K. Ravi Kumar, V.Ambika, K.Suri Babu / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 2, Issue 5, September- October 2012, pp.1797-1799

Emotion Identification From Continuous Speech Using Cepstral Analysis

K. Ravi Kumar[1], V.Ambika[2], K.Suri Babu[3]
[1]M.tech(DECS), Chaitanya Engineering College, Visakhapatnam, India.
[2]Dept of ECE, Chaitanya Engineering College, Visakhapatnam, India.
[3]Scientist, NSTL (DRDO), Govt. of India, Visakhapatnam, India

ABSTRACT

Emotion plays a major role in the area of psychology, human computer interaction, and robotics and BPO sectors. With the advancements in the field of communication technology, it is possible to establish the channel within few seconds across the globe. As most of the communication channels are public data transmission may not be authenticated. In such situation, before interacting, it is essential to recognize speaker by the unique features in the speech. A speaker can modulate his/her voice can changes his/her emotion state. Hence emotion recognition is required for the applications like telemetry, call centers, forensics and security. In our project the main emotion consider happy, angry, boredom, and sad. In this work we dealt with speaker recognition with different emotion. The basic emotions for this study include angry, sad, happy, boredom and neutral. The features we modeled using Gamma Distribution(GD) and data base generated with 50 speakers of both genders with the above basic emotions Considering feature vector combinations MFCC-LPC.

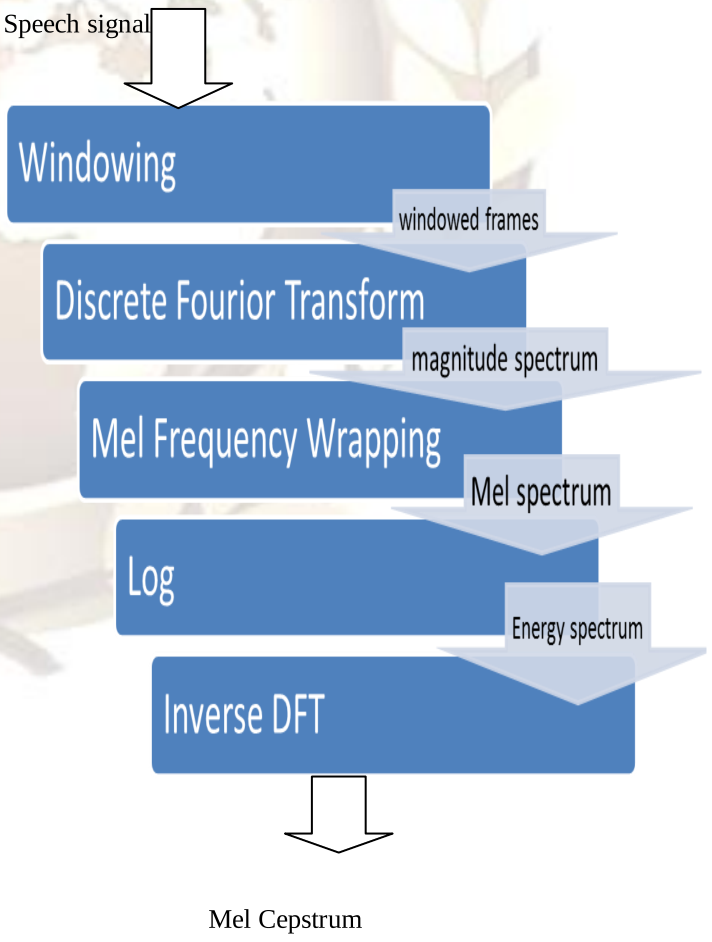
Keywords: MFCC, LPC, Gamma Distribution

1. INTRODUCTION

In some specific situations, such as remote medical treatment, call center applications, it is very much necessary to identify the speaker along with his/her emotion. Hence in this paper a methodology for emotion identification in continuous speech, the various emotions consider are happy, angry, sad, boredom and neutral. Lot of research is projected to recognize the emotional states of the speaker using various models such as GMM, HMM, SVM, neural networks [2][3][4][5]. In this paper we have considered an emotional data base with 50 speakers of both genders. The data is trained and for classification of the speakers emotions generalized gamma distribution is utilized. The data is tested with different emotions .The rest of the paper is organized as follows. In section-2 of the paper deals with feature extraction, .section-3 of the paper generalized gamma distribution is presented.section-4 of the paper deals with the experimentation and results.

2. FEATURE EXTRACTION

In order to have an effective recognition system, the features are to be extracted efficiently. In order to achieve this, we convert these speech signals and model it by using Gamma mixture model. Every speech signal varies gradually in slow phase and its features are fairly constant .In order to identify the features, long speech duration is to be considered. Features like MFCC and LPCs are most commonly used, The main advantage of MFCC is, it tries to identify the features in the presence of noise, and LPCs are mostly preferred to extract the features in low acoustics, LPC and MFCC[8]coefficients are used to extract the features from the given speech.



3. GENERALIZED GAMMA MIXTURE MODEL

Today most of the research in speech processing is carried out by using Gaussian mixture model, but the main disadvantage with GMM is that it relies exclusively on the the approximation and low in convergence, and also if GMM is used the speech and the noise coefficients differ in magnitude [9]. To have a more accurate feature extraction maximum posterior estimation models are to be considered [10].Hence in this paper generalized gamma distribution is utilized for classifying the speech signal. Generalized gamma distribution represents the sum of n-exponential distributed random variables both the shape and scale parameters have non-negative integer values [11]. Generalized gamma distribution is defined in terms of scale and shape parameters [12]. The generalized gamma mixture is given by

$$f(x,k,c,a,b) = \frac{c(x-a)^{ck-1} e^{-\left(\frac{x-a}{b}\right)^{c}}}{b^{ck}\Gamma(k)} \quad (1)$$

Where k and c are the shape parameters, a is the location parameter, b is the scale parameter and gamma is the complete gamma function [13]. The shape and scale parameters of the generalized gamma distribution help to classify the speech signal and identify the speaker accurately.

