# SPEAKER RECOGNITION MODEL BASED ON GENERALIZED GAMMA DISTRIBUTION USING COMPOUND TRANSFORMED DYNAMIC FEATURE VECTOR

K Suri Babu[1], Srinivas Yarramalle[2], Suresh Varma Penumatsa[3]

[1]Scientist, NSTL (DRDO),Govt. of India, Visakhapatnam, India
[2] Dept. of IT, GITAM University, Visakhapatnam. India.
[3] Aadikavi Nannaya University, Rajahmundry, India..

*{[1]suribabukorada2000@gmail.com, [2]sriteja.y@gmail.com }*

## ABSTRACT

*In this paper, we present an efficient speaker identification system based on generalized gamma distribution. This system comprises of three basic operations, namely speech features classification and metrics for evaluation. The features extracted using MFCC are passed to shifted delta cepstral coefficients (SDC) and then applied to linear predictive coefficients (LPC) to have effective recognition. To demonstrate our method, a database is generated with 200 speakers for training and around 50 speech samples for testing. Above 90% accuracy reported.*

## KEYWORDS

*Speaker identification, MFCC, LPC, Generalized Gamma, Shifted Delta coefficients*

## 1. INTRODUCTION

With the recent advancements in Technology, lot of information can be stored in the databases, in any of the format such as audio, video or text. Therefore, searching the exact information is difficult task [1]. Automatic indexing to the multimedia content can solve this problem. To retrieve speech signal from this Meta data is a crucial task.

The speech signal to be retrieved is considered and is divided into small streams (segments) and the features are to be extracted. In order to extract features, MFCC are mostly proffered [3], [4] since they are less vulnerable to noise and give less variability. In order to have effective recognition it is needed to extract the first and second order time derivatives of cepstral features, that is delta and delta-delta features[5], but these features will be effective for short term speech samples, for long term features shifted delta coefficients (SDC) are well proffered [6], [7], [8].

Hence in this paper, we develop a model for speaker identification, where the features obtained from MFCC are converted to shifted delta coefficients and also by converting MFCC to delta coefficients. It is observed that the features obtained from MFCC followed by SDC outperform MFCC followed by delta.

DOI : 10.5121/ijesa.2012.2308

The paper is organized as follows, the section-2 of the paper discuses about feature extraction, in section-3 generalized gamma distribution is proposed. Section -4 deals with experimental results. Finally, in section-5 conclusions are presented.

## 2. FEATURE EXTRACTION

In the proposed work we have considered the speech signals with frame amount of 20 to 30 Ms. and the window analysis is shifted by 10ms. Each frame is transformed using cepstral coefficients such as linear prediction coding and Mel frequency cepstral coefficients (MFCC) MFCC s are considered as they are based on the known variation of human ear critical band width with the frequency, each frame is transformed into 12 MFCC and a normalized energy parameter each frame consist of 39 columns including first and second derivatives, i.e. Delta and double Delta. In feature extraction the speech waves stored in wav format each converted to a parametric form. The speech signals remains stationary between the time intervals 5 Ms. to 100 Ms. and the changes observed over long periods i.e. 0.2sec or more. Therefore to identify the speech variation in short time sequence, cepstral analysis is mostly preferred hence MFCC are considered Linear prediction coding (LPC) coefficient helps to extract signal more effectively in the presence of noise and when the speech signal is of very short duration .So in this thesis we have exploited MFCC combined with LPC to have effective feature vector identification.

In speech analysis, significant information spread over few 100s of milliseconds there may be overlaps and the speech signals are not completely separated in-time. These overlaps may result in to ambiguities at the time of classification to overcome this it is assumed to extract the features between the frequencies 2 to 16 Hz, a maximum of 4 Hz.

In order to distinguish these signals in the overlapping situations Delta features are mostly preferred. In delta coefficients we obtained the derivative to estimate the differences in the speech trajectories. Delta-Delta coefficients are also considered for every longer temporal context. But these features will be effective for short term speech samples, for long term features shifted delta coefficients (SDC) are well proffered. The features obtained from MFCC are converted to shifted delta coefficients. It is observed that the features obtained from MFCC followed by SDC outperform MFCC followed by delta. SDC reflects the dynamic cepstral features along with pseudo-Prosodic feature behavior.

## 3. SPEAKER RECOGNITION ALGORITHM

The steps to be followed for recognizing the speaker effectively are given under

Step1:

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

Step2:

Pre-emphasis the speech signals to remove silence and noise.

Step3:

Identify the compound feature vectors feature vector of these speech signals by using MFCC, LPC, SDC, Delta, and Delta-Delta.

Step4:

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

Step5:

Same procedure is followed for test sequence.

Step6:

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

Step7:

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

## 4. 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 Gaussian mixture model is that it relies exclusively on the approximation and low in convergence, and also if Gaussian mixture model 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, a 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

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

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 parameter of the generalized gamma distribution helps to classify the speech signal and identify the speaker accurately.

