Speech Emotion Recognition Based on MFCC Using MATLAB

Saumya Bhautmange[1], Prof. Amit Kolhe[2]

M.Tech. Scholar, Digital Electronics, Department of ETC, Rungta College of Engineering & Technology, Bhilai, Chhattisgarh, India[1]

Assistant Professor, Digital Electronics, Department of ETC, Rungta College of Engineering & Technology, Bhilai, Chhattisgarh, India[2]

ABSTRACT: In this paper methodology for emotion recognition by speech signal is presented. Here some of acoustic features are extracted from speech signal to analyze the characteristics and behavior of speech. The system is used to recognize basic emotions like Anger, Happiness, Sadness and Neutral. It can serve as basis for further designing an application for human like interaction with machines through natural language processing and improving efficiency of emotion. In this formant, energy, Mel Frequency Cepstral Coefficients (MFCC) has been used for feature extraction from speech signal. Energy classifications are used for recognition of emotional states. Multilingual datasets are used for analysis of emotions. Using this analysis energy is trained and designed for detecting emotions in real time speech.

KEYWORDS: Human Computer Interaction, Mel Frequency Cepstral Coefficients, Speech signal, Emotion recognition.

I. INTRODUCTION

Emotion Recognition is a recent research topic in field of Human Computer Interaction Intelligence and mostly used to develop wide range of applications such as stress management for call centre employee and learning & gaming software, In E-learning field, identifying students emotion timely and making appropriate treatment can enhance the quality of teaching. Main aim of HCI is to achieve a more natural interaction between machine & humans. HCI is an emerging field using which we can improve interactions between users and computers by making computers more respond able to the user’s needs. Today’s HCI system has been developed to identify who is speaking or what he or she is speaking. If in the HCI system the computers are given an ability to detect human emotions then they can know how he or she is speaking and can respond accurately and naturally like humans do. The goal of Affective computing is to recognize emotions like Anger, Happiness, Sadness and Neutral from speech. Automatic emotion recognition and classification on the voice signals can be done using different approaches like from text, voice and from human face expressions andgestures.

During present scenario for the human emotion recognition an extensive research is made by using different speech information and signal. Many researchers used different classifiers for human emotion recognition from speech such as Hidden Markov Model (HMM), Neural Network (NN), Maximum likelihood bayes classifier (MLBC), Gaussian Mixture Model (GMM), Kernel deterioration and K-nearest Neighbours approach (KNN), support vector machine (SVM) and Naive Bayes classifier.

In proposed system basic features of speech signals like formant, Energy, and MFCC are extracted from both offline and real time speech and they are classified into different emotional classes by Energy classification. Here energy classification is used since it has a better classification performance than other classifiers. Energy classification is a supervised learning algorithm which addresses general problem of the learning to discriminate between positive and negative members of given n-dimensional vectors. The Energy classification can be used for both classification &regression purposes. Using Energy classification can be done linearly or nonlinearly. Here kernel functions of Energy classification are used to recognize emotions with more accuracy. In human-machine interaction, the emotion recognition &classification ability is very useful. It is useful for various types of communication system such as automatic answering system, dialogue system and human like robot which can apply emotion recognition and classification techniques so that a user feels like the system as a human.

II. LITERATURE REVIEW

Applications of emotion classification based on speech have already been used to facilitate interactions indaily lives. For example, in call centers apply emotion classification to prioritize impatient customers. As another example warning system has been developed to detect if a driver exhibits anger or aggressive emotions. Emotion sensing has also been used in behavior studies acoustic features have been extensively explored in both the time domain (energy, speaking rate, duration of voiced segments, zero crossing rate, etc.) and frequency domain (pitch, formant, Mel-frequency cepstral coefficients, etc.). In our work we only choose most basic features: energy, formants, and MFCC. This reduces the computational complexity of approach and can lead to both energy and bandwidth savings when the voice is captured on mobile devices. Commonly used a classifiers for human emotion recognition from speech such as Hidden Markov Model (HMM), Neural Network (NN), Maximum likelihood bayes classifier (MLBC), Kernel deterioration and K-nearest Neighbors approach (KNN), support vector machine (SVM), Naive Bayes classifier, Gaussian Mixture Model (GMM). We choose Energy classificationas our basic classifier because of its ease of training and its ability to work with any number of attributes.

In SVM kernel functions are used to map data to a higher dimensional feature space without losing the originality. This conventional method of using kernel functions in SVM is to run simulations on training sets and find kernel function which attains the highest averaged classification accuracy for the given problem. The most commonly used kernel function for SVM is Linear, Polynomial, radial basis function (RBF). The contributions of the Speech emotion recognition are as follows: 1) To obtain the maximum efficiency using the performance of SVM kernel method for each individual technique 2) Consideration of cut-off value in each technique so classification having better confidence level is selected and those with lesser confidence value are discarded as ‘not classified’. We have used multilingualspeechdatabase in this approach of emotion recognition and classification. The accuracy of emotion recognition can be made better by increasing the value of min confidence cut-offvalue.

