ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

Speech Emotion Recognition Based on SVM Using MATLAB

Ritu D.Shah[1], Dr. Anil.C.Suthar[2]

M.E Student, Dept. of Communication System Engineering, L.J. Institute of Engineering & Technology, Himmatnagar, India[1]

Director, L.J. Institute of Engineering & Technology, Ahmedabad, India[2]

ABSTRACT : In this paper methodology for emotion recognition from 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 the basic emotions: Anger, Happiness, Sadness and Neutral. It can serve as a basis for further designing an application for human like interaction with machines through natural language processing and improving the efficiency of emotion. In this, formant, energy, Mel Frequency Cepstral Coefficients (MFCC) has been used for feature extraction from the speech signal. Support Vector Machine (SVM) are used for recognition of emotional states. English datasets are used for analysis of emotions with SVM Kernel functions. Using this analysis the machine is trained and designed for detecting emotions in real time speech.

KEYWORDS : Support vector Machine, Mel Frequency Cepstral Coefficients, Speech signal, Emotion recognition.

I. INTRODUCTION

Emotion Recognition is a recent research topic in the 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 and humans. HCI is an emerging field using which we can improve the 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/she is speaking. If in the HCI system, the computers are given an ability to detect human emotions then they can know how he/she is speaking and can respond accurately and naturally like humans do. The goal of Affective computing is to recognize the emotions like Anger, Happiness, Sadness and Neutral from speech. Automatic emotion recognition and classification on voice signals can be done using different approaches like from text, voice and from human face expressions and gestures.

During present scenario, for human emotion recognition an extensive research is made by using different speech information and signal [1]. Many researchers used different classifiers for human emotion recognition from speech such as Hidden Markov Model (HMM)[2], Neural Network (NN), Maximum likelihood bayes classifier (MLBC), Gaussian Mixture Model (GMM), Kernel deterioration and K-nearest Neighbours approach (KNN), support vector machine (SVM)[2] [3], Naive Bayes classifier.

In proposed system, basic features of speech signals like formant, Energy, and MFCC[4][5] are extracted from both offline and real time speech and they are classified into different emotional classes by using SVM classifier. Here, SVM is used since it has better classification performance than other classifiers. SVM is a supervised learning algorithm which addresses general problem of learning to discriminate between positive and negative members of given n-dimensional vectors. The SVM can be used for both classification and regression purposes. Using SVM the classification can be done linearly or nonlinearly. Here the kernel functions of SVM are used to recognize emotions with more accuracy. In human-machine interaction, The emotion recognition and 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 the emotion recognition and classification techniques so that a user feels like the system as a human.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

II. RELATED WORK

Applications of emotion classification based on speech have already been used to facilitate interactions in our daily lives. For example, In call centres apply emotion classification to prioritize impatient customers. As another example, a warning system has been developed to detect if a driver exhibits anger or aggressive emotions. Emotion sensing has also been used in behaviour studies acoustic features have been extensively explored in both the time domain (energy, speaking rate, duration of voiced segments, zero crossing rate, etc.) and the frequency domain (pitch, formant, Mel-frequency cepstral coefficients, etc.). In our work, we only choose the most basic features: energy, formants, and MFCC. This reduces the computational complexity of the approach and can lead to both energy and bandwidth savings when the voice is captured on mobile devices. Commonly used 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 Neighbours approach (KNN), support vector machine (SVM), Naive Bayes classifier[6], Gaussian Mixture Model (GMM)[8]. We choose SVM as 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 the 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 a cut-off value in each technique so the classification having better confidence level is selected and those with lesser confidence value are discarded as 'not classified'. We have used English speech database in this approach of emotion recognition and classification. The accuracy of emotion recognition can be made better by increasing the value of minimum confidence cut-off value.

