# Emotion Recognition in Speech Using Back Propagation Algorithm (BPA)

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

**Master of Engineering**
in
**Information Security**

*Submitted By*
**Rohit Katyal**
**(Roll No. 801233017)**

Under the supervision of:
**Ms. Rupinderdeep Kaur**
**Lecturer**

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR UNIVERSITY
PATIALA – 147004

**July 2014**

# Certificate

I hereby certify that the work which is being presented in the thesis entitled, "Emotion Recognition in Speech using Back Propagation Algorithm", in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Information Security* 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 *Ms. Rupinderdeep Kaur* and refers other researcher's work which are duly listed in the reference section.

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

(Rohit Katyal)
801233017

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

Ms. Rupinderdeep Kaur
Lecturer, CSED
Thapar University

**Countersigned by**

**(Dr. Deepak Garg)**
Head
Computer Science and Engineering Department

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

# Acknowledgement

No volume of words is enough to express my gratitude towards my guide, **Ms. Rupinderdeep kaur**, Lecturer, Computer Science and Engineering Department, Thapar University, who have been very concerned and have aided for the entire material essential for the preparation of this thesis report. He helped me to explore this vast topic in an organized manner and provided me with all the ideas on how to work towards a research oriented venture.

I am also thankful to **Dr. Deepak Garg**, Head of Department, CSED, **Dr. Maninder Singh,** Associate Professor, **Dr. Jhilik Bhattacharya**, P.G. Coordinator, for the motivation and inspiration that triggered me for the thesis work.

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

Most importantly, I would like to thank my **Parents** and the **Almighty** for showing me the right direction out of the blue, to help me stay calm in the oddest of the times and keep moving even at times when there was no hope.

**Rohit Katyal**
**801233017**

# Abstract

Speech emotion detection refers to discovering the speech category based on the training and testing to the database provided. This research work has been classified in four sections namely SAD, HAPPY, FEAR and AGGRESSIVE. There are two major sections in this research work namely Training and Testing. The training has been done on the basis of wave files provided for every group. Features have been extracted for all groups and have been saved into the database. The testing section classifies the training set of data with the help of BACK PROPAGATION NEURAL NETWORK (BPN) classifier and SEQUENTIAL MINIMAL OPTIMIZATION (SMO) classifier. The results of the BACK PROPAGATION NEURAL NETWORK CLASSIFIER have been found superior in terms of classification accuracy.

# Table of Contents

# List of Figure

# List of Tables

# Abbreviations

| | |
|---|---|
| SVM | Support Vector Machine |
| SMO | Sequential Minimization Optimization |
| GUI | Graphical User Interface |
| RAM | Read Only Memory |
| Matlab | Matrix Laboratory |

# Chapter 1 Introduction

Emotion recognition with speech has now a day's getting attention of engineers in field of pattern recognition and speech signal processing. As computers is one of the important part now a days, so the requirement for communication between computer and humans. Through voice signals Automatic emotion recognition recognise the emotional state of the speaker. In humans life emotions plays a vital role. Humans show's their perspective or feelings or their mental state to others through emotions. Human have a natural gift to sense the emotion through speech on the other hand if we take machines to understand emotions through speech is bit difficult as it lacks the basic intelligence to monitor emotions. Understanding emotions in speech is a convoluted task as there is no obvious solution to what the right emotion for a given speech sample. Through speaker identification and speech recognition techniques machines can understand what is said but if we use emotion recognition system then machine can understand how it is said. As in human being emotions have a vital role for their different actions so there is a favourable requirement for a system interface through which we can have a good human machine communication and decision making. Emotion recognition through speech defines disclosure of emotional state of humans through feature extracted from his or her voice signal. Some other applications which require this interface such as Interactive movie, storytelling, and electronic machine pet, remote teach school & E-tutoring application. Other applications where it is been used are lie detection ,in the psychiatric diagnosis, intelligent toys, , In aircraft cockpits ,in call centre and in the car board system. Different intelligent systems have been developed on the basis of different emotions such as anger, happiness, sadness, surprise, neutral, disgust, fearful, stressed etc [25]. For emotion recognition from speech signal Prosodic features and spectral features can be used as these feature contains lot amount of emotional feature i.e. Pitch ,energy, Fundamental frequency, loudness, and speech intensity and glottal parameters are the prosodic features. Also some of the linguistic and phonetic features also used for detecting emotions through speech. There are several types of classifiers are used for emotion recognition such as Hidden Markov Model (HMM), k-nearest neighbours (KNN), Artificial Neural Network (ANN) , GMM super vector based SVM classifier , Gaussian Mixtures Model (GMM) and Support Vector Machine (SVM).

Speech is a voice sample which user can use for analysis of data. Speech is recorded either through software or by a voice recorder. Signals are mostly processed in digital forms so it can also be considers as the digital signal processing applied to speech signal. Aspects of speech processing include the acquisition, manipulation, storage, transfer and output of digital speech signals.

### 1.1 Data Mining

Data Mining is a term of searching and researches of data. The Mining of the data means searching out a piece of data from a bulk data block. As we are fetch knowledge and hence this is also called as the knowledge discovery from data (KDD) [26]. The basic work in the data mining can be categorized in two subsequent ways. First is called classification and the other is called making cluster. Even though all refer to different kind of same area but still there is difference in both the material. Clustering is the methodology of making groups of a set of data objects into multiple groups or clusters so that the objects within the cluster have high similarity, but are very dissimilar to objects in other cluster. Alternatively, it may serve as the pre processing step for other algorithms, such as characterization, attribute subset selection, and classification, which would then operate on the detected clusters and the selected attribute or feature [25]. The classification of the data is only possible if you have modified and identified the clusters. In the presented research work, our aim is to find out the maximum number of clusters in a specified region by applying the area searching algorithms.

Classification is always based on two things:

a) The area which you choose for the classification *i.e.* the cluster region.

b) The kind of dataset which you are going to apply on the selected region.

To increase the accuracy of the searching technique, any one would need to focus on two things:

a) Whether the data set has been cluster zed in proper manner or not.

b) If the clusters are defined, whether they fit into the appropriate classified area or not.

### 1.1.1 Terms related with data mining:

**a) Data**

Data are any facts, numbers, or text that can be processed by a computer. Today, organizations are accumulating vast and growing amounts of data in different formats and different databases. This includes:

- operational or transactional data such as, sales, cost, inventory, payroll, and accounting
- nonoperational data, such as industry sales, forecast data, and macro economic data
- meta data - data about the data itself, such as logical database design or data dictionary definitions

**b) Information**

The patterns, associations, or relationships among all this *data* can provide *information*. For example, analysis of retail point of sale transaction data can yield information on which products are selling and when.

**c) Knowledge**

Information can be converted into *knowledge* about historical patterns and future trends. For example, summary information on retail supermarket sales can be analyzed in light of promotional efforts to provide knowledge of consumer buying behaviour. Thus, a manufacturer or retailer could determine which items are most susceptible to promotional efforts.

## 1.2 Speech Processing

A speech is a sample of the voice of the user which he can use for the classification of the data. The speech can be recorded either by a voice recorder or by the software where the entire work has been done.

There are different properties of the speech signal. Before we move on to the speech processing, let's get to know what exactly is the speech mining. There are several terms which are required to be known .They are illustrated as following.

a) **Database**: A data base is the collection of data .In our proposed work we have used speech samples for the database. In the database we find properties of the speech signals and then we store them into the database. The question comes

that how we are going to store hundreds of files in the database. The procedure would be as follows. First of all we would fetch the properties of the voice samples. All those properties which are required would be computed and then it would be stored into an array. The array would move on as the files would move. We would fetch the features and would take the average by the end and then store them into the database for each category of the voice which we have taken *i.e*. **HAPPY, SAD, AGGRESSIVE AND FEAR [3].**

b) **Voice files**: The voice files are the files which would be processed for the feature extraction.

c) **Properties**: When we would process the voice files their properties would be fetched .For the feature extraction there are several algorithms which can be used. In our approach we have used HMM algorithm for the training purpose.

Emotional speech recognition aims at involuntarily identifying the emotional or physical condition of a human being via his or her voice. A speaker has dissimilar stages throughout speech that are recognized as emotional aspects of speech and are integrated in the so named paralinguistic aspects. The linguistic content cannot modify by emotional state; in communication of individual this is a significant factor, since feedback information is provided in numerous applications. Speech is possibly the generally proficient way to correspond with each other [18]. This too means that speech could be a helpful boundary to cooperate with machines. A few victorious examples based on it throughout the past years, while we have awareness about electromagnetism; includes the development of the megaphone, telephone. Even in the previous centuries people were researching on speech fusion. Von Kempelen developed an engine talented of 'speaking' words and phrases. At the present time, it has happened to achievable not only to expand examination and execute speech recognition systems, but also to have systems competent to real-time alteration of text into speech. Regrettably, in spite of the high-quality development made on that area, there are countless applications that are the speech recognition procedure facing dig now; speech is a very prejudiced experience that is added by the majority of them. It is very complex task to recognise the words from speech. In this system first we have to change the representation of speech signal. First task is to convert the speech file into phonemes or syllable then used the dictionary database to match speech input to phonemes or syllable. This dictionary database is used to train the system. If the

phonemes are matched then corresponding word is recognised. Step by step procedure is shown in the Figure 1.1. There are some factors that make difficult the speech recognition [19] and are discussed as:

1. Orator Sound- Identical word is pronounced another way by diverse people since gender, age, swiftness of speech, expressiveness of the speaker and vernacular variations.
2. Surrounding Noise- It is the disturbance added because of environment or surrounding noise as well as speakers voice too add to this facto

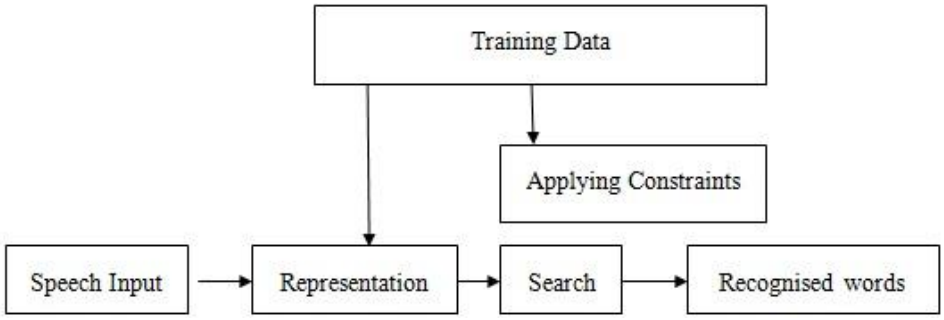


Figure 1.1 Speech Recognition System

## 1.3 Speech Features of Emotion Detection

Even though sound is a single conduit, there are two types of features that can be extracted and examined: paralinguistic features and linguistic features [3]. The paralinguistic features can be classified in prosodic, spectral, and voice quality features.