III. ACOUSTIC FEATUREEVALUATION

Speech: Primary means of communication between humans is speech. It is a complex signal which contains information about message, speaker, language, emotional state and so on.

Emotions: Emotions are defined as changes in physical &psychological feeling which influences behavior and thought of humans. It is associated with temperament, personality, mood, motivation, energy etc.

Emotional Speech Databases: In evaluation of Emotion recognizer from speech main task is to check quality, naturalness and noise level of the database used in performance and efficient result estimation. When we can use lower quality database for emotion recognition then there can be possibility of incorrect conclusion and result. Task of Classification also include detecting stress of speech and it also define the type of emotion included in the database like angry, surprised, fear, happy, disgust, sad and neutral. Databases can be different types as under.

1) As Database we can consider speech samples recorded by speaking with pre-definedemotionfrom speaker.
2) We can obtain Database from real life system like call centre, learning andgamingsoftware.
3) We can also include Database with self-explanatorysentiments.

In this paper multilingualemotional speech Database in which voice samples is recorded by female and male speakers in five types of sentimental moods. Subsequently determine different audio parameter like MFCC, Formant, Energy features and stored these features vectors in database which we use for emotion recognition from speech.

Audio Feature Extraction: Speech signal contains various type of parameters from which the properties of speech are defined. Speech features generally does not very much easy to understand because of the changing behavior and temporal adjustments make this task very tedious. In this Paper MFCC Formant and Energy features are used. Usually the speech signal is recorded with a sample rate of 16000 Hz by microphone. The steps for calculating MFCC are shown below.

IV. EXTRACTION OF MEL-FREQUENCY CEPSTRUM COEFFICIENTS(MFCC)

In speech recognition, Mel Frequency Cepstral Coefficients are most widely used feature. The main purpose of using MFCC is to mimic the behaviorofthe human ears. The block diagram for MFCC is shown in Fig. 1.

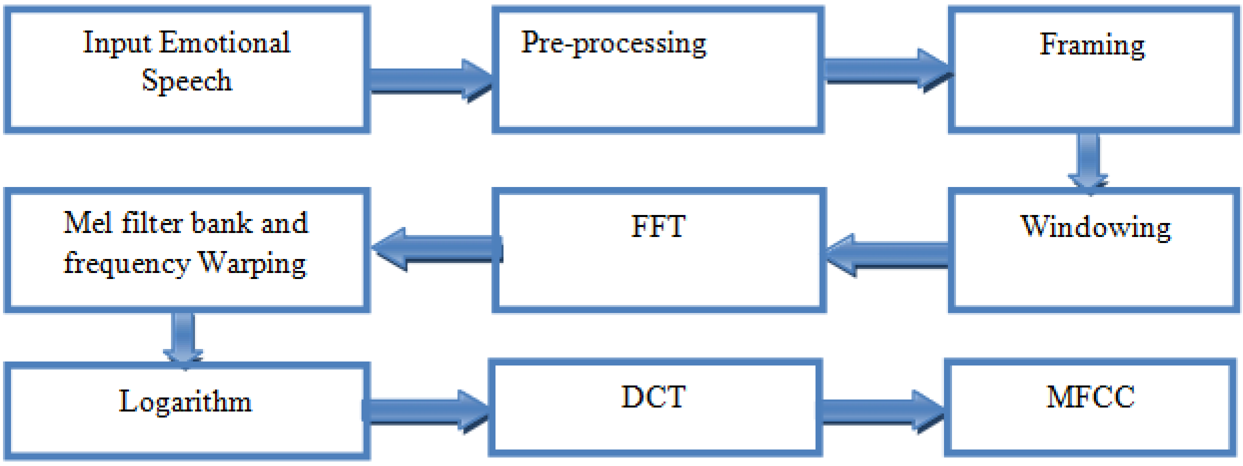

Fig-1 Block Diagram for MFCC feature extraction

Framing: In Framing the continuous input speech is segmented into N sample per frames. The first frame consists of N samples, second frame consists of M samples after N and third frame contains 2M and so on. Here we frame the signal with time length of 20-40ms. So the frame length of 16 KHz signal will have0.025*16000=400samples.

Windowing: Windowing is used to window each individual frame in order to remove discontinuities at the start and end of the frame. Hamming window is mostly used due to its relatively narrow main lobe width hence remove distortion.

Fast Fourier Transform: FFT algorithm is used for converting N samples from time domain to frequency domain. It is used to evaluate he frequency spectrum ofspeech.

