IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735. Volume 7, Issue 5 (Sep. - Oct. 2013), PP 22-27
www.iosrjournals.org

Text Independent Speaker Identification Using Imfcc Integrated With Ica

P.Suryakumari[1] , N M Ramaligeswararao[2], K.Jyothi[3], Dr.V.Sailaja[4]
[1,2,3,4,](ECE, Godavari Institute of Engineering & Technology/ Jntuk, India)

Abstract: *Over the years, more research work has been reported in literature regarding text independent speaker identification using MFC coefficients. MFCC is one of the best methods modeled on human auditory system. Murali et al (2011) [1] has developed a Text independent speaker identification using MFC coefficients which follows Generalized Gaussian mixer model. MFCC, because of its filter bank structure it captures the characteristics of information more effectively in lower frequency region than higher region, because of this, valuable information in high frequency region may be lost. In this paper we rectify the above problem by retrieving the information in high frequency region by inverting the Mel bank structure. The dimensionality and dependency of above features were reduced by integrating with ICA. Here Text Independent Speaker Identification system is developed by using Generalized Gaussian Mixer Model .By the experimentation, it was observed that this model outperforms the earlier existing models.*

Keywords: *Independent Component Analysis; Generalized Gaussian Mixer Model; Inverted Mel frequency cepstral coefficients; Bayesian classifier; EM algorithm.*

I. Introduction:

From the past knowledge it is evident that physiological structure of a vocal tract is different for different persons, each one having different features. Based on these features, we can differentiate one persons voice from others(Ben gold (2002)[2]). The Mel frequency cepstral coefficients features was first proposed for speech recognition by S.B.Davis and P.Mermelstein . Over the decades though it was most widely used, there was a drawback in its filter bank structure, which is able to capture the information in low frequency regions than higher frequency regions.As the information contained in the higher frequency regions may be missed, proportionally the average efficiency of identification of speaker may be reduced. To overcome the this problem Mr.S.Chakraborty (2007) [16] proposed a flipped MFCC filter bank using Inverted Mel scale by considering the High frequency region coefficients.

Here the main objective is to capture the information which has been missed by the MFCC by using a new filter bank structure when it is obtained by flipping or inverting the MFCC filters bank structure (s.chakroborty(2007))[16]. In this feature extraction only few inverted Mel cepstral coefficients are considered and the remaining coefficients are dropped as insignificant due to high dimensionality problems. Ignoring the dependences and dropping some of the Mel cepstral coefficients may lead to falsification of model. It is needed to reduce the dimensionality and avoid dependence among others to having a robust model for speakers. This can be achieved by integrating the Inverted Mel frequency cepstral coefficients of each speaker's speech spectra with Independent component analysis (ICA).The integration of ICA with Inverted Mel frequency cepstral coefficients will make the feature vector more robust.

The main aim of the ICA is to find out linear representation of non-Gaussian data so that the components are statistically independent. The required assumption in applying ICA is that the observed signals are linearly mixed. It helps in capturing some of the essential features of data in many applications including Automatic Speaker

Recognition Systems for high dimensional multivariate analysis (Hyvarinen(2001))[3]. ICA is a powerful tool to separate mixed signals mainly find in speech analysis results in computational and conceptual simplicity.

Hence in this paper a text independent speaker identification method is developed and analyzed by integrating independent component analysis with Inverted Mel frequency cepstarl coefficient and using Generalized Gaussian Mixer model. The procedure for extracting the feature vectors of the speaker speech spectra using inverted Mel frequency cepstral coefficients and independent component analysis is covered in Section II. Assuming that each individual speech spectra in feature vector follows a Generalized Gaussian Mixer Model and the estimation of the model parameters is obtained by using EM algorithm. The speaker identification algorithm under Bayesian frame work is presented in the section IV.

