ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

TEXT INDEPENDENT SPEAKER IDENTIFICATION WITH PRINCIPAL COMPONENT ANALYSIS

D. Vijendra Kumar[1], K.Jyothi[2], Dr.V.Sailaja[3], N. M. Ramalingeswara rao[4]

PG student, ECE Department, Godavari Institute of Engineering & Technology, Rajahmundry, India[1]

Associate Professor, ECE Department, Godavari Institute of Engineering & Technology, Rajahmundry, India[2]

Professors, ECE Department, Godavari Institute of Engineering & Technology, Rajahmundry, India[3]

Assistant Professor, ECE Department, Godavari Institute of Engineering & Technology, Rajahmundry, India[4]

Abstract- Principal Component analysis (PCA) is useful in identifying patterns in data, and expressing data in a manner which highlights their similarities and differences. This concept was extracted to reduce high dimensional Mel's Frequency Cepstral Coefficients (MFCC) into low dimensional feature vectors. Since MFCC's are high in dimensions and truncation of these dependent coefficients may lead to error in identification of speaker's speech recognition. In this paper text independent speaker identification model is developed by combining MFCC's with PCA to obtain compressed feature vectors without losing much information. Generalized Gaussian Mixture Model (GGMM) was used as modeling techniques by assuming the new feature vectors follows (GGMM) [Reynolds, (1995)] [7]. The experiment was done with 40 speakers with 10 utterances of each speaker locally recorded database.

I. INTRODUCTION

The performance of speaker identification and recognition has achieved high accuracies over past years under clean speech and well matched conditions. But it is still a challenge for designer who designs systems for complex work environment and for voice operated service in many commercials areas. To increase the performance of the system we need to concentrate on feature extraction, so that they retain the characteristics of speech while making their size compact.

In general speaker reorganization system consists of two key modules: Feature extraction and classification. The classification module needs the extracted features from the feature extraction module. The feature extraction module is crucial in any speaker recognition systems [Smith, 2002][13].

Mel-Cepstral coefficients are dependent on each other and are large in number. Truncation of some of these coefficients may lead to an error during modeling. So, to have a robust model for speech spectrum it is necessary to reduce the dimensions and avoid dependencies. It can be achieved by processing MFCC's each speaker speech spectra with PCA. Integration of PCA with MFCC will makes feature vectors robust and compact [Cardoso, 1996] [Ding, 2001][3][4].

PCA identifies patterns in data, and expresses data in such a way as to highlights their similarities and differences. Since the patterns in data are hard to find in data of high dimensions, where luxury of graphical represents is not available, PCA is powerful tool for analyzing data [Qin Jin et al 2011][9]. The major advantage of PCA is that one pattern were found in data, it compress the data, by reducing the number of dimensions, without much loss of information's.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

In this paper a text independent speaker identification method is developed and analyzed by integrating PCA with MFCC using GGMM. It was also provides the detail description of how MFCC and PCA is implemented. Here, it is assumed that feature vector of speech spectrum (MFCC) follows GGMM, Estimations of model parameters is done by using the updated equation of EM algorithm. Bayesian frame work is used as speaker identification algorithm. The performance of the present model is observed by conducting experimentation with 40 speakers with 10 utterances of locally recorded database.

II. FEATURE VECTOR EXTRACTION

Here, it explains how extraction of feature vectors from speaker speech spectra was obtained. MFCC have been the most popular low level features for speaker recognition and speech recognition system. MFC is a representation of short-term power spectra of a signal, based on a linear cosine transform of a long power on a nonlinear Mel-Scale frequency Cepstral is that in the MFC, the frequency bands are equally spaced on Mel-scale, which approximate the human auditory systems response.[3][4]

$$fmel = 2595\log[1+(f/700)] \quad \ldots\ldots(1)$$

Where, fmel is the subjective pitch in Mel's corresponding to f, the actual frequency in HZ. This leads to the definition of MFCC, a base acoustic feature for speech and speaker recognition applications. Principal component analysis is an approximation of Karhunen-Loeve Transform (KLT) algorithm used to extract few first eigenvectors which mostly retain the variations presents in all original variables. It is a mathematical method used to orthogonally project the features of high dimensional space into low dimensional subspace

