Speech Emotion Recognition Using EEMD, SVM & ANN

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering
in
Computer Science and Engineering

Submitted By
Manisha
(801232012)

Under the supervision of:
Dr. Shivani Goel
Assistant Professor

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR UNIVERSITY
PATIALA – 147004

June 2014

Certificate

I hereby certify that the matter which is being presented in the thesis entitled, "***Speech Emotion Recognition using EEMD, SVM and ANN***", in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Computer Science and Engineering* submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of Dr. Shivani Goel and refers other researcher's work which are duly listed in reference section.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

Manisha

(Manisha)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

Shivani Goel

(Dr. Shivani Goel)
Assistant Professor
Computer Science and Engineering Department

Countersigned by

(Dr. Deepak Garg)
Head Computer Science and Engineering Department
Thapar University
Patiala

(Dr. S. K. Mohapatra)
Dean (Academic Affairs)
Thapar University
Patiala

Acknowledgement

I take this opportunity to express my deep sense of gratitude to my supervisor Dr. Shivani Goel for her valuable guidance, encouragement and valuable discussion for this thesis work. Words are inadequate to express the great care and interest taken by her in all aspects of my thesis work.

I express my gratitude to Dr. Deepak Garg, Head, Computer Science & Engineering Department for the motivation and inspiration that triggered me for the thesis work.

I would also like to thank all staff members and my colleagues who were always there at the need of hour and provided with all help and facilities, which I required, for the completion of my thesis work.

Last, but not least I am thankful to all those people who have directly or indirectly helped me during my thesis work.

Manisha

Manisha

801232012

ME (CSE)

Abstract

Emotion recognition system from speech is one of most advanced topics in the electronic media. Emotion detection helps the security system to prevent the data from various attacks at the cyber world. A lot of research work has already been done into this contrast but the problem of accuracy is always there. This work has been done to categorize three emotions namely HAPPY, FEAR AND SAD using the EEMD, SVM and ANN algorithms. In this work, noise levels are taken so that the emotion can be identified even though if the voice signal is highly noised. The aim of this work is to check the accuracy of the EEMD algorithm with noisy signals in contrast to the emotion detection. We proceed as detecting the noise level and segmenting the signal for the further processing. There are two segments: first part is the training part in which the system is trained to identify the further proceedings. In this part, samples of each voice category are taken and their features are fetched after successful segmentation of the voice file and further on saved into the database. The second part is the testing part in which a voice sample is taken and all the required properties are fetched and matched with the saved database values. The closest match comes out as the category of the voice file.

Table of Content

List of Figures

Abbreviations

EEMD-Ensemble Empirical Mode Decomposition

SVM-Support Vector Machine

ANN-Artificial Neural Network

GNB- Gaussian Naïve Bayes

HMM-Hidden Markov Model

SER-Speech Emotion Recognition

Chapter 1

Introduction

This chapter contains introduction to affective computing and various algorithms used in speech emotion recognition (SER) followed by thesis outline.

1.1 Affective Computing

Affective computing is the study and field of research in Artificial Intelligence that deals with recognizing, interpreting and processing emotions and simulating human affects. It is an integrative field spanning computer science, psychology, and cognitive science. This branch of computer science originated with Rosalind Picard's 1995 paper on affective computing [1]. An encouragement for the research is the ability to understand and share the feelings of another. The machine must have the capability of understanding and interpreting human's emotional state and should be able to generate an appropriate response for that particular emotion. This artificial intelligence field is a new way by which humans and machine can communicate with each other and machine can understand human emotion. By affective computing computer will be able to recognize human emotion and respond according to the emotion. It is the study of emotions and all about giving emotional abilities to machines. Emotions are not limited to certain fields; they are the one what affects our decision making and thinking. Recognition of emotions can be done by facial expression, text, speech, by mouse or keyboard click and by physiological factors. Humans can recognize emotions very well than machines. Humans learn by experience likewise we give training to machines for learning so that when communicating with a human it can detect in what state of mind that person is and how machine should behave in that situation.

By Rosalind Picard's in 1995 [1]

Affective Computing can be defined as "computing that relate to, arises from or deliberately influence of emotion".

1.1.1 Field of Affective Computing

i. Emotional information detection and recognition

Detection of emotional information starts with sensors which catch data about the human's physical state or behaviour without translating the input. The gathered data has analogy to the cues as humans detect to judge emotions in other humans. For example, a camera may catch facial expressions, body posture and gestures and a microphone may catch speech. Some sensors might detect emotional cues by measuring directly biological data such as skin temperature.

Recognizing emotional information feels the necessity for the extraction of purposeful pattern from the gathered data. This work is done using machine learning techniques which use different techniques, such as speech recognition, NLP, or facial expression detection, and produce either features or coordinates in a valence-arousal space.

ii. Emotion in machines

Design of computational device, which was proposed for exhibiting innate capabilities or which can simulate emotions convincingly, is another area of affective computing. Another approach which is based on current technological capabilities is the simulation of emotions in conversational agents between the human and the machine in order to enrich and facilitate the interactivity. Here human emotions are often associated with surges in hormones and other neuron peptides and emotions in machines might be associated with the abstraction of states associated with progress (or lack of progress) in autonomous learning systems.

1.2 Speech Emotion Recognition System

Speech emotion recognition system is a system which recognizes emotion from user's speech. The structure of speech emotion system is illustrated in Figure 1.1. The main elements for speech emotion recognition system are same as they are for any classic pattern recognition system. It has speech with emotions as input, then features are extracted and then classification is done by taking algorithm for classification.

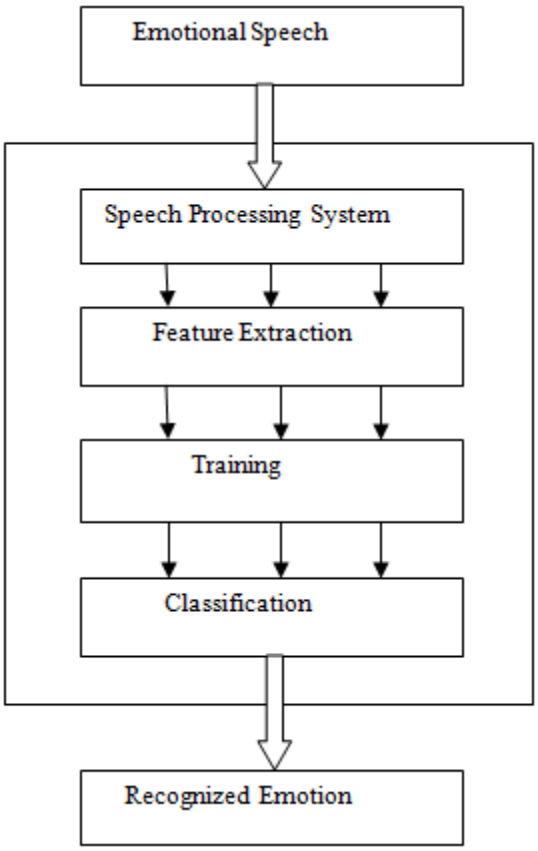

Figure 1.1: Structure of Speech Emotion Recognition System

The efficiency of the speech emotion recognition system is highly dependent upon the naturalness of database used in the system [1]. The data collected for training may be the real world situations data or recorded data by acting for emotions. After collection of the database sample, then necessary features are extracted from the speech signal. Then these extracted features are provided to the classification algorithm. Then file under test is presented to the classifier then classifier recognizes the emotion of the tested data.

Speech recognition is a computer technology that enables a device to recognize and understand spoken words, by digitizing the sound and matching its pattern against the stored patterns.The main task of speech emotion recognition is appropriate voice processing emotions. Emotions play an extremely important role in human life. It is a medium of expression of one's perspective and his/her mental state to others. Speech Emotion Recognition (SER) can be defined as the extraction of the emotional state of the speaker from his or her speech signal [2]. SER has wide range of applications such as in robotics to recognize human emotions, expert system can be

designed to help psychologist, in cars to alert that driver is not in good mood, even in the autism for autistic people. Emotion recognition solutions depend on which emotion is required to be recognized by a machine and for what purpose. In general there are six basic emotional states: neutral, happy, fear, sad, angry and disgust (or surprise). In this research the focus is on three emotional states: happy, fear and sad. Speech Emotion Recognition has different classifiers like Kernel Regression and k-nearest neighbours (KNN), Support Vector Machine (SVM), Maximum Likelihood Bayesian Classifier (MLBC), Hidden Markov Model (HMM) , Artificial Neural Network (ANN) [3].

The significance of emotion recognition from speech has risen in size in recent days to improve both the quality and efficiency of human –machine interactions. To correctly determine feature for recognition of emotion from speech a number of studies have been conducted. Even after so much of work in this field, researchers not gained that much success and accuracy for determining is still less.

1.3 Algorithms Used

In this section, all algorithms which are used in this research work are explained exclusively.

1.3.1 Support Vector Machine (SVM)

The Support Vector Machine (SVM) was first proposed by Vapnik and has attracted a high degree of interest in the machine learning research community [4]. SVM is used for the pattern recognition and for classification of patterns and it is efficient and simple computation for machine leaning algorithm. SVM has the advantage that for limited training data, it gives good classification for samples. Main idea behind Support Vector Machine is that it transforms the set of inputs to a high dimensional space using kernel function. By doing this transformation, non linear problems can be solved [5]. Input samples become linearly separable when they are converted to the high dimensional feature space. Support vector machine is based on risk minimization principle and using limited sample of information it finds good balance between the learning ability and model complexity. Optimal separable plane has to be found that separate not only two classes without error, but also make the largest difference

between them in terms of interval. It uses non-linear transformation to transform input vectors into high dimension space. SVMs are used for classification and regression and are supervised learning methods [4].

It is trained to separate the feature of one type from other types. Single SVM is created for each type. Features are given as input for testing to the SVM model. Then SVM model find the distance between the features and hyper plane. Then average distance is calculated between each type. The type to which sample belong is decided on the basis of distance.

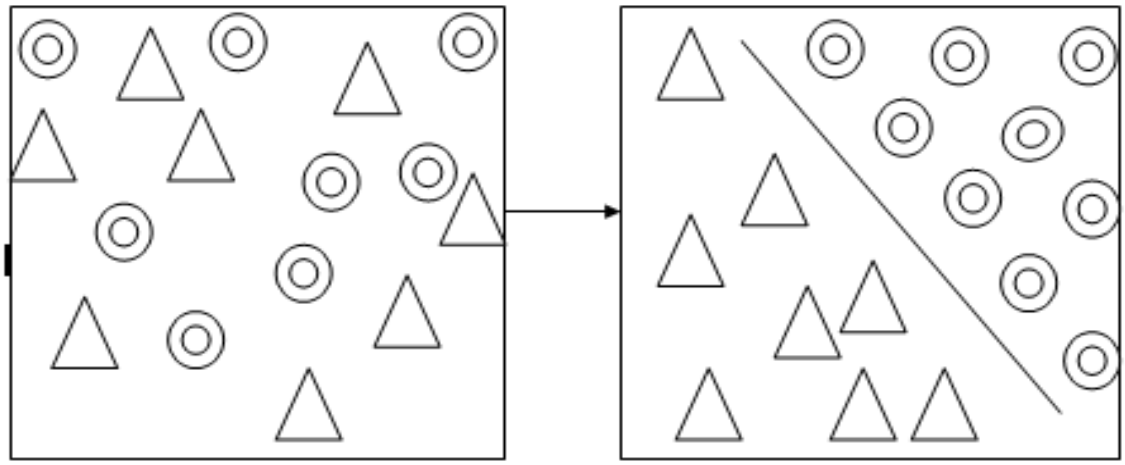

Figure 1.2: Support Vector Machine Separating Inputs [2]

Figure 1.2 shows the support vector machine in input samples in input space are converted into high dimensional space, Because of this input samples become separable [1]. SVMs are also known as Maximum Margin Classifiers because they simultaneously minimize empirical classification error and maximize the geometric margin. A maximal dimensional space is constructed by mapping input samples to the high dimensional space. On each side of the hyper plane two parallel hyper plane are constructed which separate data [4].

1.3.2 Ensemble Empirical Mode Decomposition (EEMD)

For solving the mode mixing problem in EMD, Huang proposes EEMD [4]. It is an noise assisted data analysis method that is why it is also known as noise assisted data analysis method. EEMD utilizes the noise fact. Noise fact means that addition of

noise in the time frequency space could provide uniformly distributed scale. The most common noise assisted signal processing method is the pre-whitening [6]. This pre-whitening is used in every kind of signal analysis field in which white noise is added for smoothing of pulse disturbance. Mode mixing problem in EMD is due to the uneven distribution of signals.

