International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 5– No. 5,April 2013 – www.ijais.org

Dimensionality Reduction in Feature Vector using Principle Component Analysis (PCA) for Effective Speaker Recognition

Suri Babu Korada[1] Anitha.Y[2] Anjana.K.K.V.S[3]

[1]Scientist, NSTL (DRDO), Gov. of India
Visakhapatnam, Andhra Pradesh, India

[2]Department of IT, MVGR college of Engineering,
Vizianagaram, Andhra Pradesh, India

[3]Department of MCA, MR PG COLLEGE
Vizianagarm, Andhra Pradesh, India

ABSTACT
This paper describes analysis of a speaker recognition model based on Generalized Gamma Distribution (GGD) using PCA. The proposed work mainly concentrates on the feature vectors that are generated from the speech signals contain high dimension data, but to model a speech and recognize a speaker finite speech samples which plays significant role in speech analysis are sufficient, hence it necessary to reduce dimension of the data. The PCA is considered for this purpose, it converts high dimension speech signal in to a low dimension speech signal by transforming the un-correlated components of the speech signal. PCA not only reduces the correlation among feature vectors but also the speech signal. The feature vectors are modeled by extracting MFCC followed by PCA for dimensionality reduction.

Keywords
GGD,PCA,MFCC, Dimensionality reduction;

1. INTRODUCTION

Principal component analysis is unsupervised model of dimensionality reduction which takes the linear data for reducing the dimensionality. Speech[1] signal data is converted into numerical data using cepstral coefficients (MFCC)[2][3]. These values are stored in a matrix. The PCA extracts the orthogonal principal components from this matrix and the dimensionality is reduced by calculating the 2nd order moments for the low frequency values and decorrelating these values. The PCA is based on the approximation of karhunen-loeva transformation (KLT). This process highlights the principal components by calculating the Eigen values which are of high dimension and convert then to low dimensional values by retaining the original values.

Principal component analysis (PCA) is based onMean Square Error, Co-variance matrix and Time complexity.Mean square error is a methodology of compressing the high dimensional features to low dimensional features retaining the originality, and the process is done in linear fashion. The co-variance matrix is obtained directly from the original matrix by computing the Eigen values, and under these two features, the computational time reduction and space reduction are integral part.

The steps involved in the calculation of PCA are given below:

The speech signal extracted from the microphone is converted into .wav format is processed to eliminate noise and silence. This processed signal is given to MFCC to obtain the numeric equivalent of the speech signal. The speech signal in numeric form is represented as $[x_1,x_2,\ldots\ldots x_N]$ as a N-dimensional feature vector[4] . The mean of the speech signal is to be obtained. This can be done by considering the feature vector and for the speech signals the mean, $\overline{x}$ are computed as

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \text{---1}$$

The co-variance is computed using the formula

$$C = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})(x_i - \overline{x})^T \text{---2}$$

The Eigen values and Eigen vectors can be obtained by decomposing the co-variance matrix denoted in equation 2 and are computed as

$$CE_K = \lambda_i E_k, k = 1,2, \ldots.(N-1)\text{---3}$$

Where $E_k$is the kth Eigen vector and λ_i is Eigen value corresponding to the Eigen vector.

For the purpose of dimensionality reduction, Eigen vector is selected because, it is largest among the calculated Eigen values, and it is obtained by using

$$M_{PCA} = E^T\text{--4}$$

Where $E^T = E_1, E_2, \ldots., E_{N-1}$

And the required reduced feature set is obtained by transforming the feature vector space to M_{PCA}.

2. GENERALIZED GAMMA DISTRIBUTION

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 approximation and low in convergence, and also if GMM is used the speech and the noise coefficients differ in Magnitude . To have a more accurate feature extraction maximum posterior estimation models are to be considered. 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. Generalized gamma distribution is defined in terms of scale and shape parameters. The generalized gamma mixture is given by

The probability density function of generalized gamma distribution [5] is given by

$$f(x,k,c,a,b) = \frac{c(x-a)^{ck-1} e^{-\left(\frac{x-a}{b}\right)^{c}}}{b^{ck}\,\Gamma(k)} \text{--5}$$

Where a, b, c, k are called the gamma variants and c, k are called shape parameters such that

c, k >0.a is called location parameter , b is called shape parameter with a ,b >0.

3. ALGORITHM FOR SPEAKER IDENTIFICATION

Step 1: Obtain the speech signals, using Microphone and store in .wav format.

Step 2: Calculate MFCC and obtain numeric coefficients.

(a) The speech signal is dependent of tone and to understand the spectral properties of these signals, obtains the Fourier transform.
(b) Map these coefficients using Mel scale with triangular overlapping windows.
(c) Take the logarithm of 2(b).
(d)
(e) Obtain the amplitude sequences of MFCC coefficients.

Step 3: Apply PCA, to reduce dimensionality.

Step4: Apply Generalized gamma distribution, to model the parameters.

4. EXPERIMENTATION

During the training phase, the signal must pre- processed 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 given for principle component analysis (PCA). We experimented with various feature vector combinations. 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 given 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 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. Feature vector combinations we considered for comparing with MFCC-PCA are: MFCC-DELTA-LPC, MFCC-LPC, and MFCC-SDC-LPC.