II. Feature Vectors Extraction

In this section we describe the feature vector extraction of each individual speech spectra. Since a long time MFCC is considered as a reliable front end for speaker identification because it has coefficients that related to psychophysical studies human perception of the frequency content of sounds in a nonlinear scale called Mel scale (Ding(2001))[4]. This can be mathematically expressed as

$$\mathbf{f_{mel}=2595\ log[1+(f/700)]} \qquad \mathbf{(1)}$$

Where f_{mel} is the subjective pitch in Mel's corresponding to f in Hz.
But Mel cepstral coefficients are obtained by capturing lower frequency regions and neglecting higher regions. This leads to loss of more information from the recorded data base. By considering high frequency regions by using flipping or inverting Me3l bank structure (S.chakroborty (2007)) [5], we retrieve more information. The flip of the MFCC Mel bank is expressed as

$$\widehat{\Psi i}\mathbf{(k) = \Psi_{Q+1-i}\ (M_s/2 +1-k)} \qquad \mathbf{(2)}$$

$\widehat{\Psi i}$(k) is the inverted Mel scale filter response while Ψ_i(k) is MFCC filter bank response and Q is the number of filters in the bank. Here inverted Mel-scale is defined as follows

$$\hat{f}\mathbf{mel\ (f)=2195.2860-2595\ log_{10}(1+(4031.25-f/700))} \qquad \mathbf{(3)}$$

$\hat{f}$mel (f) is subjective pitch in the new scale corresponding to f, actual frequency in Hz.

The block diagram of Inverted Mel frequency cepstral coefficients is shown in below figure 1.From the block diagram of IMFCC the frequency bands are positioned logarithmically (on Inverted Mel scale) by inverted Mel filter bank structure, which approximated the human auditory systems response (Ruche chaudhary (2012) [6]). In this IMFCC captures high frequency bands of the Inverted Mel scale, getting more efficiency than MFCC. These vector coefficients follow Independent component analysis. ICA is linear and not necessarily orthogonal. It extracts independent components even with smaller magnitudes. Analysis of Text independent speaker identification model is a challenging task because this system comprises highly complex model with huge number of feature vectors. Under this circumstance dimension reduction of data plays a major role for obtaining better identification results. This can be achieved by ICA algorithm.

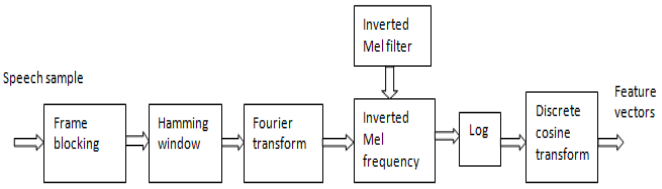

Fig 1

ICA aims at extracting a set of statistically independent vectors from the matrix of training data the Inverted Mel cepstral feature vectors derived from the original voice sample. It leads to find directions of minimal mutual information by capturing certain correlations among the frequencies present in the spectral based representation of the speech signal. All these aims are achieved by ICA in the form of linear combination of basic filter functions specific to each person.

The signal X is used as proper Mel-cepstral based representation of original signal and the data can be observed as a set of multivariate time series resulting from a linear mixing process A of independent functions S (Hyvarinen, (2001))[7][8].
Linear combination of such sources or functions can be summarized as (Cardoso, (1996)) [9]

$$\mathbf{Ax=S} \qquad \mathbf{(4)}$$

There is one problem in ICA to determine both the excitation signal S and the scalars A and the only one known component is the matrix of IMFCC coefficients of the input speech signal. S can be computed as follows (hyvarinen, (1997)) [10]

$$\mathbf{S=Ax\text{-}1} \qquad \mathbf{(5)}$$

A can be computed by considering X as a vector of observations, where each observation is expressed as a linear combination of independent component s. In order to estimate one of the independent components. A linear combination of x_i is chosen such that (Hyvarinen, (1997)) [10], (Michael, 2002[11])

$$Y=W^TX \quad (6)$$

With respect to these conditions stated in equation (4) and equation (5), the linear combination represented in equation (6) is a true combination of independent components if w was one of the columns of the inverse of A. After pre processing and whitening, the final equation is given as (Michael,(1999))[12]

$$S = Y=W^TX = \widetilde{W} p_x \quad (7)$$

Fast ICA algorithm is used to estimate W_i which constitutes the rows of W. since the components are considered to be statistically independent, the variance between them is high. The following steps are used to estimate W

