Volume 4, Issue 6, June 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Emotion Detection Independent of User Using MFCC Feature Extraction

Jagvir Kaur
MTECH,CSE,RIMT-IET,Mandigobindgarh
India.

Abhilash Sharma
CSE,RIMT-IET, Mandigobindgarh
India.

Abstract*—Emotion is a mental and physiological state associated with a wide variety of feelings, thoughts, and behavior.Speech processing is emerged as one of the important application area of digital signal processing. Various fields for research in speech processing are speech recognition, speaker recognition, speech synthesis, speech coding etc. The objective of automatic speaker recognition is to extract, characterize and recognize the information about speaker. Emotions are subjective experiences, or experienced from an individual point of view.This process includes a proper training and testing pattern. The classification method involves the processing of the saved data against the uploaded data with the same extracted features. In this paper we proposed MFCC as feature extraction technique and neural network as a classifier for detection user and its emotion.

Keywords*—Automatic Emotion Recognition (AER), MFCC, Neural network, Training, Testing.

I. INTRODUCTION

Affective computing is the study and development of systems and devices that can recognize, interpret, process, and simulate human affects. Affective computing is concerned with emotions and machines. It is an interdisciplinary field spanning computer science, psychology, and cognitive science.[1]. The machine should interpret the emotional state of humans and adapt its behaviour to them, giving an appropriate response for those emotions.Detecting and recognizing emotional information is one of the area of affective computing.

Emotions are well rooted to perception and the human neural system. Only through emotions, the communication will be effective. Emotions can be expressed and identified through speech, facial expressions, gestures etc. Since speech is the most effective medium for communication, speech emotionrecognition attains greater importance [2]. Recognizing human emotions is a very complex task in itself because of the ambiguity in classifying the acted and natural emotions. Human speech is a combination of linguistics and emotions.

Automatic Emotional Recognition (AER) finds applications in speech recognition systems, text to speech synthesis systems, forensics, medical domains and humanoid robots. Emotions are closely related topsychological, linguistics, performance of the emotionrecognition system depends on the features that are used.

Like any other recognition systems, emotion recognition systems also involve two phases namely, training and testing. Training is the process of familiarizing the system with the emotions characteristics of the speakers. Testing is the actual recognition task.The speech emotion recognition system has the emotional speech as an input and the classified emotion as an output. The system contains four main stages, preprocessing, feature extraction, feature selection and finally the classifier. Fig. 1 shows a flowchart for a typical emotional recognition system.

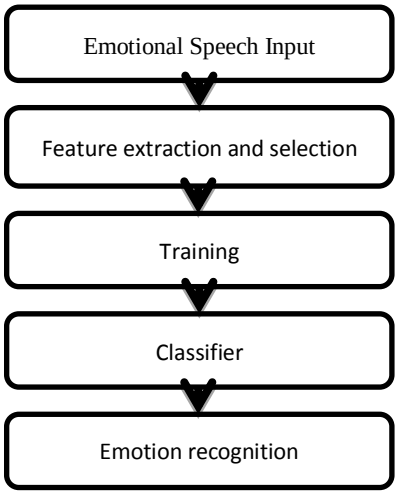

Fig.1 Block Diagram for Emotion Recognition System through Speech

The input of the system is the speech files, which first undergoes some preprocessing. The next step is to extract the main features of the input speech that will differentiate between the different motions.Then the feature selection, removal and standardization algorithms are applied to get the optimum feature vectors. The vector is then presented to the classifier in training and testing scheme. The final output is the classified emotion according to the input speech.In this paper four kinds of emotional speech segments like happy, sad, angry, and aggressive gets recognized. For this MFCC is used as a feature extraction and neural network will be the classifier.

II. MFCC FEATURE EXTRACTION

For the purpose of feature extraction, spectral analysis algorithm such as Mel-frequency Cepstral Coefficients, MFCCs will be used. For prosody analysis, the statistics of pitch and energy will be used to determine prosodic features.

MFCC is based on the human peripheral auditory system. The human perception of the frequency contents of sounds for speech signals does not follow a linear scale. Thus for each tone with an actual frequency *t* measured in Hz, a subjective pitch is measured on a scale called the Mel Scale‘. The Mel frequency scale is linear frequency spacing below 1000 Hz and logarithmic spacing above 1kHz.As a reference point, the pitch of a 1 kHz tone, 40 dB above the perceptual hearing threshold, is defined as 1000 Mels.

