# Speech Enhancement and Speech Emotion Recognition from Features Extraction on NN Classifier

S.Muthamilselvan [1], Vuppula Srinath [2], Peddi Mahesh Kumar [3], Battina Kalidas [4], Amireddy Tharun Reddy [5]
[1] A.P(O.G), Department of CSE, SRM IST, Chennai, Tamil Nadu, India.
[2, 3, 4, 5] U.G(Scholar), Department of CSE, SRM IST, Chennai, Tamil Nadu, India.

**Abstract – The project presents speech emotion recognition from speech signal based on features analysis and NN-classifier. Automatic speech emotion recognition plays an important role in HCI systems for measuring people's emotions has Dominated psychology by linking expressions to group of basic emotions (i.e., anger, disgust, fear, happiness, sadness, and surprise).The process of recognition involves the voice ,which depends on some of the factors they are frequency and related to that ,for each distinct expression there is a sample which is available in its data base and when we give inputs to the compiler through audio file format and in form of and remembrances the input audio by analogizing to the data base files through frequency levels then it shows the audio file is an expression from the data base gives the simulation through NN classifier by which the classification is done by lesser algorithmic complexity compared to other expression recognition approaches .**

**Keywords - Speech emotion recognition, VAD algorithm, NN classifier, NN training**

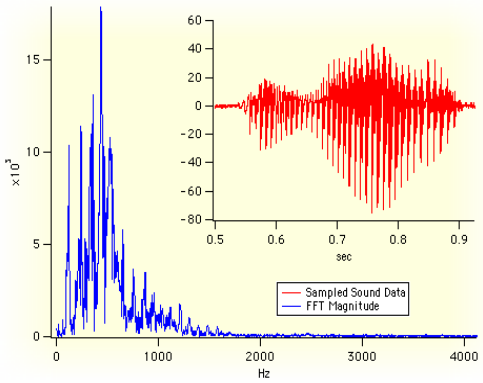
## 1. INTRODUCTION

Speech emotion recognition is all about checking the emotion of a particularly selected audio file which is inferred through remembrance of data base of audio files. Then gives the declaration of particular emotions which are Sad, Joy, fear, anger and disgust these are the basic emotions . The required details to check a particular voice is sad or happy this can be done by matching the audio files to the data base audio files and the comparison is done by NN classifier which helps the audio file to get audio file required until it checks all the files and comes to conclusion depending upon frequencies of the of the input audio file and the audio files in database and then displays which emotion has in the audio file

Another interesting area to explore in Emotion recognition is VAD (voice activity detector) algorithm it is the technique to identify the frequencies because the algorithm declares a sound according as frequency which has high frequency and low frequency and according to the fluctuations in the audio the algorithm test what type emotion in the audio.

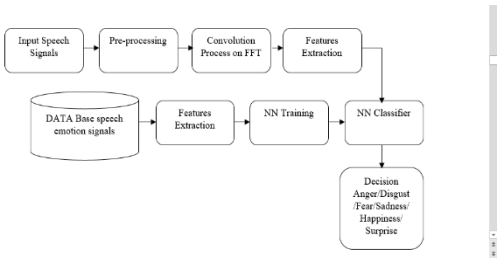
- **Capacity:** It refers to the amount of information that can be remembrance to the data base in an efficient way
- **Perceptibility:** The ability to check the emotions through Data base audio files
- **Robustness:** The amount of modification that is used to check audio files through the data base

## 2. SPEECH SIGNAL




(a)

The speech signal involves the frequencies, and they have variable to determine frequencies which is hertz .the graph indicates two variables and one of them id FFT magnitude. The **FFT** returns a two-sided spectrum in complex form (real and imaginary parts), which you must scale and convert to polar form to obtain **magnitude** and phase. The frequency axis is identical to that of the two-sided power spectrum. The **amplitude** of the **FFT** is related to the number of points in the time-domain signal




(b)

## 3. FEATURES EXTRACTION

The Speech Emotion recognition for transform features system through textural analysis and NN classifier. The block diagram explains how a feature extraction of audio and the preprocessing of the audio file. The explanation of process how it is connected to NN classifier and the data base .The process of the project starts from the input speech signal and the preprocessing stage starts which in details the audio tape which is the input to the NN classifier and recognizes the signal inform of frequency graph and the checks the input audio is remembrance the audio in the data base and then matches the audio file through the help of NN classifier which is gained through NN training and the decision is displayed through the pop and displays the result as angry or sad or remaining one of the expression .

