International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 2, February 2013)

# Emotion Recognition from Speech

Ankur Sapra[1], Nikhil Panwar[2], Sohan Panwar[3]
*Jaypee Institute of Information Technology, Noida*

***Abstract*— an emotion is a mental and physiological state associated with a wide variety of feelings, thoughts, and behavior. Emotions are subjective experiences, or experienced from an individual point of view. Emotion is often associated with mood, temperament, personality, and disposition. Hence, in this paper method for detection of human emotions is discussed based on the acoustic features like pitch, energy etc. The proposed system is using the traditional MFCC approach [2] and then using nearest neighbor algorithm for the classification. Emotions has been classified separately for male and female based on the fact male and female voice has altogether different range [1][4] so MFCC varies considerably for the two.**

***Keywords*— Emotion Recognition from Speech, Fourier Transform, MelFilterBank, MFCC, Modern MFCC Approach, Nearest Neighbor Algorithm**

## I. Introduction

The importance of emotion recognition of human speech has increased in recent days to improve both the naturalness and efficiency of human - machine interactions. Its varied range of applications includes automatic dialog systems and camera less mobile phones. Recognizing human emotions is a very complex task in itself because of the ambiguity in classifying the acted and natural emotions. A number of studies have been conducted to extract the acoustic features which would result in correct determination of emotions. But even after so much of research in the field researchers have not gained much of success and the accuracy in determination is still less than 89.2% [3]. Emotions can be classified as Natural and Artificial emotions and further can be divided into emotion set i.e. anger, joy, sadness, neutral, happy, disgust. [3][4] In this paper we will try to identify the emotion using the emotion set Anger, Happy and Neutral. However certain emotions have similar characteristics based on the set of features. Hence, systems based on these features for emotion or stress classification are unable to accurately distinguish more than a couple of stress or emotion categories.

An experimental study has been conducted to determine how well people recognize emotions in speech. Based on the results of the experiment the most reliable utterances were selected for feature selection and for training recognizers. Several machine learning techniques have been applied to create recognition agents including k-nearest neighbor, neural networks, and ensembles of neural networks. The agents can recognize five emotional states with the following accuracy: normal or unemotional state - 55-75%, happiness - 60-70%, anger - 70-80%, sadness - 75-85%, and fear - 35-55%. The total average accuracy is about 70%. The agents can be adapted to a particular environment depending on parameters of speech signal and the number of target emotional states. For a practical application an agent has been created that is able to analyze telephone quality speech signal and distinguish between two emotional states ('agitation' which includes anger, happiness and fear, and 'calm' which includes normal state and sadness) with the accuracy 77%. The agent was used as a part of a decision support system for prioritizing voice messages and assigning a proper human agent to response the message at call center environment.

Aiming at emotion deficiency in present E-Learning system, a lot of negative effects were analyzed and corresponding countermeasures were proposed. Basing on it, we combined affective computing with the traditional E-Learning system. The model of E-Learning system based on affective computing was constructed by using speech emotion, which took speech feature as input data. Our simulation experiment results showed that neural networks was effective in emotion recognition, and we achieve a recognition rate of approximately 50% when testing eight emotions .besides, other key techniques of realizing the system such as tracking the change of emotion state and adjusting teaching strategies were also introduced.

## II. Standard MFCC Approach[2]

Steps involved in Algorithm (refer fig. 1):

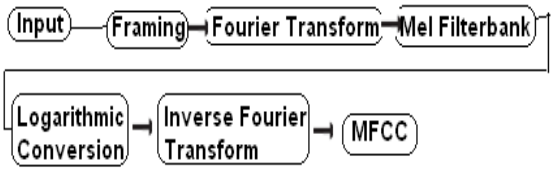
### *A. Frame Level Break Down*

Input human voice sample is first break down into frames of frame size 16 ms [3] each. This is done for frame level classification in further steps.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 2, February 2013)

*B. Frame Level Feature Extraction*

For each frame we got in 'A' we will calculate MFCC as the main feature for emotion recognition.