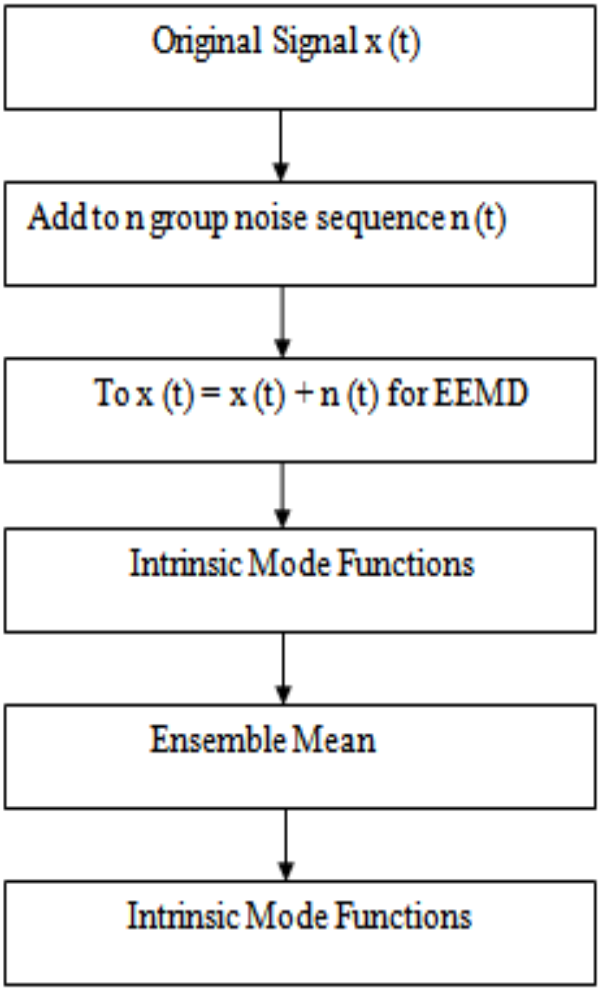

Figure 1.3: Flow Diagram of EEMD process

To decompose signals Haung added noise and uniform distribution of noise spectrum is used. A noise which is consistent for the whole time frequency distribution is added in the noise to obtain different time scales distribution. Noise will be cancelled out by each other and integration mean can be found [7]. Figure 1.3 gives flow diagram of EEMD process.

Main steps of EEMD:

i. Noise series is added to the data.
ii. Decompose signal with noise using EEMD.
iii. As the final result obtain mean of corresponding intrinsic mode function of the decomposition.

1.3.3 Artificial Neural Network (ANN)

A neural network can be defined as a model of reasoning based on the human brain. The brain consists of a densely interconnected set of nerve cells, or basic information processing units, called neurons [8, 9]. It learns from examples. **It** is configured through a learning process for a specific application. In biological systems, for learning synaptic connections are adjusted that are between the neurons. **It** is true for ANN as well. Neural networks are based on biological brains parallel architecture.

Artificial representation of human brain that tries to simulate learning process of human brain is known as Neural Network. It is inspired by biological system. ANN is made of interconnected artificial neurons. It learns by example as people. It is configured through a learning process for a specific application like as pattern recognition or data classification. In biological systems, for learning synaptic connections are adjusted that are between the neurons. It is true for ANN as well. Neural networks are based on biological brains parallel architecture. The Neural networks where we have examples for the training of the behaviour we require and where an algorithmic solution cannot be applied. Large number of interconnected neurons (processing elements) work in union to solve problems.

Figure.1.4 gives the architecture of typical neural network. It shows it has three types of layers input layer, middle or hidden layer and output layer. Input layer takes inputs and output layer produces outputs. Middle layers are layers that are hidden.

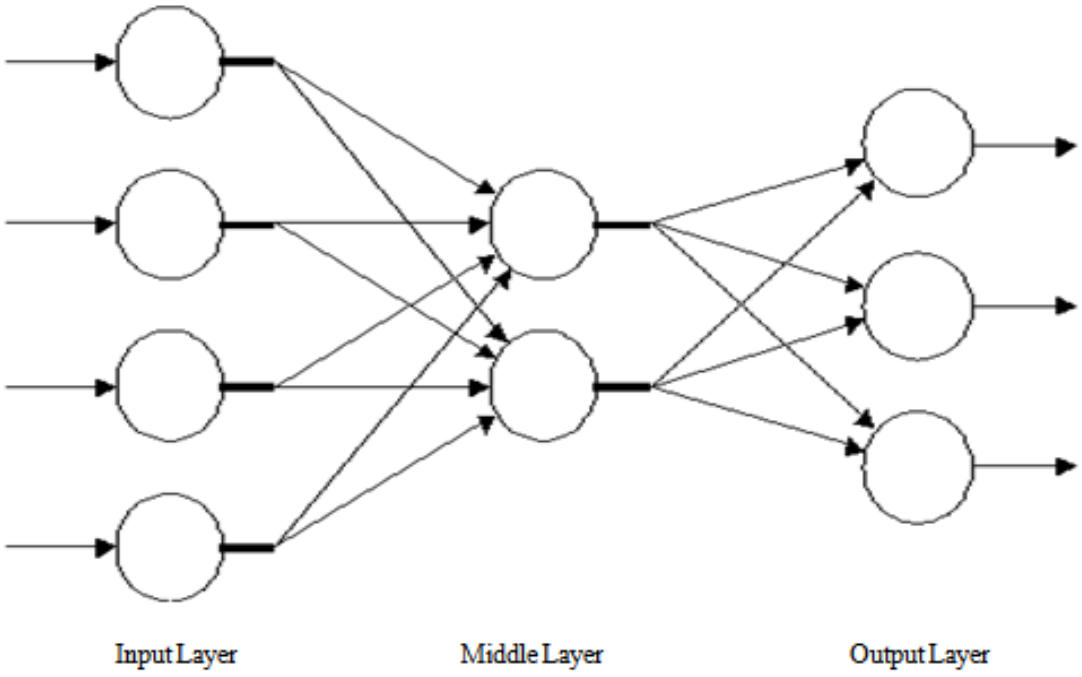

Figure 1.4: Architecture of typical Neural Network

Sum of inputs are calculated by processing units and then activation function is applied. Result to the neurons is transmitted by output line. Basic elements of artificial neuron are weights, activation function. Weights are used to determine the strength of input vectors. Each coming input is multiplied with the associated weight of the neuron connection. Activation function performs the mathematical operation on the output. Depending on the type of problem to be solved activation functions for mathematical calculation are chosen.

Single layer feed forward neural network, Multi layer feed forward neural network and recurrent neural network are the three main types of artificial neural network.

The neuron computes the weighted sum of the input signals and compares the result with a threshold value, q. The neuron uses the following transfer or activation function:

$$X=\sum_{i=1}^{n} x_i w_i$$

Y= {+1 if X>=Θ and -1 if X<Θ}

This type of activation function is called a Sign Function.

1.4 Thesis Outline

The organization of thesis is as follows:

Chapter 2- This chapter consist of literature survey done to understand the concepts of various classification algorithms and problems in emotion recognition using these algorithms.

Chapter 3-This chapter describes the problem statement which is addressed in this work.

Chapter 4- This chapter gives description of the work implemented, details of the work and snapshots of the work done.

Chapter 5- This chapter contains confusion matrices for each feature showing accuracy in percentage and comparative analysis.

Chapter 6- This chapter has conclusion of work done and future work that may be possible.

Chapter 2

Literature Survey

Lots of work has been done in speech emotion recognition and it is still going. Work done in some approaches is discussed below.

2.1 Neural Network and Ensemble Empirical Mode Decomposition

Kashi Dai et al. has used acoustic and landmark features to identify various types of emotional states in speech [10]. They analyzed 2442 utterances from the Emotional Prosody Speech and extracted 62 features. To recognize different emotional states neural network classifier is applied and 10 fold techniques was used to evaluate the performance of classification. They also stated accuracy in emotions like hot anger, sadness, and neutral states and in cold anger etc. Their work is to aid people who have difficulty in identifying emotion. They used resilient back propagation training algorithm in neural network. Based on their experiment they achieved 90% accuracy for identifying hot anger and neutral, 80% of accuracy for recognizing happy and sadness and 62% and 49% accuracy for identifying 4 and 6 emotions respectively. They also had 80% accuracy in finding emotions like cold anger and hot anger that is emotions with different emotions. And found that there are many confusing emotion pairs like happy and interest, happy and panic, interest and sadness, panic and hot anger. The accuracy in identifying these pairs of emotions was relatively low.

Aishah Abdul Razak et al. discussed an approach of speech emotion recognition that is used in a system Voice Driven Emotion Recognition Mobile Phone (VDERM) [11]. They used 18 speech features for classification. Linear predictive coding analysis was used for extraction of feature from speech. Two classifier methods were used one is Neural Network and other is Fuzzy Model. They found that both of these methods of classification have their own advantage and disadvantage. They found that Neural Network performs well with large training set where as fuzzy model performs well with small training set. It means neural networks accuracy increases as the training set increases. And neural network can work well, that is can give good accuracy rate even if small number of features are extracted and taken for classification where as fuzzy

model requires large number of feature extraction. By both methods recognition rate of up to 60% can be achieved.

Kyung Hak Hyun et al. proposed a feature which gives better results than existing feature [12]. This feature is useful in amplifying the signal so that results due to external interference do not degrade. This feature is log frequency power ratio (LFPR). And it is compared with feature LFPC. It give better results in signal magnification but not any clear improvement. Classifier used was Neural Network. They have used 45 sentences. They did their experiment on four emotions and total 5400 sentences were used, 78.2% accuracy is attained for the four features. By using this new feature 10% improvement is achieved than existing feature.

Work done by Stefan Scherer et al., presented emotion recognition not happy, sad etc. but in a different manner that is amount of stress. They recognized stress level in speech signal [13]. To do this experiment a setup was made so that different voices have different stress level. Fifteen subjects were taken i.e. voices of fifteen people have been recorded. To record speech with different stress levels, a game was played by subjects. And three questions were asked at different levels of game so that voices with different stress levels can be recorded. First question was asked at starting of level two, then at starting of level five and after that before starting of level ten. For classification between different stress levels echo state network of neural network was used. Stress levels were labelled between 0 and 100 to each speech signal. This classifier can have better recognition even than the human.

Kun Lu et al. utilized features of three type's frontal view facial expression, profile view facial expression and audio [14]. This paper had recognized emotion from speech as well as from facial expression. It used multi model fusion classifier. This classifier is based on neural network. It recognizes seven basic emotions. In facial frontal view 20 points were positioned which incorporate eyes, chin, brow and nose. In facial profile view, only six points were situated which include lips, nib and the nose, eyes and chin. And for audio speech signals higher sampling rate was used. They extracted fifty speech frames from each speech. Obviously first we need to train the neural with hints. Neural used here has total three layers. Input layer for feature sequence, output layer has seven and four additional components. Each of seven layer belongs to each emotion. And later four features are for hints.

Emotion Recognition using Neural Network from speech by J.Nicholson et al. , used one-class-in-one neural [15]. They have attained 50% accuracy rate with eight emotions. It has eight sub-neural networks. It used one sub – neural network for one emotion. All sub networks incorporate of four layers. Output layer of each sub network has single node and can have value of only 0 or 1. Fifty male and fifty female subjects were used for making the training database. For both male and female networks were trained separately. After that phase which is training, testing is done on networks. Testing done was open testing and closed testing. In closed, set of same data which is used for training is used for the testing. Whereas in open testing different files were used when we are checking the system.

Nermine Ahmed Hendy used feature of type pitch, formants, jitter, shimmer and temporal features [16]. Classifier chosen was Artificial Neural Networks. In this seven different emotions were categorized. Speech files count taken in the experiment was 535 for all 8 emotions. Total features of different types extracted were 175. Training data is divided into two parts. Out of 540 files, 480 speech files were used for training and 121 speech samples for testing. In this paper, authors compared different Artificial Neural Network approaches to find that different type of Neural Network taken with different set of features will give different accuracy rate. Results show that accuracy rate is effected by how much features are selected, format of features, sorting of training data and ANN type and architecture. Which format of feature should be used row feature or standardized feature. Standardized feature give approximately same rate with all Artificial networks. Sorting of training data affects the average accuracy rate. When dealing with online systems to decrease the effort and time delay use reduced number of features.