III. ACOUSTIC FEATURE EVALUATION

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

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

Emotional Speech Databases: In evaluation of Emotion recognizer from speech the Main task is to check quality, naturalness and noise level of the database used in performance and efficient result estimation. When we use lower quality database for emotion recognition then there can be possibility of incorrect conclusion and result. Task of Classification also include detecting the 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 defined emotion from actor.
2) We can obtain Database from real life system like call centre, learning & gaming software.
3) We can also include Database with self explanatory sentiments.

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

Audio Feature Extraction: The 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 their changing behaviour 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 through microphone. The steps for calculating MFCC are described below.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

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

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

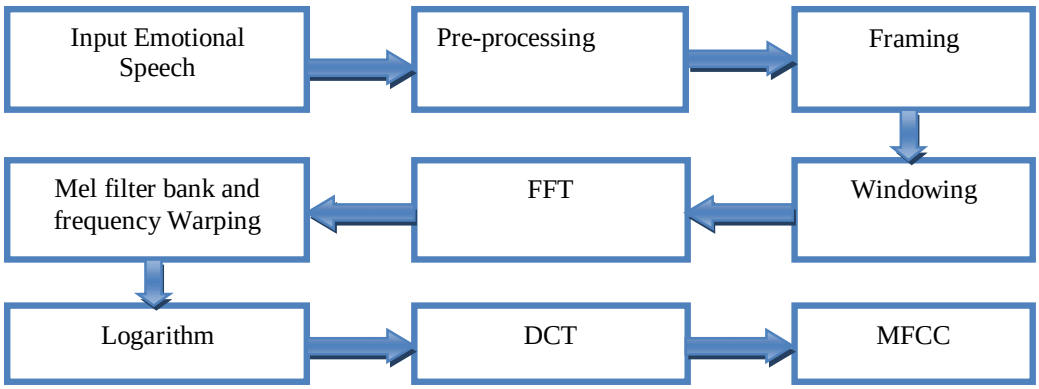


Fig-1 Block Diagram for MFCC feature extraction

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 have 0.025*16000=400samples.

Windowing: Windowing is used to window each individual frame in order to remove the 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 the N samples from time domain to frequency domain. It is used to evaluate frequency spectrum of speech.

Mel Filter Bank: In This step 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 frequencies which is determined by centre frequency of two filters. The Mel filters are graphically shown in Fig. 2.