.Principal component analysis exhibits three important features: (1) it is optimal in terms of mean squared error, i.e. it is a linear scheme used for compressing a set of high dimensional vectors into low dimensional vectors and then reconstructing them. (2) The parameters of the model can be directly obtained from the data by diagonlizing the covariance matrix. (3) Using PCA, operations used to compute the model parameters require only matrix multiplications reducing complexity and time consumed. In spite of all these advantages, PCA however has some shortcomings. It is a naive method used to compute the principal component direction and ends up having trouble with large number of data points and high dimensional data [Somervuo, 2003]. Principal component of the data set can be obtained by computing the covariance matrix of the data set and then finding the eigenvectors corresponding to the largest eigenvalues. Suppose there are N feature vectors given as {x1,x2,……, xN}. The mean of the feature vectors is represented by $\bar{x}$ and is calculated as [Smith, 2002][13].

$$\bar{x}=\frac{1}{N}\sum_{i=1}^{N} x_i \quad \ldots\ldots(2)$$

The covariance matrix C is a square and symmetric matrix of order N*N and can be computed as [Smith, 2002][13], [Shlens, 2003][12].

$$C=\frac{1}{N}\sum_{i-1}^{N} \tilde{x}_i\, \tilde{x}_i^{\,T}, \quad \ldots\ldots (3)$$

Where $\tilde{x}_i = x_i - \bar{x}$.Covariance matrix C is also observed to correlation and data dispersion. Eigen value decomposition of the covariance matrix results in eigenvalues and eigenvectors [Rabiner, 1993][10]. Eigenvectors can be computed from the following equation [Smith, 2002][13], [Shlens, 2003][12]

$$CV_k=\lambda_k V_k,\ k=0, 1\ldots N-1 \quad \ldots\ldots (4)$$

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

Where V_k is the k[th] eigenvector and λ_k is the corresponding eigenvalue. Eigenvectors corresponding to M (M<N-1) largest eigenvalues are selected to reduce the dimensions of the data set. The transformation or projection matrix is defined as the transpose of thus obtained eigenvector matrix and is given as [Smith, 2002][13], [Shlens, 2003][12].

$$W_{PCA}=V \quad \text{..... (5)}$$

Where $V^T= V_0, V_1,\ldots\ldots,V_{M-1}$

The final step is to derive the new data set, the projection of the feature vectors on to the space formed by PCA. This is simply established by multiplying the projection matrix with the original dataset (mean adjusted data). This can be represented as [Smith, 2002][13].

$$\text{New Dataset} = W_{PCA} * \text{MeanAdjustedoriginaldata} \quad \text{..... (6)}$$

Extraction procedure of feature vectors is shown in two steps.Step1: It shows how MFC coefficients were obtained from speaker utterance by a flow chart.

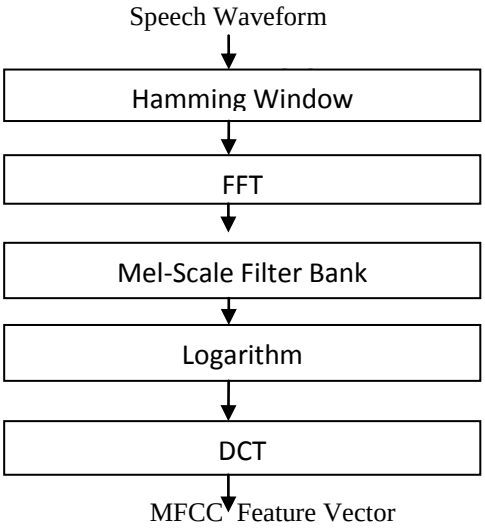

Fig 1 MFC Coefficient Block Diagram

Step 2: By applying PCA to Mel's frequency Cepstral coefficients new feature vectors of each speaker is obtained.

III. SPEAKER IDENTIFICATION MODEL WITH GENERALIZED GAUSSIAN DISTRIBUTION

In this section we describe the speaker identification process. Fig.2 & 3 represents the block diagram of the proposed text independent speaker identification system with Generalized Gaussian distribution using integrating PCA in the system after feature extraction. Here it is assumed that the feature vector (after processing the MFCC with PCA) follows a multivariate Generalized Gaussian mixture model. The motivation for considering the Generalized Gaussian mixture models is that the individual component densities of a multi model density like the mixture model may model some underlying set of acoustic processes. It is reasonable to assume the acoustic space corresponding to a speaker voice can be characterized by acoustic classes representing some broad phonetic events such as vowels, nasals or fricatives. These acoustic classes reflect some general speaker dependent vocal tract configurations that are useful for characterizing speaker identity.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

The spectral shape of its acoustic class can in turn be represented by the mean of its component density and the variation of the average spectral shape can be represented by the co-variance matrix. Therefore the entire speech spectra of the each individual speaker can be characterized as a M component Finite Multivariate Generalized Gaussian mixture distribution

The probability density function of the each individual speaker speech spectra is

$$p(\overrightarrow{x_t}|\lambda) = \sum_{i=1}^{M} a_i b_i(\overrightarrow{x_t}|\lambda) \quad \text{.....(7)}$$

Where, $\overrightarrow{x_t} = (x_{tij})$ j=1,2,…,D; i=1,2,3…,M; t=1,2,3,…,T is a D dimensional random vector representing the MFCC vector. λ is the parametric set such $\lambda = (\mu, \rho, \Sigma)$

a_i is the component weight such that $\sum_{i=1}^{M} a_i = 1$.

