ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 9(39), DOI: 10.17485/ijst/2016/v9i39/95585, October 2016* ISSN (Online) : 0974-5645

Selection of Features for Emotion Recognition from Speech

Puja Ramesh Chaudhari and John Sahaya Rani Alex*

School of Electronics Engineering, VIT University, Chennai - 632014, Tamil Nadu, India; chaudharipuja7@gmail.com, jsranialex@vit.ac.in.

Abstract

Background/Objective: Speech is one of the modes for Human Computer Interface (HCI). Speech contains message to convey as well as the speaker characteristics such as speaker identity and emotional state of the speaker. Recently, researchers are taking more interest in the emotional parameters of speech signals which helps to improve the functionality of HCI. This research focus on selecting features which helps to identify the emotion of the speaker. **Methods/Statistical Analysis:** Mel Frequency Cepstrum Coefficient (MFCC), Linear Prediction Cepstrum Coefficient (LPCC) and Perceptual Linear Predictive (PLP) methods are used to extract the features. Each emotion is modeled as one Hidden Markov Model (HMM) using Hidden Markov Tool Kit (HTK tool kit). The Beagle Bone Black (BBB) board is chosen for the implementation because of the form factor. **Findings:** The results indicate that MFCC features gives 100% accuracy for surprise emotion, PLP features gives 100% accuracy for anger emotion and LPCC features give 100% accuracy for fear emotion. **Conclusion/Improvement:** A hybrid feature extraction method should be devised to detect all emotions with 100% accuracy.

Keywords: BBB, Emotion recognition, HCI, HMM, LPCC, MFCC

1. Introduction

The speech is a unique parameter for representing the information. Each spoken word is created using the phonetic combination of a set of vowel, semivowel and consonant speech sound units. The emotion recognition plays an important role in identifying and verifying the emotional state from his/her speech signal.Emotion includes six basic parameters such as happy, sad, fear, anger, surprise and natural. The emotion recognition system is used for analyzing driver emotion state while driving the car in the city or in the automated customer care. This is very useful in terms of safety and controlling the state of mind. The development in speech analysis has largely defeat the milestone of intelligibility, dynamic research attempts in the area of ingenuousness and blandness. The emotion contains the different parameters such as pitch, energy, intensity of voice and word utterance. Basically emotion recognizing is complex for the long and complicated sentences of human. Also, it is recognizing the emotion from speech and voice intonation. Emotion recognition used for different applications in the field of Medicine, E-learning, Monitoring, Entertainment, Law, Marketing such as in Health centres for monitoring the patients emotional state after the treatment and also used for application demand natural man machine interaction, such as web movies and computer tutorial applications, where the responses to the user depends on the detected emotions.

Various literatures have been referred for this work which was helpful for the progress. The stress of person measured by asking and collecting different parameters to get particular marked values and also for trained model contained the amount stress signal which was measured by neural network[1].

In this paper[2], technologies related to emotion recognition are discussed which through threw more light on the project as it gave an idea regarding emotion

* *Author for correspondence*

recognition from speech more clearly. It is also found that emotion recognition of the speech signal helps to ensure naturalness in the performance of existing speech systems and recent works with the ideas of the emotion recognition from emotional databases, speech features and classification models. This paper[3] presents how to develop classification scheme and an emotional speech dataset for system functionality. Emotional interaction of a Thinking Robot[4] is identified, focussing on emotion recognition from speech signals and focuses on the independent emotion recognition systems. Speech emotion recognition system that combines with facial features and gestures included in multimodal interactions. This paper[5] discussed about the formant frequency, pitch methods which add to recognize the happy, sad, natural emotions.In this paper[6], researcher implemented the system for the speech recognition in Xilinx based on the Neutral Network (NN) for hybrid mechanism which is beneficial for the decreasing area and power. In this paper[7], the Self-Organising Feature Map (SOFM) isused to reduced large dimensions of features vector with same recognition accuracy. Low power small embedded board is not used in literature to implement emotion recognition system. The objective of this paper is to use the conventional feature extraction methods such asmel frequency cepstrum coefficient (MFCC)[8,9], Linear Prediction Coefficient (LPC) and Perceptual Linear Predictive (PLP) emotion recognition. Each emotion modeled using Hidden Markov Model (HMM)[10,11] and then each emotion detected using Viterbi decoder. The proposed emotion recognition system is implemented on an embedded board. Beagle Bone Black[12] is chosen as the embedded platform to implement because of its low power and small size. The rest of the paper is organized as follows; Section 2 discusses the FCC, LPC, PLP, HMM modeling of emotion: Section 3 gives the detail description of implementation and Result obtained: Section 4 presents the conclusion.

