Volume 5, Issue 1, January 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Extracting GFCC Features for Emotion Recognition from Audio Speech Signals

Shaveta Sharma, Parminder Singh
Electronics & Communication Engineering
GCET/PTU, Punjab, India

Abstract— A major challenge for automatic speech recognition (ASR) relates to significant performance reduction in noisy environments. This paper presents our implementation of the Gammatone frequency cepstral coefficients (GFCCs) filter-based feature along with BPNN and the experimental results on English speech data. By some thorough designs, we obtained significant performance gains with the new feature in various noise conditions when compared with traditional approaches. In proposed work three emotions SAD, HAPPY and ANGRY has been used to show the implementation results. The whole simulation has been taken place in MATLAB.

Keywords— Speech, MATLAB, Recognition, GFCC, Neural Network

I. INTRODUCTION

Speech [1] is a complex signal which contains information about the message, speaker, language and emotions. Emotion on other side is an individual mental state that arises spontaneously rather than through conscious effort. The database for the speech emotion recognition system is the emotional speech samples. Features for emotion recognition are extracted from these speech samples. 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 speakers voice too add to this facto.

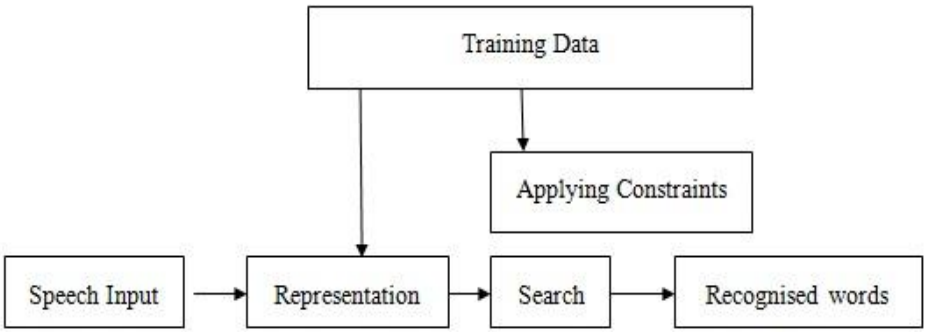

Fig.1 Speech Recognition System

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 based on extracted features. There may be type of emotions during speech like sad, happy, angry, aggressive etc. In this work three emotions are going to detect. i.e. SAD, HAPPY, ANGRY using GFCC. On basis of these emotions finally we achieve on a conclusion with accuracy results that will be achieved by classifying using BPNN method.

Remaining part of the paper is discussed as following: Section I shows 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.

II. 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. The paralinguistic features [5, 6] can be classified in prosodic, spectral, and voice quality features.

© 2015, IJARCSSE All Rights Reserved

A) Paralinguistic's Features

1) 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 caused because of movement of vocal cord. It carries information about emotions because of stress of vocal cords. There is a term pitch period , it I the time between the openings of the vocal cords. Sound energy is also produced due to the pitch movement.

2) 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. The energy operator is used to measure the flow of air.