1. **Paralinguistic Features**

**b) Prosodic features:**

Prosody studies the rhythms of speech and aims on bigger segments of speech, like words, phonemes .There are strong association between the prosodic features. The pitch signal is cussed because of movement of vocal cord. It carries information about emotions because of stress of vocal cords. There is a term pitch period; it is the time between the openings of the vocal cords. Sound energy is also produced due to the pitch movement.

- **Spectral features:**

These features are produced because of air flow from the vocal cord, as in case of anger the air flow is very fast and in case of calm mood the air flow is very slow. So it all depends on the flow of air. Teager energy operator is used to measure the flow of air.

- **Voice quality features**

There is very strong connection between voice quality features and emotions. Voice quality depends on the emotions. There are many types of voice like calm, harsh, high pitch it all depends on the mood.

**2. Linguistic Features**

Most of the times convey of sentence give the emotion quality. Most frequently used are bigrams and unigrams.

## 1.4 Generic Schema of Speech Recognition System

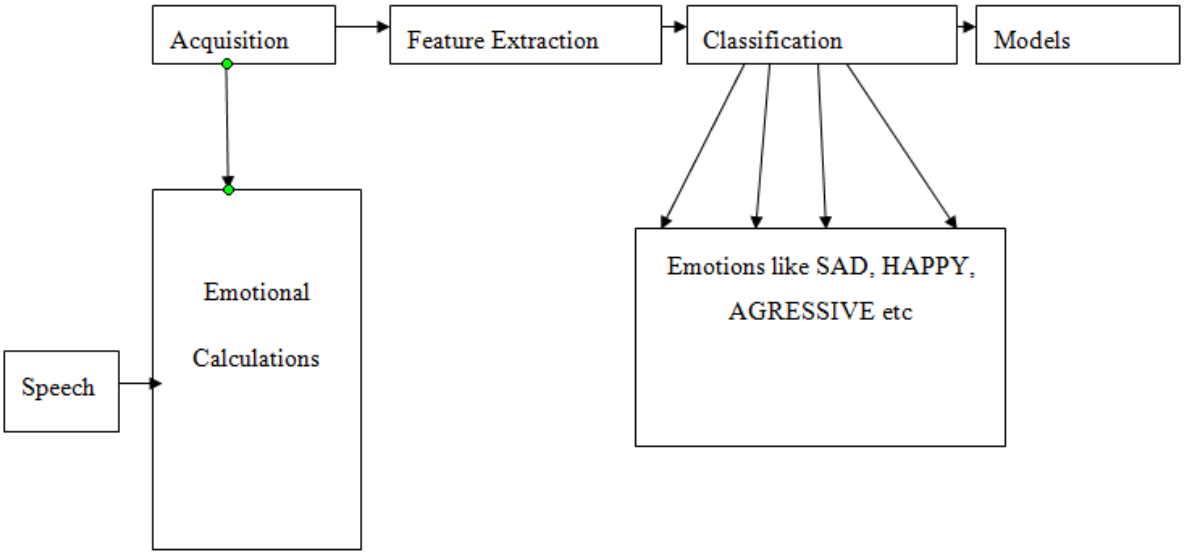


Figure 1.2 General Model of speech recognition system

Research of emotion recognition is mainly helpful for computers, robots, car industry *etc.* In the Human Computer Interaction it has been become the main role. This system comprised of two basic operations feature extraction and emotion classification. In the feature extraction features are extracted from speech using

different extraction techniques like hidden markov model or data mining techniques. There are different steps involved in emotion recognition system that are showed in Figure 1.2. Then these features are introduced to classifier for classifying the speech category. Classifier is used for the identification of emotions of different states from speech files. It will match the features of testing file with the features of trained files and computed the best match. The accuracy of the system depend upon the selection of features that are extracted and also depend upon the type of classifier is used.

## 1.5 Motivation

From the past years, researchers have showed interest in the speech recognition systems based on the emotions. Mainly the research has been done with the aim to bring closer both humans and computer with each other by recognising mood swings. In the recognising process we must know how to represent the emotions on the basis of some features like sad, happy etc. When the verbal content is well recognized on the speaker's emotion, a promising enhancement of such systems would come. So recognition is speech is a crucial step.

## 1.6 Research Questions

The objective of this proposed work is to develop a skeleton for real-time emotion recognition from speech that allows for easy incorporation into applications and thus improves the design of affective applications. For that reason the following questions are researched:

1. What is appropriate attitude for emotion recognition when taking into consideration real-time constraints?
2. To what extent is it probable to construct application reliant recognizers? How can these recognizers be trained?

## 1.7 Significance of the Study

1. As Speech Recognition Systems leads to enhanced communication between computers and human being. So developing of such systems will do the task frequently in the way of information generation, searching, transmission of data, communications and processing of information.

Speech is simple to generate than write, also it consumes time in writing to search data. So speech Recognitions Systems are very good for computer-human communication.

### 1.8 Tools Used

***Table** Error! No text of specified style in document.**-1 Tools Used***

| | |
|---|---|
| Computer | Intel Core Duo |
| RAM | 3 MB |
| Platform | Windows 7 Ultimate |
| Other hardware | Keyboard, mouse |
| Software | Matlab 7.0.4 |

The acronym of MATLAB is Matrix Laboratory. Due to ease of development of system as it has built in libraries made it enormous popular for implementation. Below are the main reasons why I opted this software:

- MATLAB may act as a calculator or as a training language
- MATLAB unite effectively computation and graphic plotting.
- MATLAB is somewhat easy to learn
- MATLAB is interpreted (not compiled), faults are simple to remove.
- MATLAB is optimized to be comparatively rapid when performing matrix operation
- MATLAB does have various object-oriented fundamentals

## 1.9 Thesis Outline

This report will mainly be separated into seven chapters.

**First Chapter: Introduction**

It is an introduction and brief overview of the thesis including the motivation, objectives, methodology, significance of the study and outline of thesis.

**Second Chapter: Literature Survey**

In this chapter we will discuss about the literature survey. In this chapter we will discuss about the literature survey of speech recognition and classifier which we use. Firstly we should survey about the basic of speech emotion like what expression a person can perform during speech like angry, sad, happy etc., which type of classifier we can apply and what already work done on emotion detection and BPA classifier.

We also survey work on SMO that is use for feature extraction. The basic classifier is BPA.

**Third Chapter: Problem Statement**

This Chapter will specify the problem statement.

**Fourth Chapter: Proposed Method**

This Chapter will describe the resources that we used to achieve our results with brief summary of each and every resource. This includes database, classifiers used.

**Fifth Chapter: Experiments and Results**

Moving on Fifth Chapter is the one in which experiments are shown with pictures, graphs, accuracy charts.. The snapshot shown in this chapter gives the step by step procedure to run the classifier and how we will get result. How speech data is represented in MATLAB platform and how it works. The results generated with classifier will identify best accuracy results.

**Sixth Chapter: Conclusion and Future Scope**

In this chapter we are going to describe that what we conclude about speech recognition using BPA (Back Propagation Algorithm) that belongs to neural networks. Classifier perform better it conclude by testing of speech data with classifier.

**Bibliography**

We will provide the references used in the dissertation and the other papers or any material or any web link that have been read out from which we get some idea or any function that is useful to us.

# Chapter 2 Literature Survey

In this chapter, the literature survey of speech recognition and classifier which we used are illustrated. Firstly, the basic of speech emotion like what expression a person 1can carry out during speech like angry, sad, happy etc., should be surveyed and then which type of classifier we can apply and what work has already been done on emotion detection and BPA classifier. We also survey work on SMO that is used for feature extraction. The basic classifier is BPA.

J. C. Platt [1] provides a new method to train the support Vector Machines efficiently. SVM is used to solve the QP problems. The memory used in SVM is linear, so it can handle large databases.

Md. Ali Hossain [2] presented the back propagation neural network for Bangla speech recognition. Features of speech signal are extracted using MFCC algorithm. These methods are implemented in Turbo C and C++.

Dimitrios Ververidis and Constantine Kotropoulos [3] gave a description on three goals that hit strike our mind when think about emotional speech recognition. Our first job is to collect data and update record where collection of emotional speech data is available. Record contains data about states of emotions, number of speakers, speech kind etc. In second step, goal features are symbolized that are used to extract features for emotional speech recognition and to measure how the emotion affects them. Mainly features that exist in market are like pitch, the vocal-tract cross-section areas, the intensity of the speech signal, and the speech rate. In the last goal, a suitable algorithm is used that will classify speech into emotional states. Here different classification techniques are examined in which timing information is been exploited. Basis for these classification techniques as Hidden Markov Models (HMM), ANN (artificial neural networks), k-nearest neighbours, SVM (support vector machines) are reviewed.

John C. Platt [4] proposed a new technique for training support vector machines: Sequential Minimal Optimization, or SMO. Requirement of a solution for the problem of training a support vector machine is to be used on a very large quadratic programming (QP) optimization problem. Firstly large QP problem is divided by SMO into a series of small possible QP problems. Solution for these small QP

problems is critically examined, in which a time-consuming numerical QP optimization as an inner loop is avoided. Requirement of memory for SMO is linear in size for the training set, in which SMO is allowed to handle large training sets. Because of the avoidance of matrix computation, for various test problems SMO scales somewhere lie between linear and quadratic in the training set size. SVM evaluation represents SMO's computation for time; hence as compared to linear SVMs, SMO are faster and bare data sets. In real-world sparse data sets, it is possible that, as compared chunking algorithm, SMO can be 1000 times faster.

Wouter Gevaert, Georgi Tsenov, Valeri Mladenov [5] described in this paper an investigation that is done on performance for classification of speech recognition. There are two standard neural networks structures that are used for performance evaluation as classifiers. Feed-forward Neural Network (FFNN) type is included for standard utilization with back propagation algorithm and Basis Functions Neural Networks is Redial used.

Mirza Cilimkovic [6] presented method for classification and clustering in data mining. Neural Networks (NN) as a classifier is used. The proposed system is capable of mimic brain activities and is able to learn. Learning of NN is made from examples. If more examples are provided to NN, then it has capability to knob those examples and classifies that data with representation of patterns in data. There are three layers in basic NN that are as input, output and hidden layer. There are numerous nodes existing in each layer and nodes of input layer need to be attached with nodes from hidden layer. Then to obtain output there should be connections between nodes of hidden layer to nodes from output layer. Weights between these nodes will show the connections.

Here algorithm of NN is described, Back Propagation (BP) Algorithm. The main purpose is to show logic behind this algorithm. The basic idea behind BP algorithm is quite simple; output of NN is evaluated against desired output. If we do not get results as expected then there is requirement of modification in connection weights and process is repeated again and again till the error value is small. Some parameters are changed to improve results for new implementation.