1. Choose an initial random guess for W
2. Iterate: $W <= E\{xg(W^TX)\} - E\{g'(W^TX)\}W$
3. Normalize: $W <= \frac{W}{||W||}$
4. If not converged , go back to step 2

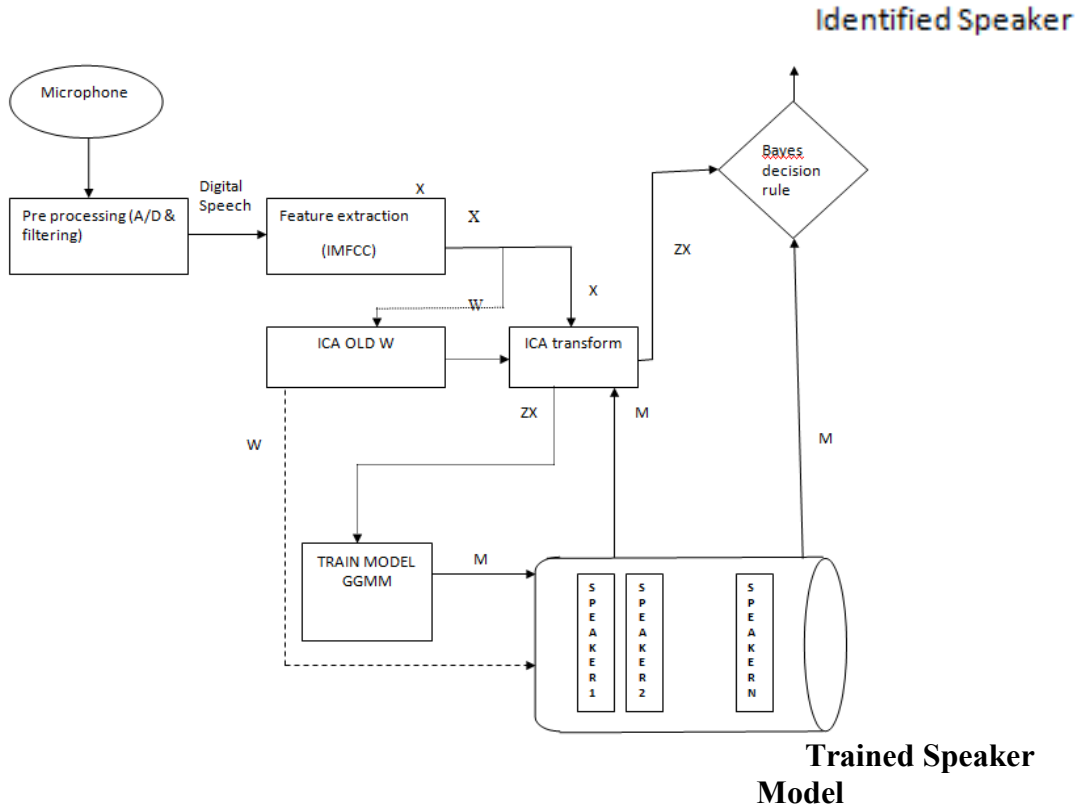

Fig 2

Once W is estimated the final step is to project the signal in to the space created by ICA

New dataset=W_{ica}* mean adjusted original data

Where , W_{ica} is the transformation matrix obtained from Fast ICA algorithm.

The extraction of Feature vectors for speaker identification is done in two steps:

1. Compute IMFCC's and
2. Apply ICA to transform them to get new feature vectors

The computation steps for extracting the new feature vectors as follows

Step1:

i. Take the fourier transform of a signal
ii. Map the powers of the spectrum obtained above on to the inverted mel scale(IMFCC), using hamming window
iii. Take the logs of the powers at each of the Inverted Mel frequencies
iv. Take the discrete cosine transform of the list of the inverted Mel log powers, as it were a signal

v. The IMFCCs are the amplitudes of the resulting spectrum

Step2:

vi. Apply ICA transform to Mel frequency cepstral coefficients to get the new feature vectors of each speaker speech spectra

III. Speaker Identification Model With Generalized Gaussian Distribution:

In this section we briefly describe the speaker identification process. Figure 1 depicts the Text Independent Speaker Identification model. In this dotted lines represents training phase and bold line represents testing phase.

