

## Emotion Recognition based on audio signal using GFCC Extraction and BPNN Classification

Shaveta Sharma<sup>1</sup>, Parminder Singh<sup>2</sup>

<sup>1</sup> Assistant Professor, Department of Electronics and Communication, GCET, PTU

<sup>2</sup> Assistant Professor, Department of Electronics and Communication, DIET, PTU

### ABSTRACT:

For automatic speech recognition (ASR) there is a big challenge which deals with momentous presentation reduction in high noisy environments. This paper presents our emotion classification based on Gammatone frequency cepstrum coefficient used for feature extraction along with Back propagation neural network and the experimental results on English speech data. Eventually, we obtained significant presentation gains with the new feature in various noise conditions when compared with traditional approaches. In our proposed work we considered two emotions SAD and HAPPY which are used to show the implementation results. The simulation environment taken to implement the whole process is in MATLAB

**Keywords:** Back propagation neural network, GFCC, audio signals, filter banks

### I. INTRODUCTION

Emotional speech recognition aims at involuntarily identifying the emotional or physical condition of a human being via his or her voice [1]. 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 database considered for emotion recognition is based on audio signals based on emotions. The features extracted from these speech samples are, the energy, pitch, linear prediction cepstrum coefficient (LPCC), Gammatone frequency cepstrum coefficient (GFCC) [2] etc. Among them GFCC is widely used for speech related studies with a simple calculation and good ability of the distinction. So, in the proposed work GFCC will be used. There are some factors that make difficult the speech [3] recognition 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 speaker's voice too adds to this facto.

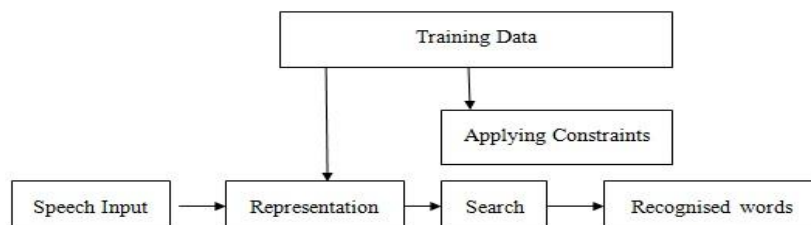


Figure 1. Speech Recognition System

In this proposed work, we are going to use Back Propagation neural network technique [4] on speech that is extract by GFCC. The classification performance is features extraction method using GFCC extraction method. In this paper we have taken two emotions i.e. SAD and HAPPY. Eventually we achieve on a conclusion on the basis of emotion having accuracy that will be achieved by classifying using Back Propagation Neural Network method.

Lasting part of this paper is discussed as following: Section I defines the introduction to the basic topic, Section II shows the introduction to speech features of emotion detection, Section III shows the proposed flowchart of the proposed methodology, Section IV shows the pseudo code of the implementation algorithm, Section V shows the results and discussion. Finally Section VI contains the conclusion part of the proposed work.

### 1. GFCC (Gamma Tone Frequency Cepstrum Coefficient)

Gammatone cepstral coefficients computation process is equivalent to MFCC extraction scheme. The audio signal is first windowed into short frames; usually of 10–50 ms. This process is based on two processes: The typical audio signals which are non stationary can be assumed to be stationary for such a short interval, thus facilitating the spectro temporal signal analysis. The efficiency of the feature extraction process is increased. Next, the GT filter bank which is composed of the frequency responses of the several Gamma Tone filters is applied to the signal's fast Fourier transform (FFT), emphasizing the perceptually meaningful sound signal frequencies. The design of the Gamma Tone filter bank which is composed of array of band pass filters is the object of study in this work, taking into account characteristics such as: total filter bank bandwidth, GT filter order, ERB model and number of filters.