Peng Peng, Qian-Li Ma, Lei-Ming Hong [7] presented a novel technique for solving method of Support Vector Machine algorithm that is SMO that is a parallel algorithm. According to this algorithm, primitive training sets are dispensed by master CPU to

slave CPUs. Slave CPU run serial SMO on the relevant training sets. As buffer and shrink methods are also selected, increment in speed of the parallel training algorithm is done, which is represented in the results of parallel SMO based on the dataset of MNIST. The results of this work proved that by using SMO performance of solving large scale SVM is good.

Rong-En Fan, Pai-Hsuen Chen, Chih-Jen Lin [8] presented a new algorithm for selection of working set in SMO type decomposition method. It discussed that in training support vector machines (SVMs), selection of working set in decomposition process is important. Fast convergence is achieved by using information of second order. Theoretical properties such as linear convergence are established. It is proved in results that proposed method provided better results in contrast to existing selection methods using first order information.

Xigao Shao, KunWu, and Bifeng Liao [9] proposed an algorithm for selection of working set in SMO-type decomposition. It showed that in training part, least square support vector machines (LS-SVMs) the selection of working set in decomposition process is important. In the proposed method a single direction is selected to achieve the convergence of the optimality condition Experimental results represented that speed of training is faster than others but classification accuracy is not better than existing ones, it's almost same with others.

Francis R. Bach, Gert R. G. Lanckriet, Michael I. Jordan [10] showed combination of kernel matrices for SVM and that combination reduces a convex optimization problem known as quadratic ally strained quadratic program n (QCQP). While classical kernel-based classifiers are based on a single kernel, mostly base of classifier are developed by a combination of multiple kernels. Unfortunately, the problem for small number of kernels can be solved with current convex optimization toolboxes and a small number of data points; besides, due to cost function the sequential minimal optimization (SMO) techniques which are essential in large-scale implementations of the SVM cannot be applied. A novel dual formulation of the QCQP as a second-order is proposed for cone programming problem, and shows how to exploit the technique of Moreau-Yosida regularization to succumb a formulation to which SMO techniques can be applied. SMO- based algorithm is much better and efficient than general purpose methods for interior points available in current optimization toolboxes.

S. K. Shevade, S. S. Keerthi, C. Bhattacharyya, and K. R. K. Murthy [11] In this paper Smola and Schölkopf's sequential minimal optimization (SMO) algorithm have some source of inefficiency which is pointed out for regression of support vector machine (SVM) that occurs by the use of a single threshold value. The KKT conditions for the dual problem is used, SMO modification is done on the basis of two threshold parameters that are employed for regression. This proposed algorithm with the modification in SMO performs faster than the original SMO.

Ibrahim Patel, Dr. Y. Srinivas Rao [12] described that recognition of speech signal can be done using frequency spectral information with Mel frequency. HMM based recognition is used to increase the results of selected approach for recognition. Speech signal is observed using Mel frequency approach in a given resolution which results in resolution feature overlapping thus resulting in recognition limit. Mapping approach for HMM is resolution disintegration with separating frequency based on speech recognition system. Results from simulation represent the improvement in the quality metrics of speech recognition with respect to computational time, also progress the accuracy for a speech recognition system.

Wei HAN, Cheong-Fat CHAN, Chiu-Sing CHOY and Kong-Pang PUN [13] In this work MFCC is extracted by using proposed algorithm for speech recognition. Computation power is reduced by 53% in new approach as compared to any other algorithm. A recognition accuracy of 92.93% is also developed that will prove results of simulation. 1.5% reduction in recognition accuracy exists as compared to the other MFCC extraction algorithm, which has an accuracy of 94.43%. However, in implementation of new algorithm requirement of logic gates is half as compared to implementation of existing algorithm, due to which the proposed algorithm is very efficient for hardware implementation.

Tobias Glasmachers and Christian Igel [14] presented a modified step for working set selection and increased results in terms of speed and accuracy. The solution of support vector machines (SVMs) has two potential advantages over the iterative learning algorithms. Firstly, online and active learning is allowed there. Secondly, the exact SVM solution may be too time consuming in case of large dataset and preference will be given to an efficient approximation. The exact SVM solution using sequential minimal optimization (SMO) is a powerful approach LASVM.

N.Pushpa, R.Revathi, C.Ramya, S.Shahul Hameed [15] presented word recognition of Tamil using multilayer feed forward network. To remove the noise amd enhance signals, four type of filters are applied on the input as pre-processing step namely, Reemphasis, Median, Average abd Butterworth band stop filter. MSE and PSNR values are used to measure the performance of these filters. Further, Feed-Forward Neural Network is used to classify the features that are extracted and recognizes Tamil spoken words. An algorithm is proposed for training deep neural networks (DNNs) as data-driven feature front-ends for large vocabulary continuous speech recognition (LVCSR) in low resource settings.

Hitesh Gupta, Deepinder Singh Wadhwa [16] described an improved algorithm Genetic Algorithm (GA) for improving the performance of speech signals recorded under noisy conditions. There are three categories for advanced speech enhancement algorithms namely; filtering/estimation based noise reduction, beam forming and active noise cancellation (ANC) technique. The features selected in this work are responsible for discriminating different words and thus apply evolutionary computation of genetic algorithm.

Pan et al. in the paper [20] defines that Emotion Recognition (SER) is presently the hottest topic in the field of Human Machine Interface. They basically monitored three emotional states i.e. sad, happy and neutral. The explored features include: Mel-energy spectrum dynamic coefficients (MEDC), energy, Mel-frequency spectrum coefficients (MFCC), pitch and. linear predictive spectrum coding (LPCC). They use two databases for training using SVM classifier. Results for multiple combinations of features over the different database are compared and explained. The final results shows that the feature combination of (Energy+MFCC+MEDC) have the best results in terms of accuracy rate of 91.3% on Chinese emotional database and 95.1% Berlin emotional database.

Utane et al. in the paper [21] defines that role of speech in human machine interface is increasing. For expressing feeling and emotions can easily express through speech thus making it an effective and attractive medium. Studies been carried out for differentiation of speakers emotional state i.e. sad , anger , surprise , happy and neutral using Hidden Markov model classifiers and Gaussian mixture model to identify the emotion of the speaker though speech. Different feature are extracted i.e.

spectral features such as Mel frequency cepstrum coefficient and prosodic features like pitch, energy. On the bases of this features emotional classification and performance of classification using Gaussian mixture model and Hidden Markov Model is discussed.

Sapra et al. in the paper [22] defines that emotions is basically a physiological and mental state in relation with different kind of feelings behaviour and thoughts. Emotions are often related with personality, deposition, mood and temperament. Basically method for detection of human emotions are discussed on the basis of acoustic features i.e. pitch, energy etc. MFCC approach is used and then nearest neighbour algorithm is used for classification. For Male and Female emotions are categorize differently on the basis that male and female different range so MFCC varies accordingly for the two.

Oudeyer in the paper [23] presents algorithms in which by modulating the intonation of its voice robot can convey their emotions. They are simple and efficient in providing a life-like speech by applying a concate native speech synthesis. These techniques have controls on the quantity of emotions and the age of a synthetic voice that are expressed. Also provide the data mining experiment regarding automatic recognition of basic emotions in everyday occurrence. We compare a set of Support Vector Machines or decision trees, machine learning algorithms, ranging from neural networks, together with 200 features, using a huge database containing many examples. Finally, explains that how this study could be applied to real world situations where the examples are few. This also describes an example of a game with a personal robot which makes teaching of examples of emotional utterances in a natural rather than unconstrained manner.

Vogt et al. in the paper [24] defines way suppress the challenges faced by automatic emotion recognition from speech in human-machine interfaces, which follows the pulling out of relevant features, segmentation of audio signals to get an appropriate units for emotions, to train database with emotional speech and classification of emotions. Have basically worked on offline evaluation of vocal emotions and as far as online is concerned processing have barely been addresses. However online processing is important for understanding human-machine interfaces which analyze and reacts to the user's emotions. With the help of sample application we indicate that

how the problems coming from online processing can be solved. Focus of this work is to make user to use the feasibility of human-machine interfaces.

# Chapter 3 **Problem Statement**

In this chapter, we will discuss the background which laid my interest in choosing this field to research. The main and crucial step of Speech Recognition Systems is to recognise the speech effectively. Here word "effectively" means to recognise the words accurately on the basis of features extracted whether it is SAD, HAPPY etc. Recognizing the emotions in speech is a complex process. Complexity of this process is about that the environment or source that generates speech is difficult to identify, source may be male or female. The problem of this research work is to identify the emotion of the speech based on four categories namely SAD, HAPPY, AGGRESSIVE and FEAR using two different algorithms namely SMO and BPA algorithm and also a comparison evaluation between SMO and BPA.

## 3.1Objectives

The aspire of this research is to build up and calculate SMO and BPNN model for recognizing emotion in speech expressed features of speech files. Meanwhile, the objectives of this research are as the following:

1) To categorize the anger, sad, happy, aggressive moods of emotional state in speech articulated by male/female using Training and Testing method.
2) To discover the threshold that performs superior in Back-propagation neural network in recognizing the anger, sad, happy, aggressive status of emotion.
3) To compare the accuracy of BPA and SMO algorithm based on feature extraction.

## 3.2 Methodology

There are number of milestones that need to be achieved in order to reach the goal, so following are the sequential steps that will led to attain the signal.

1. Ready database.
2. Train Systems using the feature extraction method. The features extracted are Maximum Frequency, Average Frequency, Minimum Frequency, Roll off, Pitch, and Loudness etc.
3. Repeat step 2 for all categories.
4. Save data to db.

5. Upload a file to test.

6. Extract same features of the uploaded file.

7. Set Target values for the evaluation.

8. Apply SMO and BPNN algorithm for classification.

9. Compare the accuracy of SMO and BPNN

# Chapter 4 Proposed Works

Speech recognition process is basically done by the Speech Recognition System. In the speech recognition process, speech input signal is processed into recognition of speech as a text form. Speech Recognition System helps the technology to bring computers and humans more closer. There is basic terminology that one must know in order to implement or develop a Speech Recognition System [17].

- Utterances- User input speech is called utterances, in simple words when user speaks something it is called utterances.
- Pronunciations- Single word has multiple meanings and multiple recognitions. It all depends on pronunciation. A single word is uttered in different means in accordance to country, age etc.
- Accuracy- It is the performance measurement tool. It is measured by number of means but in this case, if speaker utters "NO", then Speech Recognition System must recognise it as word "NO". If it is done precisely then accuracy of system is efficiently very good or else.

## 4.1 Basic Theory

The extended version of SVM is call Sequential Minimization optimization. SMO break the large QP problem into smaller QP problems and then find the solutions of smaller QP problem analytically. If the solutions are found analytically then it will consume very less time to solve the problem.