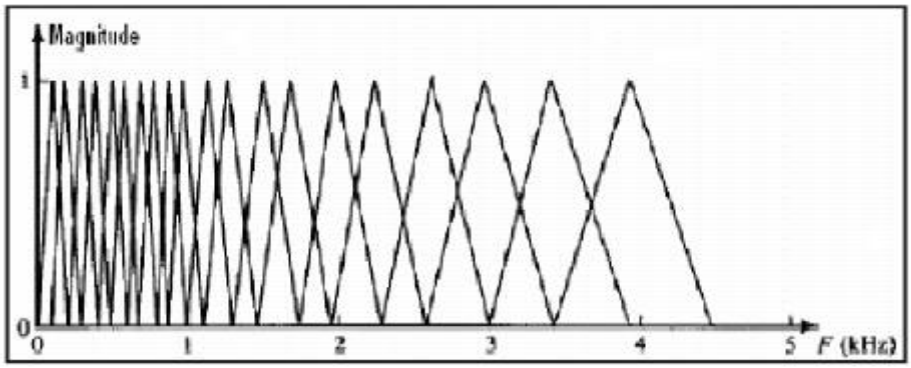


Fig-2 Mel filter bank with overlapping filters

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

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

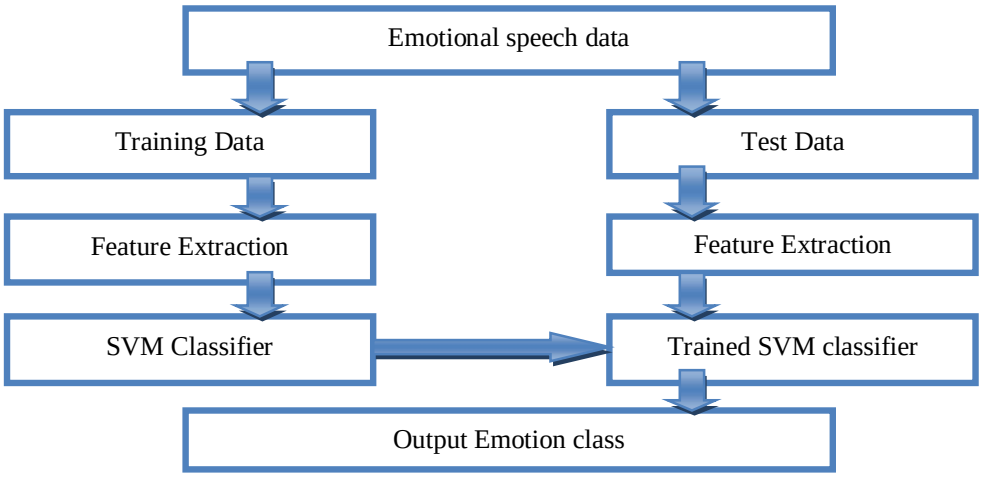
The definition of the represented of frequency in Hz to frequency in Mel scale is explained in (1) and vice versa in (2).

$$F_{Mel} = 2595 * \log_{10}(1+f_{Hz}/700) \quad \text{.....(1)}$$

$$f_{Hz} = 700 * (10^{F_{Mel}/2595} - 1) \quad \text{.....(2)}$$

V. SPEECH EMOTION CLASSIFICATION USING SVM CLASSIFIER

The SVM is a high dimensional vector supervised learning method that is based on emotion assumptions. It predicts that the 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 Fig 3.



VI. RESULT OBTAINED FOR MFCC

In this section result of Mel frequency cepstrum coefficient is obtained which is shown in fig-6. Here we have considered 22 filters in triangular filter bank and 13 MFCC values as shown in fig-6.The performance of the 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-5.