## II. PROPOSED FLOWCHART

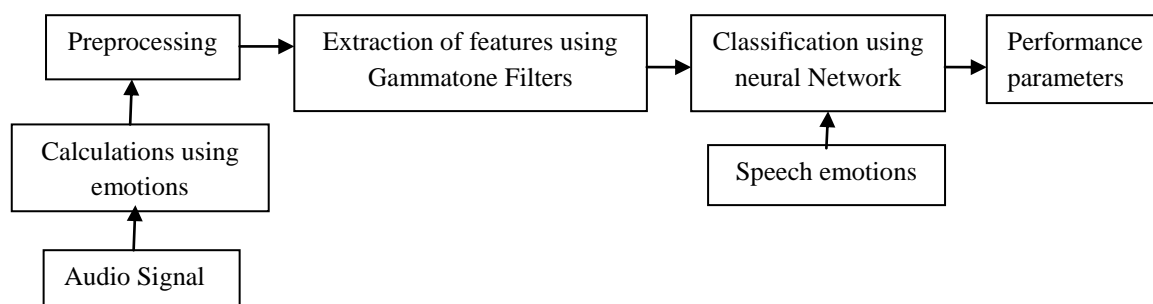


Figure 2. Flow diagram of the proposed work

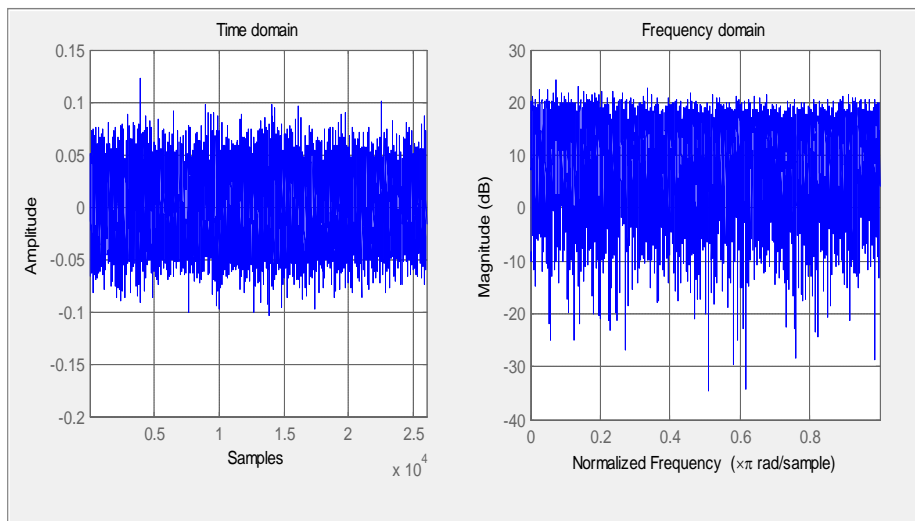
1. Firstly we need speech file for emotion detection. So the basic step is to upload speech file which is in .wav format in MATLAB software
2. Now progression of some steps to make it compatible with MATLAB platform and ready for further process through training of the audio files[7], used to familiarizing the arrangement with characteristics of the emotions of the speech.
3. Apply Gamma Tone Frequency Cepstrum algorithm for feature extraction based on audio files like min value, max value etc.
4. Now testing process will be implemented using classifier which is Back Propagation Neural Network and then accuracy [8] is measured.
5. Evaluate results.

### III. Pseudo code for GFCC Technique

1. Pass input signal through a 64-channel filter bank which contains array of band pass filters [9].
2. At each channel, fully rectify the filter response (i.e. take absolute value) and decimate it to 100 Hz as a way of time windowing.
3. Creation of time frequency (T-F) representation to convert the time domain signal to frequency domain
4. Apply logarithms for finite sequence data set
5. Apply DCT transformation for compression of speech signal and convolution computations.

### IV. Simulation and Discussion

Results simulation is taken place in MATLAB environment. Firstly we will upload the file emotion set like happy sad fear etc randomly. Then we set the noise level because we assume that the speech signal is not noise free signal. Then we will extract the features using GFCC algorithm [10] which is used for feature extraction. It includes Fast Fourier transformation used to convert the time domain signal to frequency domain for spectral analysis, filtration process like hamming window which is a type of filter to attenuate the unwanted frequencies and accepts the required frequency to boost up the frequencies and Error rectangular bandwidth which is the process of bandwidth approximation and to increase the strength of the signal in noisy environment. GFCC also includes filter bank which is an array of Band Pass Filter that separates the input signal into multiple components, each one carrying a single frequency sub-band of the original signal. Then we train the system for the emotion detection so that it will perform better for further use then we will test the uploaded file using neural network which act as a classifier to classify the speech emotion and then on the basis of that detection the accuracy will be measured.



The above figure shows the GFCC based graph in which signal is converted from time domain to frequency domain after applying fast Fourier transform because the signal is having no frequency component in the time domain so we apply FFT to boost up the frequency and for conversion of signal to frequency domain from time domain to have frequency component in the audio signal.