Christin L. Lisetti has used Neural Network to categorize emotion by using facial expression [17]. They used Back Propagation of Neural Network. Knowledge Base used to train the system have collection of different type of faces, each belonging to different individual like with glasses, beard etc. Only two emotions are categorized in this are happy and neural. They have done this using full face network and Lower face network. In the first, full face was taken and 40 points were marked on face, one per pixel and network has been given training with 40 images which has 20 different persons. From this images used for training were 30, selected from these 40 and

remaining 10 for testing. And in lower face network only mouth, chin and nose area was covered. This study shows that zooming to the particular area of face give better accuracy rate.

Two neural algorithms were compared namely Multilayer perceptron Neural Network and Generalized Feed Forward Network [18]. They applied same techniques of classification on same dataset and used to categorize seven emotions. 493 samples were taken. In this Prosodic features were analyzed and extracted. Neural Network used has seven output and fourteen input components. With Multilayer Perceptron Neural Network 89.62% accuracy rate was found. In this 10% of total files have been used for classification. At hidden layer, in Multilayer Perceptron Neural Network 10 processing elements were used. With Generalized Feed Forward Neural Network 98.8% accuracy rate was found. Five hidden layers were used to get best result. At hidden layer 19 processing elements were used.

Vidhyasaharan Setu proposed an approach which uses a new features weighted factor along with prosodic features [19]. Weighted feature was based on the instantaneous frequency. And this feature decomposes the voice file into AM-FM signals using Ensemble Empirical Mode Decomposition. These decompose signals were symmetric about zero. Three dimensional weighted features were combined with pitch, energy and zero cross by rate to obtain six-dimensional feature. 40ms frames were used for computation. Using combination of new and acoustic give 9% increase in accuracy rate as compared to when only acoustic feature were used.

Quasi-Gradient search algorithm is based on Ensemble Empirical Mode Decomposition to classify emotions in speech. First lower bound of error was calculated using Non-Linear correlation coefficient (NCC). To increase the correctness rate, low ensemble number can be chosen and then can be increased exponentially. In this work,Quassi-Gradient Search was also applied to solve the mode mixing problem of ensemble empirical decomposition by finding features from speeches of different scales. Quashi Gradient Search was applied to find what parameter can be used with EEMD to improve accuracy. This proposed algorithm was fast and highly efficient [20].

Ling Ha et al. suggested two new methods stress and emotion categorization in speech [21]. First approach used EEMD to find Intrinsic Mode Functions (IMFs). And then it finds the average entropy from the intrinsic mode functions. In second method, speech was divided into sub-bands then mean of spectral energy was found out. In this method, data with three sub bands were tested: critical, Bark and ERB. In this work, new features were compared with existing features. Accuracy rate of 77% were attained with ERB band and anisotropic filtering. But without filtering 53% accuracy rate was achieved.

2.2 Support Vector Machine (SVM) and Hidden Markov Model (HMM)

Yixiong Pan et al. have implemented speech emotion recognition using SVM. In this, they implemented three emotions happy, sad and neutral [22]. Features extracted were energy, pitch, linear predictive spectrum coding, mel-ferquency spectrum coefficients. Different combinations of features are applied for classification. In this emotion database, there were 406 speech files. Different combinations of features were applied and out of all combination of features best feature combination is MFCC +MEDC +Energy.

Chaudhari et al. used two algorithms HMM and SVM for recognizing emotions. They used HMM and SVM as classifier and classify five emotional states such as anger, Happiness, sadness, surprise and neutral state. Features related to pitch, energy and mel-frequency capstral coefficients, fundamental frequency were used. They observed that both the classifiers obtained approximately similar accuracy for the emotion recognition. And they observed that it is mandatory to record correct emotional speech database samples because accuracy of the system highly depends on emotional speech database [1].

In the work of Bjorn Schuller et al., they used two methods to recognize the emotions and were propagated and their results were compared [23]. One method Gaussian Mixture model was used to classify global statistics framework of an utterance. Second method was Hidden Markov model by which temporal complexity was increased. For both the methods same database was used and same voice files were used for testing. Total twenty features were extracted for training. Both the methods

even with the same training material and same testing material differ in their behaviour.

Tin Lay Nwe et al. proposed a method for classification of emotion from speech using HMM in a different way [24]. The Block Diagram is shown in Figure 2.1. Approach described here uses short time log frequency power coefficients for the training of the system. And for classification of emotions they have used hidden markov model. They have done classification of six emotions. In this method for the training purpose 60 emotional voice files are taken each from 12 speakers. They have attained average accurate rate of 78% and best accuracy is 98%. For the training of system a codebook is constructed using means of all features using LFPC. And then these codebooks are submitted to the hidden markov model classifier. They have shown that best performance is given by four states- hidden markov model.

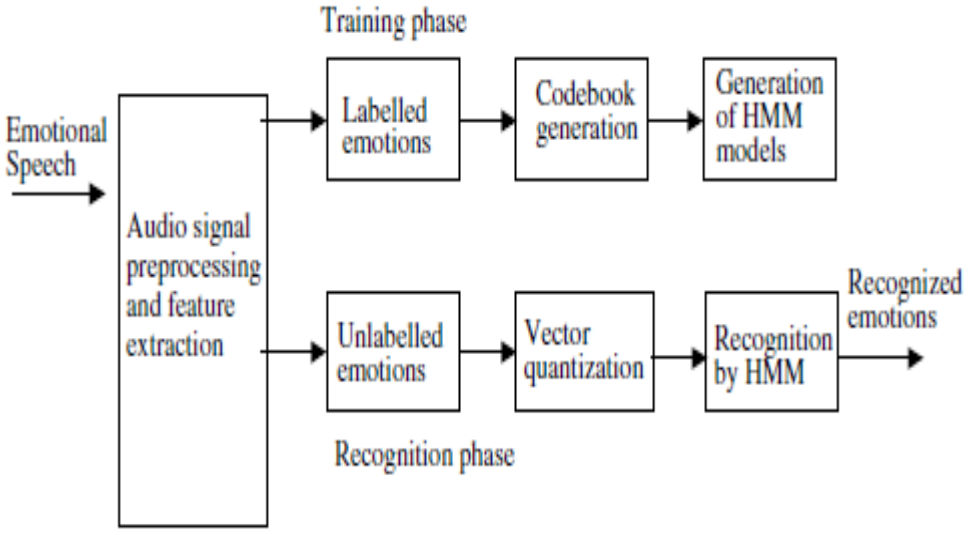

Figure 2.1: Block Diagram [24]

Kazuhiko Takahshi, proposed a system which uses bio-potential features for training the system [25]. Five emotions were recognized in this approach. Audio and video signals were gathered with equipment with three sensors and two personal computers. Support Vectors machine (SVM) was used for classifying the signal i.e. to which category it belongs. Five support vector machines were used. Each vector machine belongs to one emotion. Average recognition rate with five emotions earned is 41.7% while 66.7% is achieved for three emotions. They proved that bio-potential signals can be used to detect emotion and Support vector machine was suited for this task.

Emotion recognition from speech by using Rough set and Support Vector Machine done by Jian Zhou et al [26]. They have shown that, this approach helps in decreasing calculation cost even with high accuracy rate. Support vector machine is used for classification in this approach. Rough Set theory is used to select the set of features. In this, 1200 samples were used from which 600 randomly were selected for training. First they tested with extraction of 37 features using only SVM and attained 77.91%. Then they tested with 13 features and using Rough set and SVM and attained 74.75% accuracy rate. Though it is only 3.16% less than that with SVM but cost of this approach is less because we only have to deal with 13 features instead of 37.

Kazuhiko Takahashi used bio-potential features that are pulses and skin conductance response [27]. In this three emotions were evaluated which were positive emotion, negative emotion and normal. For classification support vector machine was used. Bio potential features were used along with the speech. By using these parameters 41.2% of accuracy is attained. By this they shown that these two bio potential signals can be used to detect emotion and support vector machine can be applied to this work.

Shashidhar G. Koolaugudi compared two algorithms Support Vector Machine (SVM) and Gaussian Markov Model to examine variety of speech signals. Vowel omit points were used to find word boundaries in fifteen samples. Experience was carried out at all by taking prosodic features, spectral features and by combining both types of these features. Accuracy rate of 36% was attained with prosodic features using only beginning words [28].

2.3 k-Means and Gaussian Naive Bayes

Abhinav Dhall et al., put forward a method which uses k-Means to shape up the images [29]. That is to normalize the image. Feature extracted for this approach were pyramid of histogram of gradients and local face quantization. It used constraint Local Model (CLM) for finding the points on face. It used two types of data, person specific that is experiment with different persons and person independent means does not depend on persons, person can be same. With person specific recognition rate of this proposal was 73%. And with person independent correctness rate was 44%. Overall system perfection rate was 55%.

Only primary emotions were recognized using k-Means for classification [30]. It uses Multimodal Classifier of k-Means. Prosodic Features were used to train the system. Dataset for this paper have been made by recording speech from television programmes specially reality shows. All speech recording was done in noisy environment and with bad sound quality. Because natural noise is never free from noise. 500 voice files were taken for the experiment. In this method first k-Means has been applied. It finds three intensities centroid. Difficulty not arises with finding opposite emotions like happy and sad but arises when dealing with similar category like happy and excited. For each of the emotions which are under classification gives average correctness rate found was 50%.

Murugappan et al., put forward a new way for extracting features from Electrocephalogram signal [31]. They did multi-resolution testing of wavelet functions. Sixty four biosensors were used. They collected data from six persons which all lie in the age between 21 to 27. In this authors had compared two clustering approaches Fuzzy C-Mean and Fuzzy k-Means for categorization of emotion. They examined the whole process with twenty four biosensors along with sixty three biosensors. Fuzzy k Means gave high correctness rate with twenty four channels instead of sixty three channels. Whereas fuzzy C-Mean performed well in categorizing with both twenty four and sixty four channels. They created different dataset for both sixty three and twenty four channels. They worked on classification of three features happy, disgust and fear. Wavelet function has been used.

An emotion classifier which used bio-signals was recommended by Taehyun Kim [32]. It studied ECG signals for categorization of emotions. Classifier used was EM algorithm which is based on the k-Means. In this , two works have been done, one to recognize emotion and other to find stress level that subject is in stress or not. And perfection rate shown for emotions is in between 55.8% to 75.9%. For stress level percentage was 83.2%.

A method was presented, with semi supervised approach. Motion energy image and motion history image were used for the features representation. Classification is done using k-Means. On the basis of correlation coefficients average correctness rate reached was above 90% [33].

An approach was proposed by Dimitrios Ververdies et al. which used Gaussian Naive Bayes for classification [34]. 500 speech files were taken over which different 87 features were calculated. Sequential forward selection method was used to select the set of features from 87 features. In this 90% of speech files from the whole dataset has been taken for training purpose and remaining 10% of files were used for testing. In this two experiments have conducted, one is to categorize gender and other without gender categorization. Correctness rate 50.6% has been found without taking gender information. And 61.1% accuracy rate has been attained with gender categorization.

Dimitrios Ververidis has done work for emotion recognition using features like pitch, energy and formants were extracted from 500 speech files [35]. First probability of signals matching was found. Class classification was done using Gaussian Naïve Bayes with different densities. Sequential floating forward selection was used to select features which give best solution. Mixtures of two or three Gaussian densities were used with naïve bayes. With single /Gaussian distribution accuracy rate 61.8% for males and 57.6% for females was attained. And with two densities 55.6% correctness rate was found. They concluded that with combination of densities, gain of 4% was obtained. No gain when tests were done individually for each gender.

Emotion Recognition using Gaussian Naïve Bayes with two densities was presented [36]. This method gave 3% gain over sequential forward selection even with gender categorization or without it. Experiments done on two datasets of 500 and 1300 samples. First set of samples contain isolated words and sentences. And second set of 1300 samples contain words, paragraphs and sentences. Correctness rate of first rate of first set was 6 to 10% greater than that of second set. Gaussian Naïve Bays improves the accuracy rate by 3% as compared with sequential forward selection.

2.4 Hybrid Approaches

Wei Zhang et al. presented a speech emotion recognition system using Ensemble Empirical Mode Decomposition and Hilbert-Haung method [6]. Amplitude feature of emotional signal marginal spectrum is extracted. It used Support Vector Machine (SVM) to classify the signal. In this ensemble empirical mode decomposition and Hilbert Haung transformation was used. By comparisons of different emotional speech signals they recognized different emotions have different gaps in speech

emotion recognition rate. It was found in this that proposed method can efficiently recognize emotion and have practical significance.