## 5. EXPERIMENTAL RESULTS

During the training phase, the signal must be preprocessed and the features are extracted using MFCC. In order to have an effective recognition system we have sampled the data into short speech samples of different time frames and the MFCC features that are extracted are converted delta coefficients and shift delta coefficients. It is observed that MFCC combined delta coefficients could not effectively recognize the speech samples as compared to that of MFCC combined with SDC. The output is then fed to LPC (linear predictive coefficients). The features extracted are then given as input to the classifier that is generalized gamma distribution, using these feature set, the generalized gamma distribution is effectively recognized. The speech samples that are obtained from MFCC-SDC-LPC, it can also be seen that as and when the sample size is increased, these features that are extracted helps to classify the speakers most effectively. The results are presented in both tabular and graphical formats.

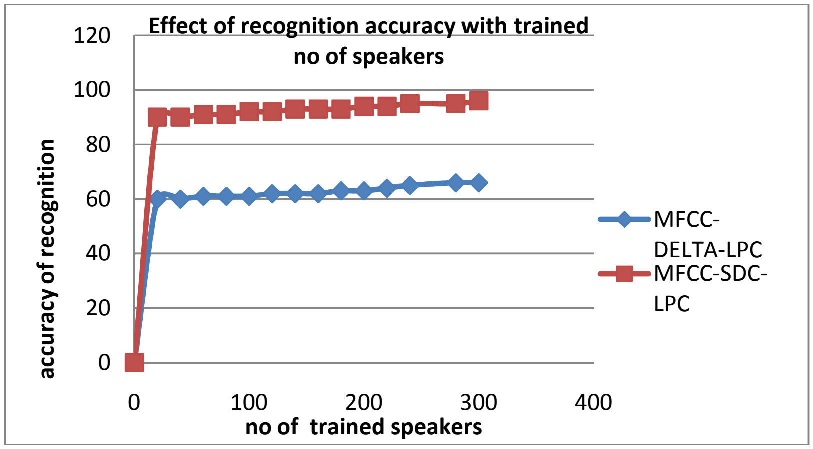

Fig.1: Effect of Recognition accuracy with trained dataset

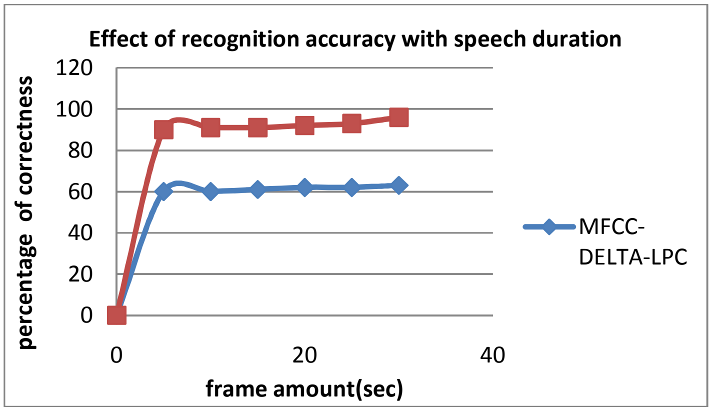

Fig.2: Effect of Recognition accuracy with Speech duration

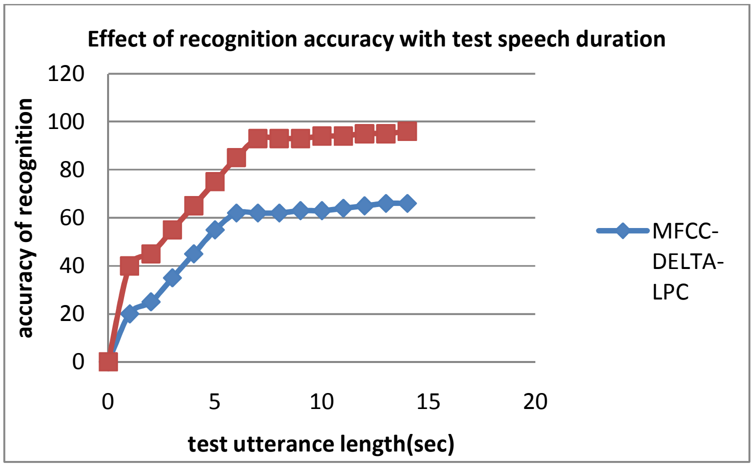

Fig.3: Effect of recognition accuracy with test Speech duration

Table 1: Statistical data showing accuracy %

| | **No of trained speakers** | **Frame amount (sec)** | **Test utterance length (sec)** | **Recognition Accuracy (%)** |
|---|---|---|---|---|
| **MFCC-DELTA-LPC** | 0 to 50 | 0 to 5 | 0 to 5 | Less than 60 |
| | 50 to 100 | 5 to 10 | 5 to 10 | Around 60 |
| | 100 to 300 | 10 to 30 | 10 to 15 | Above 62 |
| **MFCC-SDC-LPC** | 0 to 50 | 0 to 5 | 0 to 5 | Less Than 80 |
| | 50 to 100 | 5 to 10 | 5 to 10 | Around 85 |
| | 100 to 300 | 10 to 30 | 10 to 15 | Above 90 |

From the above figures and table (Fig.1 to Fig.3 and Table 1), it could be easily seen that the MFCC-SDC-LPC outperforms MFCC-Delta-LPC and over all recognition rate is above 90% is seen in the developed model.