### 4.1.1 Sequential Minimization Optimization Algorithm (SMO)

Speech Recognition System consists of training part and testing part.[1]. So, basically SMO is used for training of SVM part. Primarily SMO is designed to speed up the rate of algorithm based on optimization techniques. The SMO consists of two B constraints $B_1$ and $B_2$ and give optimize values for both Bs. This task is reviewed until Bs full exposure. Process of SMO is described below in Figure 4.1:

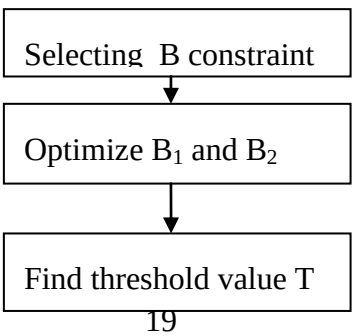


Figure 4.1 Process of SMO

### 4.1.2 Selecting B Constrained

SMO has optimization techniques that led to enhance the speed of algorithm. So there are two constraints $B_1$ and $B_2$, choose that constraint that will led to achieve the result as far as possible. To maximize the optimization function, is vital step to enhance the speed of computer. As there are n (n-1) choices of possible solutions so there is minute up gradation. SMO tries to optimize the Bs but if no $B_1$ and $B_2$ are changed then terminates the SMO.

### 4.1.3 Optimize $B_1$ and $B_2$ Constraints

$B_1$ and $B_2$ optimization is done using Lagrange operators. Then restrict these constraints. Restrictions on the constraints are shown as:

$$\text{If } x^{(1)} /= x^{(2),} \qquad L = \max(0, B_1 - B_2), \qquad H = \min(C, C + B_1 - B_2)$$

$$\text{If } x^{(1)} = x^{(2),} \qquad L = \max(0, B_1 + B_2 - C), \qquad H = \min(C, B_1 + B_2)$$

Now, find the $B_2$, to maximize the optimization of objective function [5]. If this rate finishes up lying exterior the boundaries L and H, is range for $B_2$ within which it exists, optimal $B_2$ value is given as:

$$B_2 := B_2 - x^{(2)} (E_1 - E_2) / \acute{\eta}$$

### 4.1.4 Compute the Threshold T

After optimizing $B_1$ and $B_{2,}$ we select thresholding T. If after optimization $B_1$ is not at the bounds (i.e., 0< B, < C) then following threshold $T_1$ is valid.

$$T_1 = T - E_1 - x^{(1)} (B_1 - B^{(OLD)} (y^{(1)} - x^{(2)} (B_2 - B_1^{(OLD)} (Y^{(1)}, Y(^2))$$

These thresholds are valid in which one is $0 < B_1 < C$ and second one is $0 < B_2 < C$, and they will be identical.

## 4.2 Back Propagation Algorithm

Back propagation neural network (BNN) is the for the most accepted and commonly used neural network[2-7] .Usually back-propagation neural network consists of feed forward multi-layer network in an input layer, an output layer, and at least one hidden layer. All layers are completely associated to the successive layer. There is no

interconnection or feedback between PEs in the same level. Each PE has a predisposition input with non-zero weight. The bias input is analogous to the. In a neural network consisting of P processing elements, the input/output function is defined as:

$$k = L\,(mW)$$

Where m = ***{m,)*** is the input vector to the network,
k = {Kkj} is the output vector from the network,
W is the weight matrix. It is afterward defined as

$$W = (w_i^T, wj^T, ..., Wn^T)^T$$

Where the vectors $w_i, wj, ..., Wn$ are the individual PE weight vectors, which are given as:

$$W_t = \begin{matrix} w_i \\ wj \\ wn \end{matrix}$$

The main function of back propagation neural network is that it moves from input to output layer. In the back propagation neural network error is found at output layer, so to remove error at output layer, moves backward takes place , hence name is back propagation neural network. Steps are shown in Figure 4.2.

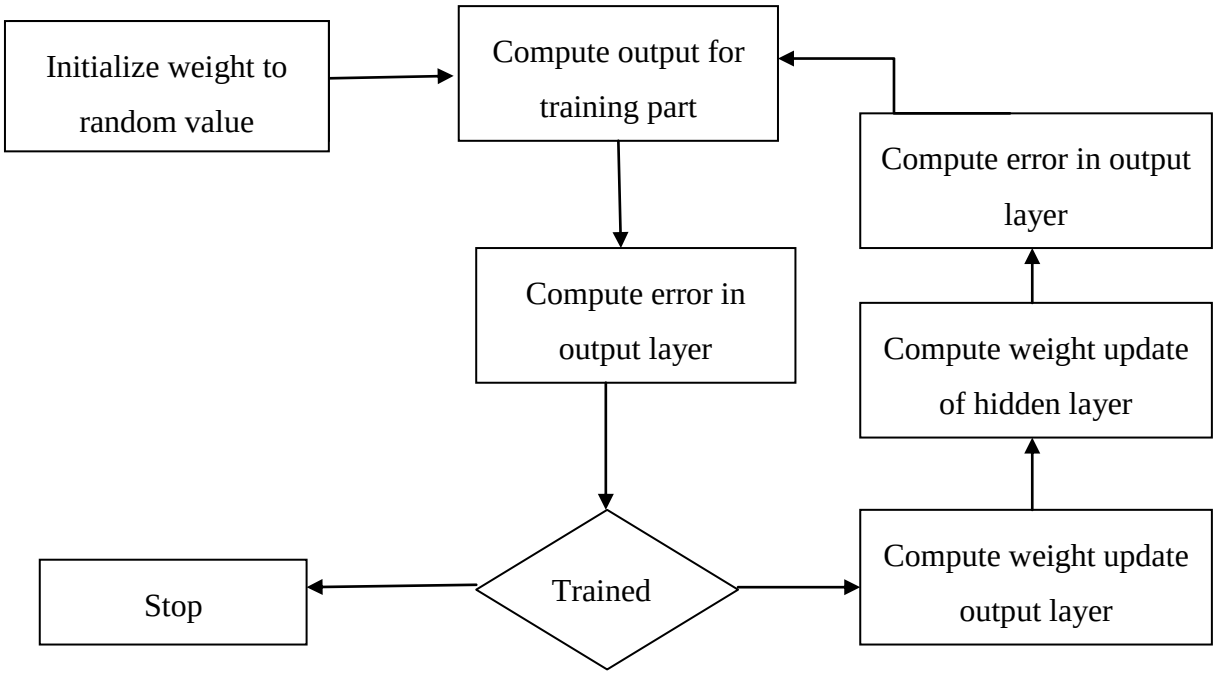


Figure 4.2 Back Propagation Training Algorithm

## 4.3 **System Design**

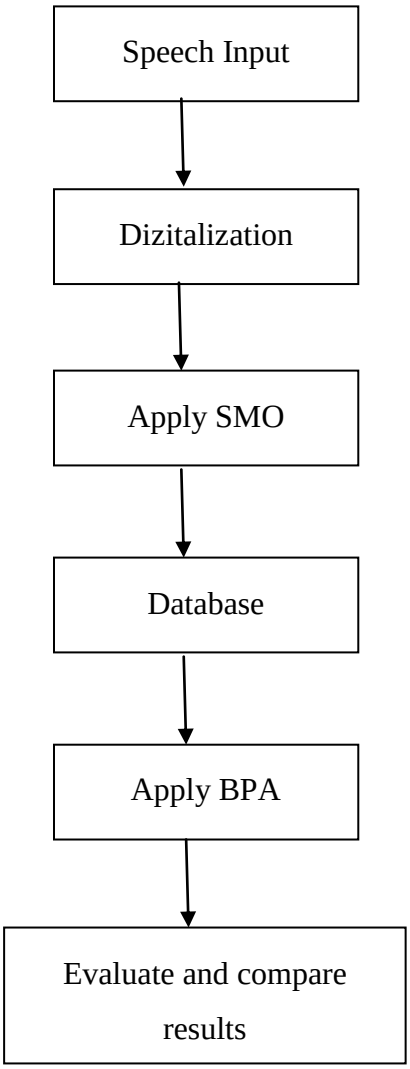


Figure 4.3 Flowchart of Proposed Method

There are two sections in our research work. The sections are explained as follows.

A. **Training:** The training section ensures that the database gets trained properly so that at the time of testing it produces wide-ranging results. The characteristics of the training are as follows.

- **Maximum Frequency:** It is the rate which we get at the tip on a frequency chart. When so ever we place a voice sample over the time and frequency model, the maximum peak is known as the maximum frequency of the voice illustration.

- **Minimum Frequency**: It is the rate which we get at the tip on a frequency chart. When so ever we place a voice sample over the time and frequency model, the minimum peak is known as minimum frequency of the voice illustration.
- **Average Frequency**: The average frequency can be considered using two techniques. The first technique is to append all the frequency samples and then separate the entire sum with the total number of frequency. The second method is an incredibly principled method in which we can insert the minimum frequency and the maximum frequency and then we can divide them by two.

**Average Frequency = (Minimum frequency + Maximum Frequency)/2**

- **Spectral Roll off**: The spectral roll off is difference between the maximum frequency differences with the adjacent frequency.
- **Noise Level:** The noise level is the additional amount of bits which has been supplementary into the voice sample.
    1. **Uniform Noise**: Uniform noise is the noise which is concurrently equivalent all over the voice sample.
    2. **Non Uniform Noise**: Non-uniform noise is the noise that does not remain constant all over the sample.
- **Pitch**: It is the standard value of the entire voice sample.
- **Spectral Frequency**: The spectral frequency is the frequency of the voice pitch subsequently to the highest voice sample.

**B. Testing:** In the testing phase, the trained network was computer-generated with unknown speech blueprint. It was observed that the trained network performs very well and more than ten words can be recognized by using the developed system [14].

## 4.4 Speech Input

A speech is an illustration of the voice of the user which is used for the categorization of the data [17]. The speech can be verified by a voice recorder otherwise by the software in which work has going to take place.

## 4.5 SMO Pseudo Code

Step -1 Input parameters and training data

Step -2 To get the output Lagrange's parameter to find solution and threshold value [5]. In each iteration, it selects two coefficients (A, B), even as the rest of the coefficients keep their existing values. Examine that this is the smallest number of coefficients that may be selected, since the constraint $\sum Ni=1\ AB = 0$.

## 4.6 Database

A data base is the gathering of data [15] .In our proposed work we have used speech samples for the database. In the database we find properties of the speech signals and then we store them into the database. The question comes that how we are going to store hundreds of files in the database. The procedure would be as follows. First of all we would fetch the properties of the voice samples. All those properties which are required would be computed and then it would be stored into an array. The array would move on as the files would move. We would fetch the features and would take the average by the end and then store them into the database for each category of the voice which we have taken i.e. **HAPPY, SAD, AGGRESSIVE AND FEAR.**

## 4.7 BPA Algorithm

BPA algorithm is one of the best algorithms of Neural Network [10]. BP algorithm can be broken into four main steps:

i) Feed-forward calculation

ii) Back propagation useful on output layer

iii) Back propagation applied on hidden layer

iv) Updating of weight performed

# Chapter 5 Experiments and Results

## 5.1 Database

The two algorithms have been trained on 200 speech signals of four categories like sad, happy, fear, Aggressive. We have saved the speech files in .wave format.