$b_i(\overrightarrow{x_t}|\lambda)$ is the probability density of the i[th] acoustic class represented by new vector of speech data and the D-dimensional Generalized Gaussian (GG) distribution [M. Bicego et al (2008)][2] and is of the form

$$b_i(\overrightarrow{x_t}|(\mu,\rho,\Sigma)) = \frac{[\det(\Sigma)]^{-1/2}}{[x(\rho)A(\rho,\sigma)]^D} \exp\left(-||\frac{\Sigma^{\frac{1}{2}}(\overrightarrow{x_t}-\overrightarrow{\mu_i})}{A(\rho,\sigma)}||_\rho\right) \quad \text{..... (8)}$$

Where $Z(p) = \Gamma(\frac{1}{p})$ and ..... (9)

$$A(\rho,\sigma) = \sqrt{\frac{\Gamma(1/\rho)}{\Gamma(3/\rho)}}$$

And $||x||_p = \sum_{i=1}^{D}|x_i|^p$ stands for l_p norm of vector x, Σ is symmetric positive define matrix. The parameter $\overrightarrow{\mu_i}$ is the mean vector, the function of A (p) is a scaling factor which allows the var(x) = σ^2 and ρ is the shape parameter when ρ=1, the Generalized Gaussian corresponds to a laplacian or double exponential Distribution. When ρ=2, the Generalized Gaussian corresponds to a Gaussian distribution. In limiting case ρ→ + ∞ Equation mentioned above Converges to a uniform distribution in (μ-√3, μ+√3) and when ρ →o +, the distribution becomes a degenerate one when x=μ.

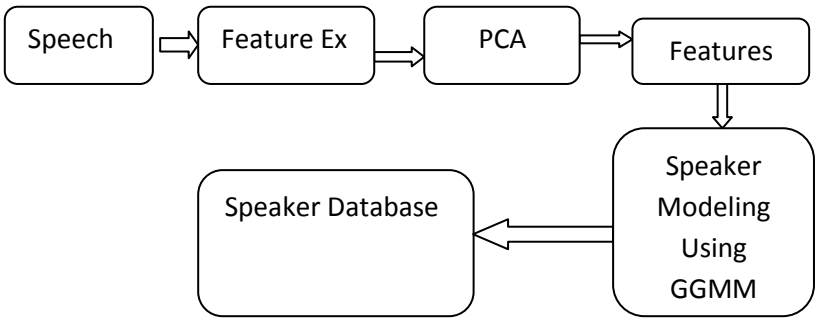

Fig 2 Training Model (Enrollments)

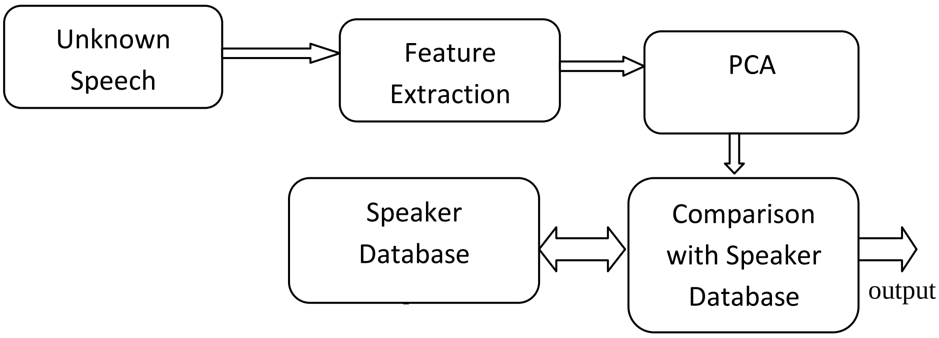

Fig 3 Testing Model (Identification)