**A** .Preprocessing:

The preprocessing stage in speech recognition systems is used in order to increase the efficiency of subsequent feature extraction and classification stages and therefore to improve the overall recognition performance.

At the end of the preprocessing the compressed and filtered speech frames are forwarded to the feature extraction stage and then the NN classifier that plays an important role and which is used for training purposes and this type of training is called NN training which has many structure to be explained as follows

**B.** Fourier transform:

The Fourier transform is a tool which helps to formulate the data into sine and cosine format which means in a wave for us it will be useful to know the audio clip is transformed to the audio frequency signal . The next processing step is the Fast Fourier Transform, which converts each frame of N samples from the time domain into the frequency domain. The FFT is a fast algorithm to implement the Discrete Fourier Transform (DFT), which is defined on the set of N samples $\{x_n\}$, as follow:

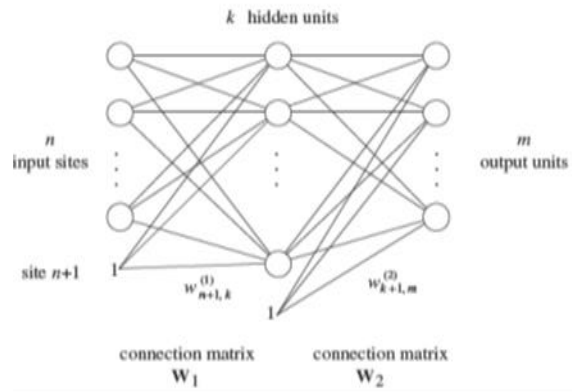
$$X_k = \sum_{n=0}^{N-1} x_n e^{-j2\pi kn/N}, \qquad k = 0,1,2,\ldots,N-1$$

In general $X_k$'s are complex numbers and we only consider their absolute values (frequency magnitudes). The resulting sequence $\{X_k\}$ is interpreted as follow: positive frequencies $0 \le f < F_s/2$ correspond to values $0 \le n \le N/2-1$, while negative frequencies $-F_s/2 < f < 0$ correspond to $N/2+1 \le n \le N-1$. Here, $F_s$ denotes the sampling frequency.

The result after this step is often referred to as spectrum or periodogram.

## 4. NN CLASSIFIER

In the classifier to explain first knowing of NN is most required NN( Neural network).the neural network can be explained as the more number of input it gets and it continues to most of the possibilities for example we can assume there are n inputs and the outputs are m and there are k hidden data and this all together is perfectly mapped this construction of neural network can be explained through structure of neural network which is given below in the diagram




(c)

Fig.3. Neural network structure

The background propagation of algorithm is used to do the necessary settings that are required to improve the efficiency of the algorithm after chosing them randomly every possible is to checked in short span of time the algorithm divide it self as four following steps

(i)Feed –forward computation

(ii)Back proportion to output layer

(iii)Back proportion of the hidden layer

(iv) Weight updates

These are steps which are required for the classified algorithm and these help to reduce time complexity as well

## 5. SIMULATION PROCESS

### *A. Extraction Process*

The extraction process is the initial and crucial part of The emotion recognition in which the input signal are converted to the frequency format through the Fourier transform tool which is present in matlab

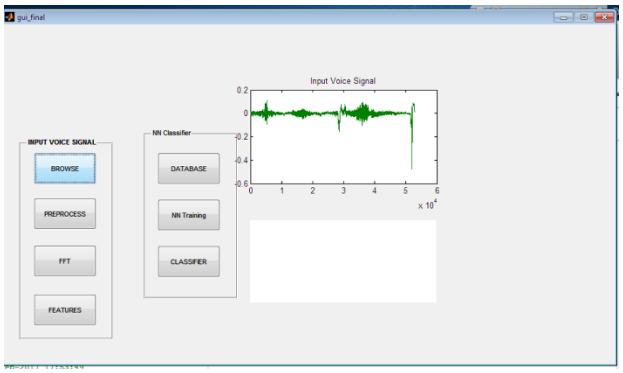



Fig.4.(a)selecting a input signal

B. Preprocessing :

As mentioned above the preprocessing which is to be done through the input signal and the data base signal which helps to identify according to the signal frequency