B) Voice quality features

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

C) Linguistic Features

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

III. PROPOSED FLOWCHART

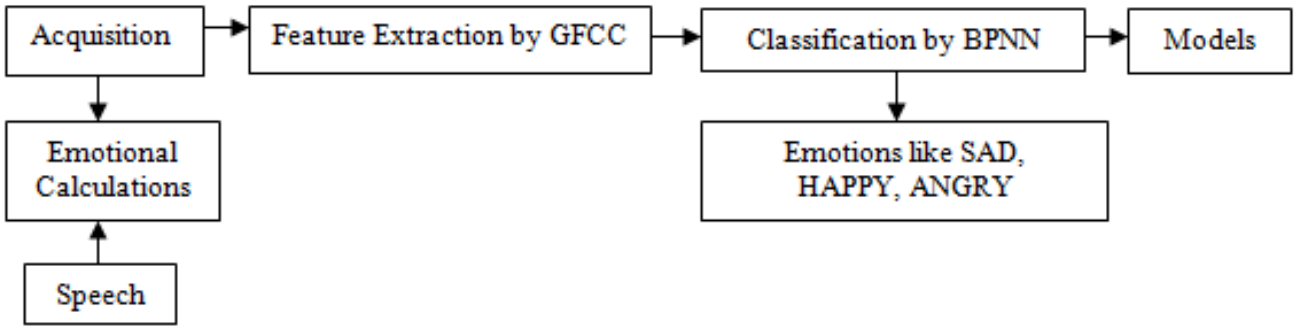

Fig.2 Proposed Flowchart

1. First of all we need speech file for emotion detection. So upload wave file of speech in MATLAB tool.
2. Now process some steps to make it compatible in our platform and ready for further process through training [7], used to familiarizing the system with the emotion characteristics of the speech.
3. Apply GFCC algorithm for feature extraction
4. Now testing process will be implemented using Back Propagation Neural Network and then accuracy [8] is measured.
5. Evaluate results.

IV. 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.

V. RESULTS 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 training the system for the emotion detection then we will test the uploaded file using neural network which act as a classifier to classify the speech emotion.

© 2015, IJARCSSE All Rights Reserved

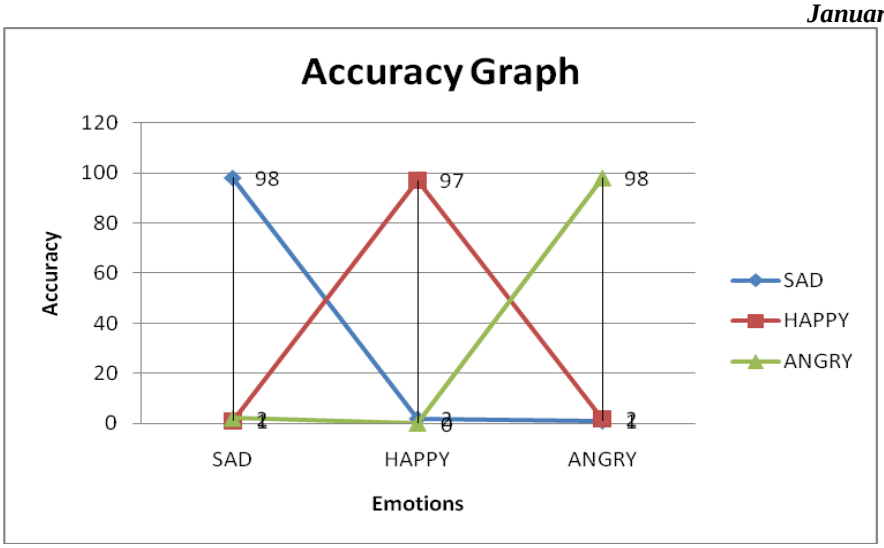

Fig.3 GFCC Performance

Above Figure shows that by using GFCC feature extraction method, accuracy rate of emotions are shown respectively.

VI. CONCLUSIONS AND FUTURE SCOPE

In proposed work, we conclude about speech recognition 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 speech emotion detection 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 Prodhan, 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 Specral Decomposion 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 Algoithm and Neural*

[7] *Network*, Published in IEEE, Print ISBN No: 978-1-61284-881-5/11/$26.00, on 2011

[8] Richard P. Lippmann, *Speech recognition by machines and humans*, Elsevier Science B.V. All rights reserved. II S0167-6393_ 97. 00021-6.

[9] 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, 2011:9.

[10] JozefVavrek, JozefJuhar and Anton Cizmar, *Emotion Recognition from Speech*, Published in IEEE Transactions on Audio Speech Vol.21,No.12,on dec2013.

[11] 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

[12] 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.

© 2015, IJARCSSE All Rights Reserved