The model can have one covariance matrix per a Generalized Gaussian density of the acoustic class of each speaker. Based on the previous studies, the diagonal covariance matrix is used for speaker model. As a result of diagonal covariance matrix for the feature vector, the features are independent and the probability density function of the feature vector is

$$b_i(\vec{x_t} \mid \lambda) = \prod_{j=1}^{D} \frac{\exp\left(-\left|\frac{x_{ij}-\mu_{ij}}{A(\rho_{ij},\sigma_{ij})}\right|^{\rho_{ij}}\right)}{\frac{2}{\rho_{ij}}\Gamma\left(1+\frac{1}{\rho_{ij}}\right)A(\rho_{ij},\sigma_{ij})} = \prod_{j=1}^{D} f_{ij}(x_{tij}) \ldots\ldots(10)$$

To find the estimate of the model parameters α_i, μ_{ij} and σ_{ij} for i= 1,2,3 …,M, j=1,2,…,D, we maximize the expected value likelihood (or) log likelihood function. Here the shape parameters 'ρ_{ij}' is estimated by the procedure given by Armando's el at (2003) [1] for each acoustic class of each speech spectra.

The updated equations of the parameters for EM algorithm are as given by Sailaja et al (2010) [11] are The updated equation for estimating αi is

$$\propto_i^{(l+1)} = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\propto_i^{(l)} b_l(\vec{x_t},\lambda^{(l)})}{\sum_{i=1}^{M}\propto_i^{(l)} b_l(\vec{x_t},\lambda^{(l)})}\right] \ldots\ldots (11)$$

Where $\lambda^{l} = (\mu_{ij}{}^{(l)}, \sigma_{ij}{}^{(l)})$ are the estimates obtained.

The updated equation for estimating μ_{ij} is

$$\mu_{ij}^{(l+1)} = \frac{\sum_{t=1}^{T} t_l(\vec{x_t},\lambda)^{(l)A(N,\rho_i)}\,(x_{tij}-\mu_{ij})}{\sum_{t=1}^{T} t_l(\vec{x_t},\lambda)^{(l)A(N,\rho_{ij})}} \ldots\ldots (12)$$

Where, A (N,P_{ij}) is some function which must be equal to unity for P_i=2 and must be equal to $\frac{1}{P_{ij}-1}$ for P_i=1, in the case of N=2. We have also observed that A (N,P_{ij}) must be an increasing function of P_{ij}.

The updated equation for estimating σ_{ij} is

$$\sigma_{ij}^{(l+1)} = \left[\frac{\sum_{t=1}^{T} t_l(\vec{x_t},\lambda^{(i)})\left(\frac{\Gamma\left(\frac{3}{\rho_{ij}}\right)}{\rho_i\,\Gamma\left(\frac{1}{\rho_i}\right)}\right)|x_{ij}-\mu_{ij}^{(i)}|^{\frac{1}{\rho_{ij}}}}{\sum_{t=1}^{T} t_l(\vec{x_t},\lambda^{(i)})}\right]^{\frac{1}{\rho_{ij}}} \ldots\ldots(13)$$

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

IV. SPEAKER IDENTIFICATION USING BAYES' DECISION RULE.

Feature vectors extracted from each test speaker were applied to the function "PCA transform" (Fig. 2) and were estimated into the space of PCA created by the associated speaker with unique speaker ID. This uses the PCA from the trained speaker model. The new feature vectors of the test utterance and the trained models were fed to a Byes classifier for identification applications which employs large group of data sets and the corresponding test speaker was identified [Dorningo's , (1997)] [5][6].

$p(\frac{i}{x_t}, \lambda)$ is called a posteriori probability for an acoustic class i and is defined by the following equation

$$P\left(\frac{i}{x_t}, \lambda\right) = \frac{p_i\, b_i\,(\widehat{x_t})}{\sum_{K=1}^{M} p_k b_k(\widehat{x_t})} \quad \text{.....(14)}$$

For a given observation sequence the main goal is to find the speaker model that has the maximum posteriori probability represented as [Reynolds, (1995)] [7]

$$\hat{S} = \max_{1<k<s} P_i(\lambda_k|X) \quad \text{..... (15)}$$

$$= \arg\max_{1<k<s} [\, P_i(\lambda_k|X)\, p_r(\lambda_k)]$$

The speaker identification system is finally computes S using the logarithms and the independence between the observations

V. EXPERMENTAL RESULTS

Developed model performance is evaluated by using a database of 40 speakers. For each speaker 10 utterances were recorded by using high quality microphone. Out of which first five are used for training data and the remaining are used for testing data.

By using front end analysis explained in section II feature vectors are calculated. The data set is divided into training set and a test set. With the test set, the efficiency of the developed model is studied by identifying the speaker with the speaker identification process given in section III.