The extraction and selection of the best parametric representation of acoustic signals is an important task in the design of any speech recognition system; it significantly affects the recognition performance. A compact representation would be provided by a set of Mel-frequency cepstrum coefficients (MFCC), which are the results of a cosine transform of the real logarithm of the short-term energy spectrum expressed on a Mel-frequency scale. The MFCCs are proved more efficient [3] [4]. Therefore, here we are using MFCC for spectral feature extraction. The calculation of the MFCC includes the following steps.

*Step1: Frame Blocking*The process of segmenting the speech samples obtained from analog to digital conversion (ADC) into a small frame with the length within the range of 20 to 40 msec. The voice signal is divided into frames of N samples. Adjacent frames are being separated by M (M<N). Typical values used are M = 100 and N= 256.

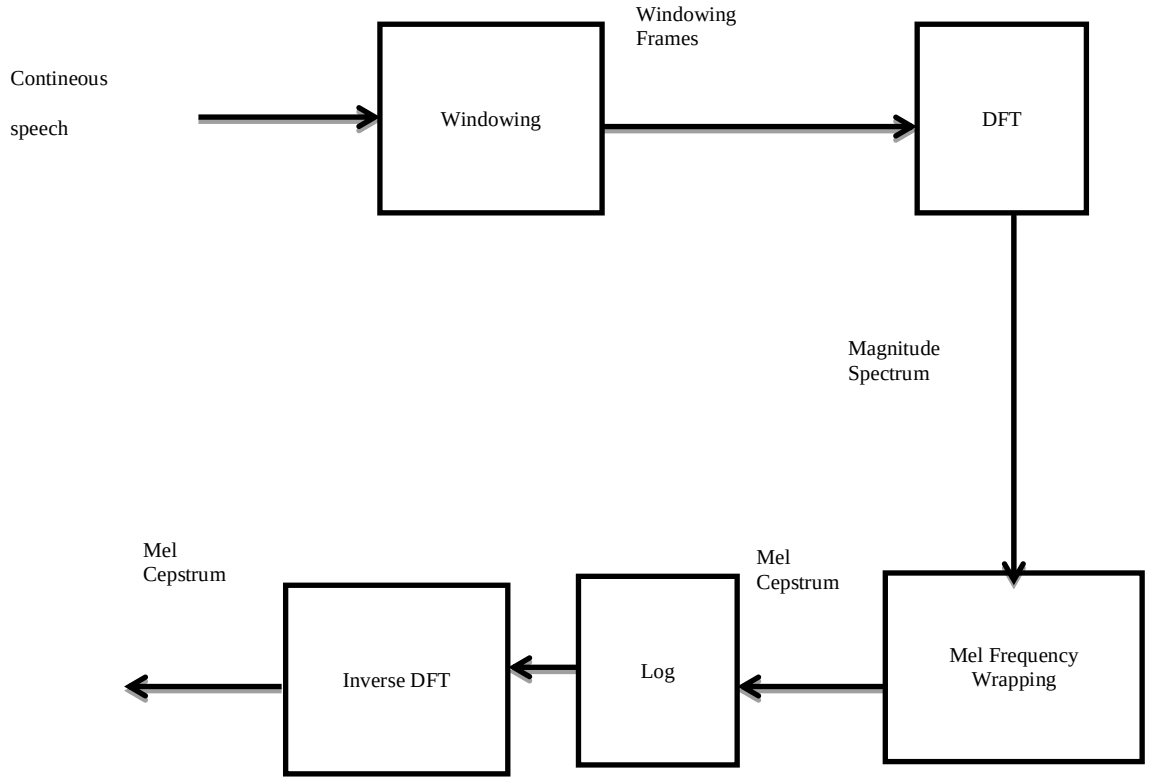

Fig.2: Complete Pipeline for MFCC

Step 2: Windowing Hamming window is used as window shape by considering the next block in feature extraction processing chain and integrates all the closest frequency lines.