## 5.2 Experiments

User Section Figure 5.1 illustrates the categories which we are going to use. The first category is sad, followed by fear, happy and aggressive.

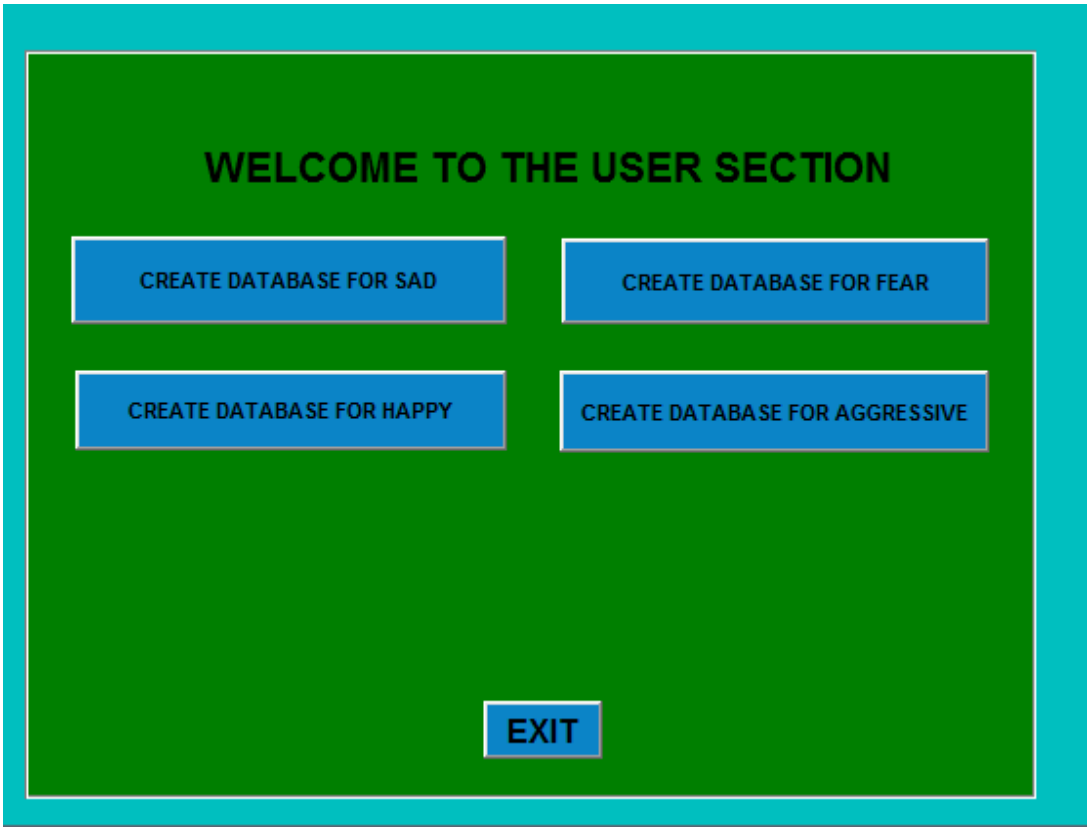


Figure 5.1 User Section

The main window is shown in Figure 5.2 where we impart training to the system. As it is clearly visible that that the features are at the bottom end of the window and the selection of the files are at the up of the window. We can proceed with this window by taking sample and it would provide us the features.

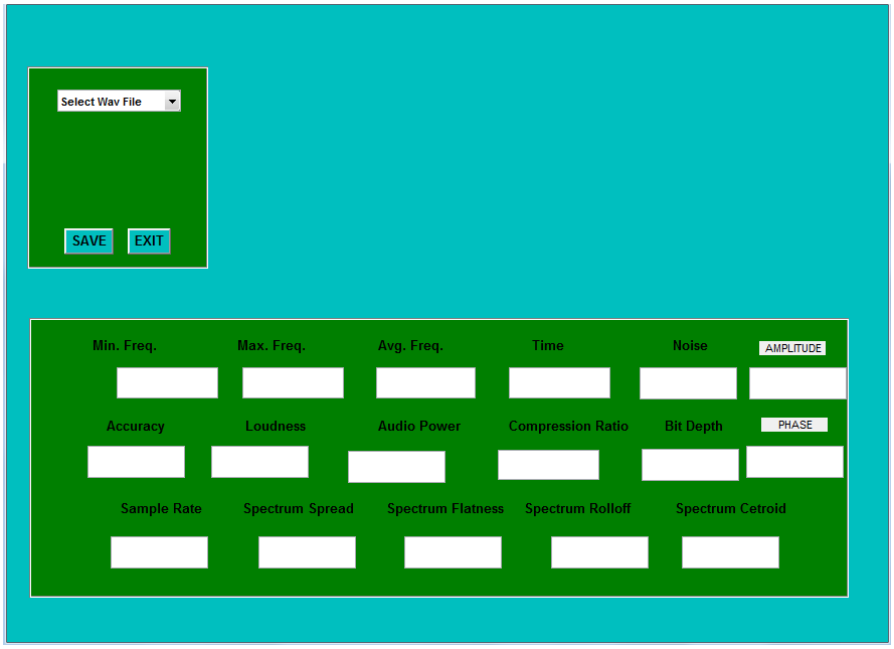


Figure 5.2 Main Window

Figure 5.3 shows the windows that get opened for the selection process when we have to choose a file and it shows four data samples as we have mentioned it in our system.

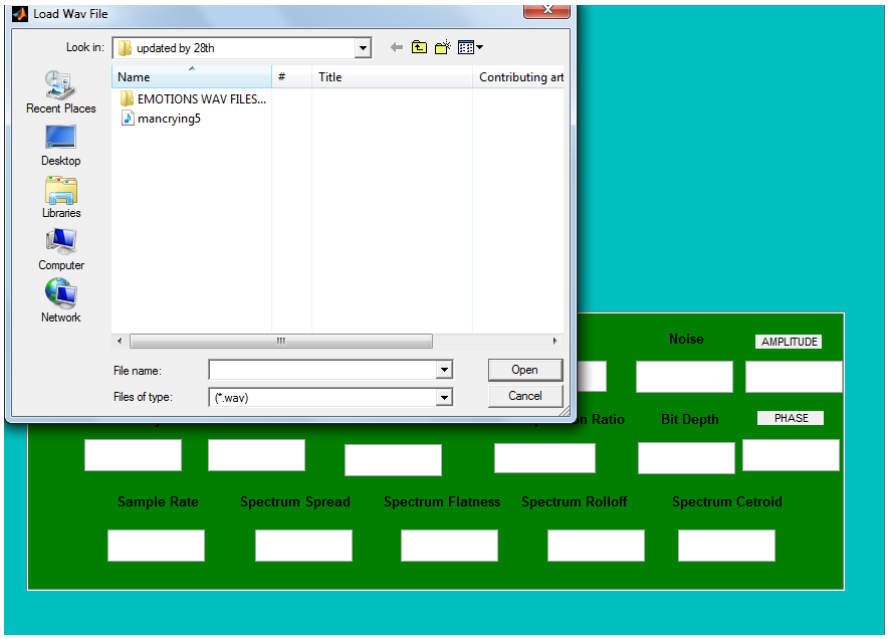


Figure 5.3 Data Samples

Once we are done with the selection procedure of a wave file it would produce results of the feature extraction as shown in Figure 5.4. As we can see that the systems have fetched a lot of properties values for the database and there is also a button of save form where we can save the category database.

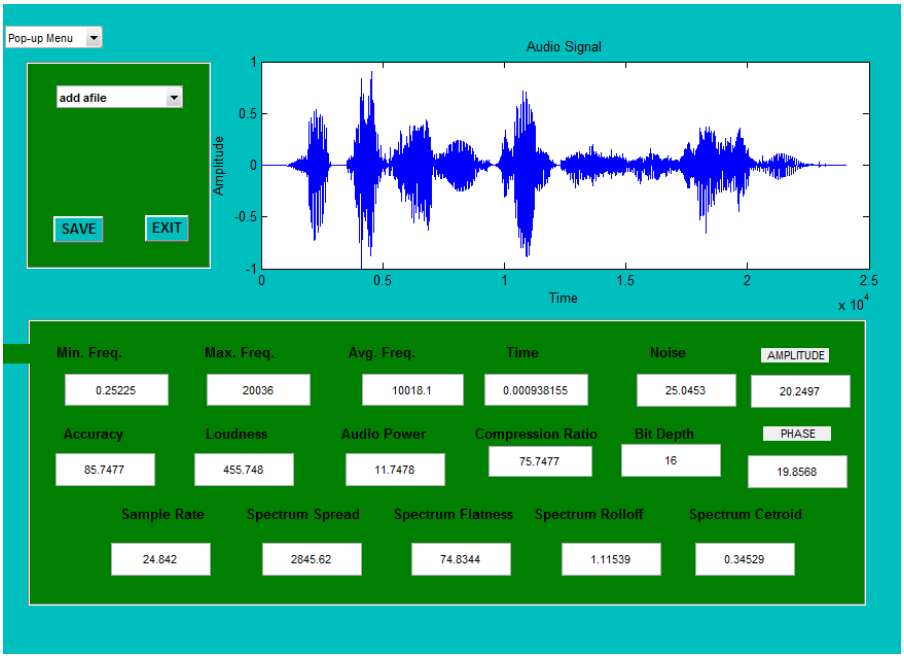


Figure 5.4 Feature Extraction

Now a database of 50 files has been saved in form of mat files as shown in Figure 5.5 which can be seen in the current folder of the MATLAB version 2010.

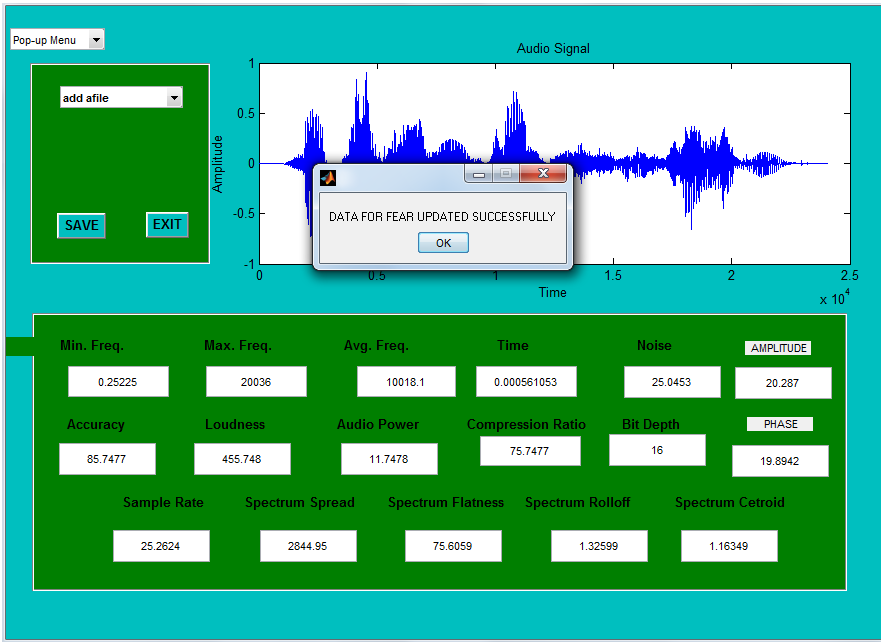


Figure 5.5 Saving of Speech signal into database

Here we can see that our sad sample data has been saved into the database as shown in Figure 5.6.