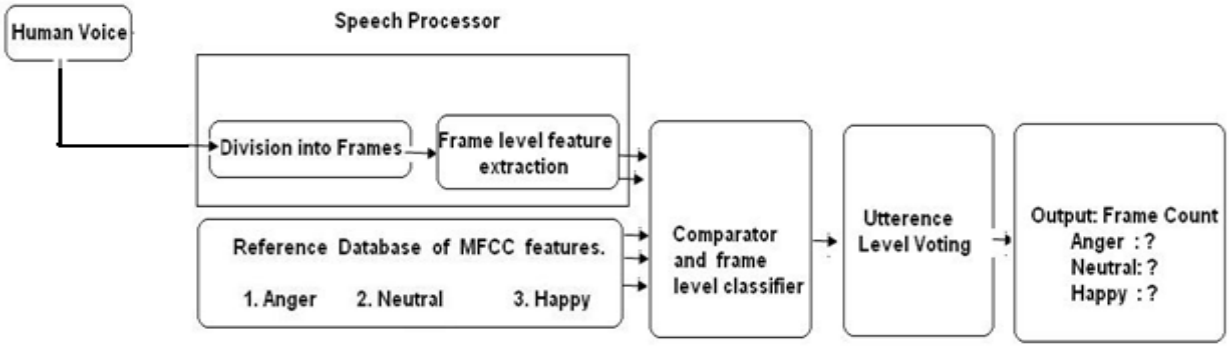




**Figure 1: Standard MFCC Approach**

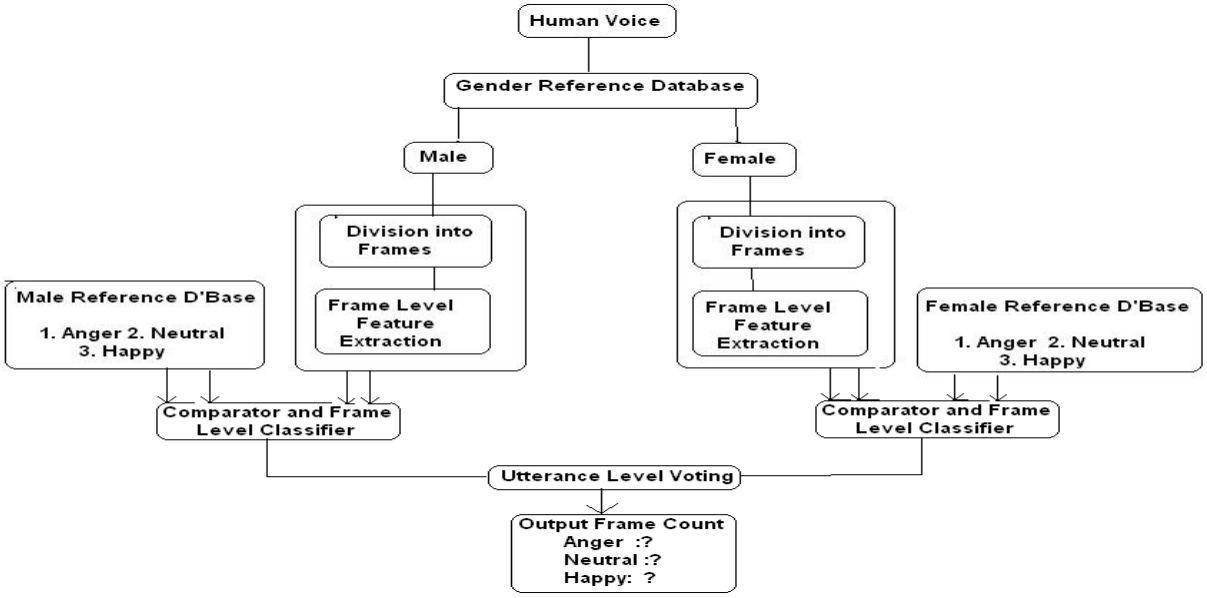
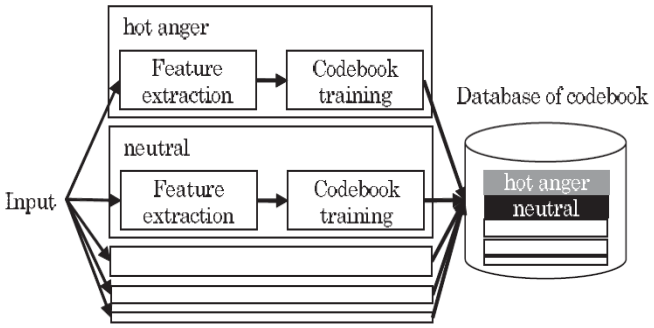