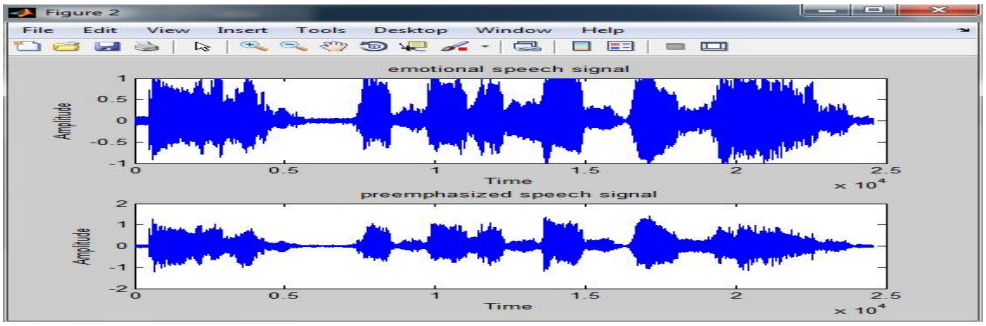


Fig-4 Emotional speech signal and pre-emphasized signal

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

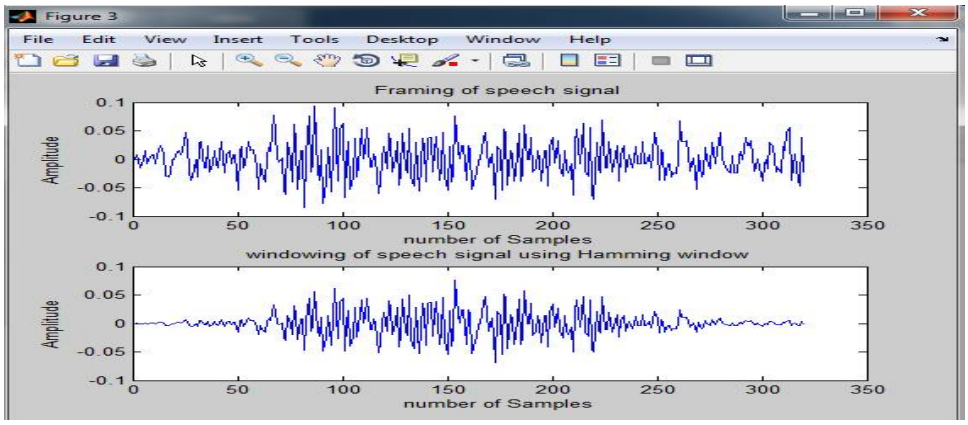


Fig-5 Framing and windowing of speech signal

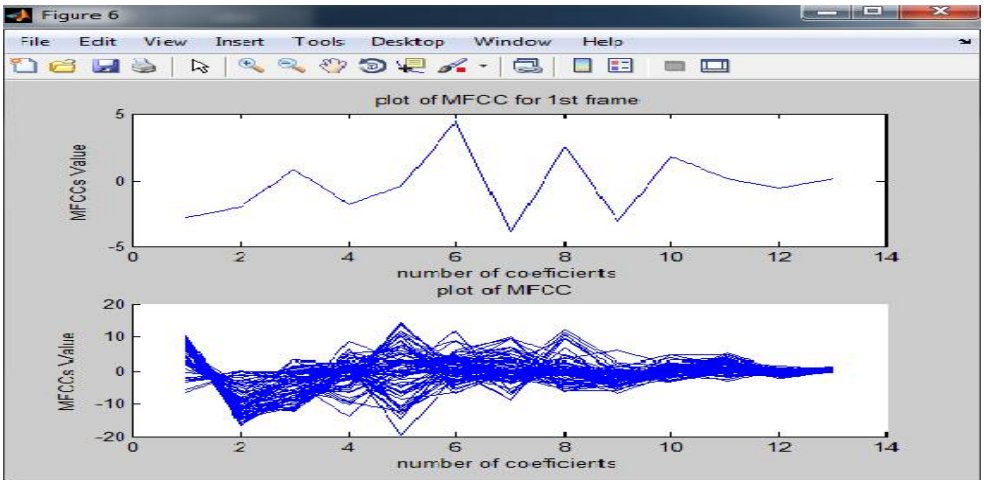


Fig-6 Obtained MFCC Coefficients for 1st frame and for all frames

VII. CONCLUSION AND FUTURE WORK

In this paper, most recent work done in the field of Speech Emotion Recognition and Most used methods of 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 the signal characteristics. MFCC reduce the frequency information of the 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 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 of real time speech and use our system in different application such as stress management for call centre employee, and learning & gaming software, In E-learning field etc. which makes our life more effective.

REFERENCES

1. Jeet Kumar, Om Prakash Prabhakar, 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, January 2014.
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 July 2008.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

3. Peipei Shen, Zhou Changjun, Xiong Chen," Automatic Speech Emotion Recognition Using Support Vector Machine" IEEE International Conference on Electronic and Mechanical Engineering and Information Technology (EMEIT) volume2 , Page(s) : 621 - 625 , 12-14 Aug. 2011.
4. Akalpita Das, Purnendu Acharjee , 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, Debananda Padhi , Bhoomika Panda, Prof. Sanghamitra Mohanty, " 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, April 2012.
6. Vinay, Shilpi Gupta, Anu Mehra,"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, Abhishek Madhavan, Bharath Bhushan, Srinivas Saketh "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 May 2015.

BIOGRAPHY

Ritu.D.shah received her B.E. degree in Electronics & Communication Engineering from Gujarat technological University. She is currently pursuing her M.E degree in Communication System Engineering from Gujarat technological University at L.J.Institute of Engineering & Technology, Himmatnagar, India.

Dr. Anil.C.Suthar is currently Director at L.J.Institute of Engineering & Technology, Ahmedabad, India. He is Ph.D. in Electronics & communication. He has completed ME in Communication systems from L.D.College of Engineering, Ahmedabad, in year 2006. He has published 12 articles in Electronics for you and EM media Pvt. limited, most popular electronics magazine in Asia. He has guided several projects at graduate and post graduate level.