Schullar et al., has used semantic signal features and prosodic features and their fusion for recognition of emotion by analyzing speech [37]. They classified seven emotions neutral, joy, anger, irritation, fear, sadness and disgust. In this, average magnitude distance function was used to extract the features and extracted 20 features. They have done classification using dynamic time warping algorithm. Features mean is calculated individually for each sample signal and these regularize to standard deviation. For classification minimum distance between two vectors of a class was chosen as parameter. They have tested on 17 speakers' voice and with 595 total voices sample i.e. 85 samples per emotion. With prosodic features, total recognition rate earned is 80.3%. Some emotions are mixed with other emotions too. With non-verbal phrases they have attained 88.1% recognition rate. And with semantic features 62.1% recognition rate is reached.

Aishah Abd. Razak have done emotion recognition using linear prediction analysis algorithm in emotion recognition for the feature extraction and for the purpose of the training of system [38]. Fuzzy is used for emotions categorization that is for classification. 22 speech features were used and extracted for the proposed approach. In this, total 1200 samples were selected. From which 240 voice files i.e. 40 files of emotions were selected to form the knowledge base for the training. Whereas 360 files i.e. 60 for each emotion were selected for the recognition i.e. files for testing. They have also compared computer recognition rate with the human recognition rate. In according to proposed method 68.59% of accuracy rate is attained by computer whereas 62.35% accuracy rate is attained by human which is less than the recognition rate of computer.

Aishah Abdul Rajak et al., used 18 speech features for the system. First they have proposed emotion classifier which is shown in Figure 2.2. Two different databases have been used one for training of the system and second for testing i.e. files under this are used for testing purposes. In this approach they compared two methods neural network and fuzzy model. Using both of these methods 60% accuracy rates was achieved. They have done two experiments for two different purposes. Samples for training the system are 840 in number from which they made three sets. First set have

all 840 samples, second set have 540 files taking from that 840 files only and set three have only 240 files taking from remaining files. First experiment was done to find that how size of training samples will affect the performance of two methods. They found that neural network gives high accuracy rate with larger size of training data whereas fuzzy performs well with small sat of data. Their second experiment was to find number of features required to have better results and their finding is that neural network gives high accuracy with limited number of features that is less than or equal to 15 whereas fuzzy give high rate with number of features exceeding 15 [39].

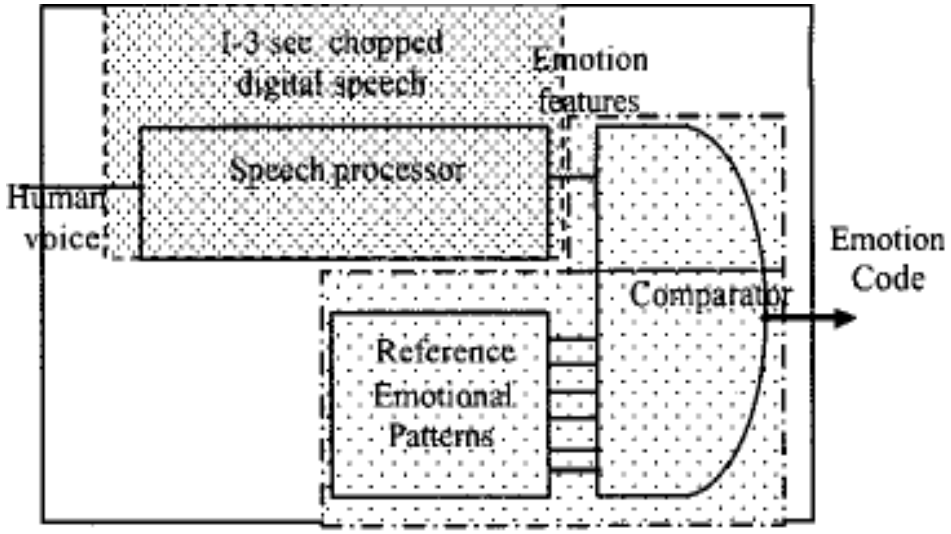

Figure 2.2: Design for Emotion Classifier [39]

Chapter 3

Problem Statement

This chapter contains the problem discussed in this research work.

Emotion Recognition using speech started from affective computing by Picard's. It has big area of applications such as in Robotics, Expert System, and psychology and for autistic people. Various classifications and clustering mechanism are there which work efficiently in the field of speech processing.

The problem statement can be more rectified and said as:

In digital world, emotion recognition is one of the recent topics. A lot of research work has been done into emotion recognition from speech but problem is to have high correctness rate. Detecting the motion of speech is not that easy as it seems to be. We tried to improve the accuracy of the system with noisy signals using EEMD. This work has been done to categorize three emotions namely 'HAPPY', 'FEAR' and 'SAD' using the EEMD, SVM and ANN algorithms. Noise levels are taken so that the emotion can be identified even though if the voice signal is highly noised.

Chapter 4

Details of Research Work

In this chapter, solution to problem we discussed in previous chapter is discussed exhaustively.

4.1 Work Explanation

Work is in Speech Emotion Recognition and it is developed by using MATLAB. In this we recognize three emotions happy, fear and sad from speech. It follows following steps:

i. Records

Voice files are taken in .wav format. Three types of voice files are taken happy, sad and fear. These all files are stored in binary form. All these three type of record are taken for training the system so that it can recognize emotion. All features of these signals are extracted.

ii. Add Noise

Noise is added to the voice signal to obtain signal with noise. Noise is added to the signal because original signal is never free from noise. Signal can have different levels of noise but it is never free from noise. Voice samples in database are recorded from microphone and they are recorded in an environment so that noise does not come in signal. But in actual signal has noise. So, that is why noise is added. Noise Level has to be selected as 10 or 20 etc. Then the noise in binary form is added to the signal which is also is in binary form. Mean of features of all signals extracted.

iii. Segmentation

For segmentation of signals EEMD algorithm is used. It decomposes signals on the basis of threshold value. In this a random Number is generated say for example

100.Then first 100 bits of first signal are in one segment and threshold value is calculated. Threshold value is calculated by calculating mean of peak values of first 100 bits (random number choose).Then 101th bit is compared with threshold and if it less than or equal to threshold, then it is added in the same segment and checking of next bit continue till a bit having peak value greater than threshold encounter. When a bit having value greater than threshold encounter, then that bit will be in new segment. Then also for next segment follow same procedure that is first find random number, threshold etc.

iv. Feature Extraction

Next step is feature extraction in which feature are extracted from each voice file according to their respective types. Seven features of different types are extracted from the files. And mean for each type of file for each feature is taken. For example say for happy file type each feature is calculated for each happy file, then mean of a single feature for whole happy database is taken and saved. Likewise mean of all feature are extracted for each file of the same type. Likewise features of sad and fear are also extracted. Maximum frequencies, Minimum frequency, Average frequency, Roll off, Loudness, Spectral efficiency, Pitch Density are features which are extracted.

v. Storage of value of features

Extracted feature are stored in binary form. For this SVM algorithm is used. And also test files features are also stored in binary form.

vi. Upload Test file

File which is to be test for finding emotion is uploaded. File must be in .wav format.

First segmentation is done on this file also and then features are extracted of this file in same way as others. Then in binary format extracted are stored by using SVM algorithm.

vii. Classification

It is a process in which we identify type of emotion on the basis of extracted feature from sample file and tested file. For this work ANN is used. Back Propagation Neural Network is used.

For neural, first input is happy record features and second input will be features of file under test. Then neural applies activation function and find the result by comparing each feature of both the inputs and give result as showing difference in each feature of both the inputs respectively. Then this is repeated by taking sad records features as first input and second input remains same .Then repeated for fear records. ANN gives different results for happy, fear, sad comparing with test file. Then the final result is giving by comparing results of ANN which is minimum and closest to the file tested.

4.2 Flowchart of Research Work Implementation

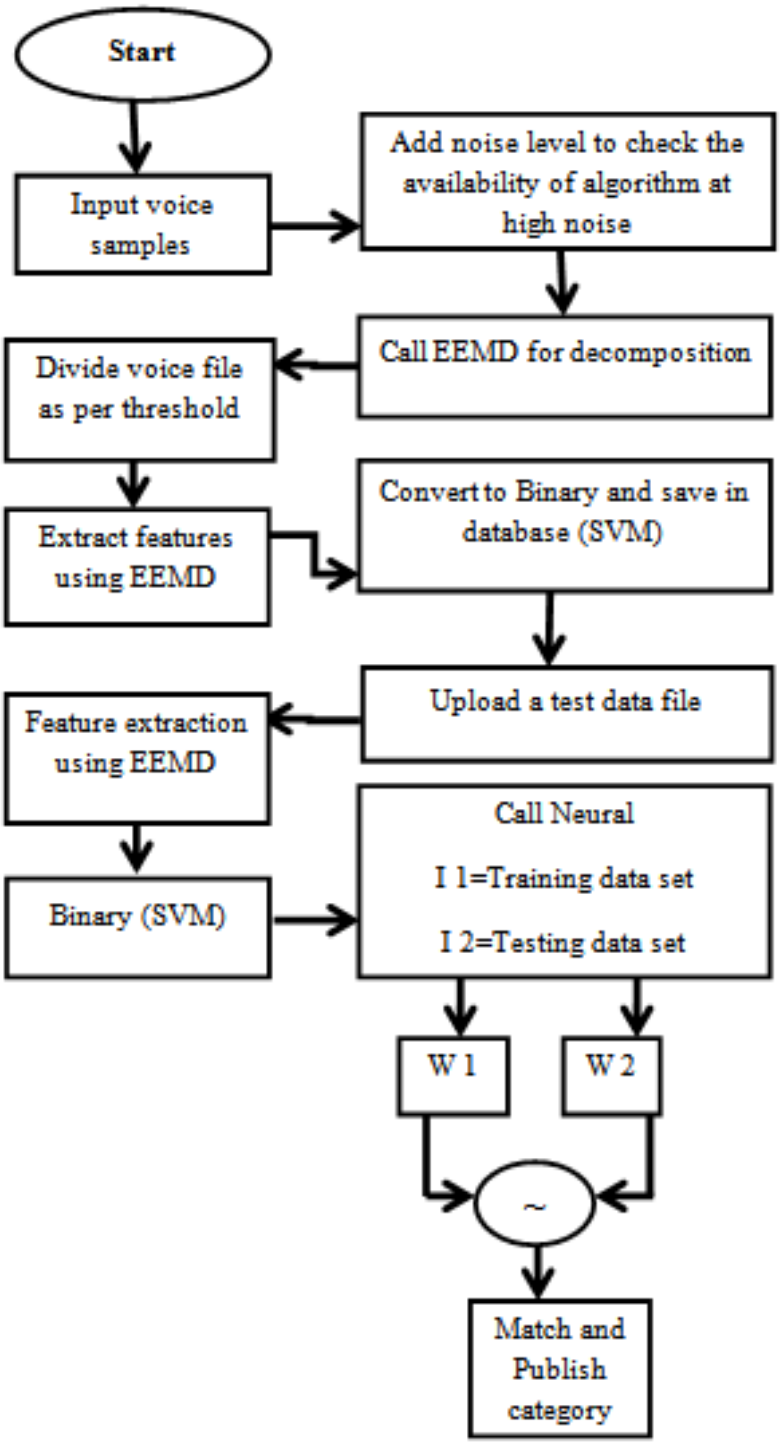

Figure 4.1: Flowchart of Implementation

4.3 SNAPSHOTS

4.3.1 Interface

Figure 4.2 shows interface. It has three parts built on it. First part is Training module which is used to train from the sample files. In this module we upload files and add noise then we decompose signal by using EEMD. A testing module which is used to upload file for testing of emotion by using k-mean, GNB and Custom Neural Network and back-propagation. Third is pre processing area in which values of extracted features are shown. Value of features changes when we upload file, when we add noise and when we decompose signal.

