Journal Of Harmonized Research (JOHR)

Journal Of Harmonized Research in Engineering
2(1), 2014, 183-186

ISSN 2347 – 7393

Original Research Article

ADI SPEECH EMOTION RECOGNITION SYSTEM

Laba Kr. Thakuria, Akalpita Das, Purnendu Acharjee, Deepankar Sarma, P.H.Thakdar

Department of Instrumentation and USIC
Gauhati University

Abstract:
An emotion is a physiological and mental state associated with a wide variety of thoughts, feelings and behavior. Emotions are subjective experiences or experienced from an individual point of view and it is often associated with personality, mood, temperament and disposition. Hence, in our research paper, we have describe the the technique for detection of human emotions based on some acoustic features like pitch, energy etc. In our proposed system we have used the traditional MFCC approach and then we use the nearest neighbor algorithm for the classification. Here the emotions are classified separately for male and female speaker based on the fact that the male and female voice has altogether different range. So MFCC varies considerably for the two.

Keywords— Adi, FFT, Mel Filter Bank, MFCC, Modern MFCC, Nearest Neighbor Algorithm.

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. A number of studies have been conducted to extract the acoustic features which would result in correct determination of emotions. 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 try to identify the emotion using the emotion set Anger, Happy and Neutral. Our study has been conducted to determine how well people recognize emotions in speech. Based on the results of the experiment the most reliable utterances are selected for feature selection and for training recognizers. The agents can recognize five emotional states with the following accuracy: normal or unemotional state - 50-75%, happiness - 65-70%, anger - 75-80%, sadness -

For Correspondence:
thakurialabaATgmail.com
Received on: February 2014
Accepted after revision: February 2014

75-80%, and fear - 30-55%. The total average accuracy is about 70%. The model of E-Learning system based on affective computing is constructed by using speech emotion. 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.

A). Adi language

The Adis have decided to adopt the Roman script with certain additions of two vowels (gaayo merey) and two consonants (merey) whose phonetics are not very common in international phonetic alphabets (IPA). The four new alphabets coined for use in Adi language (agom) are dual-letter vowels - EY & UI and consonants - NG & NY. With this restructured script, AAK, the apex literary body of Adi community, will develop study materials for use in the schools of Adii areas as third language. All books, poetries and archive writings, already in printed form, will be re-written with this script in due course of time. Describing as "historic" the reform brought in towards enriching the traditional cultural heritage of Adi community, the ABK exudes hope that this will usher in a new era of linguistic and literary development of Adi

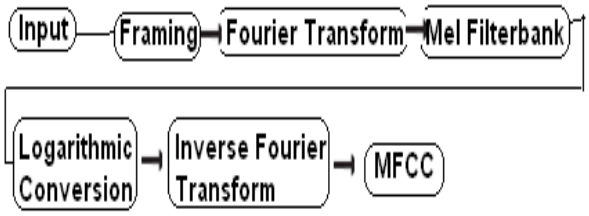
Standard Mfcc Approach

A. Frame Level Break Down

Here the input human voice sample is first break down into frames of frame size 16 ms each. This is done for frame level classification in further steps.

B. Frame Level Feature Extraction

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



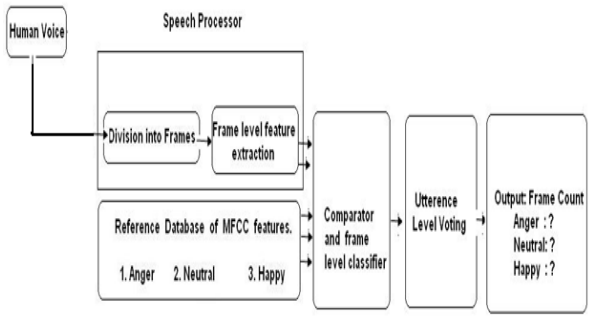


Fig 1: Standard MFCC

C. Comparator and Frame Level classifier/ Nearest Neighbor Algorithm

Here the database is maintained with emotion 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.

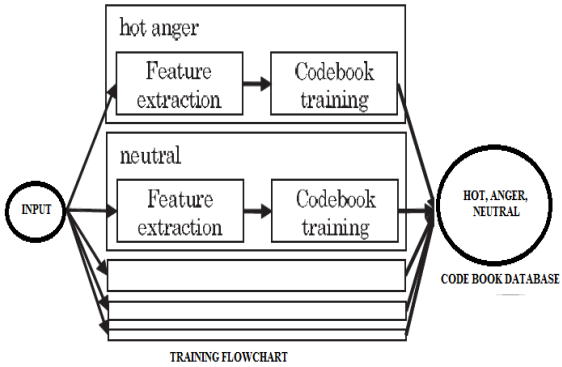


Fig 2: Training flowchart

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.