2. Implementation

2.1 Data based Collection

Emotion recognition system start with collecting the information from various speech signals for 10 samples for each emotions and recording by using the head mounted microphone for the various sampling rates such as 16 kHz, 44.1 kHz and 48 kHz as a **.wav** format.

2.2 Feature Extraction

Certain attributes of the speaker is extracted from the speech signals for each emotions. It is basically represented the each emotions by extracting the small portion of data from the voice. Feature extraction for emotion must have some specific characteristics: It should be easily determinable from the set of known voices generated naturally and frequently in speech. It should be consistent for each emotion models. Speech is the slowly time varying signal. When identified the voice signal over the short time period, then voice signal establish as an unmoving. However, when identified the voice signal symptomatic being change on long duration of time. It is showed the distinct voice vocalization existences in spoken words. Thus, small interval of spectral analysis is same manner to differentiate voice parameters.

Many algorithms are used to extract emotional parametric representation. They are Linear Predictive Coding (LPC), Perceptual Linear Predictive (PLP) and MFCC.

2.2.1 Mel Frequency Cepstral Coefficient (MFCC)

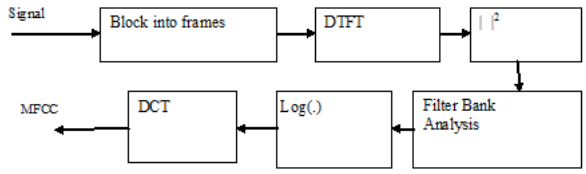

Figure 1. MFCC feature extraction method.

MFCC takes human conception sensitivity with respect to frequencies into account which is good for emotion recognition. The MFCC basically performing the Fourier analysis depend on short–term power spectrum. The MFCC spaced at low frequencies and logarithmically at high frequencies for collecting important features of voice. The scale of Mel-frequency is the spectral analysis for linear frequency spatial arrangement at a lower place 1000 Hz and for logarithmic frequency spatial arrangement over 1000 Hz.

Speech signal is framed and windowed for 25 ms with an overlapping period of 10 ms. Frequency component of the framed signal is passed through 24 triangular Mel scale filter banks. The filtered output is compressed using logarithmic and then the cepstral coefficients are de-correlated by applying Discrete Cosine Transform (DCT). For first 13 output of the DCT block is considered

as static MFCC. From the Static MFCC, derivatives and double derivatives are calculated and used for emotion recognition. Figure 1 shows the block diagram of MFCC.

2.2.2 Linear Predictive Coding (LPC)

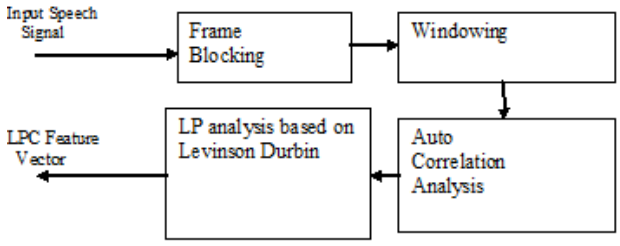

Figure 2. LPC feature extraction method.

LPC is the method to extract or compressed the speech signal. It can be consider as a subset of the filter. Sound is the biometric parameter. The speech contains the energy, pitch, frequency of each and every word to be spoken which is periodic in time. The train of impulse and the random noise can be evaluated as an irritation source and digital filter. An unexpressed signals irritation is modelled by a white Gaussian noise source. The speech signal processes through the speech analysis filter annihilate the redundancy in that speech signals. This signal passes through frame blocking window which is compare a small numbers of bit from the speech signals or else transfer the speech function to bring forth original signals. LPC is used for determining the group of predictor coefficients which will reduce the mean squared error across a small section of speech signal. Figure 2 shows the basic blocks of LPC.

2.2.3 Perceptual Linear Predictive (PLP)

It is a combination of the Discrete Fourier transform and linear predictor techniques. It is also known as all pole model. The basic idea behind this method is to depict the psychophysics of human hearing more efficiently in the feature extraction method. In this modelled the speech signal passed to the FFT which processed to the frequency wrapping. Equal loudness is used to simulate the pre-emphasis block to compress the amplitude to each signal and converted back to the time domain by using DCT and this small segment represented as the PLP feature. Figure 3 shows the block diagram of PLP.