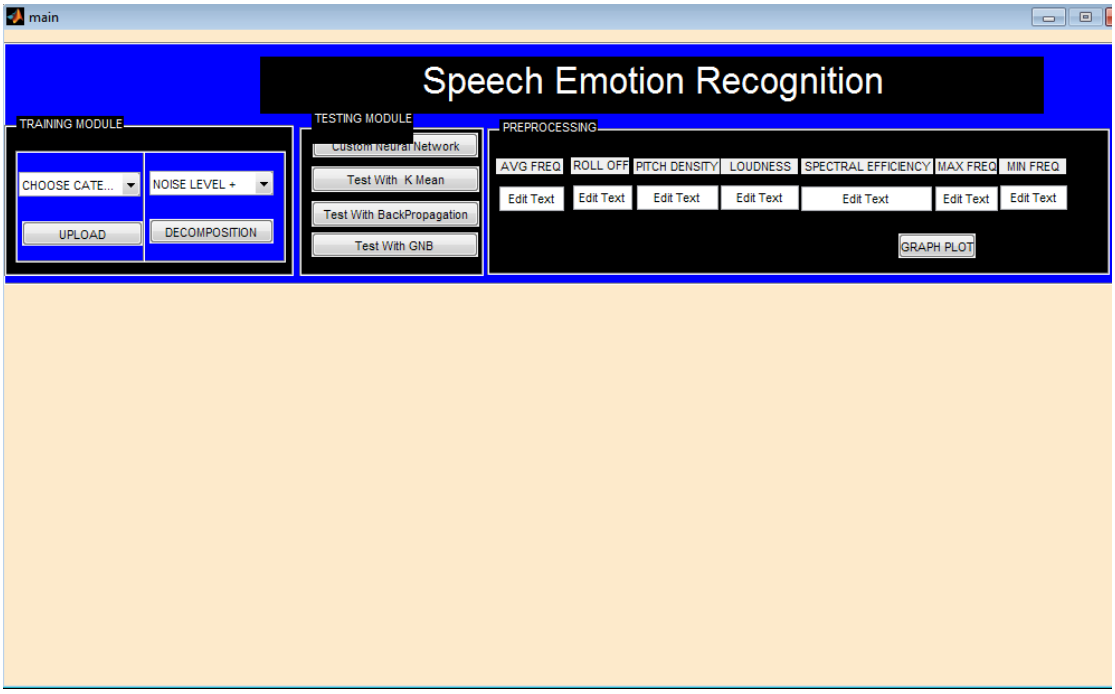

Figure 4.2: Interface

4.3.2 Training

Figure 4.3 shows how sample database can be uploaded for training. After selecting category click on Upload Button. Then happy files for training are uploaded. In the same way files for fear and sad are uploaded as shown in Figure 4.4 and Figure 4.5 respectively.

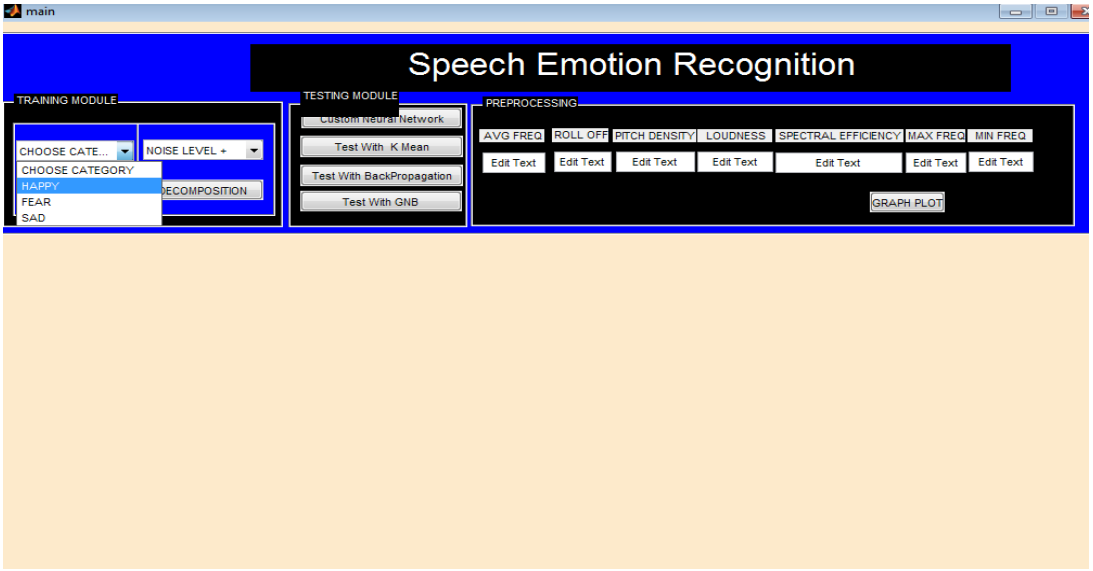

Figure 4.3: Uploading Happy Database

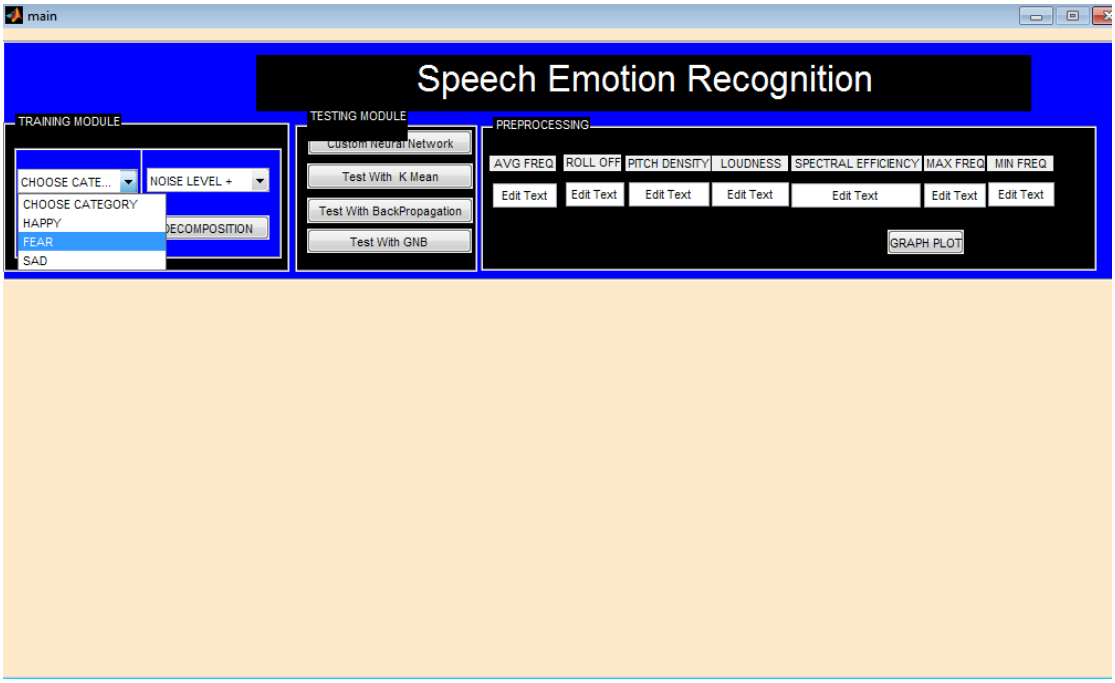

Figure 4.4: Uploading Fear Database

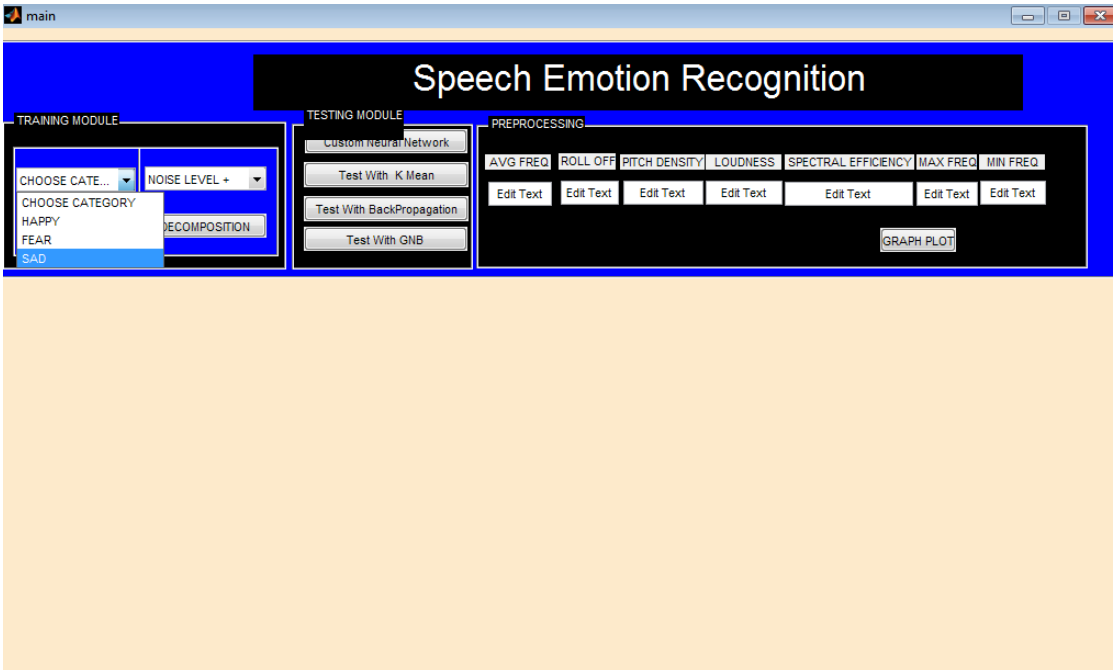

Figure 4.5: Uploading Sad Database

4.3.3 Signal Generation

Figure 4.6 shows original audio signal generated without noise of a happy file. It shows signal of any one file from all happy files. Likewise, Signal of fear and sad file can be generated.

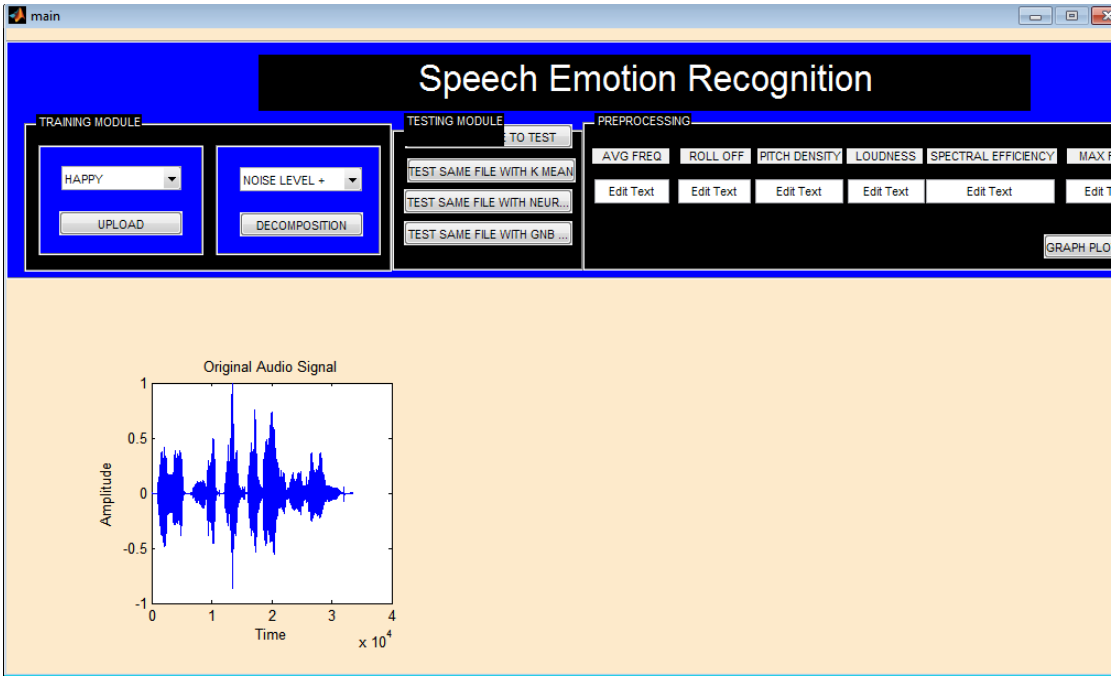

Figure 4.6: Original Audio Signals Generated

4.3.4 Inserting Noise

Figure 4.7 show addition of noise by selecting noise level. Noise level is in the percentage of original signal. Then shows signal with noise. Signal is stored in binary form so noise is also added to signal in binary form. And again signal is stored in binary form.

