Text Dependent and Gender Independent Speaker Recognition Model based on Generalizations of 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

Nagesh Vadaparthi
Department of IT,
MVGR College of Engineering,
Vizianagaram, India

ABSTRACT
Speaker recognition is one of the research potential areaswith applications in biometrics and content based retrievals, it helps to identify a speaker from the speech signal. To develop an effective speaker recognition system, it is needed to have a concrete methodology of feature extraction and a mechanism to model these features, most of the models available in the literature are more focused towards the speech rather than the speaker, a novel speaker model is developed in this article using the generalized gamma mixture model, here we have considered Mel frequency cepstral coefficients (MFCC)and linear predictive coefficients (LPC).To demonstrate our model we have generated data base with 200 speakers for training the data and 50 speech samples for testing the data, the speech samples are considered for testing are segmented into frames of both long duration and short duration of five seconds,ten seconds and fifteen seconds respectively. The accuracy of the developed methodology is calculated and above 88% of accuracy is observed.

Keywords: speaker recognition, MFCC, LPC, Generalized Gamma Distribution, feature extraction.

1. INTRODUCTION
Speech is a medium of communication in the present world it helps to recognize a speaker based on the speech signal, it has a wide applications in security for authenticating a speaker .In any speech system, it is customary to identify a speaker based on the speech , The speech recognition system, basically consists of preprocessing, feature extraction and classification.Inorder to extract the features, MFCC are mostly used [1][2],because they are less vulnerable to noise and gives less variability. LPC features helps to recognize the speech samples in low acoustics also. This article includes both MFCC and LPC coefficients for effective recognition. Any speaker identification system can be divided into text-independent and text-dependent methods[3].In text-dependent method, the speaker recognition is carried out based on utterance of the exact word from different speakers.While in text-independent method, specific characteristics of the speakersspeech are extracted, irrespective of what one is saying the choice of using text-dependent speech or text-independent speech is application specific[4].the main applications of speaker recognition are verification and identification[5]. In any speaker recognition system, the primary task is recording the speaker's voice and extracting the feature vector set, this process is called enrollment and during the identification phase a speech test signal is compared with these feature vectors. The paper is organized as follows,section-2 of the paper deals with feature extraction. in section-3, the generalized gamma mixture model is proposed ,section-4 deals with the estimation of the model parameters through E-M Alogorithm.section-5 of the paper deals with speaker identification. And finally in section-6, the experimental results are presented and section -7 describes about conclusions.

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 signalsand 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[6]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 [7]. To have a more accurate feature extraction maximum posterior estimation models are to be considered [8].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 [9]. Generalized gamma distribution is defined in terms of scale and shape parameters [10]. 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[11]. 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

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

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

The updated equation for the scale parameters is given by

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

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

5. SPEAKER IDENTIFICATION

The main purpose of any speech recognition system is to recognize the speaker based on the speech signal from the group of S-speakers = [1,2,3,….S] which is represented by the generalized gamma distribution. The speech input signal is pre-processed to eliminatethe noise and speech signals are trained. Inorder to identify a speaker a sample is drawn from the S-speakers and compared with the training set and the speaker is classified using the generalized gamma distribution using the feature vector MFCC and LPCs. Another verification method, which is carried out here is the speech signals are stratified into small frames each of 5 seconds and above and using these frames, further recognition is done.

6. EXPERIMENTAL RESULTS

In order to demonstrate our method we have considered a training set with 200 speech signals and for testing we have considered 50 samples of both genders. The test signals are considered and segmented into two parts, short frame and continuous long frames. The short frame are obtained by slicing the speech signal into samples of 5 seconds, 10 seconds and above, long sequences are of 35 sec and 45 sec duration. The accuracy of the identification system is computed using percentage of correct identification

pci=percentage of correct identification
=(no of correctly identified speakers/total no of speakers)*100

7. DISCUSSION

During the training process, the MFCC and LPC coefficients for each speaker are extracted and the data is modeled by using Generalized Gamma mixture model. In order to recognize the speaker, two different frames of the speech signal are considered, the short frame sequence and the long frame sequence, the short frames are of 5 sec, 10 sec,15sec,20 seconds duration and the long frame is above 35 sec frame duration. The percentage of identification is carried out by using the formula given above and the graphical output is presented. The percentage of correct identification results are given as a function of length of the test utterance, we have tested with short frame signals and long frame signals and it is observed that if the size of the speech signal is more then the accuracy of recognition is better. This is very useful in applications, where a decision is to be taken using short frame sequences. The accuracy of the developed model for short frame sequence is around 80% and above 88% in case of long frame durations. The recognition accuracy is also tested by varying the number of trained speakers.

