Text-Independent Speaker Recognition using Emotional Features and Generalized Gamma Distribution

K Suri Babu
Scientist, NSTL (DRDO),
Govt. of India,
Visakhapatnam, India

Srinivas Yarramalle
Department of IT,
GITAM University,
Visakhapatnam, India

Suresh Varma Penumatsa
Dept. of Computer Science,
Adikavi Nannaya University,
Rajahmundry, India

ABSTRACT
In this article, a novel methodology of text independent speaker recognition associated with emotional features is proposed. MFCC and LPC features are considered for speaker voice feature extraction. The speaker data is classified using generalized gamma distribution. The experiment is conducted on the emotional speech data base, containing 50 speakers with 5 different emotions namely happy, angry, sad, boredom and neutral. This approach is very much useful in costumer care and call centre applications. The accuracy of the developed model is presented using Confusion matrix. The results show that there is a great influence of the emotion state, while identifying the speaker.

Keywords
text-independent speaker recognition, generalized gamma distribution, MFCC, LPC emotions, confusion matrix.

1. INTRODUCTION
With the recent developments in technology, communication across the globe became easier and the communication can be established within few seconds, since most of the communication channels are public, the voice data transmission may not be authenticated ,it is nessecery to safe guard this voice data. The first step in this approach is to identify the speaker most efficiently using voice features of the speaker. To identify the features cepstral coefficients MFCC, LPC are considered. The main advantage of MFCC is, it tries to identify the features in the presence of noise. LPCs are mostly preferred to extract the features in low acoustics, LPC and MFCC [1] coefficients are used to extract the features from the given speech.

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 of text independent speaker identification associated with the speaker emotions, 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]. Very little work, except the works of Sudhir[6], Maris Vasile Ghiurcau et al [7] have been reported in literature for text-independent recognition associated with emotional features. In this paper we have consider a emotional data base with 50 speakers of both genders is considered .The data is trained and for classification of the speakers emotions generalized gamma distribution is utilized .The parameters are updated using E-M algorithm. 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 estimation of the model parameter using E-M algorithm.section-5 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. ESTIMATION OF THE MODEL PARAMETERS THROUGH EXPECTATION-MAXIMIZATION ALGORITHM

For effective speaker identification model it is mandatory to estimate the parameters of the speaker model effectively. For estimating the parameters EM Algorithm is used which maximizes the likely hood function of the model for a sequence i,

Let $xi = (x1, x2, \ldots \ldots . xt)$ be the training vectors drawn from a speaker's speech and are characterized by the probability density function of the Generalized Gamma Distribution as given in equation-1, the updated eqns for the shape parameters are given by

$$c^{(l+1)} = \frac{1}{\frac{1}{f}\frac{\partial f}{\partial c} - k \log\left(\frac{x-a}{b}\right) + \frac{(x-a)^c}{b^c \log\left(\frac{x-a}{b}\right)}} \qquad (2)$$

$$k^{(l+1)} = 1 + \frac{\left[\int_0^\infty e^{-t} (\log_e t)\ t^{k-1} dt\right]}{\Gamma(k-1)\left[c \log\left(\frac{x-a}{b}\right) - \frac{1}{f}\frac{\partial f}{\partial k}\right]} \qquad (3)$$

The updated equation for the scale parameters is given by

$$b^{(l+1)} = \frac{ck}{\frac{c}{b^{c+1}}(x-a)^c - \frac{1}{f}\frac{\partial f}{\partial b}} \qquad (4)$$

Using the equations (2) to (4) we model the parameters of the Generalized Gamma distribution.

5. EXPERIMENTATION AND RESULTS

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%.The following graph –I, shows performance of emotion recognition of all the emotions considered, table-I shows how right emotions give right results observing diagonal elements of confusion matrix.

Graph-I Bar Chart Of Emotionsrecognition Considering Mfcc-Lpc

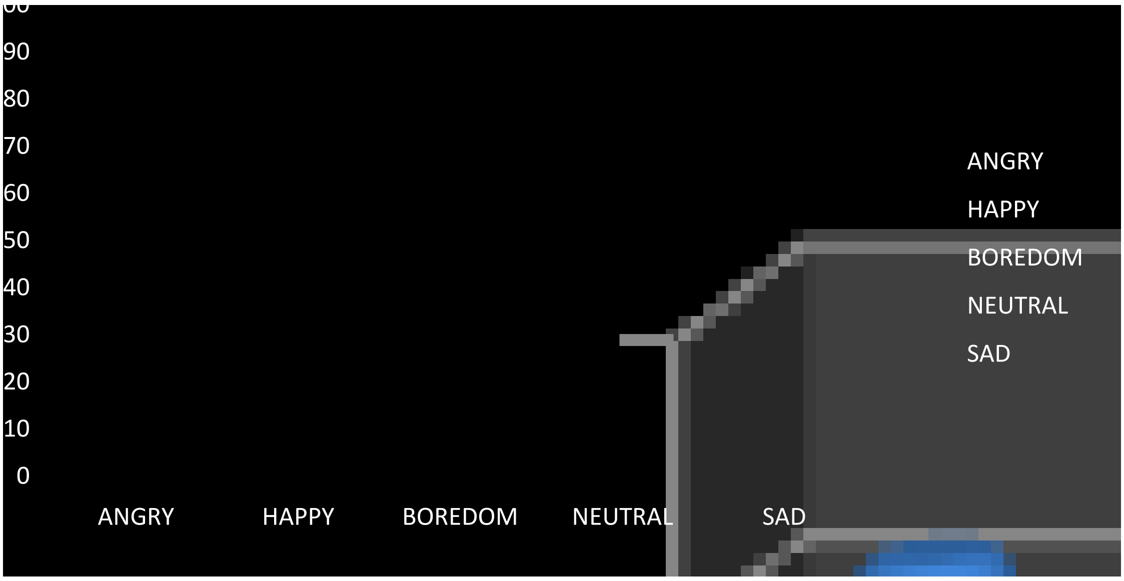

Table-I Confusion Matrix Of Considering Mfcc-Lpc

Emotions	Emotion Recognition Rate (%)				
	Angry	Happy	Boredom	Neutral	Sad
Angry	91.3	0	9.2	9.2	7.2
Happy	0	88.2	6.5	7.6	0
Boredom	7	0	91.5	12.1	8.5
Neutral	6	0	12.6	90.4	6.5
Sad	5.3	0	11.5	8.2	92.9

6. CONCLUSIONS

In this paper a novel frame work for text independent speaker identification with MFCC and LPC together with emotion recognition is presented .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.

7. 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 andComputer 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.

[12].J.Won Shin et al(2005),Statitical modeling of Speech Based on Generalized Gamma Distribution,IEEE,Signal Processing Letters,Vol.12,No.1,March(2005),pp258-261.

[13].RajeswaraRao.R.,Nagesh(2011),Source Feature Based Gender Identification System Using GMM, International Journal on computer science and Engineering,Vol:3(2),2011,pp-586-593.

[14].Christos Tzagkarakis and AthanasiosMouchtaris,(2010),Robust Text-independent Speaker Identification using short testand training sessions,18th Eropian signal Processing conference(EUSIPCO-2010).