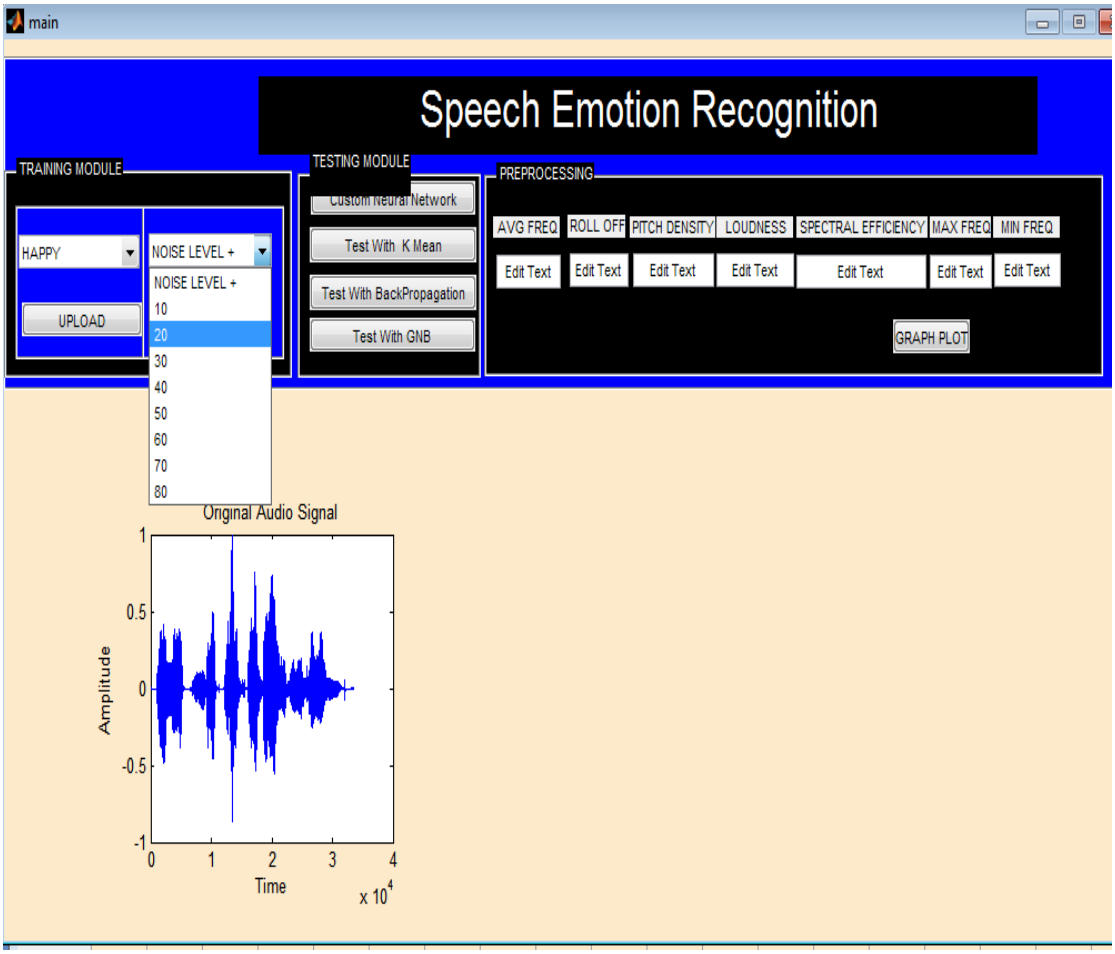

Figure 4.7: Inserting Noise into the Original Signal

Figure 4.8 is showing signal generated and extracted features after adding noise to the signal. In this, noise in binary form is added into original signal which is also in binary form. Then value of all seven features is calculated on generated signal with noise. Value of features after calculation is shown in pre-processing part.

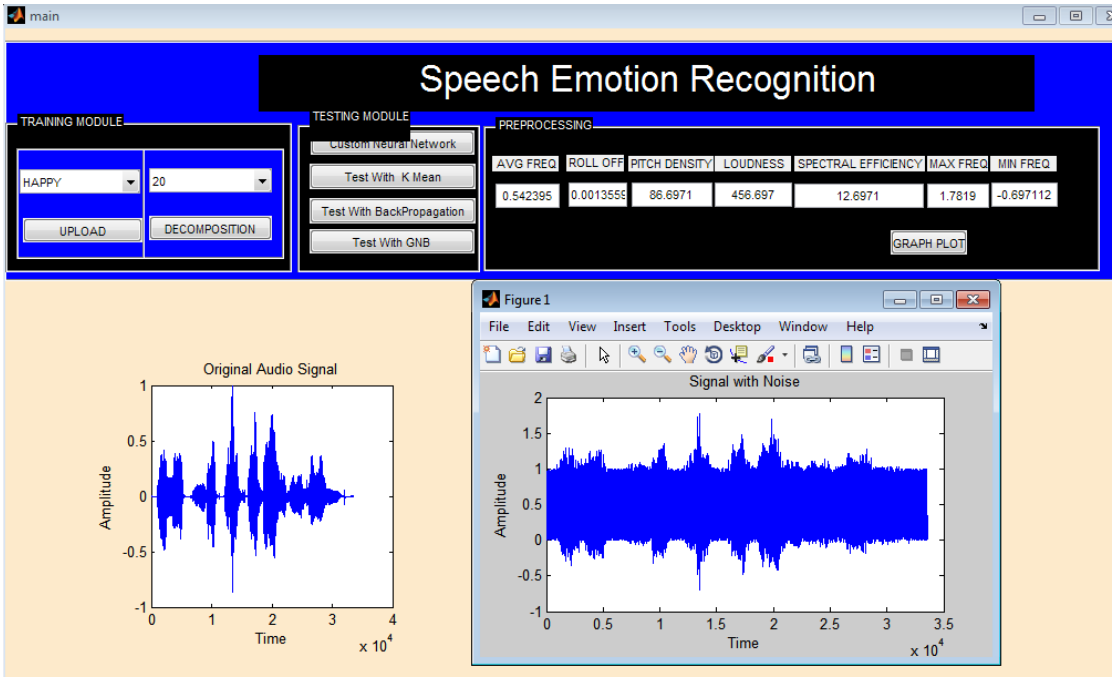

Figure 4.8: Signal Generation and Feature Extraction after Addition of Noise

4.3.5 Decomposition

Figure 4.9 showing data for happy updated and signal modified by EEMD and all the features for happy are now again extracted after decomposition. Graphs shown are smooth signal modified and modified by EEMD. Likewise, data for Fear and Sad are updated as shown in Figure 4.10 and Figure 4.11.

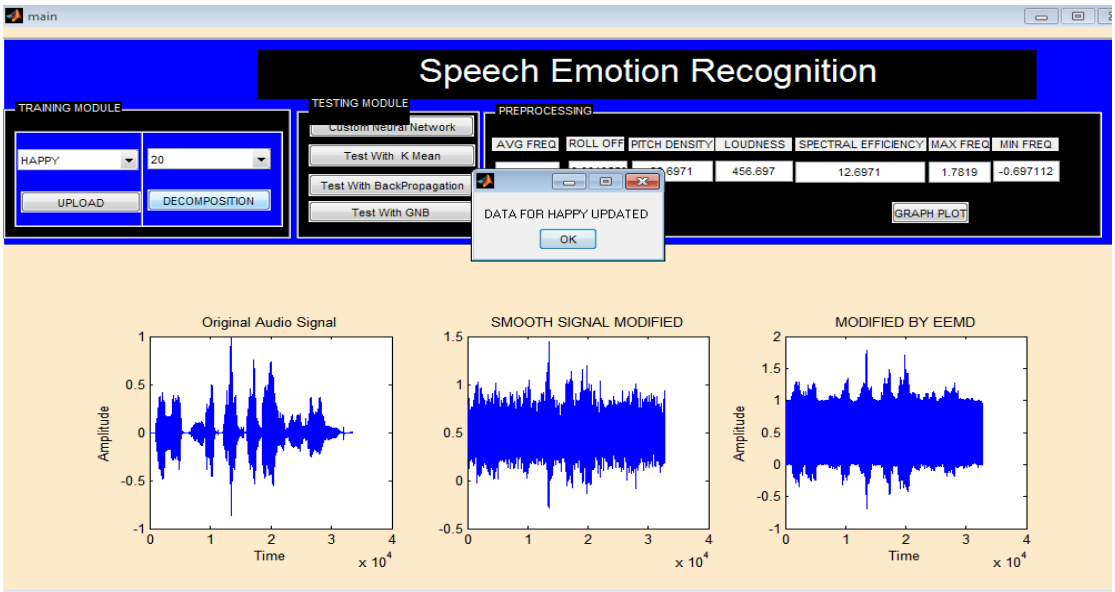

Figure 4.9: Decomposition and Data for Happy Update

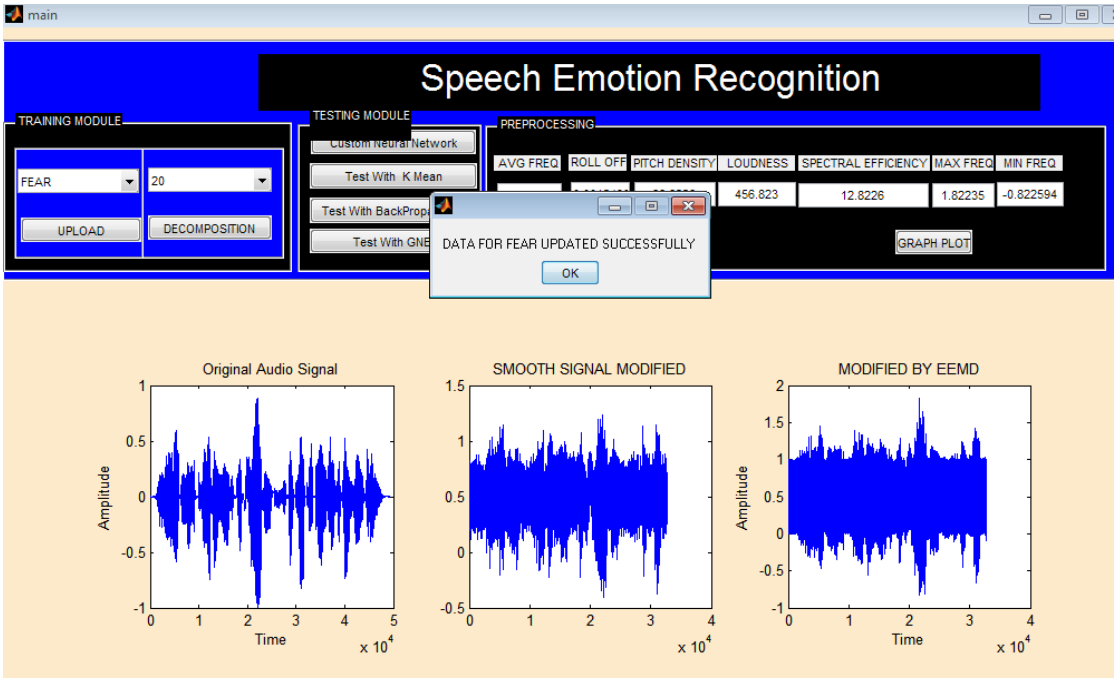

Figure 4.10: Decomposition and Data for Fear Updated

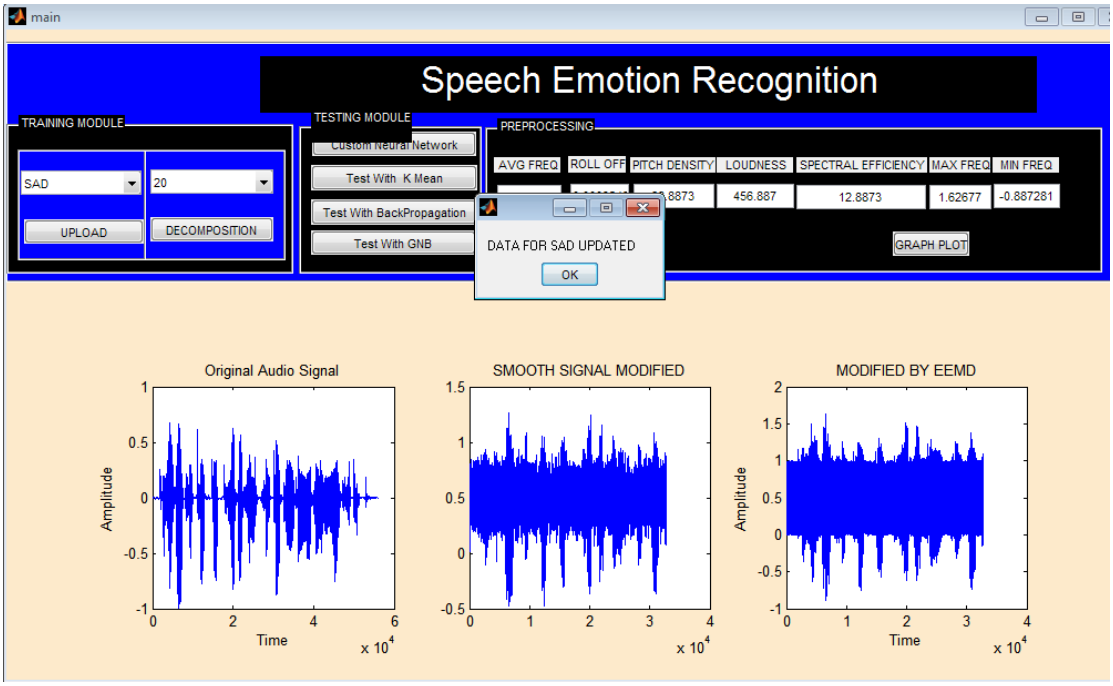

Figure 4.11: Decomposition and Data for Sad Updated

4.3.6 Testing

Figure 4.12 showing selection of a file for classification. In this, a file of wav type is added by testing module for testing. All features of this file are extracted. By comparing these features with training data features result is calculated.

