Human Emotion Recognition using ANN

Renuka S. Lodha
Electronics &Telecommunication Dept. SNJB COE
Chandwad, Dist- Nashik

M. A. Mechkul
Electronics &Telecommunication Dept. SNJB COE
Chandwad, Dist- Nashik

ABSTRACT
The aim of study of "Human Emotion Recognition By Speech Synthesis" is to make language interfaces in human-computer interaction applications more efficient. This study deals with the six emotions like anger, boredom, disgust, fear, happiness, sadness. Energy, MFCC (14), LPC (12), ZCR, Pitch, First three formant these six different features were extracted. Here Artificial Neural Network is used as classifier.

General Terms
Feature Extraction, Emotional Speech Database

Keywords
Emotion Recognition, Artificial Neural Network.

1. INTRODUCTION
Emotions play an important role in human communications.[1] An emotion is a mental and physiological state associated with a wide variety of feelings, thoughts, and behavior.[2] Speech emotion recognition is one of the latest challenges in speech processing.[3] If the emotional state of the person can be accurately identified in human–machine interaction, then the machine can be produce more appropriate responses . From human speech variety of temporal and spectral features that can be extracted. This project uses statistics relating to the prosody features like Energy, Pitch, ZCR, derived features like Mel Frequency Cepstral Coefficients(14) ,LPC(12) and quality features like Formants of speech as inputs to classification algorithms. There are many applications were emotion identification is used such as Lie detector and can be used as voice tag in different database access systems. This voice tag is used in telephony shopping, ATM machine as a password for accessing that particular account.[4]. Speaker independent emotion recognition is a difficult issue. In a survey conducted to measure human performance on emotion recognition, only 60 percent people can correctly determine the expressed emotions of unknown people [7]. The paper is structured as follows: Section 2 illustrates the feature extraction; Section 3 introduces the Artificial Neural Network, Section 4 describes the database, Section 5 present the results of comparative experiments and the conclusions.

2. FEATURE EXTRACTION
To obtain more stability the speech signal must be divided into frames. Every frame is 50% overlap with previous frame. Energy, Zero Crossing Rate , Pitch, Linear Predictive coefficient(12), Mel Frequency Cepstram coefficient(14), First three Formant frequencies these features are extracted . After that mean, maximum , minimum and standard deviation of the 32 features are obtained. Thereby 128 features are obtained.

2.1 Energy
The amplitude of unvoiced speech segments is much lower than the amplitude of voiced segments. The energy of the speech signal provides reflection of these amplitude variations [5].

2.2 Zero Crossing Rate
For voiced/unvoiced classification Zero-crossing rate is an important parameter. It is often used as a part of the front-end processing in the automatic speech recognition system. The zero crossing rate is nothing but an indicator of the frequency. Voiced speech is generated due to excitation of vocal tract because of periodic flow of air at glottis and it shows a low zero-crossing rate. The unvoiced speech is generated due to the constriction of the vocal tract to cause turbulent air flow. This results in noise and shows high zero-crossing rate [5].

2.3 Pitch
The pitch is usually taken to be the fundamental frequency F0 that is defined as the lowest frequency of a periodic wave-form which it is measured in Hz. Pitch carries information about emotion because it depends on the tension of the vocal track and the sub-glottal air pressure.

2.4 Mel Frequency Cepstral Coefficients
In many applications for spectral representation of speech including speech and speaker recognition MFCCs are the most widely used. Kim et al. argued that statistics relating to MFCCs also carry emotional information [9]. In this work for each 20ms frame of speech, fourteen standard MFCC parameters are calculated.

2.5 Formant Frequencies
Formants are nothing but the spectral peaks of the sound spectrum |P(f)| of the voice. In speech science and phonetics, formant frequencies are an acoustic resonance of the human vocal tract which is measured as an amplitude peak in the frequency spectrum of the sound. In acoustics, formants are referred as a peak in the sound envelope and/or to a resonance in sound sources, as well as that of sound chambers. Here extraction of First three Formant Frequencies is done by LPC.

2.6 Linear Predictive Coefficient
The LPC calculates a power spectrum of the signal. It is desirable to compress signal for efficient transmission and storage[4]. For encoding good quality speech at a low bit rate, LPC is one of the most useful method. In this work direct MATLAB command is used for LPC calculation.