If window is defined as W (n), $0 \leq n \leq N-1$ where

N = number of samples in each frame

Y[n] = output signal

X (n) = input signal

W (n) = Hamming window, then the result of windowing signal is

$$Y(n) = X(n) \times W(n) \quad (1)$$

$$W(n) = 0.54 - 0.46\ \cos\{2\pi n|N-1\}\ \ 0 \le n \le N-1\,(2)$$

Step 3: Fast Fourier Transform (FFT) To convert each frame of N samples from time domain into frequency domain. The Fourier Transform is to convert the convolution of the glottal pulse U[n] and the vocal tract impulse response H[n] in the time domain.

$$Y(w) = FFT[h(t) * X(t)] = H(w) * X(w)(3)$$

If X (w), H (w) and Y (w) are the Fourier Transform of X (t), H (t) and Y (t) respectively.

*Step 4: Mel-frequency wrapping*The frequencies range in FFT spectrum is very wide and voice signal does not follow the linear scale. Set of triangular filters that are used to compute a weighted sum of filter spectral components so that the output of process approximates to a Mel scale. Each filter‘s magnitude frequency response is triangular in shape and equal to unity at the centre frequency and decrease linearly to zero at centre frequency of two adjacent filters.Then, each filter output is the sum of its filtered spectral components.

$$F(Mel) = [2595 * log10[1 + f]700](4)$$

Above equation is used to compute the Mel for given frequency f in HZ.

Step 5: Cepstrum (Discrete Cosine Transform (DCT)) This is the process to convert the log Mel spectrum into time domain using Discrete Cosine Transform (DCT). The result of the conversion is called Mel Frequency Cepstrum Coefficient. The set of coefficient is called acoustic vectors. Therefore, each input utterance is transformed into a sequence of acoustic vector [5] [6].

III. NEURAL NETWORK

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyses. An artificial neuron is a device with many inputs and one output. The neuron has two modes of operation; the training mode and the using mode. In the training mode, the neuron can be trained to fire (or not), for particular input patterns. In the using mode, when a taught input pattern is detected at the input, its associated output becomes the current output. If the input pattern does not belong in the taught list of input patterns, the firing rule is used to determine whether to fire or not.

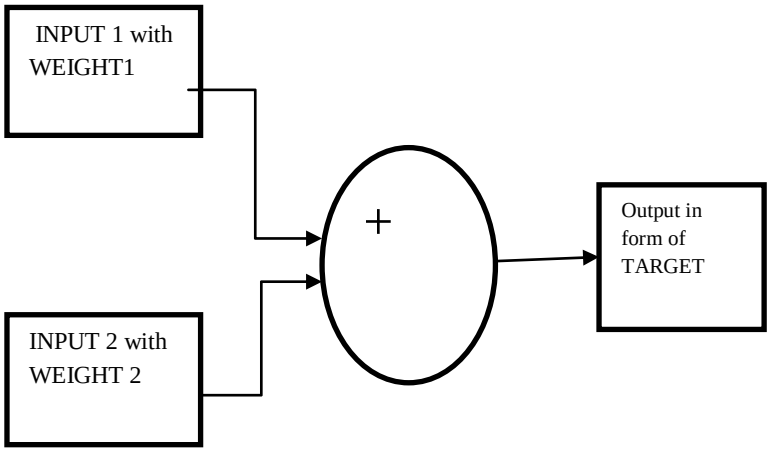

Fig3: Illustrate the general working principle of the NEURAL NETWORKS.

IV. METHDOLOGY

The methodology of this work is quite simple. Data base has to be trained for different categories. As explained above, features will be extracted for the wave files.The extracted features will be saved to the database for every category processed. The neural network works on two scenarios namely the training and the testing part. The integrated neural network of MATLAB needs to get a target set, the target set will be helpful in indentifying what exactly the files are . Then randomly a file will be uploaded to test the training scenario of the neural network. The neural network will intake the features of the uploaded file and would set the targets of the value provided earlier. It would classify the wave files according to the target set and would provide the exact result. The false predictions would be called as far.