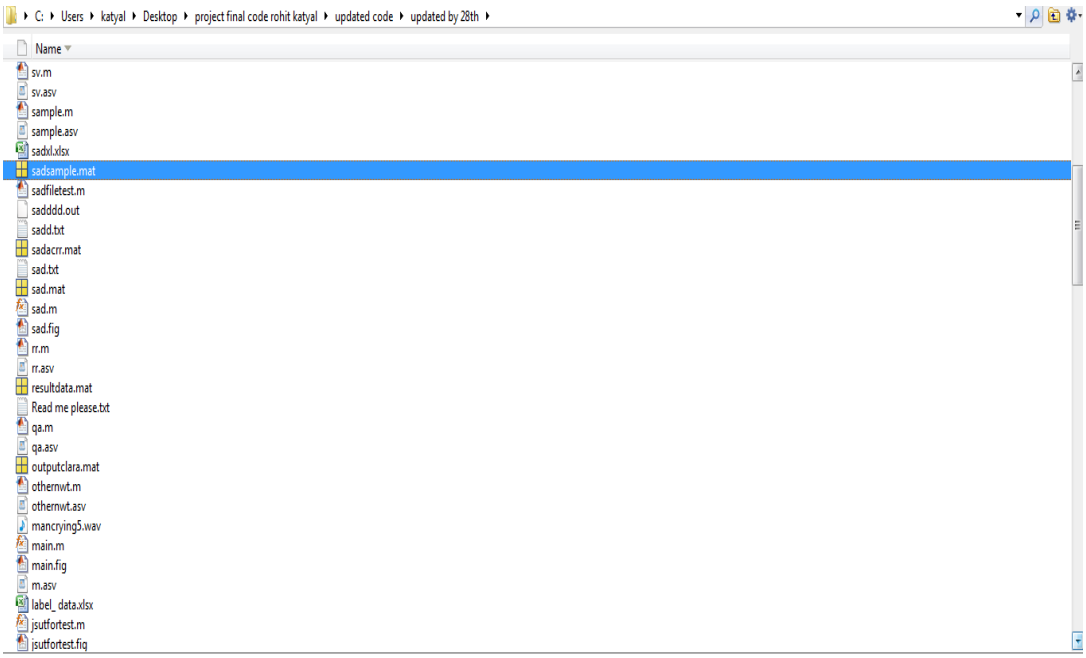


Figure 5.6 saving of speech file

TESTING SCREEN SHOTS

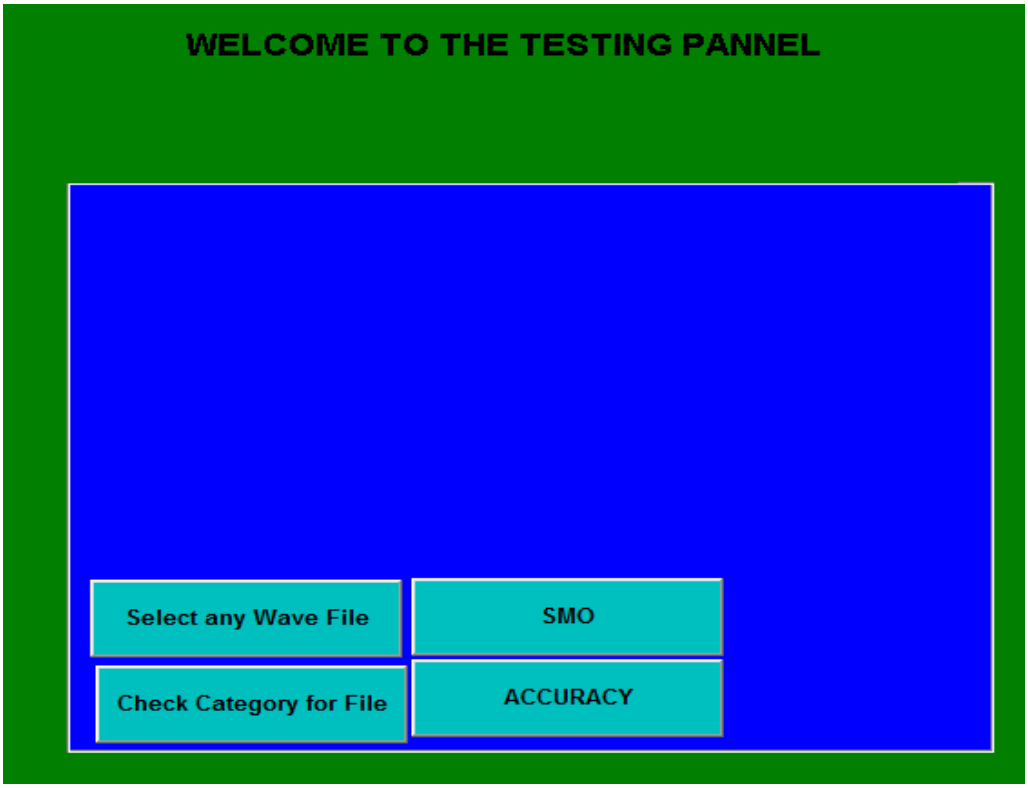


Figure5.7 testing of speech file

The above Figure 5.7 represents the main working window for the testing of the speech files for all the categories mentioned earlier. There are four major sections in this panel. The first section is the uploading of the wave file whose features has to be extracted and then has to be passed to both the classifiers to check the accuracy of the processing.

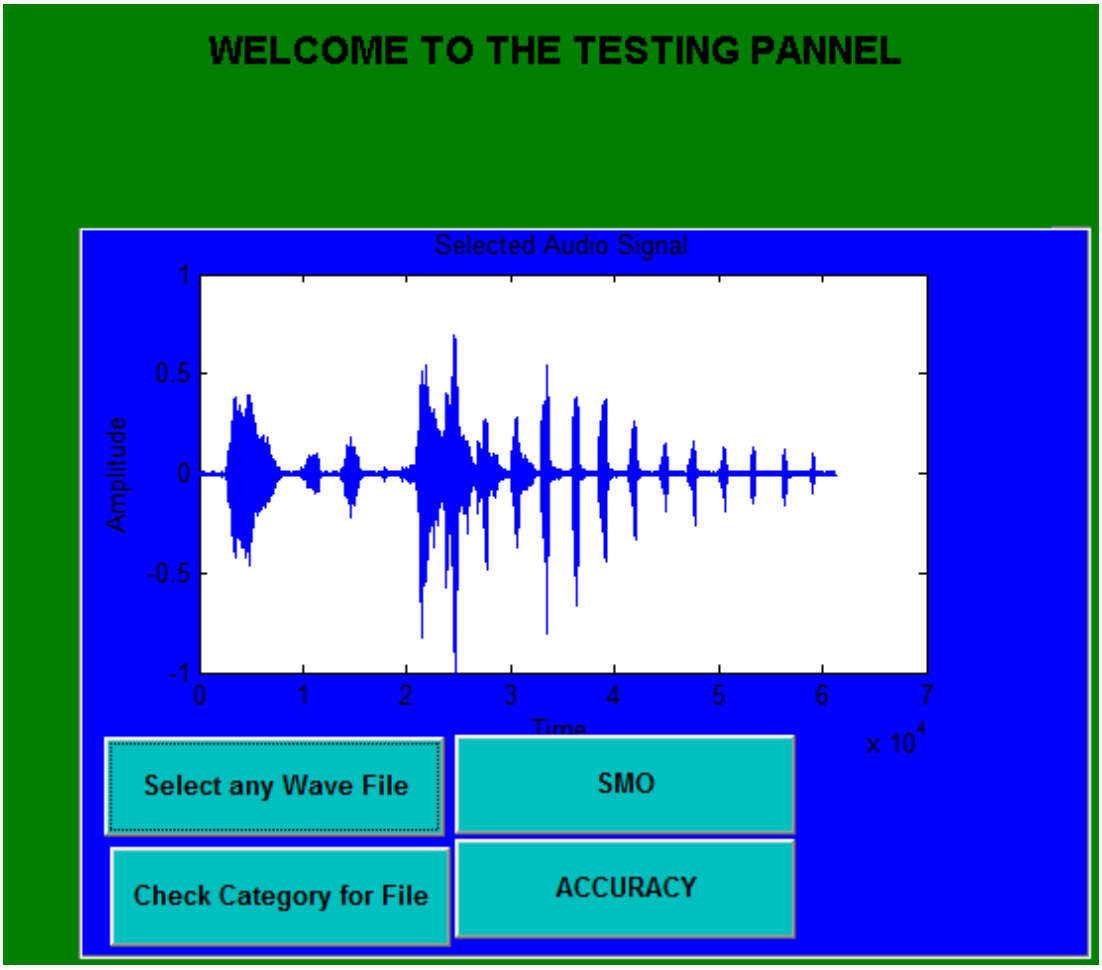


Figure 5.8 Uploading of speech file

The above Figure 5.8 represents the completion of the uploading process. The plotted graph shows the time vs. amplitude ratio of the uploaded speech signal.