EXPERIMENT -1:

The percentage of correct identifier is computed as

$$\text{PCI} = \%\text{ of correct identification} = \frac{No.of\ correctly\ \ identified\ \ speaker}{Total\ \ no.of\ \ speakers} * 100 \text{.....(16)}$$

It is observed that this model identifies the speakers correctly with 96.54%, which more than embedded PCA with GMM (90.50%).

EXPERIMENT -2:

40dB of Additive White Gaussian Noise (AGWN) is added to train signals where as test signals with varying SNR of 0, 10, 20, 30dB were used. It was observed from the table that with increase in SNR values, identification rate also increase.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

Table 1 : Performance of PCA with variations in SNR of the signal.

Train: 40dB	Test values of SNR in dB			
Transformation	0	10	20	30
Embedded PCA with GMM	40.33%	62.41%	76.70%	87.00%
Embedded PCA with GGMM	42.15%	65.41%	78.90%	89.15%

From this, it was concluded that MFCC+PCA using Generalized Gaussian Mixture Model (GGMM) improved speaker identification even in noisy conditions.

VI. CONCLUSION

In this paper test independent speaker identification model is developed by using Generalized Gaussian mixture model. The performance of designed model was evaluated with speech database of 40 speakers each with 10 utterances by using high quality microphone. New feature vectors were obtained by embedding PCA after computing MFC Coefficients which reduces high dimensional MFCC into low dimensional by retaining its identification factors. Important areas of speaker identification includes: Remote time and attendance logging, Home parole verification, Prison telephone usage, forensics.

REFERENCES

[1] Armando .J et al "A Practice Procedure to estimate the shape parameters in the Generalized Gaussian Distribution.2003
[2] Md M. Bicego , D Gonzalez, E Grosso and Alba Castro "Generalized Gaussian Distribution for Sequential Data Classification " ,2008,IEEE Trans.978-1-4244-2175-6.
[3] Cardoso, "Equariant adaptive Source Separation", IEEE Transaction on signal processing, 1996,vol.44 .No.12, pp 3017-3030.
[4] Ding , " Personal recognition using Independent Component Analysis ",2001, 8th international conference on Neural information process.
[5] Domingo "Speaker Recognition , A Tutorial" ,1997, Proceedings of IEEE, Vol.85.No.9.
[6] Domingo "on the Optimality of the sample Bayesian Classifier under Zero-One loss",1997, Machine Learning, vol.29.pp.103-130, 1997.
[7] Douglas A. Reynolds and Richard C. Rose " Robust Text Independent Speaker Identification Using Gaussian Mixture Speaker Model", 1995, Speech and Audio Processing vol.3.pp.72-83.
[8] NM Ramalingeswara Rao et al "Text Independent Speaker Identification Using Integrated Independent Component Analysis with Generalized Gaussian Mixture Model ",2011, IJCSA vol.2.No.12,.
[9] Qin Jin and Thomas Fong Zhong " Overview of Front-End Features For Robust Speaker Recognition " APSIPA ASC 2011.
[10] Rabiner , L and Juang ,B.H ," Fundamentals of Speech Recognition " Englewood Cliffs: Prentice Hall , 1993.
[11] V.Sailaja , K.Srinivas Rao & K V V S Reddy (2010) " Text Independent Speaker Identification Model With Finite Multivariate Generalized Gaussian Mixture Model & Hierarchical Clustering Algorithm " IJCA vol.11. No.11.pp 25-31.
[12] Shelns J.," A Tutorial On Principal Component Analysis Deviation , Discussion & Singular Value Decomposition ", Version 1,2003.
[13] Smith L. I," A Tutorial On Principal Component Analysis", 2002.
[14] Smitha Gangisetty , " Text Independent Speaker Recognition ,2005, MS Thesis, college of Engineering and Mineral Resource at Morgantown, West Virginia University.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

BIOGRAPHY

Mr. D. Vijendra Kumar Pursuing M.Tech in Digital Electronics and Communication Engineering In GIET afflicated to Jawaharlal Nehru Technological University, Kakinada, INDIA .

Mrs. K. Jyothi working as a Associate Professor in electronics and communication in GIET. She got M.Tech In DECS Specialization. She published 6 research papers in referred International and National journals and she guided 5 M.Tech students. She is Associate member of IETE.

Dr.V.Sailaja is professor and HOD of electronics and communication engineering in GIET. She received her Ph.D in Statistical signal processing from Andhra University. She published 20 research papers in International and National Journals and she guided 15 M.Tech students. She is the fellow of IETE and Life member of ISTE.