**Figure 2: Modified MFCC Approach**

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 2, February 2013)

*C. Comparator and Frame Level classifier/ Nearest Neighbor Algorithm*

Reference database is maintained which contains the MFCCs of emotions i.e. of Anger, Neutral and Happy. MFCC of the frames are compared with the MFCCs stored in reference database and the distance is calculated between the comparable frames.




Codebook training flowchart.

*D. Utterance-Level Voting*

Based on the distance of the analysis frame from the reference database, we classify the frame as anger, happy or normal. And the output is displayed in terms of emotional frame count.

## III. Proposed Modified MFCC Approach (Refer Fig.2)

Standard approach takes into account a mixed data set for male and female samples as reference database. But if we separate these two, accuracy of recognition of emotion increases. So before we breakdown the speech sample into frames we will first classify if the speech sample is of male or female and then compare it with appropriate database.

Differences in proposed algorithm:

*Step 1: Preprocessing/Gender Recognition.*

Speech sample is first passed through a gender reference database which is maintained for recognition of gender before It goes to step A. Statistical approach is followed taking pitch [8] as feature for gender recognition. We find a lower and upper bound for pitch for both male and female samples using the reference database.

*Steps A and B will remain same as in Standard Approach.*

*C. Comparator and Frame Level Classifier/ Nearest Neighbor Algorithm*

The difference between proposed approach and standard approach is mainly in the reference database for MFCC feature comparison when it comes to frame level classification. Here what it does this firstly it makes the overall comparison more clear and secondly since there is so much difference in the pitch range of male and female voices it helps with the accuracy of recognition as well.

*Step D will remain same as in Standard Approach*

Test Report for Modified MFCC Approach

*For Female Samples*

| Sample File Name(input) | Pitch Observed | Gender Correctly Recognized (output1) | Emotion Correctly Recognized (output 2 ) |
|---|---|---|---|
| Angry 1 | 257.59 | Yes | Yes |
| Angry2 | 130.39 | No | No |
| Angry3 | 275.28 | Yes | Yes |
| Happy1 | 303.71 | Yes | Yes |
| Happy2 | 244.72 | yes | Yes |
| Happy3 | 297.57 | Yes | No |
| Normal1 | 280.53 | Yes | No |
| Normal2 | 242.04 | Yes | Yes |
| Normal3 | 250.28 | Yes | No |
| Normal4 | 266.94 | Yes | No |
| Normal5 | 259.1 | Yes | Yes |

*For Male Samples*

| Sample File Name | Pitch Observed | Gender Correctly Recognized | ---Emotion Correctly Recognized |
|---|---|---|---|
| Normal1 | 148.38 | Yes | Yes |
| Normal2 | 154.08 | Yes | Yes |
| Normal4 | 168.83 | Yes | Yes |
| Normal5 | 148.88 | Yes | Yes |
| Normal6 | 156.82 | Yes | Yes |
| Normal7 | 152.48 | Yes | Yes |
| Angry14 | 167.55 | Yes | No |
| Angry2 | 196.69 | Yes | No |
| Happy2 | 155.17 | Yes | Yes |
| Happy3 | 179.85 | Yes | No |
| Happy4 | 266.94 | No | Yes |

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 2, February 2013)

## IV. RESULTS

Test has been performed using 22 samples database on both Standard MFCC emotion recognizer and Modified MFCC emotion recognizer. Recording is done is noise free environment using Windows Sound Recorder. Speaker is assumed to be speaking in English and the matter which the speaker has recorded is:

*For Angry:* What the hell you are talking about, you looser. I was not even there at that moment. Don't ever say that again to me otherwise I am going to kick you hard.