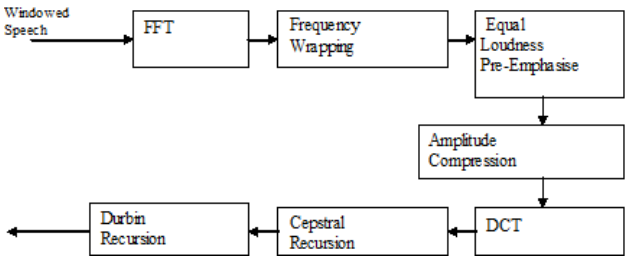

Figure 3. LPL feature extraction method.

2.3 HTK Tool Kit

In this project HTK tool kit is used for the emotion recognition. This HTK tool kit constructing HMM is used for recognition purpose. The HTK toolkit is a moveable toolkit[8,9]. HTK lie in the group of library modules and tools are available in C language. The tools render elegant adeptness for speech analysis, HMM model training, testing and results analysis done through HTK tool. The software corroborates HMMs using both continuous density mixture Gaussians and discrete distributions and can be used to build composite HMM systems.

2.4 Hardware

In this paper[12], Beagle Bone Black (BBB) embedded board is employed to implement emotion recognition. This is a low cost SitaraXAM3359AZCZ100 cortex A8 ARM processor, 1 GHZ which is shown in Figure 4.

Figure 4. Beagle bone black board.

It is uses the 512 MB memory devices in BBB. It is handling at a Clock Frequency of 303 MHz. The board is equipped with a single MicroSD connecter to work for board as sources of secondary boot and also we can select as a primary boot source.

In this hardware interfacing connect the mini USB cable provided to connect to the USB port of laptop for

the power supply purpose. HTK is open source software developed by Microsoft and Cambridge University. Viterbi decoder in HTK kit is licensed by Microsoft. All the feature extraction, Modeling and decoding is done using HTK.

3. Implementation and Results

In this research, twelve speakers of age between 20-25 are requested to read same set of documents for duration of 1 sec for each emotion. The audio file is stored as .wav format. Using HTK tool kit, MFCC, LPC and PLP features are extracted. For each type of feature method, 8 samples are used for HMM modelling of each emotion is done. Using HTK, Emotion of a person recognized by using Viterbi decoder which takes the maximum probability of the matching between emotional modelled to identify the particular emotion. Speaker dependent testing recognized happy, sad, natural, fear, surprise, and anger emotion of a person by extracting the feature with 3 methods gives 100% accuracy.

Figure 5. Emotion recognition output from the BBB.

```
user@user-HP-Pavilion-15-Notebook-PC: ~
root@beaglebone:~/ranialex/RBL_2/EMOTION# ./scripts_1/testing.scr
./scripts_1/testing.scr: line 1: !/bin/tcsh: No such file or directory
Read 6 physical / 6 logical HMMs
Read lattice with 9 nodes / 13 arcs
Created network with 17 nodes / 21 links
File: data/puja_h.mfcc
happy  ==  [499 frames] -65.4800 [Ac=-32674.5 LM=0.0] (Act=15.0)
root@beaglebone:~/ranialex/RBL_2/EMOTION#
```

Figure 6. Emotion recognition with MFCC method.

```
user@user-HP-Pavilion-15-Notebook-PC: ~
root@beaglebone:~/ranialex/RBL_2/EMOTION1# ./scripts_1/testing.scr
./scripts_1/testing.scr: line 1: !/bin/tcsh: No such file or directory
Read 6 physical / 6 logical HMMs
Read lattice with 9 nodes / 13 arcs
Created network with 17 nodes / 21 links
File: data/nandita_f.lpc
fear  ==  [499 frames] -2.4294 [Ac=-1212.3 LM=0.0] (Act=15.0)
root@beaglebone:~/ranialex/RBL_2/EMOTION1#
```

Figure 7. Emotion recognition with LPC method.

```
user@user-HP-Pavilion-15-Notebook-PC: ~
root@beaglebone:~/ranialex/RBL_2/EMOTION2# ./scripts_1/testing.scr
./scripts_1/testing.scr: line 1: !/bin/tcsh: No such file or directory
Read 6 physical / 6 logical HMMs
Read lattice with 9 nodes / 13 arcs
Created network with 17 nodes / 21 links
File: data/nandita_a.plp
anger  ==  [499 frames] 1.3350 [Ac=666.1 LM=0.0] (Act=15.0)
root@beaglebone:~/ranialex/RBL_2/EMOTION2#
```

Figure 8. Emotion recognition with PLP method.