The Literature shows that probabilistic models like GGMM for have yielded better performance results for training both text-dependent and text-independent speaker recognition applications. Due to the probabilistic property of a GGMM, it can also be applied to speaker identification applications in the presence of different noises increasing the channel robustness and therefore more suited to this research.

Here it is assumed that the feature vectors (after preprocessing the IMFCC with ICA) follow a multivariate Generalized Gaussian Mixer model. It is reasonable to assume the acoustic space corresponding to a speaker voice can be characterized by acoustic classes representing the some broad phonetic events such as vowels, nasals. These acoustic classes reflect some general speaker dependent vocal tract configurations that are useful for charactering speaker identity. The shape of each spectra in turn is represented by the mean of its component density and variation of the average spectral shape represented by co-variance matrix. Hence the entire speech spectra of the each individual speaker can be characterized as an M component Finite Multivariate Generalized Gaussian Mixer model.

The Probability density function of the each individual speaker speech spectra is represented as follows:

$$\boldsymbol{p(\vec{x}_t|\lambda) = \sum_{i=1}^{M} \alpha_i b_i(\vec{x}_t|\lambda)} \qquad \textbf{(8)}$$

Where, $\vec{x}_t = (x_{tij})$ j=1,2,….D; i=1,2,….M; t=1,2….T is a D dimensional random vector representing IMFCC vector.

λ is the parametric set such $\boldsymbol{\lambda = (\mu, \rho, \sum)}$

α_i is the component weight such that $\boldsymbol{\sum_{i=1}^{M} \alpha_i = 1}$

$b_i(\vec{x}_t|\lambda)$ is the probability density function of i[th] acoustic class represented by the new vectors of the speech data the *D*-dimensional Generalized Gaussian (GG) distribution (M.bicego et al (2008)(13)). It is in the form of the following

$$\boldsymbol{b_i(\vec{x}_t|\ (\mu, \rho, \sum)) = \frac{[\det(\sum)]^{-1/2}}{[z(\rho)A(\rho,\sigma)]^D} \exp\left(- \left\| \frac{\sum^{-1/2}(\vec{x}_t - \vec{\mu}_i}{A(\rho,\sigma)} \right\|_\rho \right)} \qquad \textbf{(9)}$$

Where $\boldsymbol{z(\rho) = \frac{2}{\rho}\Gamma\left(\frac{1}{\rho}\right)}$ and $\boldsymbol{A(\rho,\sigma) = \sqrt{\frac{\Gamma(1/\rho)}{\Gamma(3/\rho)}}}$

The parameter $\vec{\mu}_i$ is the mean vector and the summation of $A(\rho)$ is a scaling parameter which allows a variance var (σ^2) and ρ is the shape of the parameter when ρ=1, the Generalized Gaussian corresponds to laplacian or double exponential Distribution. When ρ=2, the Generalized Gaussian corresponds to a Gaussian distribution. In limiting case ρ→ + ∞. The model can have one covariance matrix per a Generalized Gaussian density of the acoustic class of each speaker. The diagonal convergence matrix is used for speaker model, based on the previous studies. This results diagonal covariance matrix for the feature vector, the features are independent and the probability density function of the feature vector is

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

To find the estimate of model parameters αi μij and ij for i= 1,2,3 …,M, j=1,2,…,D, we maximize the expected value b y using log likelihood function.the parameters are given by Armando.j el at (2003)[14] for each speech spectra

The following are the updated equations of the parameters of EM algorithm as given by sailaja et al (2010)[15]

$$\boldsymbol{\alpha_i^{(l+1)} = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\alpha_i^{(l)} b_i(\vec{x}_t,\lambda^{(l)})}{\sum_{i=1}^{M}\alpha_i^{(l)} b_i(\vec{x}_t,\lambda^{(l)})}\right]} \qquad \textbf{(11)}$$