*For Happy:* yayyy... Today is the best day of my life. I am very thankful to you for all the support. It's only because of you I have made here. Once again thank you very much my friend.

| Sample Name(input) | Pitch Observed | Emotion Recognized Correctly (output) |
|---|---|---|
| F_angry1 | 257.59 | No |
| Angry2 | 130.39 | No |
| Angry3 | 275.2809 | No |
| Happy1 | 303.71 | No |
| Happy2 | 244.72 | Yes |
| Happy3 | 297.59 | No |
| Normal1 | 280.53 | No |
| Normal2 | 242.04 | Yes |
| Normal3 | 250.28 | Yes |
| Normal4 | 266.94 | Yes |
| Normal5 | 259.1 | Yes |
| M_Angry1 | 167.55 | No |
| Angry2 | 196.69 | No |
| Happy2 | 155.17 | Yes |
| Happy3 | 179.85 | Yes |
| Happy4 | 266.94 | Yes |
| Normal1 | 148.38 | Yes |
| Normal2 | 154.08 | No |
| Normal4 | 168.83 | Yes |
| Normal5 | 148.88 | No |
| Normal6 | 156.82 | Yes |
| Normal7 | 152.48 | No |

*For Normal:* Beat the exam blues-the CBSE board exams are round the corner and most of the students are under pressure. Though it is not easy to deal with exam stress, there are many teenagers in east Delhi who manage to cope well with it.

Following Results were obtained:

*Standard Approach:*

Success Rate: 54.54%

*Modified MFCC Approach:*

| Gender | Success Rate |
|---|---|
| Female | 54.54 |
| Male | 72.72 |

Overall Success Rate: 63.63 %

## V. CONCLUSION

MFCC approach for emotion recognition from speech is a stand-alone approach which does not require calculation of any other acoustic features but if we want the accuracy to climb as high as 90-95% MFCC approach can be clubbed with another approach i.e. emotion recognition using facial expressions. For more information on this refer to [7-10]. The major disadvantage of using the proposed approach is if the gender is recognized incorrectly by the system then further processing will be all in vain but it happens rare.

## REFERENCES

[1] Chiu Ying Lay, Ng Hian James. "Gender Classification from Speech",(2005)Webreference: ttp://sg.geocities.com/nghianja/CS5240.doc

[2] Nobuo Sato and Yasunari Obuchi. "Emotion Recognition using MFCC's" Information and Media Technologies 2(3):835-848 (2007) reprinted from: Journal of Natural Language Processing 14(4): 83-96 (2007)

[3] T L Nwe'; S W Foo L C De Silva, "Detection of Stress and Emotion in Speech Using Traditional And FFT Based Log Energy Features" 0-7803-8185-8/03 2003 IEEE ( 2003)

[4] Chang-Hyun Park and Kwee-Bo Sim. "Emotion Recognition and Acoustic Analysis from Speech Signal" 0-7803-7898-9/03 Q2003 IEEE (2003)

[5] Daniel Neiberg, Kjell Elenius, Inger Karlsson,and Kornel Lskowski," Emotion Recognition in Spontaneous Speech" Working Papers 52 (2006)

[6] Kyung Hak Hyun, Eun Ho Kim, Yoon Keun Kwak, " Improvement of Emotion Recognition by Bayesian Classifier Using Non-zero-pitch Concept" , 7803-9275-2/05/2005 IEEE(2005)

[7 ] Eun Ho Kim, Kyung Hak Hyun," Robust Emotion Recognition Feature, Frequency Range of Meaningful Signal" IEEE International Workshop on Robots and Human Interactive Communication, 0-7803-9275-2/05 2005IEEE (2005)

[8 ] Quran and Rafik A. Goubran," Pitch -Based Feature Extraction for Audio Classification" 0-7803-8108-4/03/$17.00 0 2003 IEEE (2003)

[9 ] YI-LIN LIN, GANG WEI, " Speech Emotion Recognition Based On HMM AND SVM"(2005)

[10 ] Tsung-Long Puo, Yu-Te Chen and Jun - Heng Yeh, " Emotion Recognition From Madarin Speech signal, S0-7803-8678-7/04 02004 IEEE (2004)