Real time testing is shown in Figure 5, 6, 7, 8. For the speaker independent data set (4 samples), MFCC gives 100% accuracy for surprise emotion, PLP gives 100% accuracy for anger and LPC gives 100% accuracy for fear which is shown in Figure 9, 10, 11.

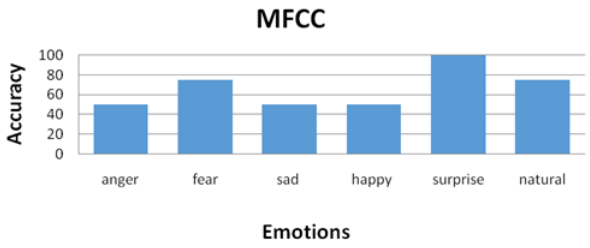

Figure 9. Speaker independent emotion recognition with MFCC.

Figure 10. Speaker independent emotion recognition with LPC.

Figure 11. Speaker independent emotion recognition with PLP.

From the above table we can say that for the modelled and trained speech signal for the three methods are MFCC, LPC, PLP gives the 100% accuracy for each speech signal of same data set. For the without modelled and trained

speech signal in MFCC method gave 100% accuracy for Surprise emotion, LPC method gave 100% accuracy for Fear emotion, PLP method gave 100% accuracy for Anger emotion.

4. Conclusion

Emotion recognition plays a vital role is determining the states of the human mind. Interactive voice based automated application performance improves by adding an emotion recognition system. It is advantageous to Health centres in monitoring the patients' emotional state after the treatment and also useful for applications requiring natural man machine interaction. This system could be employed as to determine the emotional state of the driver while driving the car which helps the safety of the driver and control the car movement. In this paper, experimental evaluation of emotion recognition on a low-cost, small footprint device BBB is carried out. The speaker dependent Emotion Recognition (ER) system gave 100% accuracy for all emotions such as sad, happy, surprise, anger and neutral. For speaker independent ER, MFCC features gives 100% accuracy for surprise emotion, PLP features gives 100% accuracy for anger and LPC features gives 100% accuracy for fear. It is also observed that not a single conventional feature extraction method gave 100% accuracy for all emotion. Because of this all feature extraction methods are implemented in the low power embedded system which increases complexity of emotion recognition system. In future, to reduce complexity of the ER system, a hybrid feature extraction method could be designed to detect all human emotions from speech.

5. Referenceses

1. Scherer S, Hofmann H, Lampmann M, Steffenrhinow M, Mschwenker F, Untherpalm G. Emotion recognition from speech: Stress experiment. Germany; 2008. p. 1–6.
2. Shashidhar G, Koolagudi K, Rao S. Emotion recognition from speech: A review. Int J Speech Technol. 2012; 3(2):1–5.
3. ElAyadi MA, Mohamed SB, Karray BF. Survey on speech emotion recognition: features, classification schemes, and databases. Pattern Recognition. 2011; 44(3):572–87.
4. Kim EH, Hyun KH, Kim SH, Kwak YK. Improved emotion recognition with a novel speaker-independent feature. IEEE/ASME Transactions on Mechatronics. 2009; 14(3):317–25.
5. Bageshree V, Sathe-Pathak S, Ashish R, Panat P. Extraction of pitch and formants and its analysis to identify 3 different emotional states of persons. IJCSI. 2012; 9(4):1–6.
6. Patel S, Alex JSR, Venkatesan N. Low-power multi-layer perceptron neural network architecture for speech recognition networks. Indian Journal of Science and Technology. 2015; 8(20):1–6.
7. Alex JSR, Mukhedkar AS, Venkatesan N. Performance analysis of SOFM based reduced complexity feature extraction methods with back propagation neural network for multilingual digit recognition networks. Indian Journal of Science and Technology. 2015; 8(19):1–8.
8. Malta L, Miyajima C, Kitaoka N, Takeda K. Analysis of real-world driver's frustration. IEEE Transactions on Intelligent Transportation Systems. 2011; 12(1):1–10.
9. Tiwari V. MFCC and its applications in speaker recognition. Int J Emerg Technol. 2010; 1(1):19–22.
10. HTK Book 3.4.1. Available from: www.speech.ee.ntu.edu.tw/homework/DSP_HW2-1/htkbook.pdf
11. Lawrence R, Rabiner R. A tutorial on hidden markov models and selected application in speech recognition. Proceeding of the IEEE. 1989; 77(2):257–68.
12. Beagle Board. Available from: http://beagleboard.org/