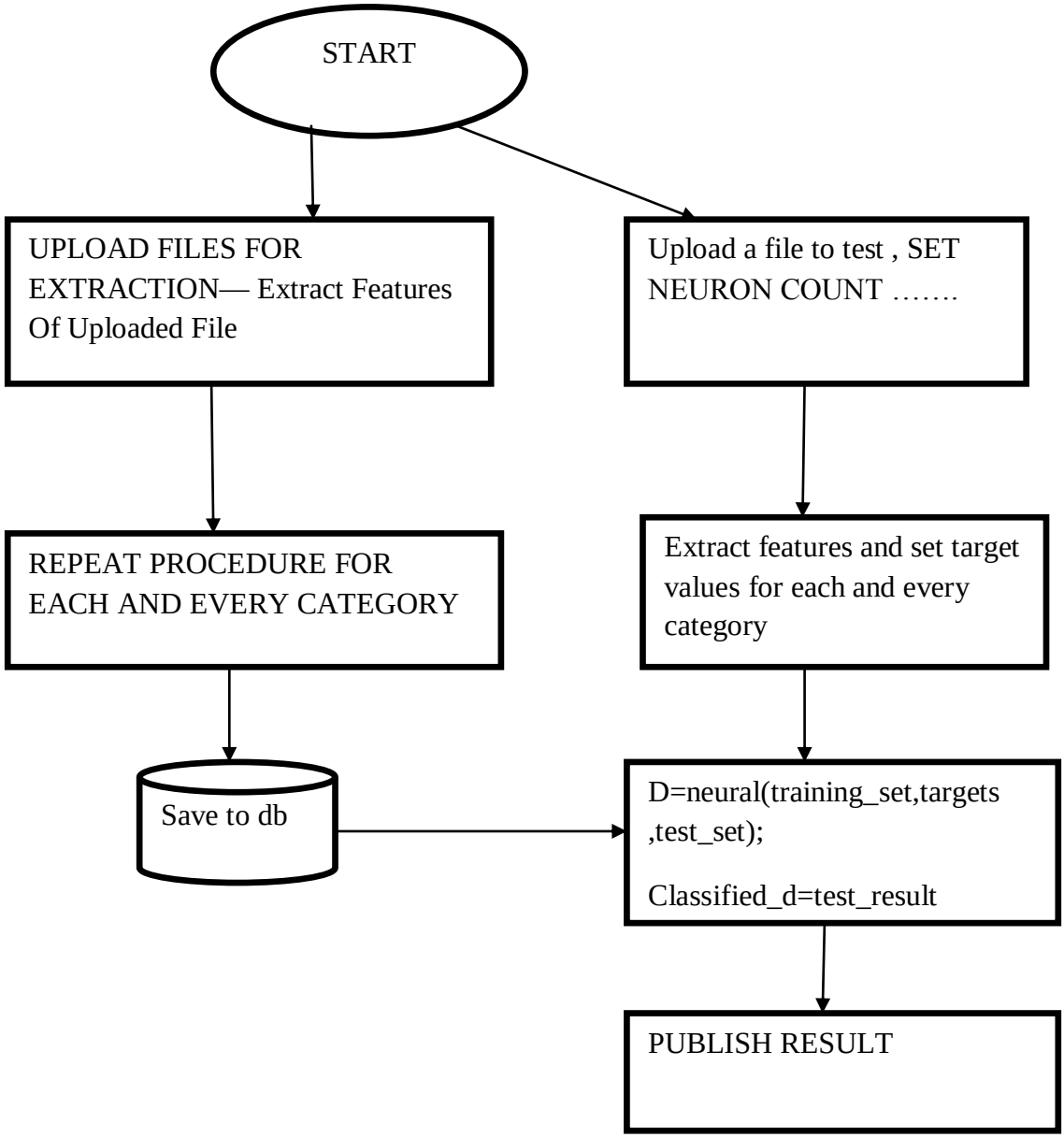

Fig4: Represents Flow Diagram of ASR

V. CONCLUSION

In this paper we have focusedspeech emotion detection. We also reviewed detailed study of feature extracting technique MFCC. MFCCs are proved to be more efficient. Furthermore we will be using different version of neural as a classifier for emotion detection and try to enhance the accuracy of the voice files of different categories.

ACKNOWLEDGEMENT

I express my sincere gratitude to my guide Mr.Abhilash Sharma, for his valuable guidance and advice. Also I would like to thank all the people who have given their heartwelling support in making this completion a magnificent experience.

REFRENCES

[1]. Tao, Jianhua; Tieniu Tan (2005). "Affective Computing: A Review". *Affective Computing and Intelligent Interaction*. LNCS 3784.Springer. pp. 981–995. doi:10.1007/11573548

[2]. Banziger and K.R. Scherer, "The Role of Intonation in Emotional Expressions," in Speech Communication, vol. 46, pp. 252-267, 2005.I.Daubechis "Orthonormal Bases of Compactly supported wavelets" Communication on pure and Applied Math.Vol.41, 909-996. 1988

[3]. S. D. Shirbahadurkar, A. P. Meshram, AshwiniKohok&SmitaJadhav, ―*An Overview and Preparation for Recognition of Emotion from Speech Signal with Multi Modal Fusion*‖ IEEE Proceedings, Vol.5., 2010.

