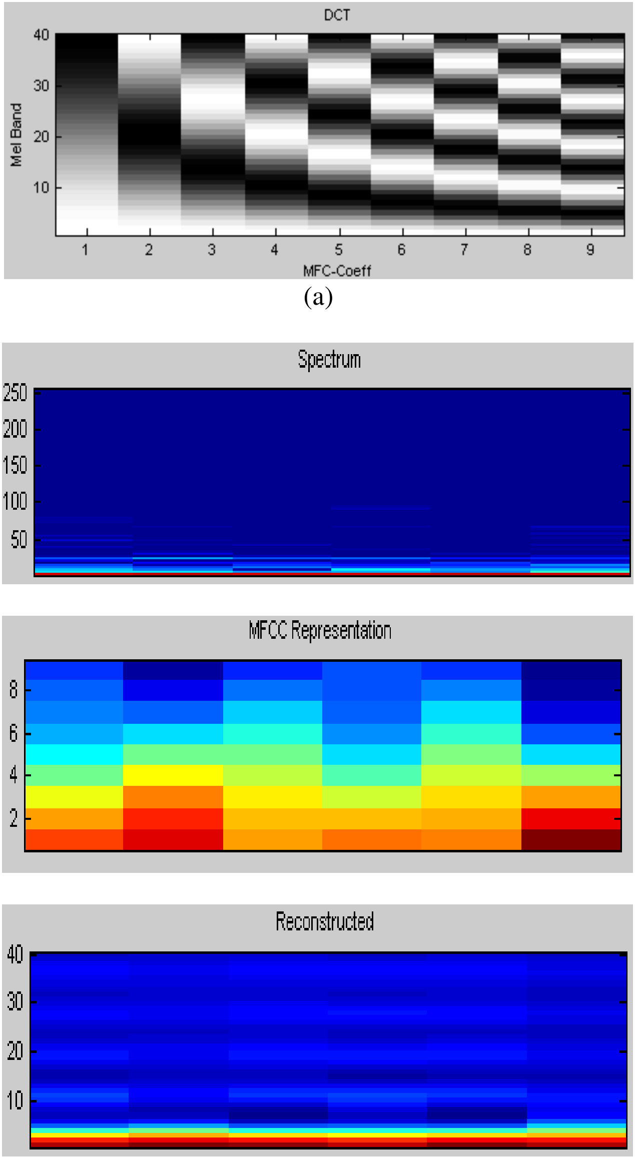
Take Logarithm: The logarithm has the effect of changing multiplication into addition. Therefore, this step simply converts the multiplication of the magnitude in the Fourier transform into addition [4].

Take Discrete Cosine Transform: It is used to orthogonalise the filter energy vectors. Because of this orthogonalisation step, the information of the filter energy vector is compacted into the first number of components and shortens the vector to number of components [4].

In this work input is a .wav file .wav file containing emotional speech utterances from Danish Emotion Database. The five emotions used are, anger, happiness, sadness, fear, and neutral state. Then Segmenting all concatenated voiced speech signal into 25.6ms-length frames. Then Estimate the logarithm of the magnitude of the discrete Fourier Transform (DFT) for all signal frames. Filtering out the center frequencies of the sixteen triangle band-pass filters corresponding to the mel frequency scale of individual segments. Estimate inversely the IDFT to get all 10-order MFCC coefficients. After that analyzing all extracted MFCC dataset for two dimension principal components and then used as an input vector for testing and training of PSO.

This all processes are implemented in Maltab program





(a)

(b)

Fig. 4: Waveforms of extracted features (a: Filtered signal; b: MFCC & Spectrum of speech signal)

5. Classifier Selection

The classification-based approaches focus on designing a classifier to determine distinctive boundaries between emotions. The process of emotional speech detection also requires the selection of a successful classifier which will allow for quick and accurate emotion identification. Currently, the most frequently used classifiers are linear discriminate classifiers (LDC), k-nearest neighbor (k-NN), Gaussian mixture model (GMM) [6], support vector machines (SVM) [4], decision tree algorithms and hidden Markov models(HMMs) [5].Various studies showed that choosing the appropriate classifier can enhance the overall performance of the system.

Gaussian mixture model allows training the desired data set from the databases. GMM are known to capture distribution of data point from the input feature space, therefore GMM are suitable for developing emotion recognition model when large number of feature vector is available. Given a set of inputs, GMM refines the weights of each distribution through expectation-maximization algorithm. GMMs are suitable for developing emotion recognition models using spectral features, as the decision regarding the emotion category of the feature vector is taken based on its probability of coming from the feature vectors of the specific model [6].

The artificial neural network (ANN) is also one of the classifier, used for many pattern recognition applications. They are known to be more effective in modeling nonlinear mappings. Also, ANN's classification performance is usually better than HMM and GMM when the number of training examples is relatively low. The classification accuracy of ANN is relatively low compared to other classifiers. The ANN based classifiers may achieve a correct classification rate of 51.19% in speaker dependent recognition, and that of 52.87% for speaker independent recognition [5].

Another classifier is Support Vector Classifier.SVM classifiers are based on the use of kernel functions to nonlinearly map the original features to a high dimensional space where data can be well classified using a linear classifier. SVM classifiers are widely used in many pattern recognition applications and shown to outperform other well-known classifiers [4].