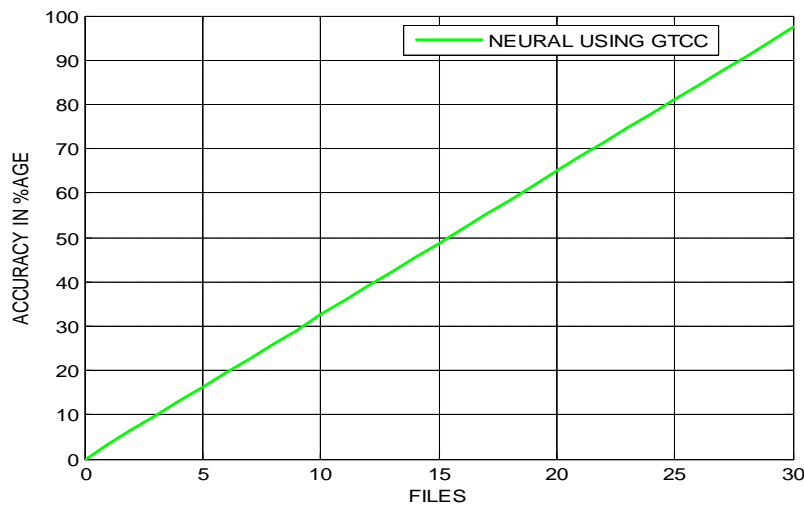
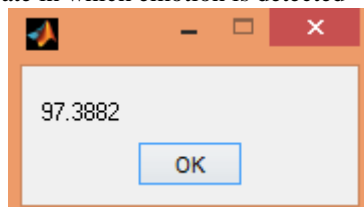


Figure.3 GFCC Performance

The above figure shows the accuracy rate in which emotion is detected



The above figure shows the message box used in MATLAB which shows the accuracy value which is 97.3882

### V. Conclusion and Future scope

In proposed work, eventually we conclude about recognition of speech using BPA (Back Propagation Algorithm) that belongs to neural networks. Classifier [11] perform better it conclude by testing of speech data with classifier. 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 detection of speech emotion by using BPA (Back Propagation Algorithm) that belongs to neural networks with GFCC feature extraction method performs better.

In future, we can apply this proposed algorithm on different languages to make this proposed algorithm more practical .i.e. Punjabi, Hindi, Bengali speech etc. We can also replace the BPA classifier by the MLP Classifier to get differentiate accuracy rate. Also we can increase the number of features extracted from speech data inputted like pitch, frequency range etc.

### REFERENCES

- [1] Md. Ali Hossain, Md. Mijanur Rahman, Uzzal Kumar Prodhon, Md. Farukuzzaman Khan “Implementation Of Back-Propagation Neural Network For Isolated Bangla Speech Recognition” In ijscit , 2013.
- [2] Wouter Gevaert, Georgi Tsenov, Valeri Mladenov, “Neural Networks used for Speech Recognition”, JOURNAL OF AUTOMATIC CONTROL, UNIVERSITY OF BELGRADE, VOL. 20:1-7, 2010.
- [3] 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, No. 5, September 2000.
- [4] Ibrahim Patel, Dr. Y. Srinivas Rao, “Speech Recognition Using Hmm With GFCC- An Analysis Using Frequency Spectral Decomposition Technique”, Signal & Image Processing : An International Journal(SIPIJ) Vol.1, No.2, December 2010.
- [5] Hitesh Gupta, Deepinder Singh Wadhwa, “Speech Feature Extraction and Recognition Using Genetic Algorithm”, International Journal of Emerging Technology and Advanced Engineering, Vol. 4, Issue 1, January 2014.
- [6] Jian Wang, Zhiyan Han, and ShuxianLun, “Speech Emotion Recognition System Based on Genetic Algorithm and Neural Network”, Published in IEEE, Print ISBN No: 978-1-61284-881-5/11/\$26.00, on 2011
- [7] Richard P. Lippmann, “Speech recognition by machines and humans”, 1997 Elsevier Science B.V.
- [8] Kuo-Hau Wu, Chia-Ping Chen and Bing-Feng Yeh, “Noise-robust speech feature processing with empirical mode decomposition”, EURASIP Journal on Audio, Speech, and Music Processing 2011.
- [9] JozefVavrek, JozefJuhar and Anton Cizmar, “Emotion Recognition from Speech”, Published in IEEE Transactions on Audio Speech Vol.21,No.12,on dec2013.
- [10] M.A.Anusuya, S.K.Katti, “Speech Recognition by Machine: A Review” (IJCSIS) International Journal of Computer Science and Information Security, Vol. 6, No. 3, 2009
- [11] Rahul.B.Lanjewar,D.S.Chaudhari,“SpeechEmotion Recognition:A Review” International Journal of Innovative Technology and Exploring Engineering, ISSN:2278-3075, Vol.2,Issue-4, March 2013.