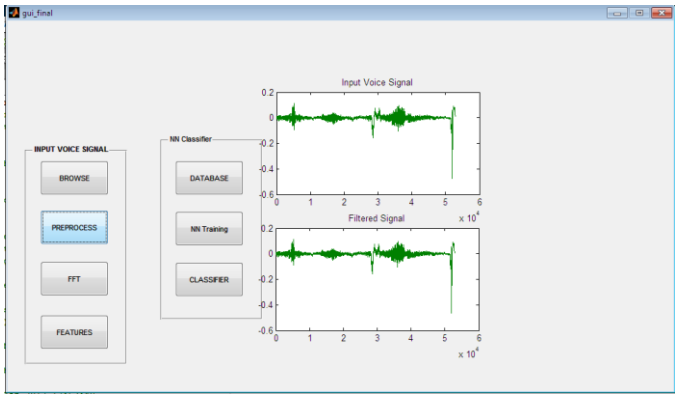



Fig.4.(b)preprocessing the voice signal

C. FFT Modulation:

The FFT modulation which helps the graph to show in two dimensional view and this helps us in to view through frequency measured unit which is hertz. To get more information and pictorial representation we can see below diagram fig 4(c) and Fourier transform tool is needed to show this diagrammatic format. This FFT is the fast fourier transformation. The spectrum produced using this transform is periodic with the sampling frequency for the real input signals.

Compared with traditional Fast Fourier Transform (FFT), Modulation FFT preserves its characteristics of low complexity when used to estimate the frequencies of real valued signals. But there are spectral leakage and picket-fence effect, frequency resolution is still difficult to achieve high accuracy. Multiple signal Classification (MUSIC) algorithm can achieve high frequency estimation accuracy. But it requires searching the spectral peak in the full range frequency domain and this takes a long time.

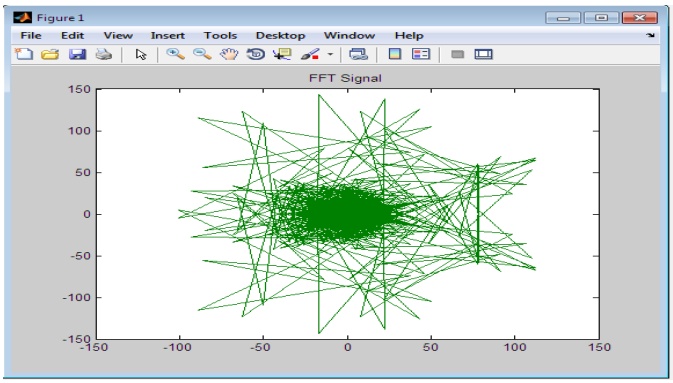
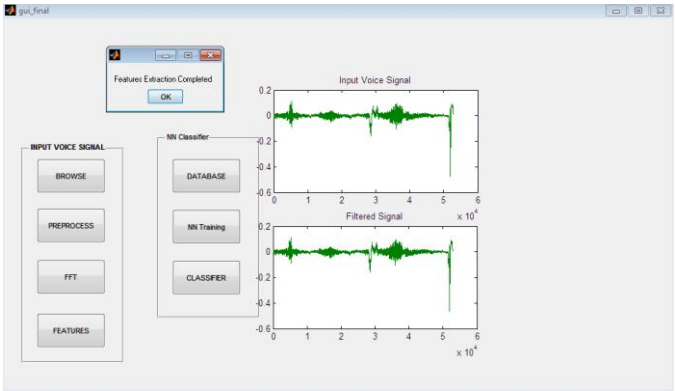



Fig.4 (c) FFT modulation

D . Feature Extraction:

This helps the data or the input signal to work effectively and signifies the data by matching and when completes it shows out the pop up of feature extraction completed




FIG(D) FEATURE EXTRACTION

The below diagrams show the step process of hoe to load data and how it is NN trained and what type of pop out we get explained through pictorial representation of our project