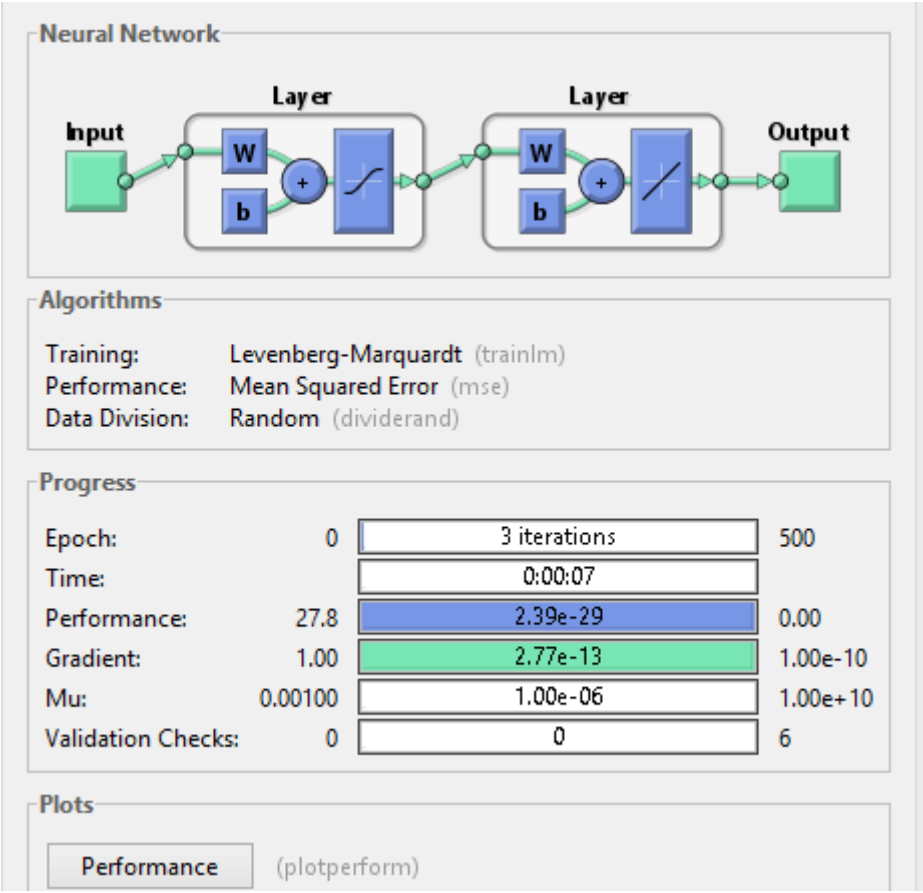


Figure 5.9 Working of Neural Network

The above Figure 5.9 represents the back propagation neural network processing in the network. The figure has the following parameters

a) **Epochs:** Epochs is the number of iterations which the neural network runs. The above diagram shows that there are three iterations out of 500 which have successfully predicted the category of the speech file. If the system runs 3 iterations, it does not predict that the best result would come out at third iteration though the network takes an average of the three iterations and depicts the best results out of time.

b) **Time:** It determines the total time consumed in the prediction.

c) **Performance Measure:** It represents the elapsed performance hit in the network architecture which can be plotted through the SOM plots of the neural network.

d) **Validation Checks:** It depicts the number of validations which the network can apply to the testing procedure.

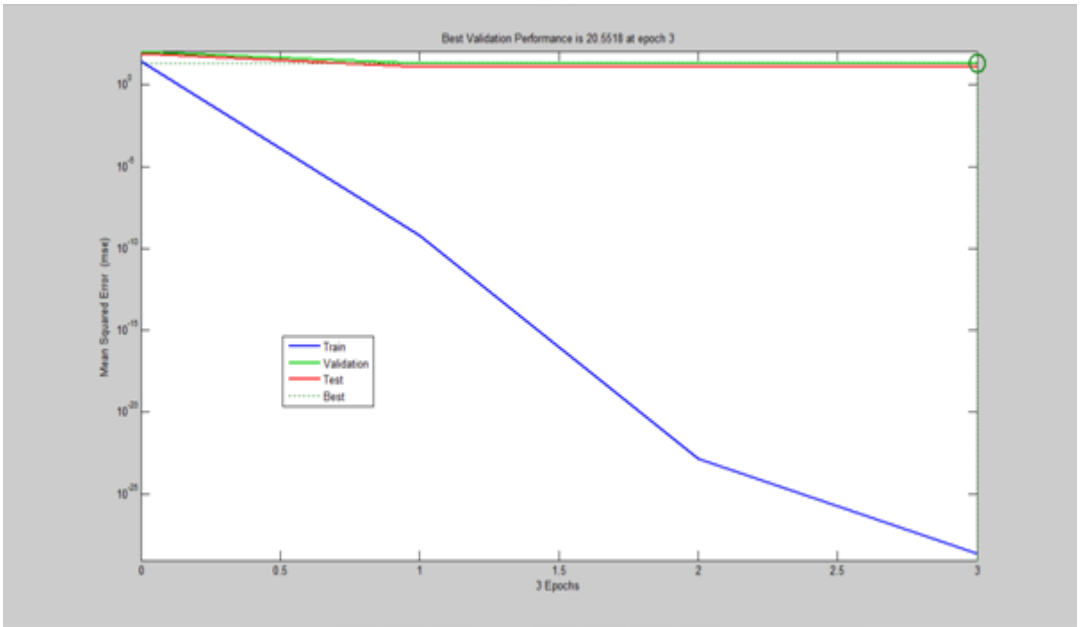


Figure 5.10 Performance of Neural Network

The above Figure 5.10 is the performance plot of the neural network. The graph shows the MSE of the test procedure with each and every iteration.

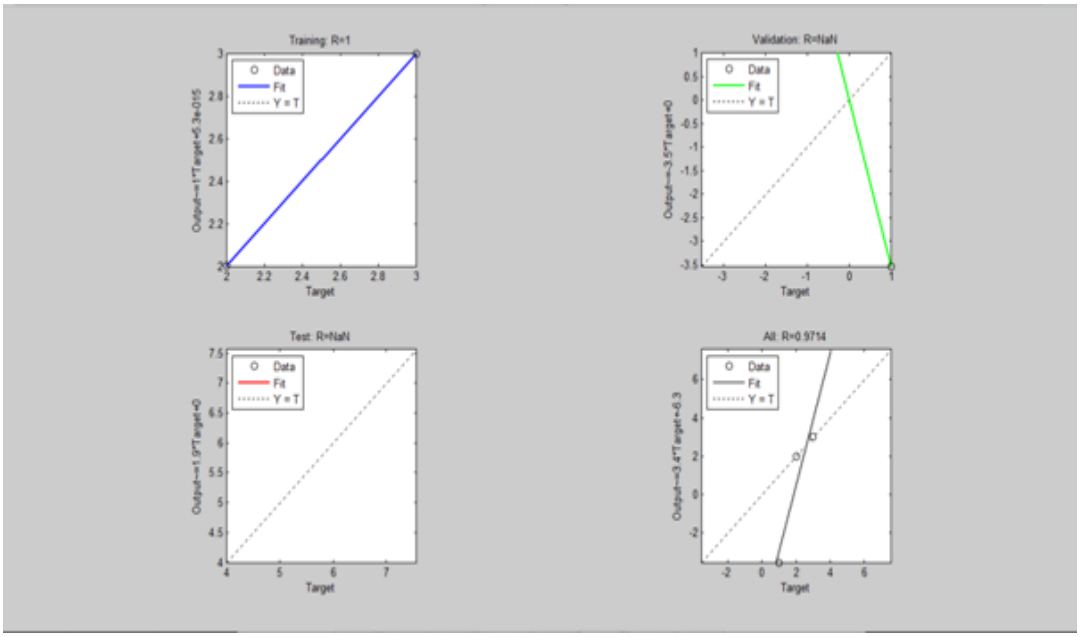


Figure 5.11 Regression of Neural Network

The above Figure 5.11 depicts the regression plot of the neural network with the target set provided to it. This graph details the system about the prediction of the category of the speech file through neural network.

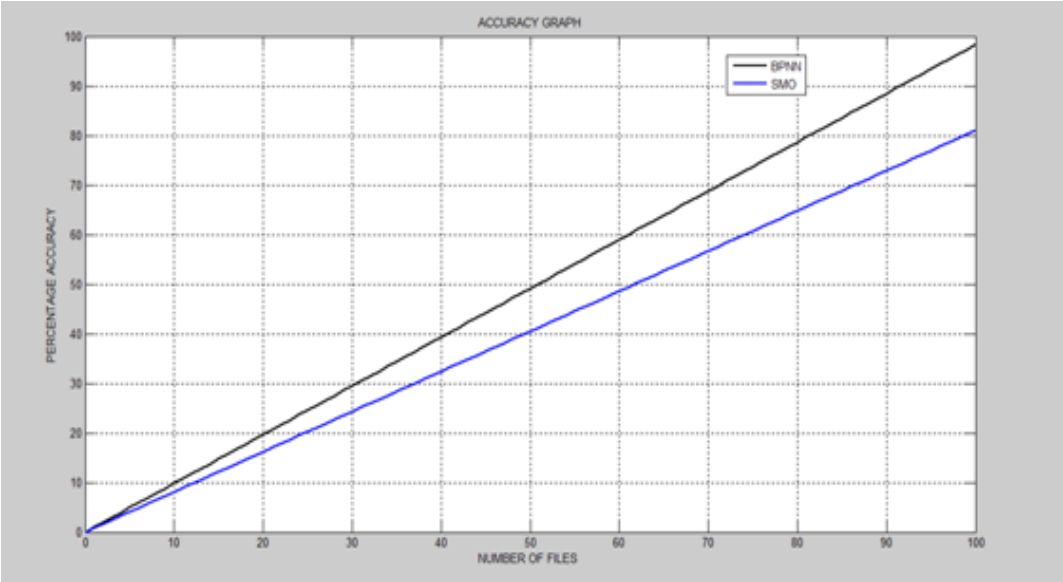


Figure 5.12 Accuracy of BPNN and SMO

The above Figure 5.12 depicts the accuracy between the BPNN and SMO algorithm. The graph explains itself that the accuracy of BPNN is more than that of SMO for the same classification and it differs from 10 to 15 percent.

## 5.3 Results

The experiment as discussed above has been done on the basis of four speech categories namely sad, happy, fear and aggressive in MATLAB 2010. The processing has been captured in screen as follows.

Accuracy percentage of mood swings like SAD, HAPPY, FEAR, AGGRESSIVE using BPNN and using SMO is shown in Figure 5.13 and Figure 5.14 respectively.