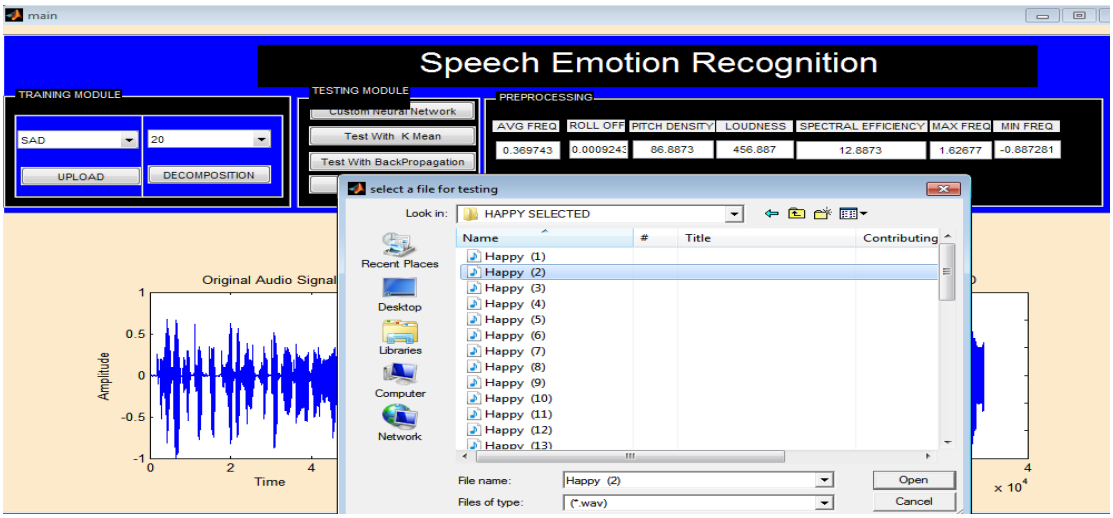

Figure 4.12: File Testing

4.3.7 Classification

Figure 4.13 shows final result as happy after applying classification. We have uploaded happy file so result is happy. In the same way we can upload files for fear and sad for recognition. It uses Back Propagation Neural Network provided by MATLAB.

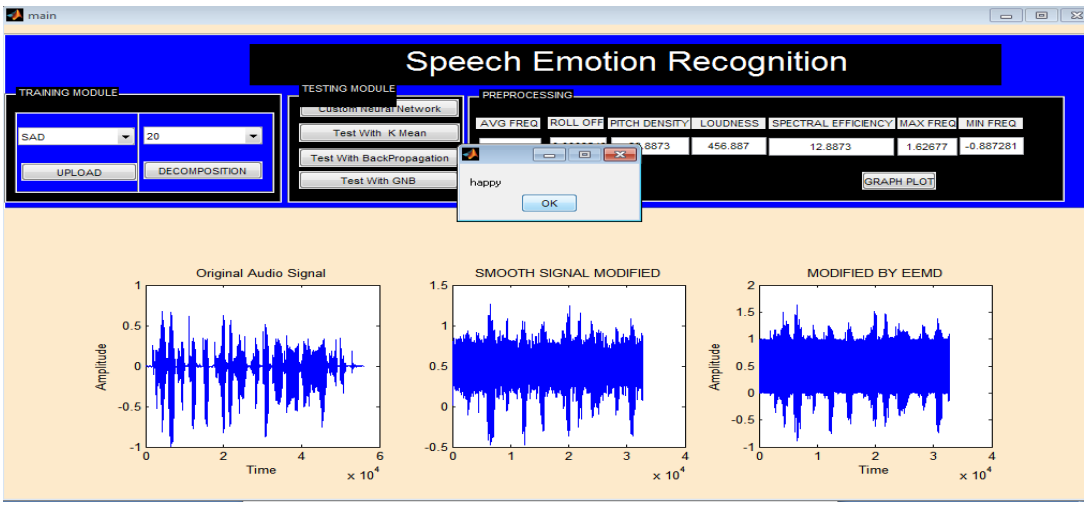

Figure 4.13: Final Results for Happy File

Figure 4.14 Shows result for fear file. Fear file was selected to check and it was correctly recognized as fear type.

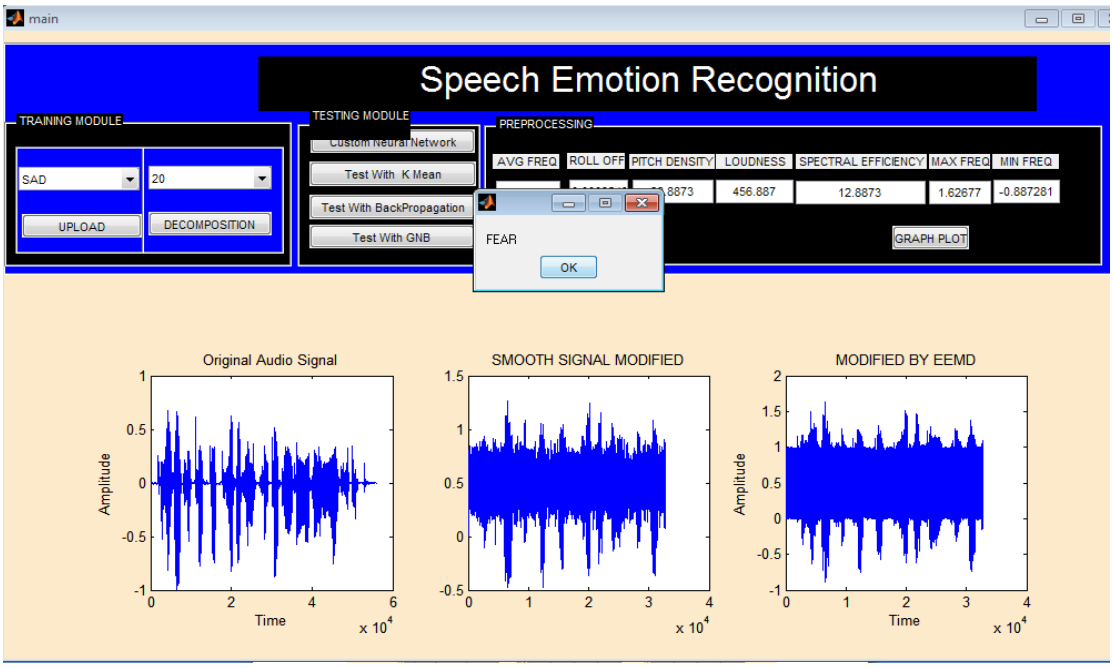

Figure 4.14: Final Result for Fear File

Figure 4.15 Shows result for sad file. Sad file was selected to check and it was correctly recognized as sad type.

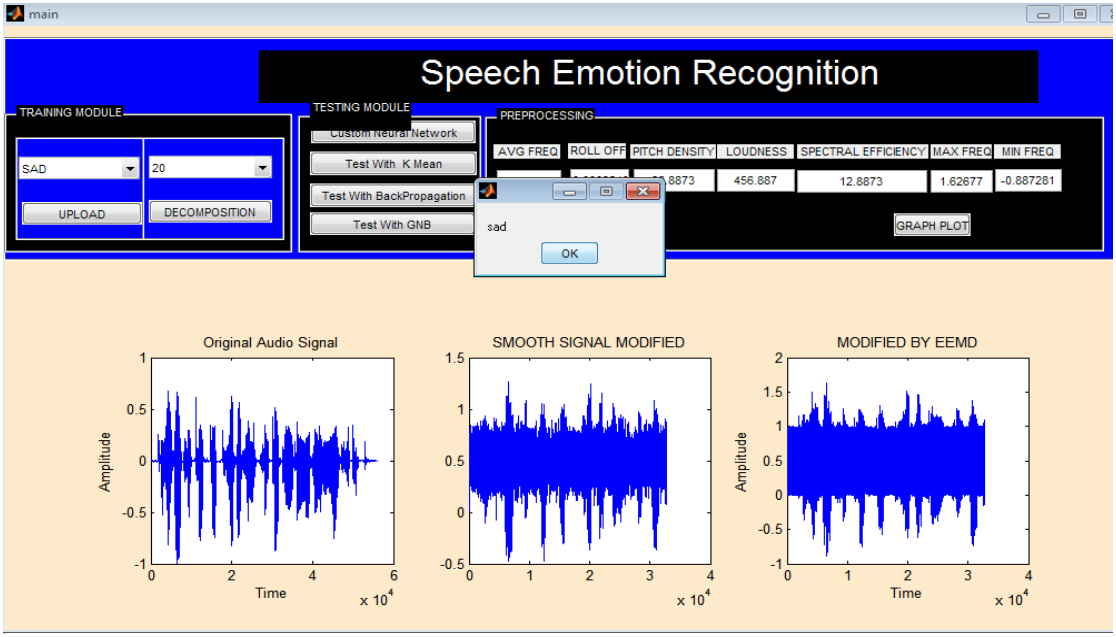

Figure 4.15: Final Result for Sad File

Chapter 5

Results and Analysis

This chapter is focussed towards results for various emotions, followed by comparative analysis.

5.1 Confusion Matrix for Fear

Figure 5.1 shows confusion matrix for Fear. It shows that 56.67% of fear type files are correctly recognized as fear and 26.55% of fear types are recognized wrong i.e. of other type. And 31.33 % of other types of files are wrongly recognized as fear. And 75.46% of other types of files are correctly recognized as others.

CONFUSION MATRIX IN PERCENTAGE		
	FEAR	OTHERS
FEAR	56.6700	26.5500
OTHERS	31.3300	75.4500

Figure 5.1: Confusion Matrix for Fear

5.2 Confusion Matrix for Sad

Figure 5.2 shows confusion matrix for Sad. It shows that 75.20% of Sad type files are correctly recognized as sad and 22.53% of fear types are recognized wrong i.e. of other type. And 14.80 % of other types of files are wrongly recognized as Sad. And 85.80% of other types of files are correctly recognized as others.

CONFUSION MATRIX IN PERCENTAGE		
	SAD	OTHERS
SAD	75.2000	22.5397
OTHERS	14.8000	85.8000

Figure 5.2: Confusion Matrix for Sad

5.3 Confusion Matrix for Happy

Figure 5.3 shows confusion matrix for Happy. It shows that 65% of happy type files are correctly recognized as happy and 27.17% of happy type are recognized wrong i.e. of other type. And 25 % of other types of files are wrongly recognized as Happy. And 87.50% of other types of files are correctly recognized as others.

CONFUSION MATRIX IN PERCENTAGE		
	HAPPY	OTHERS
HAPPY	65	27.1739
OTHERS	25	87.5000

Figure 5.3: Confusion Matrix for Happy

5.4 Confusion Matrix for all three Emotions

Figure 5.4 shows confusion matrix for all three emotions in percentage. It shows that 75.20% files of type 'fear' are correctly recognized, 3.34 % files of type 'fear' are wrongly recognized as 'sad' and 2.5% of files are recognized as 'happy' and remaining files are not recognized. Similarly, 68.67% files of 'sad' are correctly recognized, 8.38% files of type 'sad' are wrongly recognized as 'fear' and 9.67% of files are recognized as 'happy' and remaining files are not recognized. Likewise, 85.20% files of 'happy' are correctly recognized, 3.12% files of type 'happy' are wrongly recognized as 'fear' and 15.67% of files are recognized as 'sad' and remaining files are not recognized.

CONFUSION MATRIX IN PERCENTAGE			
	FEAR	SAD	HAPPY
FEAR	75.2000	3.3400	2.5000
SAD	8.3800	68.6700	9.6700
HAPPY	3.1210	15.6700	85.2000

Figure 5.4: Confusion Matrix for all three Emotions

5.5 Comparative Analysis

Performance of proposed technique is measured with most recent techniques i.e. K-Means, Custom Neural Network and GNB classifier. Figure 5.5 shows that our technique outperforms in every case. Among all techniques Custom Neural Network gives best result and New Approach is giving better result than Custom Neural Network. After doing experiment it has been found that proposed technique is best suited for large number of files.