Mel Filter Bank: In mapping of each frequency from frequency spectrum to Mel scale is performed. The Mel filter bank will usually consist of overlapping triangular filters with cut off the frequencies which is determined by center frequency of two filters. The Mel filters are graphically is shown in Fig.2.

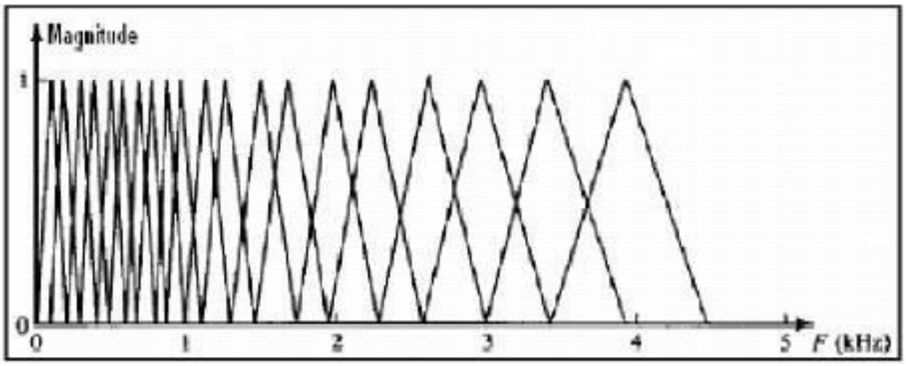

Fig-2 Mel filter bank with overlapping filters

Cepstrum: The obtained Mel spectrum is converted back to time domain with help of DCT algorithm.

V. SPEECH EMOTION CLASSIFICATION USING ENERGYCLASSIFIER

The energy classifier is high dimensional vector supervised learning method that is based on emotion assumptions. It predicts that presence (or absence) of a specified feature of a class is not related to the presence (or absence) of all other features. It is very simple to program and execute it. Its parameters are simple to assume, even on very large databases learning or training is very fast and effective and its accuracy is comparatively better in comparison to the other techniques. The emotion recognition process along with training and testing phases is shown in the Fig 3.

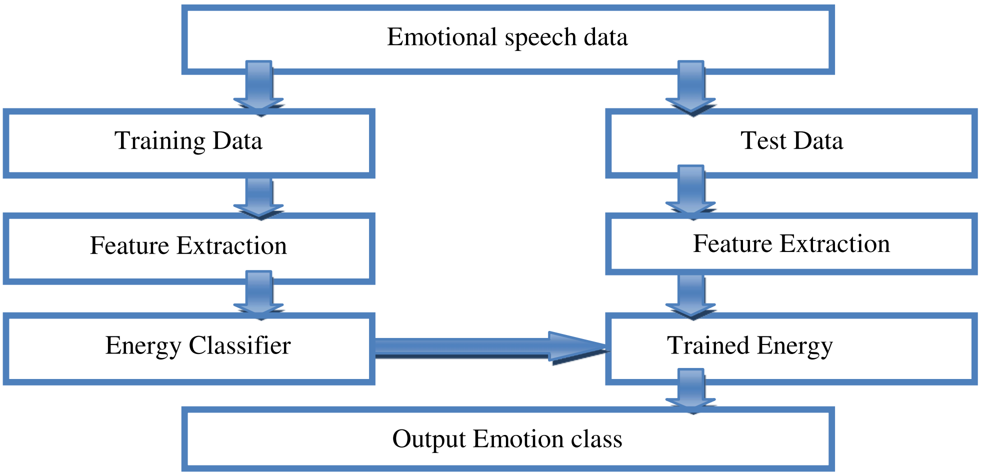

Fig-3 Emotion classification process model

VI. RESULT AND DISCUSSION

In this section result of Mel frequency cepstrum coefficient is obtained which is shown in fig-3. Here we have considered 22 filters in triangular filter bank and 13 MFCC values as shown in fig-3.The performance of Mel-frequency Cepstrum coefficients is affected by the number of filters and type of window used. In this paper we have shown result for applied emotional speech signal and pre-emphasized signal as shown in Fig-4 and we have used hamming window as shown in fig-4.