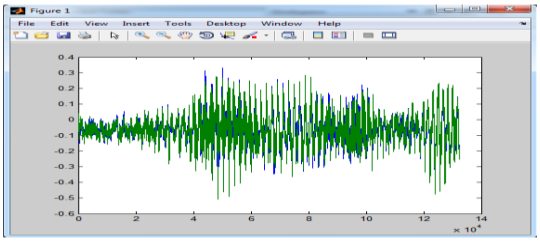

Fig 1: Recorded speech in .wav

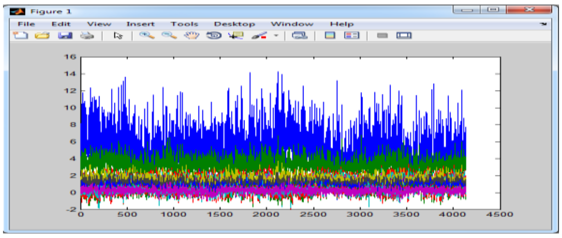

Fig 2: MFCC feature vector

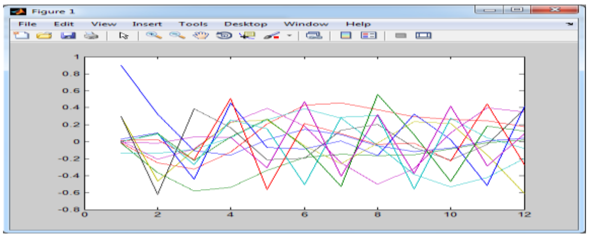

Fig 3: MFCC- PCA combination signal

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.

$$Acceptance\ rate = \left(\frac{total\ \ no\ of\ speakers}{total\ \ no\ of\ accepted}\right) * 100 \qquad \text{--6}$$

$$False\ acceptance\ rate = \left(\frac{total\ \ no\ of\ speakers}{total\ \ no\ of\ speakers + total\ no\ of\ sccepted}\right) * 100 \qquad \text{--7}$$

$$Misssed\ detection\ rate = \left(\frac{total\ \ no\ of\ missed\ \ to\ recognize}{total\ \ no\ of\ speakers}\right) * 100 \qquad \text{--8}$$

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

The results obtained are presented in fig 4 and fig .5

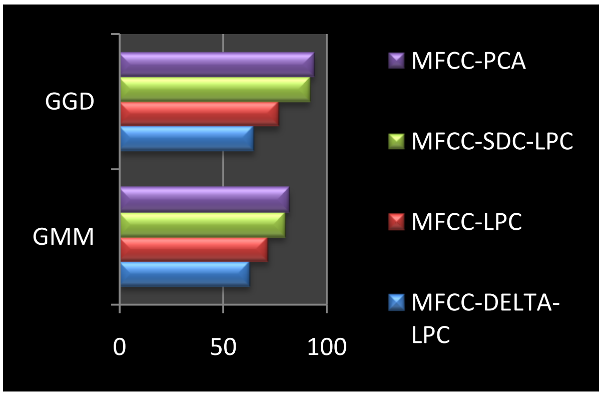

Fig.4

The above Fig.4 shows the performance of Gaussian mixture model (GMM)and generalized gamma distribution(GGD) with acceptance rate as metric with various combinations of feature vectors as MFCC-DELTA-LPC[6],MFCC-LPC,MFCC-SDC-LPC,MFCC-PCA and it shows performance comparison of feature vectors and all generalized gamma distributions cases are compared with existing model Gaussian mixture model (GMM)[7]. From the above Fig.4 itcan be clearly seen that MFCC-PCA feature vector out performs than all the combinations of feature vectors. And generalized gamma distribution with MFCC-PCA performs superior than all distribution-feature vector combinations. From the above fig.4, it can be clearly seen that Generalized Gamma Distribution out performs than GMM.

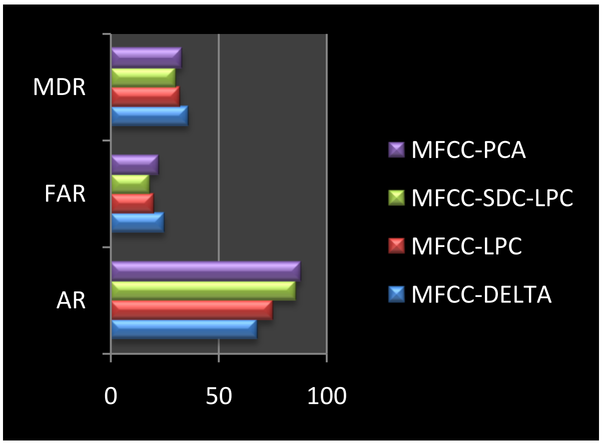

Fig.5

The above Fig.5 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-LPC,MFCC-SDC-LPC,MFCC-PCA and it shows performance comparison of feature vectors using generalized gamma distribution with three metrics . From the above Fig.5, It can be clearly seen that MFCC-PCA feature vector out performs than all the combinations of feature vectors.

6. FUTURE SCOPE:

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, delata-delata and shifted delta coefficientsfollowed 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.further we can extend thiswork to recognize speaker in teleconference environment.

7. REFERENCES:

[1] K.Suri Babu et al "Speaker recognition model based on generalized gamma distribution using compound transformed dynamic feature vector", communicated to Asian journal of science and technology"International Journal of Embedded Systems and Applications (IJESA) Vol.2, No.3, September 2012.

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

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

[4] Suribabukorada et al(2011), "Text Independent Speaker Recognition Model Based On Gamma Distribution Using Delta, Shifted Delta Cepstrals" published in Springer link conference (SPPR-2012).

[5] CorneliuOctavian.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.

[6] suribabukorada et al,(2011), "Text Dependent and Gender Independent Speaker Recognition Model based on Generalizations of Gamma Distribution" International Journal of Computer Applications (0975 – 8887) Volume 35– No.6, December 2011

[7] 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.

[8] 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.

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

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

[11] DayanaRibasGonzalez,JoseR.Calvo de Lara(2009),"Speaker verification with shifted delta cepstral features:Its Pseudo-Prosodic Behaviour"proc I Iberian SLTech 2009.