In addition to this proposed approach is capable of achieving accuracy close to 90% which itself shows the advantage over all other technique. There is drastic difference in performance of New Approach with that of GNB classifier.

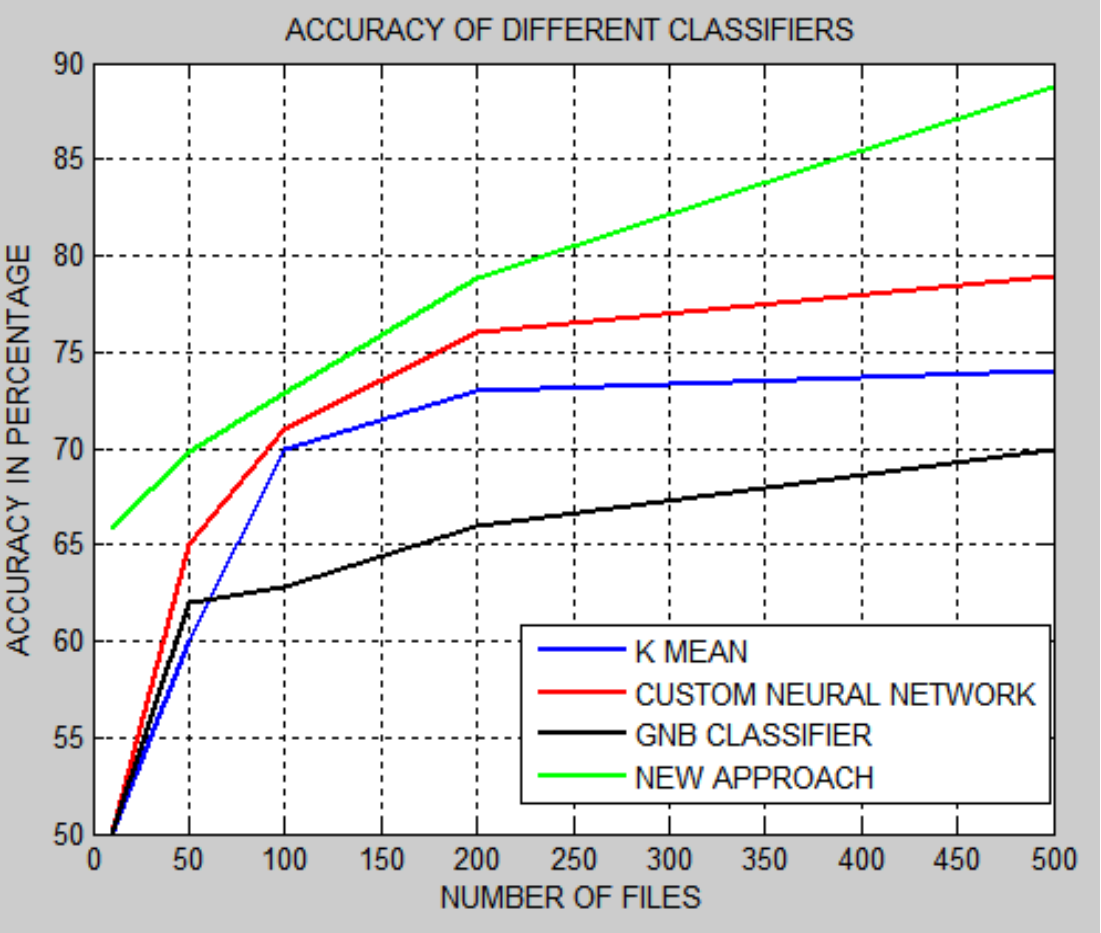

Figure 5.5: Comparative Analysis

Chapter 6

Conclusion and Future Work

In this chapter, our work is concluded along with future work.

6.1 Conclusion

After the successful implementation of this project we conclude that EEMD algorithm is an effective algorithm when it comes to the segmentation of the wav signals and also it is very helpful in finding out the sufficient features on the basis of which we can proceed further to test a wave file .We also conclude that the accuracy with the EEMD algorithm is approximately 90% in contrast of the EMOTION DETECTION which is good enough.

6.2 Future Work

Although the results are quite satisfactory but in future a combinational algorithm of EEMD , HMM can also be applied to check out whether there is any improvement in the result or not. We have already found that EEMD performs well in case of segmentation hence if it can be combined with HMM, it may produce unexpected results.

References

[1] A. B. Ingale, D. S. Chaudhari , "Speech Emotion Recognition using Hidden Markov Model and Support Vector Machine", International Journal of Advanced Engineering Research and Studies, vol. 1, 2012, pp. 316-318.

[2] R. Lawrence, Fundamentals of speech recognition. Pearson Education India, 2008.

[3] L. R. Rabiner and B.-H. Juang, Fundamentals of speech recognition. PTR Prentice Hall Englewood Cliffs, 1993, vol. 14.

[4] D. K. Srivastava and L. Bhambhu, "Data classification using support vector machine." , Journal of Theoretical & Applied Information Technology, vol. 12, 2010, pp. 1-7.

[5] P. Shen, Z. Changjun, and X. Chen, "Automatic speech emotion recognition using support vector machine" , in IEEE International Conference on Electronic and Mechanical Engineering and Information Technology, vol. 2, 2011, pp. 621–625.

[6] W. Zhang, X. Zhang, and Y. Sun, "Based on eemd-hht marginal spectrum of speech emotion recognition", in IEEE International Conference on Computing, Measurement, Control and Sensor Network, 2012, pp. 91–94.

[7] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method", Advances in adaptive data analysis, vol. 1, no. 01, 2009, pp. 1–41.

[8] M. D. Richard and R. P. Lippmann, "Neural network classifiers estimate bayesian a posteriori probabilities", Neural computation, vol. 3, no. 4, 1991, pp. 461–483.

[9] D. Doye, U. Kulkarni, and T. Sontakke, “Speech recognition using modified fuzzy hypersphere neural network”, in Proceedings of the International Joint Conference on Neural Networks, vol. 1, 2002, pp. 65–68.

[10] K. Dai, H. J. Fell, and J. MacAuslan, “Recognizing emotion in speech using neural networks”, Telehealth and Assistive Technologies, 2008, pp. 31–38.

[11] A. A. Razak, R. Komiya, M. Izani, and Z. Abidin, “Comparison between fuzzy and nn method for speech emotion recognition”, in IEEE International Conference on Information Technology and Applications, vol. 1, 2005, pp. 297–302.

[12] K.-H. Hyun, E.-H. Kim, and Y.-K. Kwak, “Robust speech emotion recognition using log frequency power ratio”, in IEEE International Joint Conference on SICE-ICASE, 2006, pp. 2586–2589.

[13] S. Scherer, H. Hofmann, M. Lampmann, M. Pfeil, S. Rhinow, F. Schwenker, and G. Palm, “Emotion recognition from speech: Stress experiment”, in LREC, 2008.

[14] K. Lu and X. Zhang, “Audio-visual emotion recognition using neural networks learned with hints”, in IEEE International Conference on Image and Graphics, 2013, pp. 515–519.

[15] J. Nicholson, K. Takahashi, and R. Nakatsu, “Emotion recognition in speech using neural networks”, Neural computing & applications, vol. 9, no. 4, 2003, pp. 290–296.

[16] N. A. Hendy and H. Farag, “Emotion recognition using neural network: A comparative study”, in World Academy of Science, Engineering and Technology, vol. 7, 2013, pp. 1149-1155.

[17] C. L. Lisetti and D. E. Rumelhart, “Facial expression recognition using a neural network”, in FLAIRS Conference, 1998, pp. 328–332.

[18] K. Khanchandani and M. A. Hussain, “Emotion recognition using multilayer perceptron and generalized feed forward neural network”, Journal of Scientific and Industrial Research, vol. 68, no. 5, 2009, pp. 367.

[19] V. Sethu, E. Ambikairajah, and J. Epps, “Empirical mode decomposition based weighted frequency feature for speech-based emotion classification”, in IEEE International Conference on Acoustics, Speech and Signal Processing, 2008, pp. 5017–5020.

[20] Z. Shen, Q. Wang, Y. Shen, J. Jin, and Y. Lin, “Accent extraction of emotional speech based on modified ensemble empirical mode decomposition”, in IEEE International Conference on Instrumentation and Measurement Technology Conference (I2MTC), 2010, pp. 600–604.

[21] L. He, M. Lech, N. C. Maddage, and N. B. Allen, “Study of empirical mode decomposition and spectral analysis for stress and emotion classification in natural speech”, Biomedical Signal Processing and Control, vol. 6, no. 2, 2011, pp. 139–146.

[22] Y. Pan, P. Shen, and L. Shen, “Speech emotion recognition using support vector machine”, International Journal of Smart Home, vol. 6, no. 2, 2012, pp. 101-108.

[23] B. Schuller, G. Rigoll, and M. Lang, “Hidden markov model-based speech emotion recognition”, in IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, 2003, pp. II–1.

[24] T. L. Nwe, S. W. Foo, and L. C. De Silva, “Speech emotion recognition using hidden markov models”, Speech communication, vol. 41, no. 4, 2003, pp. 603–623.

[25] K. Takahashi, “Remarks on svm-based emotion recognition from multimodal bio-potential signals”, in IEEE International Workshop on Robot and Human Interactive Communication, 2004, pp. 95–100.

[26] J. Zhou, G. Wang, Y. Yang, and P. Chen, “Speech emotion recognition based on rough set and svm”, in IEEE International Conference on Cognitive Informatics, vol. 1, 2006, pp. 53–61.

[27] K. Takahashi, “Remarks on computational emotion recognition from vital information”, in Proceedings of 6th International Symposium on Image and Signal Processing and Analysis, 2009, pp. 299–304.

[28] S. G. Koolagudi, N. Kumar, and K. S. Rao, “Speech emotion recognition using segmental level prosodic analysis”, in International Conference on Devices and Communications (ICDeCom), 2011, pp 1–5.

[29] A. Dhall, A. Asthana, R. Goecke, and T. Gedeon, “Emotion recognition using phog and lpq features”, in IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011, pp. 878–883.

[30] D. Kaminska, T. Sapinski, and A. Pelikant, “Recognition of emotional states in natural speech”, in IEEE International Conference on Signal Processing Symposium (SPS), 2013, pp. 1–4.

[31] M. Murugappan, M. Rizon, R. Nagarajan, S. Yaacob, I. Zunaidi, and D. Hazry, “Eeg feature extraction for classifying emotions using fcm and fkm”, International Journal of Computers and Communications, vol. 1, no. 2, 2007, pp. 21–25.

[32] T. Kim, D. Shin, and D. Shin, “Towards an emotion recognition system based on biometrics”, in IEEE International Joint Conference on Computational Sciences and Optimization, vol. 1, 2009, pp. 656–659.

[33] H. Yuan and C. Wang, “A human action recognition algorithm based on semi-supervised kmeans clustering”, in Transactions on edutainment VI. Springer, 2011, pp. 227–236.

[34] D. Ververidis and C. Kotropoulos, “Automatic speech classification to five emotional states based on gender information”, in Proceedings of 12th European on signal Processing conference, 2004, pp. 341–344.

[35] D. Ververidis and C. Kotropoulos, “Emotional speech classification using gaussian mixture models”, in IEEE International Symposium on Circuits and Systems, 2005, pp. 2871–2874.

[36] D. Ververidis and C. Kotropoulos , “Emotional speech classification using gaussian mixture models and the sequential floating forward selection algorithm”, in IEEE International Conference on Multimedia and Expo, IEEE, 2005, pp. 1500–1503.

[37] B. Schuller, M. Lang, and G. Rigoll, “Automatic emotion recognition by the speech signal”, Institute for Human-Machine-Communication, Technical University of Munich, vol. 80290, 2002.

[38] A. A. Razak, M. H. M. Yusof, and R. Komiya, “Towards automatic recognition of emotion in speech”, in proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology, 2003, pp. 548–551.

[39] A. A. Razak, R. Komiya, M. Izani, and Z. Abidin, “Comparison between fuzzy and nn method for speech emotion recognition”, in IEEE Third International Conference on Information Technology and Applications, vol. 1, 2005, pp. 297–302.

List of Publications

[1] Manisha, S Goel , “Speech Emotion Recognition using EEMD, SVM and ANN”, accepted for publication in second International Conference in Emerging Research in Computing, Information, Communication and Applications – ERCICA 2014, Bangalore, August, 2014(Sponsored by Elsevier).