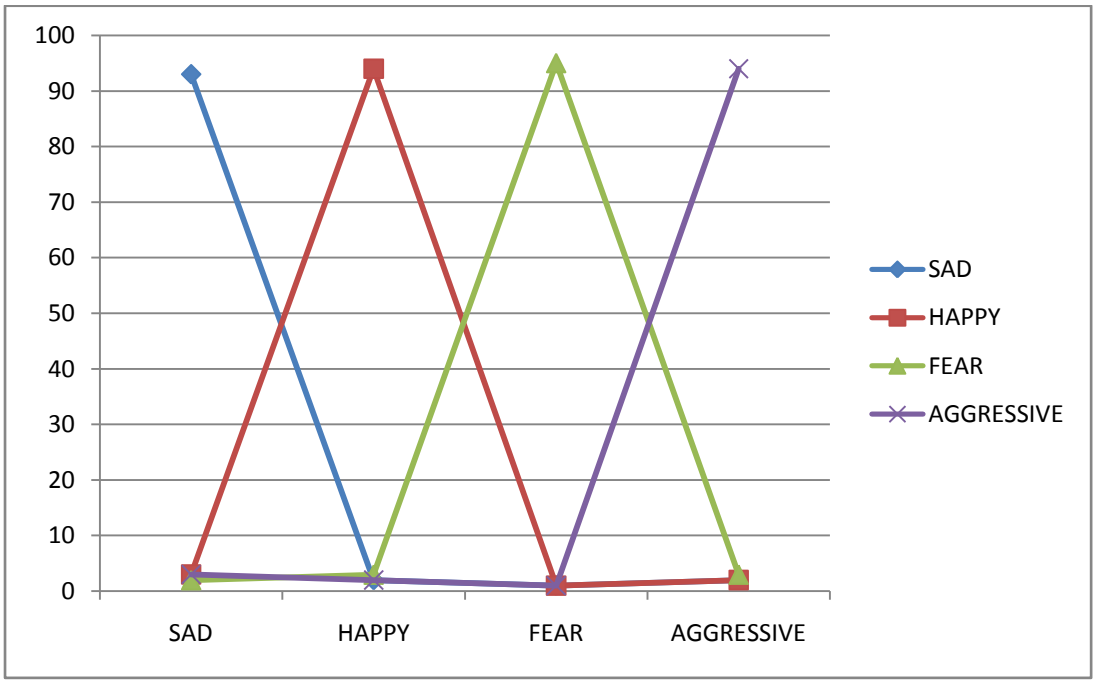


Figure 5.13 Graph representing BPNN accuracy

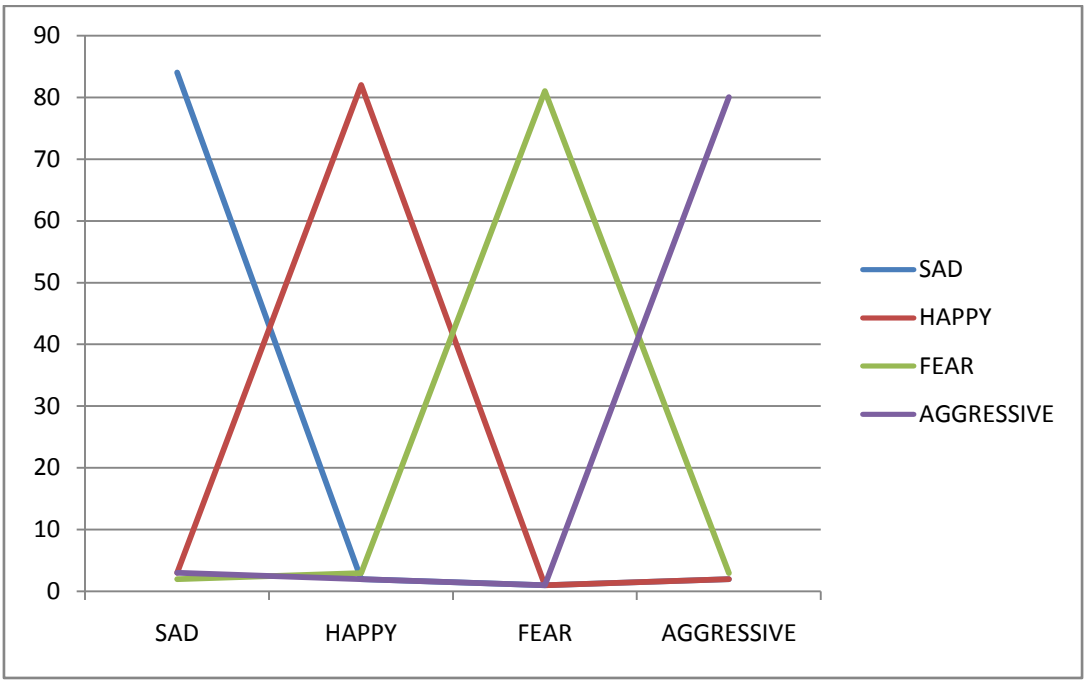


Figure 5.14 Graph representing accuracy for SMO

| CATEGORY | SAD | HAPPY | FEAR | AGGRESSIVE |
|---|---|---|---|---|
| **SAD** | 93 | 3 | 2 | 2 |
| **HAPPY** | 2 | 94 | 3 | 1 |
| **FEAR** | 1 | 1 | 95 | 3 |
| **AGGRESSIVE** | 2 | 2 | 3 | 93 |

Figure 5.15 Accuracy percentages of different moods

The above Figure 5.15 shows the percentage graph of four categories.

# Chapter 6 Conclusion and Future Work

In this chapter we are going to describe that what we conclude about speech recognition using BPA (Back Propagation Algorithm) that belongs to neural networks. Classifier perform better it conclude by testing of speech data with classifier. The results have been also categorized for SMO algorithm

## 6.1 Conclusion

We apply BPA classifier on selected speech data. Each classifier has different theory for implementation. From all the above calculations we come to the conclusion that speech emotion detection by using BPA (Back Propagation Algorithm) that belongs to neural networks perform better with high accuracy. The classification has been done using two algorithms whose details have been already added to the upper sections. The accuracy of the system varies for BPNN or BPA and SMO. The classification accuracy for SMO has come out to be 75-83 percent where as the accuracy for BPA algorithm varies from 85 to 95 percent.

## 6.2 Future Work

The current research work emphasises on the speech emotion detection over mentioned categories and the accuracy of the specified files are good. The current research work opens a lot of gates for the future research workers. This research work is limited for the users that mean the current system has to be trained for the every user you want to identify. The other aspect of this research work is that it does not provide any gender information in terms of the classified data set that means it does not specify that the classified voice is for male or for female. For this purpose, the data base has to be trained with a huge data set. In the similar manner, no segmentation has been performed in the analysis of this work to identify the spoken words. The future research workers can also try their hands on combining the values of BPNN and SVM.

.

# References


[1] J. C. Platt, "Fast Training of Support Vector Machines using Sequential Minimal Optimization," *In Advances in Kernel Methods*, pp. 185-208, 1999.

[2] M. A. Hossain, M. M. Rahman, U. K. Prodhan and M. F. Khan, "Implementation Of Back-Propagation Neural Network For Isolated Bangla Speech Recognition," *International Journal of Information Sciences and Techniques*, vol. 3(4), 2013.

[3] R. Rojas," Neural Networks: A Systematic Introduction," *Springer, 1996*.

[4] D. Ververidis and C. Kotropoulos "Emotional speech recognition: Resources, features and methods," *Speech Communication* , vol. 48(9), pp. 1162-1181, 2006.

[5] J. C. Platt, "Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines", *Advances in Neural Information Processing Systems*, 1999.

[6] W. Gevaert, G. Tsenov and V. Mladenov, "Neural Networks used for Speech Recognition," *Journal of Automatic Control*, Vol. 20 (1), pp. 1-7, 2010.

[7] M. Cilimkovic, "Neural Networks and Back Propagation Algorithm", Institute of Technology Blanchardstown, Blanchardstown Road North Dublin 15, Ireland.

[8] P. Peng, Q. L. Ma and L. Hong, "The Research Of The Parallel Smo Algorithm For Solving Svm," *Proceedings of the Eighth International Conference on Machine Learning and Cybernetics*, vol. 3, pp. 1271-1274, 2009.

[9] R. Fan, P. Chen and C. Lin, "Working Set Selection Using Second Order Information for Training Support Vector Machines," *Journal of Machine Learning Research* Vol. 6 , pp. 1889–1918, 2005.

[10] X. Shao, KunWu, and B. Liao, "Single Directional SMO Algorithm for Least Squares Support Vector Machines," *Computational Intelligence and Neuroscience* Vol., 2013.

[11] F. R. Bach, G. R. G. Lanckriet and M. I. Jordan, "Multiple Kernel Learning, Conic Duality, and the SMO Algorithm", *Proceedings of the 21st International Conference on Machine Learning*, 2004.

[12] S. K. Shevade, S. S. Keerthi, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to the SMO Algorithm for SVM Regression," *IEEE Transactions on Neural Networks*, Vol. 11(5), pp. 1188-1193, 2000.

[13] Y. S. Rao and I. Patel, "Speech Recognition Using Hmm With Mfcc- An Analysis Using Frequency Specral Decomposion Technique," *Signal & Image Processing : An International Journal (SIPIJ)* Vol.1(2), pp. 101-110, 2010.

[14] W. Han, C. Chan, C. Choy and Kong-Pang Pun, "An Efficient MFCC Extraction Method in Speech Recognition," *International Symposium on Circuits and Systems Proceedings*, 2006.

[15] T. Glasmachers and C. Igel, "Second Order SMO Improves SVM Online and Active Learning," *Journal Neural Computation*, Vol. 20 (2) pp. 374-382, 2008.

[16] N.Pushpa, R.Revathi, C.Ramya and S.Shahul Hameed, "Speech Processing Of Tamil Language with Back Propagation Neural Network and Semi-Supervised Training," *International Journal of Innovative Research in Computer and Communication Engineering*, Vol.2 (1), pp. 2718-2723, 2014.

[17] H. Gupta and D. S. Wadhwa, "Speech Feature Extraction and Recognition Using Genetic Algorithm," *International Journal of Emerging Technology and Advanced Engineering*, Vol. 4 (1), pp. 363-369, 2014.

[18] J. Bachorowski, "Vocal Expression and Perception of Emotion," *Current Directions in Psychological Science*, vol. 8 (2), pp. 53–57, 1999.

[19] D. A. Sauter; F. Eisner, A. J. Calder and S. K. Scott, "Perceptual cues in nonverbal vocal expressions of emotion," *The Quarterly Journal of Experimental Psychology*, vol. 63 (11), pp. 2251–2272, 2010.

[20] Y. Pan, P. Shen , and Liping Shen," Speech Emotion Recognition Using Support Vector Machine," *International Journal of Smart Home,* vol. 6(2) , pp.101-107, 2012.

[21] A. Utane, S.L Nalbalwar, "Emotion Recognition Through Speech Using Gaussian Mixture Model And Hidden Markov Model," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, pp. 742-746, 2013.

[22] A. Sapra, N. Panwar, and S. Panwar , "Emotion Recognition from Speech", *International Journal of Emerging Technology and Advanced Engineering*, vol. 3, pp. 341-345, 2013.

[23] P. Yves and Oudeyer, “The production and recognition of emotions in speech: features and algorithms” *International Journal of Human-Computer Studies*, vol. *59*(1), pp. 157-183, 2003.

[24] T. Vogt, E. Andr´e, and J. Wagner, “Automatic Recognition of Emotions from Speech: A Review of the Literature and Recommendations for Practical Realisation,” *In Affect and emotion in human-computer interaction*, pp. 75-91, 2008.

[25] K.A. Senthildevi and E. Chandra, “Speech Data Mining & Document Retrieval,” publication of the IEEE signal processing. www.asru2007.org/submission/EDICS.txt-2

[26] R. Agrawal, M. Mehta, J. Shafer, and R. Srkant, “The Quest Data Mining System,” *Proceedings of International Conference on data mining and knowledge Discovery*, vol. 96, pp. 244-249, 1996.

# List of Publications

[1] Rohit katyal and Rupinderdeep kaur," Efficient Emotion Recognition System based on Classifier", Second International Conference on Emerging Research in Computing, Information, Communication and Application (ERCICA-2014), August 01-02, 2014, NMIT, Bangalore.