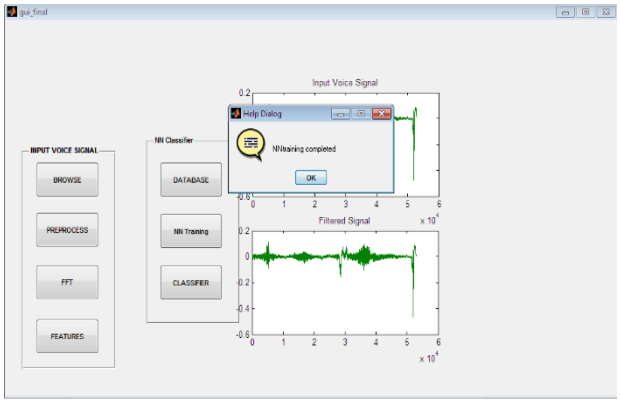



FIG (e) Data loading of the audio Files

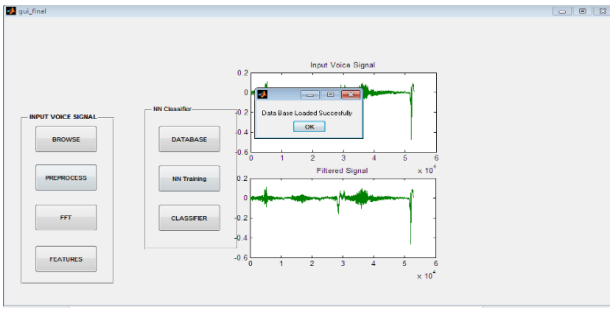
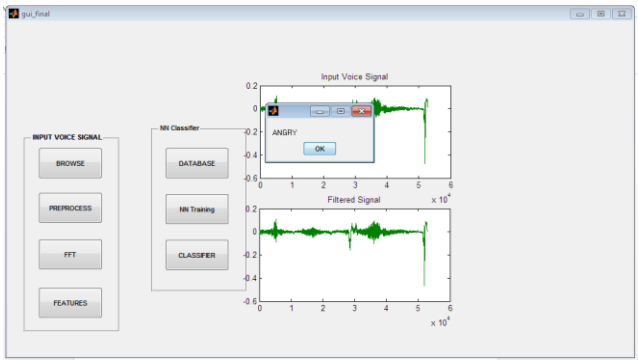



FIG4 ( F) ,NN TRAINING




FIG(G) OUTPUT

## 6. CONCLUSION

In this project, most recent work done in the field of Speech Emotion Recognition is discussed. Most used methods of feature extraction and several classifier performances are reviewed. Success of emotion recognition is dependent on appropriate feature extraction as well as proper classifier selection from the sample emotional speech. It can be seen that Integration of various features can give the better recognition rate.

Classifier performance is need to be increased for recognition of speaker independent systems. The application area of emotion recognition from speech is expanding as it opens the new means of communication between human and machine. It is needed to model effective method of speech feature extraction so that it can even provide emotion recognition of real time speech. This speech emotion recognition has various application in robotics and medical field. The sole purpose of this process is to help in man machine interaction. In robotics, Robot can recognize emotional information as contained in the human speech signals for friendly interaction with human beings, and eventually satisfactory performance and effect are realized. In call-center application. The experiments results indicate that the proposed method provides very stable and successful emotional classification performance and it promises the feasibility of the agent for mobile communication services.

## REFERENCES

[1] A Comparative Study of Traditional and Newly Proposed Features for Recognition ofSpeech Under StressSahar E. Bou-Ghazale, Member, IEEE, and John H. L. Hansen, Senior Member, IEEE

[2] E. C. Lin, K. Yu, R. A. Rutenbar, and T. Chen, “In silico vox: Towards speech recognition in silicon,” in Proc. HOTCHIPS.

[3] L. M. Gruppet al., “Characterizing flash memory: Anomalies, observations, and applications,” in Proc. 42nd Annu. IEEE/ACM Int. Symp. Microarchitecture (MICRO).

[4] F. Jelinek, Statistical Methods for Speech Recognition. Cambridge, MA, USA: MIT Press, 1997.

[5] M. Price, J. Glass, and A. P. Chandrakasan, “A 6 mW, 5,000-word real-time speech recognizer using WFST models,” IEEE J. Solid-State Circuits, vol. 50, no. 1, pp. 102–112, Jan. 2015.

[6] S. B. Davis and P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,” IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-28, no. 4, pp. 357–366, Aug. 1980.