3. ARTIFICAIL NEURAL NETWORK

3.1 Resemblance with brain
Neural networks resemble the human brain as follows :

1. A neural network acquires knowledge through learning.
2. A neural network's knowledge is stored within inter-neuron connection strengths known as synaptic weights .
3. Neural networks modify own topology just as neurons in the brain can die and new synaptic connections grow.

3.2 What is the ANN?

A neural network is a powerful data modeling tool which is able to capture and represent complex input/output relationships.[1]. Neural Network technology performs "intelligent" tasks similar to those performed by the human brain [2]. ANN is composed of a large number of highly interconnected processing elements called as neurons. "Figure 1" shows the general structure of neural network. ANN is One of the most successful classifiers yet.

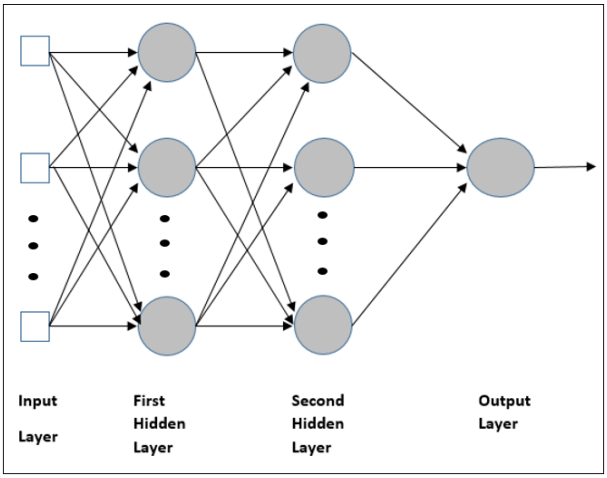

Fig 1: Artificial Neural Network

3.3 Advantages of ANN

1. Adaptive learning : It is an ability to learn how to do given task based on the data given for training or for initial work.
2. Self-Organization : An ANN creates its own representation of the information which is receives during learning time.
3. Parallel Structure : If any failure come in network, whole system never fails due to parallel nature of ANN.

3.4 Feed Forward Neural Network

In this work we use the feed-forward neural network. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes and to the output nodes. There are no cycles or loops in the network. "Figure 2" shows the structure of feed forward neural network. Nodes(presented as Circle) represent the neurons and arrows represent the links between them. It is also called as perceptron. Hidden nodes in hidden layer are not connected directly to the environment. They are hidden inside the network.

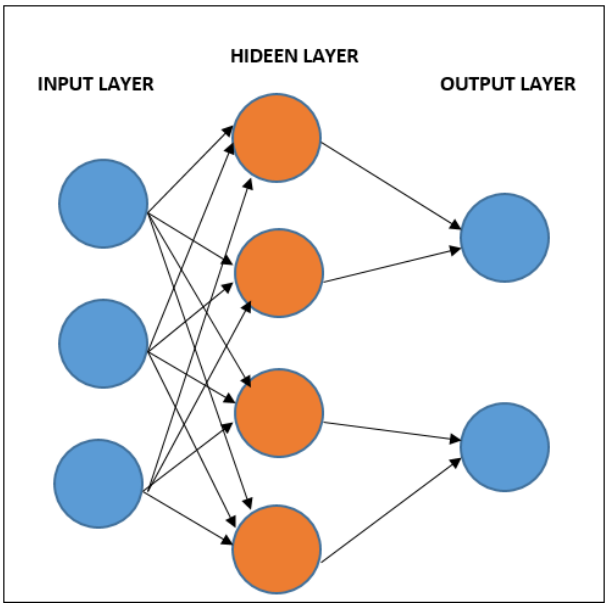

Fig 2 : Feed Forward Neural Network

Feed forward neural network can be deal with both real and discrete domains. It has very fast classification time and it requires slow training time.

4. DATABASE DISCRIPTION

In this work "Berlin Database Of Emotional Speech" is used. It contains about 500 utterances spoken by actors in a anger, .boredom, disgust, fear , happiness, sadness way as well as in a neutral version. Except neutral version six emotions are selected for this work All these utterances are performed by five actors and five actresses whose ages are between 21 and 35. There are 10 different texts . All these utterances have a sample frequency of 16000 Hz and a mean duration of 1.2 s. For this work we made the two groups of database. One for training and another for testing. 50 utterance with six emotions are used for testing and 64 utterance with six emotions are used for training.