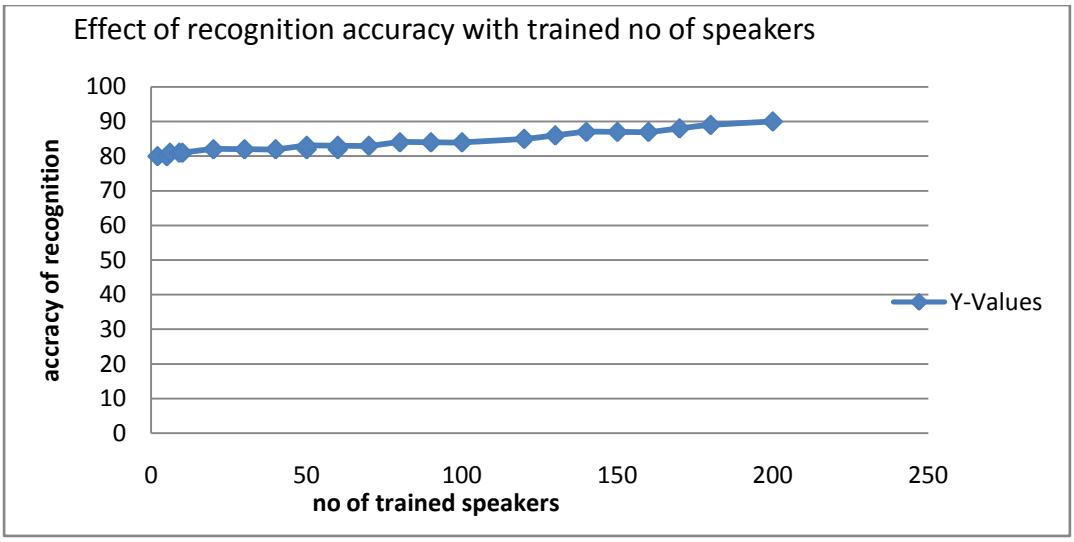

Figure 1: Effect of Recognition accuracy with trained no. of Speakers

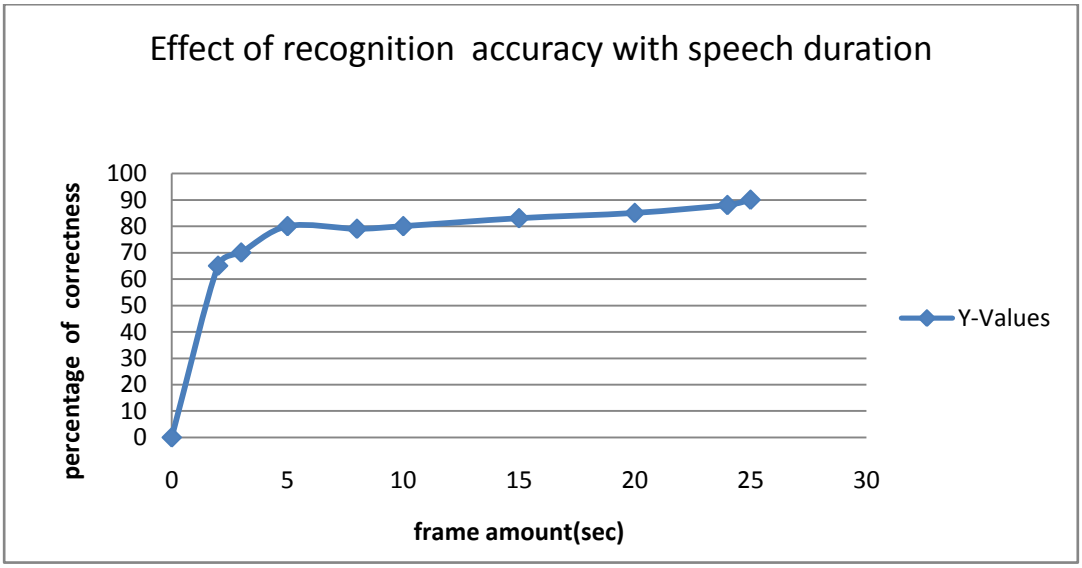

Figure 2: Effect of Recognition accuracy with Speech Duration

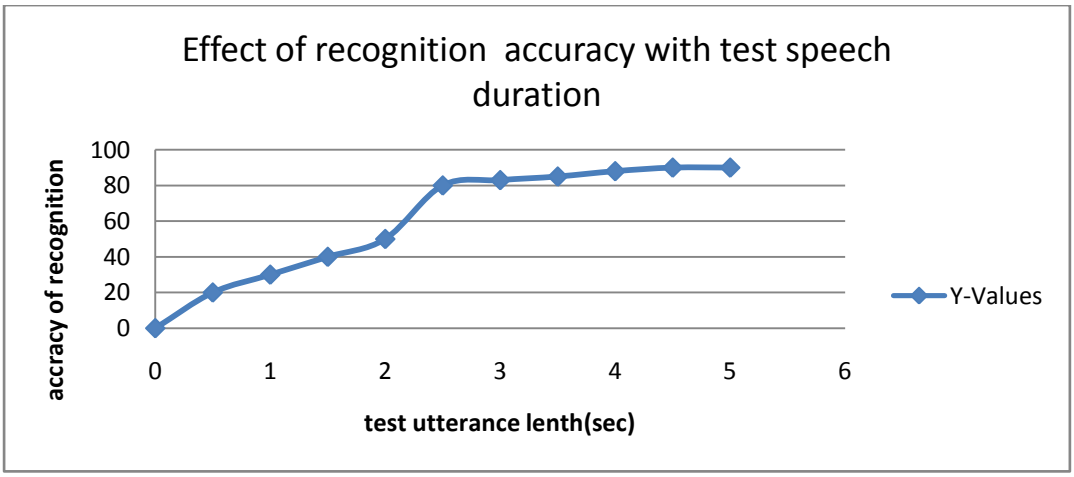

Figure 3: Effect of Recognition accuracy with Test Speech Duration

Table 1: Accuracy Statistics based on Speech Samples

No of speech samples	Trained speech duration/ frame amount(sec) (experimented fixed and variable sizes)	Test utterance duration(sec)	Recognition accuracy (%)
5	5	3	80
		5	82
10	10	5	80
		10	82
20	15	3	82
		5	82
30	20	5	84
		10	84
50	25	10	83
		12	83

The results examined are presented in graphical form and tabular form, and from the above table it can be easily seen that if the sample size is more the accuracy is better. In all the situations, it is observed that this model identifies the speaker correctly with above 80% accuracy rate.

8. CONCLUSIONS

In this paper we have introduced and analyzed a new model for speaker identification and recognition based on generalized gamma distribution .the experimentation to validate the model is constructed with speech samples of short duration and long duration and the identification is carried out by splitting the speech samples with different time duration. The classification accuracy is calculated by using ***pci***.The experiment results show that the developed method could identifies speaker accurately and the recognition rate is observed to be above 88%.

9. REFERENCES

[1] Lawrence R.Rabiner,(1989), A Tutorial on HMM & Selected Applications in speech Recognition, proceedings of IEEE vol-77,No-2,feb-1989,pp257-284.