Proposed Modified Mfcc Approach

We take 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.

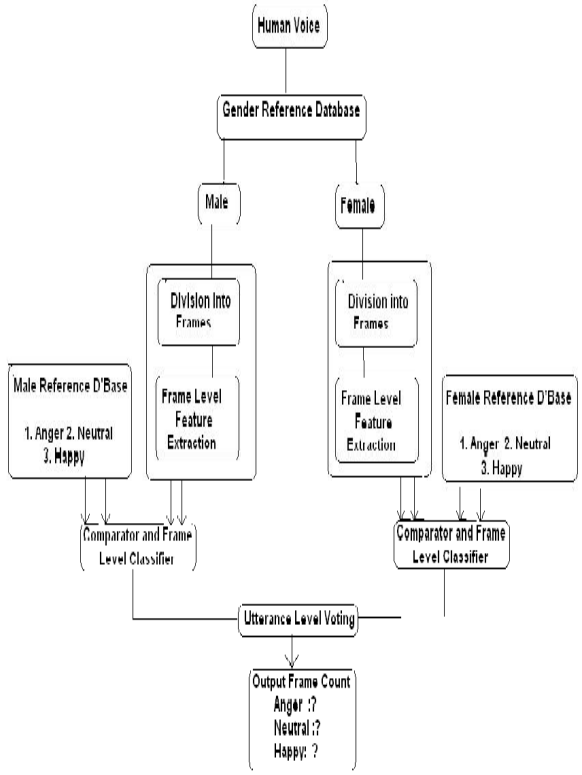


Fig 3: Modified MFCC

Step 1: Preprocessing/Gender Recognition.

Before going to step A, speech sample is first passed through a gender reference database which is maintained for recognition of gender. Then statistical approach is followed taking pitch as feature for gender recognition. The pitch for both male and female samples using the reference database, we find a lower and upper bound.

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

C. Comparator and Frame Level Classifier

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

Table 1: Test Report for Modified MFCC Approach of female samples

Name (Input)	Pitch Observed	Gender Correctly (Output1)	Emotion Correctly (Output2)
Anger1	255.60	Yes	Yes
Anger2	135.50	No	No
Anger3	270.20	Yes	Yes
Happy1	300.40	Yes	Yes
Happy2	250.30	Yes	Yes
Happy3	198.50	Yes	No
Normal1	290.20	Yes	Yes
Normal2	240.45	Yes	Yes
Normal3	150.40	Yes	Yes

Table 2: Test Report for Modified MFCC Approach of female samples

Name(Input)	Pitch Observed	Gender Correctly (Output1)	Emotion Correctly (Output2)
Anger1	175.25	Yes	Yes
Anger2	197.35	yes	yes
Anger3	200.30	Yes	Yes
Happy1	156.20	Yes	Yes
Happy2	180.41	Yes	Yes
Happy3	150.45	No	No
Normal1	155.30	Yes	Yes
Normal2	150.65	Yes	Yes
Normal3	170.21	Yes	Yes

Results

Test has been performed using 20 samples database on both Standard MFCC emotion recognizer and Modified MFCC emotion recognizer.

Through this experiment we have obtained as follows

Standard Approach:

Success Rate: 55.64%

Modified MFCC Approach:

Gender	Success Rate
Female	55.64
Male	74.82

Table 3: Test Report for Modified MFCC Approach of female & Male samples

Name(Input)	Pitch Observed	Emotion Correctly (Output)
FAnger1	255.80	No
FAnger2	135.56	No
FAnger3	270.28	Yes
MAnger4	275.67	No
MAnger5	147.50	No
MAnger6	290.35	Yes
MAnger7	276.46	yes
FHappy1	300.40	Yes
FHappy2	250.30	Yes
FHappy3	198.50	Yes
MHappy4	198.61	Yes
MHappy5	270.40	Yes
MHappy6	208.57	No
FNormal1	290.20	Yes
FNormal2	240.45	Yes
FNormal3	150.41	Yes
MNormal4	295.27	No
MNormal5	276.56	No
MNormal6	187.42	Yes
MNormal7	168.34	Yes

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 85-95% 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.

Acknowledgment

This work is done under the department of Instrumentation & USIC, Gauhati University. We are very much thankful to Prof. P. H. Talukdar, supervisor speech research centre, Gauhati University for his constant cooperation and guidance for the completion of the paper.

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).