5. RESULT

Training and the testing of neural network is done by using train dataset. Then used the test dataset which contains the utterance of different speakers . "Figure 3 " gives the result of train data set. The confusion matrix shows the result of test data set. Result is near about 86%. "Figure 4 " gives the result of test data set. The confusion matrix shows the result of test data set. Result is near about 89%. In this work 15 iterations for testing data set and training dataset have been taken. 92 accuracy is getting after averaging it .

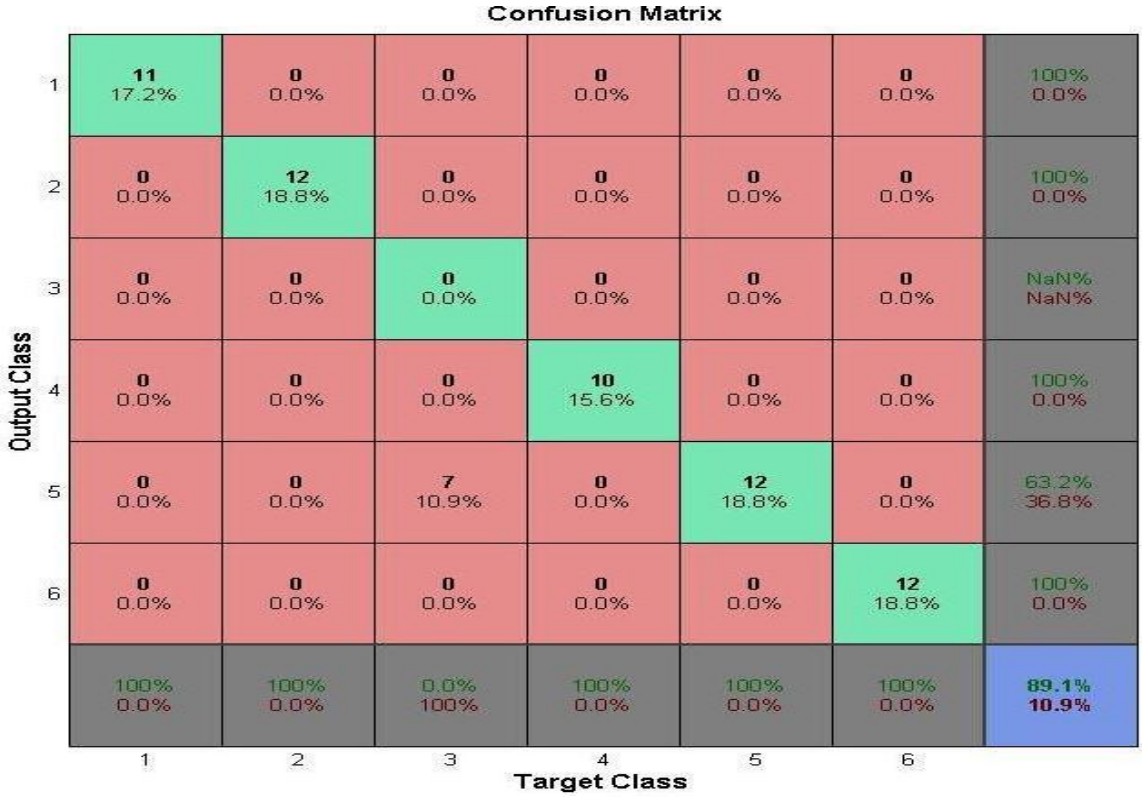

Fig 3: Confusion Matrix for Train Data set

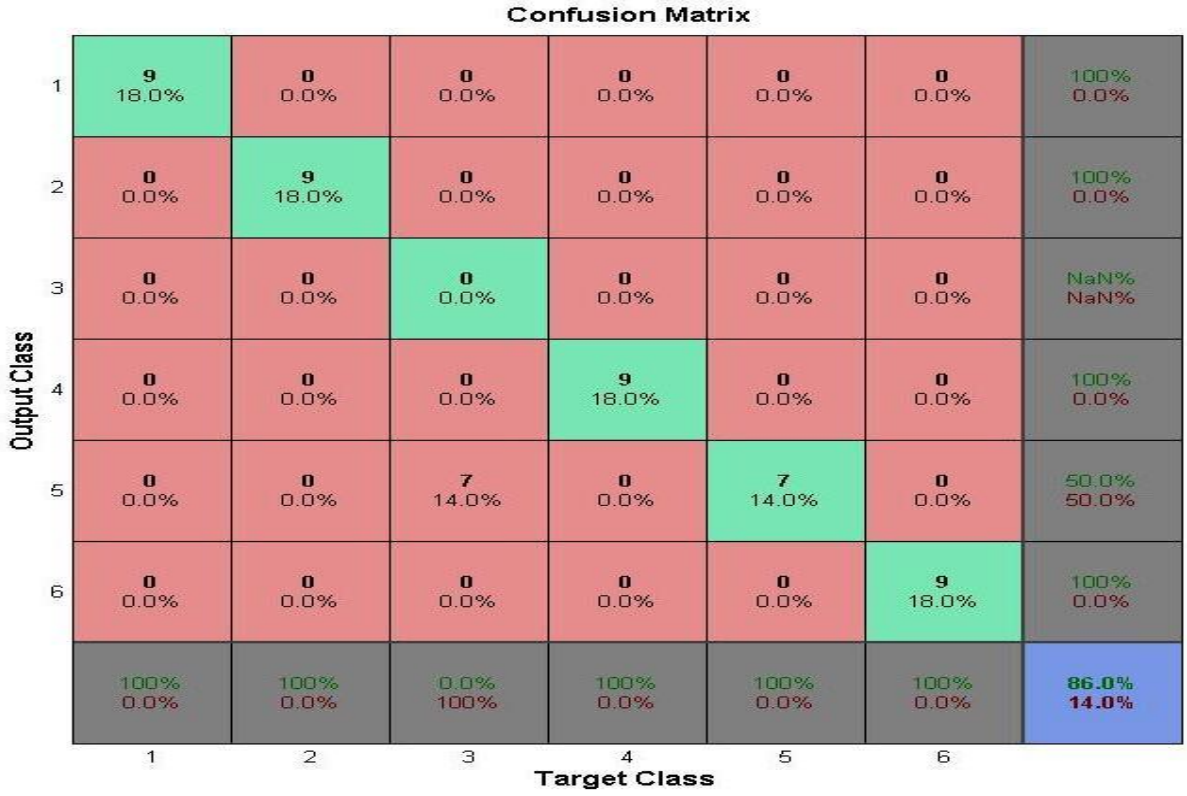

Fig 4: Confusion Matrix for Test Data set

6. CONCLUSION

The designed system is speaker independent. Six different human emotions are recognized. Six different features of emotional speech are recognized. And performance of classifier is studied accordingly.

7. REFERENCES

[1] Sindhu S Pandya, I/C Principal, Laxmi Institute of Computer Applications (BCA), Sarigam, Valsad, Gujarat),"Real-World Application of Neural Network" Automated Paper Assessment, International Journal of IT, Engineering and Applied Sciences Research (IJIEASR) ISSN: 2319-4413 Volume 2, No. 7, July 2013.