## 5. PERFORMANCE EVALUATION

In order to evaluate the performance of the developed model various metrics such as Acceptance Rate (AR), False Acceptance Rate (FAR), and Missed Detection Rate (MDR) are considered. The various formulas for evaluating the metrics are given below.

$$Acceptance\ rate = \left(\frac{total\ no\ of\ accepted}{total\ no\ of\ speakers}\right) * 100$$

$$False\ acceptance\ rate = \left(\frac{total\ no\ of\ speakers - total\ no\ of\ sccepted}{total\ no\ of\ speakers}\right) * 100$$

$$Misssed\ detection\ rate = \left(\frac{total\ no\ of\ missed\ to\ recognize}{total\ no\ of\ speakers}\right) * 100$$

The developed model is tested for accuracy using the above metrics mentioned in equations

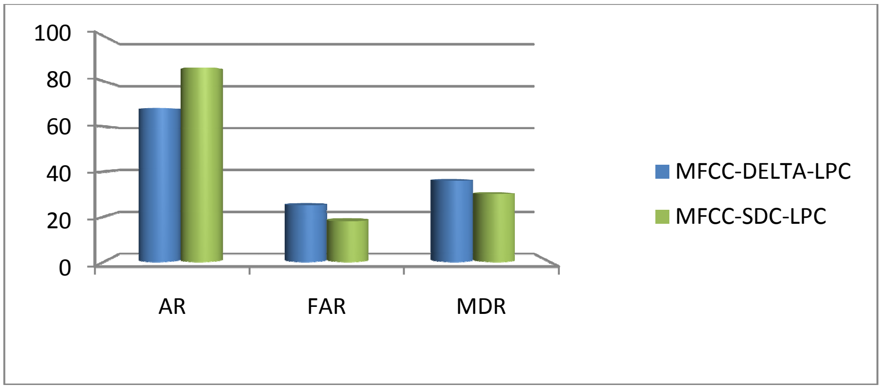

Fig .4: feature vector performance evaluation

The above Fig.4 shows the performance with metrics as Acceptance rate(AR) ,False Acceptance Rate(FAR) and Missed Detection Rate(MDR) by considering Various combinations of feature vectors as MFCC-DELTA-LPC, ,MFCC-SDC-LPC and it shows performance comparison of feature vectors using generalized gamma distribution with three metrics . From the above Fig.4, It can be clearly seen that MFCC-SDC-LPC feature vector out performs than all the combinations of feature vectors.

## 5. CONCLUSIONS

In this paper, we have developed a new model for speaker identification based on generalized gamma distribution. The speeches are extracted using MFCC are combined with delta coefficients followed by LPC and also MFCC combined with SDC followed by LPC. The model is demonstrated a database of 200 samples and tested with 50 samples, the accuracy is around 90% and proved to be efficient model.

## REFERENCES

[1] Marko kos, Damjan Vlaj,Zdravko Kacic,(2011) "Speaker's gender classification and segmentation using spectral and cepstral feature averaging", 18th International Conference on Systems, Signals and Image Processing - IWSSIP 2011.

[2] J.Razik,C.SEnac,D.Fohr,O.Mella and N Parlangeau-Valles,(2003) "comparision of two speech/Music segmentation systems for audio indexing on Web",in Proc WMSCI'03,Florida,USA,July2003.

[3] 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.

[4] Sunil Agarwal et al,(2010), "Prosodic Feature Based Text-Dependent Speaker Recognition Using machine Learning Algorithm",International Journal of Engg.sc &Technology, Vol:2(10), 2010,pp5150-5157.

[5] Dayana Ribas Gonzalez,Jose R.Calvo de Lara(2009), "Speaker verification with shifted delta cepstral features:Its Pseudo-Prosodic Behaviour", proc I Iberian SLTech 2009.

[6] P.A.Torres-Carrasquillo and E.Singer and M.A.Kohlerand.R.J.Greene and A.Reynolds and J.R.Deller Jr.(2002) "Approches to language Identification Using GAusian Mixture Models and Shifted delta cepstral features", Proc of ICSLP2002,pp89-92.

[7] T.Kinnunen.C.W.E.Koh,L.Wang.H.Li,E.S.Chang,(2006) "Temporal discrete cosine trans-form: Towards longer term temporal features for speaker verification", Proc of ICSLP 2006.

[8] J.Calvo andR.Fernndez and G.Hernndez,(2007) "Channel/Handset Mismatch Evaluation in Biometric Speaker Verification using Shifted Delta Cepstral Features".Proc of CIARP 2007.LNCS 4756.PP96-105.