[2] Corneliu Octavian. D, I. Gavat, (2005), Feature Extraction Modeling &Training Strategies in continuous speech Recognition For Roman Language, EU Proceedings of IEEE Xplore,EUROCN-2005,pp-1424-1428.

[3] Sunil Agarwal et al,(2010),Prosodic Feature Based Text-Dependent Speaker Recognition Using machineLearningAlgorithm,InternationalJournalofEngg.sc &Technology,Vol:2(10),2010,pp5150-5157.

[4] .Md. RashidulHasan, et al(2004),Speaker identificationusing Mel Frequency Cepstral Coefficients,**3**rd International Conference on Electrical & Computer Engineering,ICECE 2004, 28-30 December 2004, Dhaka, Bangladesh.

[5] Douglas.A.Reynolds,member,IEEE and Richard.C.Rose,member,IEEE, Robust text-Independent Speaker Identification Using Gaussian Mixture Speaker Models,IEEE Transactions on speech and audio processing,vol.3No.1,january1995.

[6] Eddie Wong and SridhaSridharan ,(2001),Comparison of Linear Prediction Cepstrum Coefficients and Mel-Frequency Cepstrum Coefficients for Language Identification,lnternational Symposium on Intelligent Multimedia, Video and Speech Processing. May 24 2001 Hong Kong.

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

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

[9] 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.

[10] EwaPaszek, The Gamma & chi-square Distribution, conexions, Module-M13129.

[11] 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.

[12] Rajeswara Rao. 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.

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