[4]. VibhaTiwari,"MFCC and its applications in Speaker Recognition", Published in International Journal on Emerging Technology, ISSN No: 0975-8364, April 2010,page 19-23.

[5]. Kshamamayee Dash* , Debananda Padhi2 , Bhoomika Panda3, Prof. Sanghamitra Mohanty4, " Speaker Identification using Mel Frequency Cepstral Coefficient and BPNN ", Published in IJARCSSE,ISSN No. 2277 128X,Volume 2, Issue 4, April 2012

[6] K. J. Patil P. H. Zope S. R. Suralkar," **Emotion Detection From Speech Using Mfcc&Gmm**", Published in IJERT, ISSN No. 2278-0181, Vol. 1 issue 9, November- 2012.

[7]. FirozShah.A, RajiSukumar.A, BabuAnto.P, "Automatic Emotion Recognition from Speech Using Artificial Neural Networks With Gender Dependent Databases", Published in IEEE, Print ISBN No: 978-0-7695-395-7/09/$26.00, on 2009.

[8]. BjörnSchuller, Gerhard Rigoll, and Manfred Lang, "Hidden Markov Model- Based Speech Emotion Recognition" Published in IEEE, Print ISBN No. 0-7803-7663-3/03/$17.00 ©2003 ,pp 401-404.

[9]. WiqasGhai, S. Navdeep, "Analysis of Automatic Speech Recognition Systems for Indo-Aryan Languages: Punjabi A Case Study", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-1, March 2012.

[10]. Jian Wang, Zhiyan Han, and ShuxianLun, "Speech Emotion Recognition System Based on Genetic Algoithm and Neural Network", Published in IEEE, Print ISBN No: 978-1-61284-881-5/11/$26.00, on 2011

[11]. EliathambyAmbikairajah ," Emerging Features for Speaker Recognition", 1-4244-0983-7/07/$25.00 ©2007 IEEE ICICS 2007.

[12]. Dr. Joseph Picone, "FUNDAMETALS OF SPEECH RECOGNITION: A Short Course", Institute for Signal And Information Processing.

[13]. MohitDua, R.K.Aggarwal, VirenderKadyan and ShelzaDua, "Punjabi Automatic Speech Recognition Using HTK",IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 1, July 2012 ISSN (Online): 1694-0814.

[14]. Richard P. Lippmann, "Speech recognition by machines and humans", 0167-6393r97r$17.00 q 1997 Elsevier Science B.V. All rights reserved. II S0167-6393_ 97. 00021-6.

[15]. 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.

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

[17]. 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.

[18]. S.Bhupinder, S. Parminder, "Voice Based user Machine Interface for Punjabi using Hidden Markov Model,"JCST Vol. 2, Issue 3, September 2011 I S S N : 2 2 2 9 - 4 3 3 3 (P r i n t) | I S S N : 0 9 7 6 - 8 4 9 1.

[19]. Rahul.B.Lanjewar, D.S.Chaudhari, "Speech Emotion Recognition:A Review" International Journal of Innovative Technology and Exploring Engineering, ISSN:2278-3075, Vol.2,Issue-4,March 2013.

[20] Santosh K.Gaikwad, Bharti W.Gawali and PravinYannawar, "A Review on Speech Recognition Technique," International Journal of Computer Applications (0975 – 8887) Volume 10– No.3, November 2010.

[21] W. M. Campbell, D. E. Sturim W. Shen D. A. Reynolds and J. Navratily, "The MIT- LL/IBM Speaker recognition System using High performance reduced Complexity recognition", MIT Lincoln Laboratory IBM 2006

[22] N. Mikael, E. Marcus, "Speech Recognition using Hidden Markov Model, Performance evaluation in noisy environment", Degree of master of science in Electrical Engineering, Department of Telecomminications and Engineering, Blekinge Institute of Technology, March 2002.

[23] T. Nagarajan and H. A. Murthy, "Subband-Based Group Delay Segmentation of Spontaneous Speech into Syllable-Like Units," in Eurasip Journal on Applied Signal Processing , Hindawi Publishing Corporation 2004:17, pp. 2614–2625.