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

The updated equation for estimating μ_{ij} is

$$\boldsymbol{\mu_{ij}}^{(l+1)} = \frac{\sum_{t=1}^{T} \boldsymbol{t_1}(\vec{x}_t \boldsymbol{\lambda})^{(l)^{A(N,\rho_1)}} (x_{tij} - \mu_{tij})}{\sum_{t=1}^{T} \boldsymbol{t_1}(\vec{x}_t, \boldsymbol{\lambda}^{(l)})^{A(N,\rho_{ij})}} \qquad \textbf{(12)}$$

where $A(N, \rho_{ij})$ is a function which must be equal to unity for ρ_i = 2and must be equal to $\frac{1}{\rho_{ij}-1}$ for $\rho_i \neq 1$ in the case of N=2, we have also observed that $A(N, \rho_{ij})$ must be an increasing function ofρ_{ij}.
The updated equation for estimating α_{ij} is

$$\boldsymbol{\sigma_{ij}}^{(l+1)} = \left[\frac{\sum_{i=1}^{N} \boldsymbol{t_1}(\vec{x}_t \boldsymbol{\lambda}^{(l)}) \left(\frac{\Gamma\left(\frac{3}{\rho_{ij}}\right)}{\rho_t \Gamma\left(\frac{1}{\rho_t}\right)} \right) |x_{tij} - \mu_{ij}^{(l)}|^{\frac{1}{\rho_{ij}}}}{\sum_{t=1}^{T} \boldsymbol{t_1}(\vec{x}_t \boldsymbol{\lambda}^{(l)})} \right]^{\frac{1}{\rho_{ij}}} \qquad \textbf{(13)}$$

IV. Speaker Identification Using Bayes' Decision Rule:

From each test speaker, we obtained the extracted feature vectors were applied to the function "ICA TRANSFORM" and were estimated into the space of ICA created by the associated speaker with unique speaker ID. The new feature vectors of the test utterances and the trained models were fed to a Byes classifier for identification applications which employ large group of data sets and the corresponding test speaker was identified. (Domingo's, (1997))[16][17].
$p(^{i}/_{x_t}, \lambda)$ is the posteriori probability for an acoustic class i and is defined by the following equation (Reynolds,(1995))[6].

$$\mathbf{p}(^{\mathbf{i}}/_{\mathbf{x_t}}, \boldsymbol{\lambda}) = \frac{\mathbf{p_i b_i}(\mathbf{x_t}^{\hat{}\bar{}})}{\sum_{\mathbf{k=1}}^{\mathbf{M}} \mathbf{p_k b_k}(\mathbf{x_t}^{\hat{}\bar{}})} \qquad \textbf{(14)}$$

$$\hat{\boldsymbol{s}} = \mathbf{max}_{\mathbf{1<k<s}}\, \boldsymbol{p_l}(\boldsymbol{\lambda_k}|\boldsymbol{X}) = \mathbf{arg\,max}_{\mathbf{1<k<s}} [\boldsymbol{p_l}(\boldsymbol{\lambda_k}|\boldsymbol{X}) \boldsymbol{p_r}(\boldsymbol{\lambda_k})] \qquad \textbf{(15)}$$

Finally we computes S using the logarithms and independence between the observations.

V. Expermental Results:

The developed model, IMFCC with integration of ICA with GGMM performance is evaluated. An experiment was conducted on 30 speakers, for each speaker 10 utterances with each of duration 6sec are recorded by high quality microphone. Out of this voice corpus, first, fifth and seventh sessions are used for testing and remaining sessions are used for training. By using front end process explained in section II, IMFCC feature vectors were obtained which were again refined by using ICA. The global model for each speaker density is estimated by using the derived parameters. With the test data set, the efficiency of the developed model is studied by identifying the speaker with the speaker identification algorithm given in section IV. The average percentage of identification was computed and the results are tabulated.