In this work now we have a set of features available with us we will take use of the classifiers to distinguish between the different emotional states. Now we will go to use the Particle Swam Optimization as classifier for classification purpose. First we train the classifier with some inputs of different emotional states. After training the classifier, we will use it for recognizing the new given input. The process of PSO training contains labeling of extracted features and training PSO.

6. Particle Swarm Optimization

The bio-inspired computation algorithms are the excellent tools for global optimization [1]. Particle swarm optimization is global optimization algorithm that originally took its inspiration from the biologically examples by swarming, flocking and herding phenomena in vertebrate. It uses a number of agents (particles) that constitute a swarm moving around in the search space looking for the best solution. Each practical keeps track of its coordination in the solution space. Particle swarm optimization is a population based optimizing tool which work out by cooperation of each individual. The convergence ability in PSO is faster than other evolutionary algorithms (GA, SA and ACO) and also requires the less parameter for calculating the optimizing value [11].

7. Proposed Work

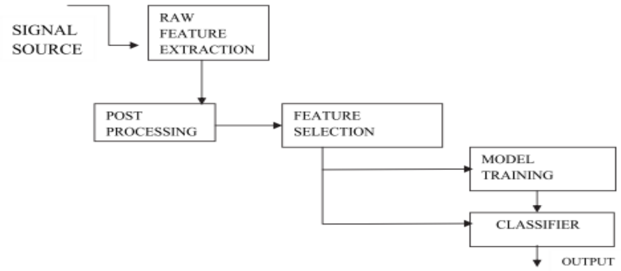


Fig5:-Generalized Block Diagram of Speech Emotion [10]

In this research work, The speech signal is used as input signal. Statistical and, Mel frequency cepstrum coefficient (MFCC) feature are extracted. These features will be then given to the Particle Swarm Optimization as a model of training and as a classifier, which classify the different emotion.

8. Flow Chart

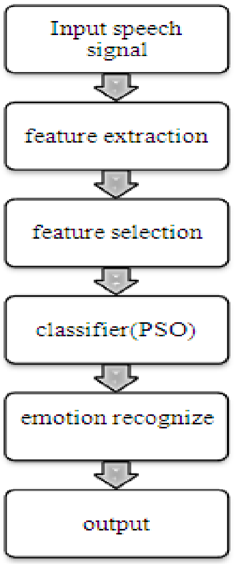


Fig6:-Proposed Flow chart of Emotion Recognition from Speech using Particle Swarm optimization

8. Conclusion

From all the reference paper, most recent work done in the field of Speech Emotion Recognition is discussed. The several classifier performances are reviewed. Success of emotion recognition is dependent on appropriate feature extraction as well as proper classifier selection from the sample emotional speech. Classifier performance is needed to be increased for recognition of speaker independent systems. In this research work we extracted the feature which are now going to give as input to the Particle swarm optimization which we are going to use as a classifier, so it may give good result as compare to other classifier. The application area of emotion recognition from speech is expanding as it opens the new means of communication between human and machine.

References

[1] Sonali Navghare,Shubh-angi Giripunje, Preeti Bajaj,” Biologically Inspired Computing System for Facial Emotion Detection: a Design Approach ” International Journal of Computer Applications (0975 – 8887)

[2] Simina Emerich , Eugen Lupu , Anca Apatean ,” Bimodal Approach in Emotion Recognition using Speech and Facial Expressions ” 2009 IEEE

[3] Shubhangi Giripunje,Mr.Ash-ish Panat ,” Speech recognition For Emotion With Neural Network: A Design Approach ” 8th International Conference, KES 2004,

[4] Vaishali M. Chavan, V.V. Gohokar ,” Speech Emotion Recognition by using SVM-Classifier ”, International Journal of Engineering and Advanced Technology (IJEAT)

[5] Firoz Shah.A, Raji Sukumar.A, Babu Anto.P,’ Automatic Emotion Recognition from Speech using Artificial Neural Networks with Gender- Dependent Databases’ 2009 IEEE DOI 10.1109/ACT.2009.4

[6] Angeliki Metallinou, Sungbok Lee and Shrikanth Narayanan,’ Audio-Visual Emotion Recognition using Gaussian Mixture Models for Face and Voice’ 2008 IEEE DOI 10.110!/ISM.2008.40

[7] Mandar Gilke , Pramod Kachare , Rohit Kothalikar , Varun Pius Rodrigues and Madhavi Pednekar ,’ MFCC-based Vocal Emotion Recognition UsingANN’(ICEEI 2012)

[8] Björn Schuller, Gerhard Rigoll, and Manfred Lang,’ HIDDEN MARKOV MODEL-BASED SPEECH EMOTION RECOGNITION’ 2003 IEEE

[9] Hyoun-joo go,keun-chang-kwalk,dae-jong-lee,hyung-genu-chun“ Emotion Recowition From the Facial Image and Speech Signal’ SICE Annual Conference in Fukui, August 4-6,2003 Fukui University, Japan

[10] S. Ramakrishnan,’ Recognition of Emotion from Speech: A Review.

[11] Hemlata s. Urade,Prof. Rahila Patel,'Study and Analysis of Particle Swarm Optimization:A Review',2nd National Conference on Information and Communication Technology(NCICT)2011.

[12] Kamran Soltani, Raja Noor Ainon,'SPEECH EMOTION DETECTION BASED ON NEURAL NETWORKS' 2007 IEEE.

[13] Yongjin Wang and Ling Guan,AN INVESTIGATION OF SPEECH-BASED HUMAN EMOTION RECOGNITION' 2004 IEEE.