[2] Meghana Nagori, Sarita T. Sawale, V. P. Kshirsagar, "Emotions and Strategies for Preparation of Emotional Speech Database", International Journal Of Computer Science And Applications Vol. 3, No. 1, January / February 2010 ISSN: 0974-1003.

[3] Fatin B. Sofia, Sahar K. Ahmed &Abdul-basit K. FaeqMosul University, "Emotion Recognition in Speech Using NeuralNetwork", J.Edu.& Sci. , Vol. (21) No.(1) 2008.

[4] A. A. Khulage and Prof. B. V. Pathak., "ANALYSIS OF SPEECH UNDER STRESS USING LINEAR TECHNIQUES AND NON-LINEAR TECHNIQUES FOR EMOTION RECOGNITION SYSTEM "

[5] D.S.Shete , Prof. S.B. Patil ,Prof. S.B. Patil , "Zero crossing rate and Energy of the Speech Signal of Devanagari Script" , IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 4, Issue 1, Ver. I (Jan. 2014), PP 01-05 e-ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197.

[6] Christos Stergiou and Dimitrios Siganos ,"NEURAL NETWORKS".

[7] Tin Lay Nwe , Say Wei Foo , Liyanage C. De Silva, "Speech emotion recognition using hidden Markov models", Speech Communication 41 (2003) 603–623.

[8] Amiya Kumar Samantaray, "Development of a Real-time Embedded System for Speech Emotion Recognition".

[9] Namrata Dave, 1G H Patel College of Engineering,Gujarat Technology University, INDIA, "Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition", Volume 1, Issue VI, July 2013.