From the table 1, it is observed that the average percentage of correct identification of speaker for the developed model is 98.3±10. The percentage correctness for Generalized Gaussian Mixer Model with Integrated ICA using MFCC as a feature extraction is 97.8 ±12. This clearly shows that the speaker Identification model with IMFCC in a front end processing is having higher average percentage of correct identification than the other models.

Table1.Average percentage of correct Identification using MFCC and IMFCC feature vectors

Model	MFCC	IMFCC
Embedded ICA with GGMM	97.8±12	98.3±10

V. Conclusion:

In this paper we proposed and developed a Text Independent Speaker Identification with IMFC coefficients as a Feature vectors. Experimental Results shows that this method gives tremendous Improvement and it can detect the correct speaker among 30 speakers with 10 utterances. Therefore this
developed model will be applicable to real time applications.

References:

[1]. N M ramaligeswararao, Dr.V.Sailaja & Dr.k.Srinivasa Rao(2011) "Text Independent Speaker Identification using Integrated Independent Component Analysis with Generalized Gaussian Mixture Model"International Journal of Advanced Computer Science and Applications vol.2, No12, 2011pp.85-91

[2]. Ben Gold and Nelson Morgan (2002), "speech and audio processing "part 4 chapter 14 pp.189-203 John wills and sons

[3]. Hyvarinen,(1999), "Fast and Robust Fixed Point Algorithm for ICA" proceedings of the IEEE transactions on neural networks,Volume.10,no.3,pp.626-634.

[4]. Heidelberg (2003), "ICANN/ICONIP'03 of the 2003 joint international conference on Artificial neural networks/neural information processing".

[5]. S.chakroborty. a. roy and g. saha , " improved closed ser text independent speaker identification by combining MFCC with evidence from flipped filter banks " International journal of signal processing vol.4 , no.2 ,pp.114-121,2007.ISSN 1304-4478

[6]. Ruche chaudhary, National technical research organization ,"performance study of Text Independent speaker Identification system using MFCC and IMFCC for telephone and microphone speeches" International Journal of computing and Business Research(IJCBR),vol. 3 issue 2 My 2012

[7]. Hyvarinen, (2001), "Independent component analysis, john Wiley and sons.

[8]. Hyvarinen,(1999), "Fast and Robust Fixed Point Algorithm for ICA" proceedings of the IEEE transactions on neural networks,Volume.10,no.3,pp.626-634.

[9]. Cardoso,(1996) "Equivarant adaptive source separation IEEE Transaction on signal processing"vol.44.no.12.pp-3017-3030

[10]. Hyvarinen.(1997) "A family of fixed point algorithms for independent component analysis" proceedings of the IEEE international conference on acoustics speech and signal processing (ICASSP),VOLUME5,pp.3917-3920.

[11]. Heidelberg (2003), "ICANN/ICONIP'03 of the 2003 joint international conference on Artificial neural networks/neural information processing".

[12]. Ning Wang , P. C. Ching(2011), "Nengheng Zheng, and Tan Lee, "Robust Speaker Recognition Using Denoised Vocal Source and Vocal Tract Features speaker verification," IEEE Transaction on Audio Speech and Language processing, Vol. 1, No. 2, pp. 25-35.

[13]. Md M. Bicego, D Gonzalez, E Grosso and Alba Castro(2008) "Generalized Gaussian distribution for sequential Data Classification" IEE Heidelberg (2003), "ICANN/ICONIP'03 of the 2003 joint international conference on Artificial neural networks/neural information processing".

[14]. Armando. J et al (2003), "A practical procedure to estimate the shape Parameters in the Generalized Gaussian distribution

[15]. V Sailaja, K Srinivasa Rao & K V V S Reddy(2010), "Text Independent Speaker Identification Model with Finite Multivariate Generalized Gaussian Mixture Model and Hierarchical Clustering Algorithm " International Journal of Computer applications Vol. 11, No. 11, , pp. 25-31.

[16]. Domingo(1997), "Speaker Recognition", A Tutorial", Proceedings of the IEEE, Vol. 85, no. 9

[17]. Domingo's, (1997), "On the optimality of the simple Bayesian classifier under zero-one loss," Machine Learning, vol. 29, pp. 103-130, 1997