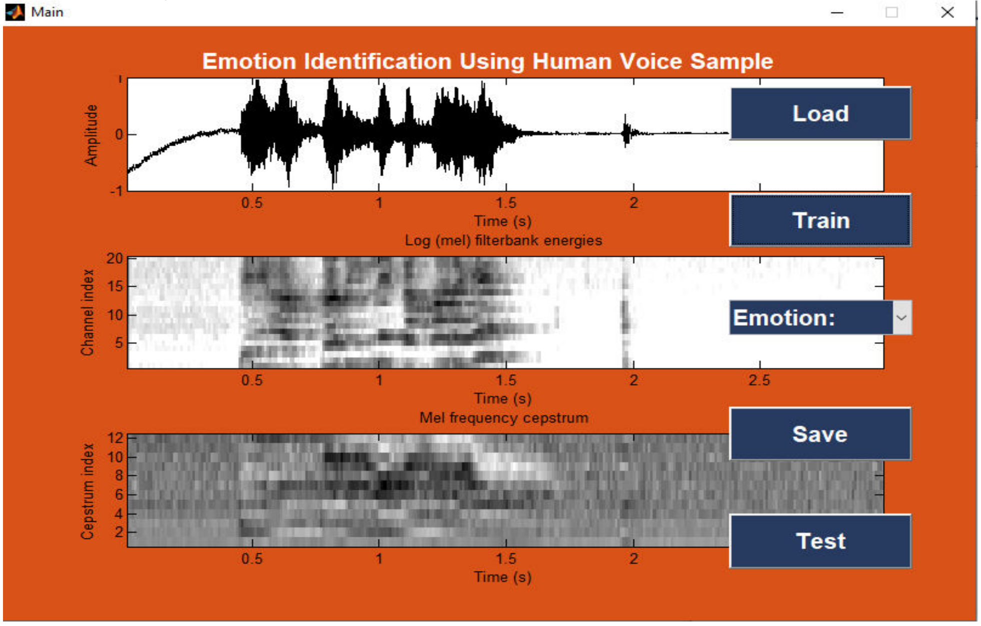

Fig-4 Emotional speech signal and pre-emphasized signal.Obtained MFCC Coefficients.

VII. CONCLUSION AND FUTUREWORK

In this paper most recent work done in field of Speech Emotion Recognition and most used methods of the feature extraction and several classifier performances are reviewed. In this paper we discussed about MFCC which is well known techniques used in speech recognition to describe signal characteristics. MFCC reduce the frequency information of speech signal into small number of coefficients which is easy and fast to compute. Success of emotion recognition is dependent on appropriate feature extraction as well as the proper classifier selection from the sample emotional speech. In Future work It is needed to work on Emotion classification process model with SVM using different kernel functions so that it can provide better emotion recognition ofreal time speech and use our system in different application such as stress management for call center employee and learning & gaming software.InE-learning field etc. which makes our life more effective.

REFERENCES

1. Jeet Kumar, Om PrakashPrabhakar, Navneet Kumar Sahu, “Comparative Analysis of Different Feature Extraction and Classifier Techniques for Speaker Identification Systems: A Review”, International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE),Vol. 2, Issue 1, pg- 2760-2769, January2014.
2. Liqin Fu, Xia Mao, Lijiang Chen “Speaker Independent Emotion Recognition Based on SVM/HMMs Fusion System” IEEE International Conference on Audio, Language and Image Processing(ICALIP), pages 61-65, 7-9 July2008.
3. PeipeiShen, Zhou Changjun, Xiong Chen,“ Automatic Speech Emotion Recognition Using Support Vector Machine” IEEE International ConferenceonElectronicandMechanicalEngineeringandInformationTechnology(EMEIT)volume2 , Page(s):621-625,12-14Aug.2011.
4. Akalpita Das, PurnenduAcharjee ,Laba Kr. Thakuria , “ A brief study on speech emotion recognition” , International Journal of Scientific & Engineering Research(IJSER), Volume 5, Issue 1,pg-339-343,January-2014.
5. Kshamamayee Dash, DebanandaPadhi , Bhoomika Panda, Prof. SanghamitraMohanty, “ Speaker Identification using Mel Frequency Cepstral Coefficient and BPNN”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 4, pg.- 326-332, April2012.
6. Vinay, Shilpi Gupta, AnuMehra,“Gender Specific Emotion Recognition Through Speech Signals”, IEEE International Conference on Signal Processing and Integrated Networks (SPIN), 2014 , Page(s):727 – 733, 20-21 Feb.2014.
7. Norhaslinda Kamaruddin, Abdul wahab Rahman,Nor Sakinah Abdullah, “Speech emotion identification analysis based on different spectral feature extraction methods”, IEEE Information and Communication Technology for The Muslim World, 2014 The 5th International Conference, Pages:1-5,2014.
8. A. D. Dileep, C. Chandra Sekhar, “GMM Based Intermediate Matching Kernel for Classification of Varying Length Patterns of Long Duration Speech Using Support Vector Machines”, IEEE Transactions on Neural Networks and Learning Systems, Volume: 25, Issue: 8,Pages: 1421 - 1432,2014.
9. S.Lalitha, AbhishekMadhavan, BharathBhushan, SrinivasSaketh “Speech Emotion Recognition” IEEE International Conference on Advances in Electronics, Computers and Communications (ICAECC), Page(s): 1-4, 2014.
10. S.Sravan Kumar, T.RangaBabu , Emotion and Gender Recognition of Speech Signals Using SVM, International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 4, Issue 3, pg.- 128-137 May2015.