4. OUR APPROACH OF SPEAKER EMOTION RECOGNITION

For identifying emotion in a continuous speech, our method considers MFCC-LPC as feature vector. Emotions of speaker identified using various sets of recorded. Unknown emotion recognition of speaker considering continuous speech. Every speech consists of the basic emotions: Angry, happy, Boredom, neutral, sad to identify emotional state of speaker, we should train the speaker's speech properly i.e. selection of feature vector is crucial in emotion identification. Recorded a continuous speech with unknown emotion .Known emotions are used for training. The steps to be followed for identification of emotion in continuous speech effectively are given under

Step1:

Obtain the training set by recording the speech voices in a .wav form

Step2:

Identify the feature vector of these speech signals by application of transformation based compound feature vector MFCC-LPC.

Step3:

Generate the probability density function (PDF) of the generalized gamma distribution for all the trained data set.

Step4:

Cluster the recorded speech of unknown emotion for testing using fuzzy c-means clustering technique and perform step2 to step 3.

Step5:

Find the range of speech of test signal in the trained set.

Step6:

Evaluation metrics such as Acceptance Rate (AR), False Acceptance Rate (FAR), Missed Detection Rate(MDR) are calculated to find the accuracy of speaker recognition.

5. EXPERIMENTATION

In general, the Emotion signal will always be of finite range and therefore, it needs to truncate the infinite range. Hence it is always advantage to consider the truncations of GMM into finite range; also it is clearly observed that the pitch signals along the right side are more appropriate. Hence, in this paper we have considered Generalized Gamma Distribution with acted sequences of 5 different emotions, namely happy, sad, angry, boredom, neutral. In order to test the data 50 samples are considered and a database of audio voice is generated in .wav format. The emotion speech data base is considered with different emotions such as happy, angry, sad, boredom and neutral. The data base is generated from the voice samples of both the genders .50 samples have been recorded using text dependent data; we have considered only a short sentence. The data is trained by extracting the voice features MFCC and LPC. The data is recorded with sampling rate of 16 kHz. The signals were divided into 256 frames with an overlap of 128 frames and the MFCC and LPC for each frame is computed .In order to classify the emotions and to appropriately identify the speaker generalized gamma distribution is considered .The experimentation has been conducted on the database, by considering 10 emotions per training and 5 emotions for testing .we have repeated the experimentation and above 90% over all recognition rate is achieved. The experimentation is conducted by changing deferent emotions and testing the data with a speaker's voice .In all the cases the recognition rate is above 90%.

6. CONCLUSIONS:

In this paper a novel frame work for emotion identification in continuous speech with MFCC and LPC together as feature vectors. This work is very much useful in applications such as speaker identification associated with emotional coefficients .It is used in practical situations such as call centers and telemedicine. In this paper generalized gamma distribution is considered for classification and over all recognition rate of above 90% is achieved.

REFERENCES:

[1]. K.suri Babu, Y.srinivas, P.S.Varma,V.Nagesh(2011)"Text dependent & Gender independent speaker recognition model based on generalizations of gamma distribution"IJCA,dec2011.

[2]. Xia Mao, Lijiang Chen, Liqin Fu, "Multi-level Speech Emotion Recognition Based on HMM and ANN", 2009 WRI World Congress,Computer Science and Information Engineering, pp.225-229, March2009.

[3]. B. Schuller, G. Rigoll, M. Lang, "Hidden Markov model-based speech emotion recognition", Proceedings of the IEEE ICASSP Conference on Acoustics,Speech and Signal Processing, vol.2, pp. 1-4, April 2003.

[4]. Yashpalsing Chavhan, M. L. Dhore, Pallavi Yesaware, "Speech Emotion Recognition Using Support Vector Machine", International Journal of Computer Applications, vol. 1, pp.6-9,February 2010.

[5] Zhou Y, Sun Y, Zhang J, Yan Y, "Speech Emotion Recognition Using Both Spectral and Prosodic Features", ICIECS 2009.International Conference on Information Engineering and Computer Science, pp. 1-4, Dec.2009.

[6]. I. Shahin, "Speaker recognition systems in the emotional environment," in Proc. IEEE ICTTA '08, Damascus,Syria, 2008.

[7]. T. Giannakopoulos, A. Pikrakis, and S. Theoridis, "A dimensional approach to emotion recognition of speech from movies," in Proc. IEEE ICASSP '09, Taipei, Taiwan,2009.

[8]. Md. AfzalHossan, SheerazMemon, Mark A Gregory "A Novel Approach for MFCC Feature Extraction" RMIT University978-1-4244-7907-8/10/$26.00 ©2010 IEEE.

[9]. George Almpanidis and Constantine Kotropoulos,(2006)voice activity detection with generalized gamma distribution, IEEE,ICME 2006.

[10]. XinGuang Li et al(2011),Speech Recognition Based on K-means clustering and neural network Ensembles,International Conference on Natural Computation,IEEEXplore Proceedings,2011,pp614-617.

[11]. EwaPaszek,TheGamma&chi-square Distribution,conexions,Module-M13129.