ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 6, June 2015**

# Speaker Recognition by Combining Gaussian Mixture Model (GMM) Spectral Representation and Phase Information

Anto Felix Shaji P[1], G.Kanagavalli*[2]

Assistant Professor, Dept. of ECE, Jerusalem College of Engineering, Chennai, Tamil Nadu, India [1]

Assistant Professor, Dept. of ECE, Bharath University, Chennai, Tamil Nadu, India[2]

* Corresponding Author

**ABSTRACT:** In conventional methods for text-independent speaker recognition, Gaussian mixture model (GMM) is known for its effectiveness and scalability in modeling the spectral distribution of speech. A GMM-super vector characterizes a speaker's voice by the GMM parameters such as the mean vectors, covariance matrices and mixture weights. Besides the first-order statistics, it is generally believed that speaker's cues are partly conveyed by the second-order statistics. In conventional speaker recognition methods based on Mel-frequency cepstral coefficients (MFCCs), phase information has been ignored. In this paper, we propose a phase information extraction method that normalizes the change variation in the phase according to the frame position of the input speech and combines the phase information with MFCCs. Along with phase information we measure Bhattacharyya-based GMM-distance between two GMM distributions.

**KEYWORDS:** Gaussian Mixture Models (GMM), Mel-frequency cepstral coefficients (MFCCs), speaker recognition, Bhattacharya distance

## 1. INTRODUCTION

Speaker recognition can be classified into identification and verification. Speaker identification is the process of determining which registered speaker provides a given utterance. Speaker verification, on the other hand, is the process of accepting or rejecting the identity claim of a speaker. The system that we will describe is classified as text-independent speaker identification system since its task is to identify the person who speaks regardless of what is saying. Speaker recognition systems can be characterized as text-dependent or text independent. Speaker recognition system [1] may be considered to consist of four stages. variation of the human ear's critical bandwidth with suitable frame size and shift for the feature extraction.

This project encompasses the implementation of a speaker recognition program in MATLAB. Speech analysis involves analyzing the speech signal using frequency. MFCC has two types of filter which are spaced linearly at low frequency below 1000 Hz and logarithmic spacing above 1000Hz. A subjective pitch is present on Mel Frequency Scale to capture important characteristic of phonetic in speech. The overall process of the MFCC is shown in Fig.1

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 6, June 2015**

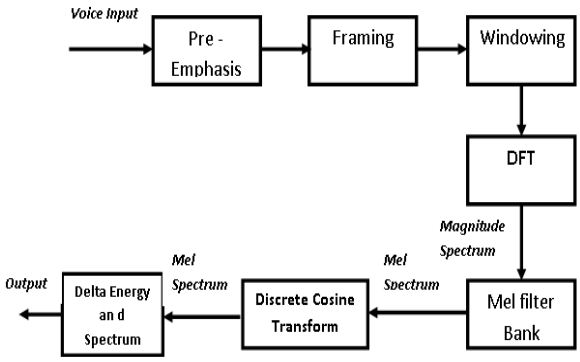


Fig.1 MFCC BLOCK DIAGRAM

in III, overview of MIMO System in Section IV, Section V gives the importance of Channel estimation and Sphere decoding algorithm, design and implementation using FPGA and results in the subsequent sections.

## II. CHALLENGES IN WIRELESS COMMUNICATION

One of the performance degrading effect arises from multipath that exists between wireless transmitter and receiver. Multipath refers to the presence of multiple copies of a transmitted signal at the receiver end due to the presence of multiple radio channel between them. These multiple copies arise due to reflections from various interacting objects present. Adverse effects of multipath and its level is manifested in several forms in communications receivers and it depends on the path difference relative to the wavelength of propagation, the of path difference relative to the signaling rate, and the relative speed between the transmitter and receiver. Multipath from interacting objects that are spaced very close together will cause a random change in the amplitude of the received signal. signals arriving at the receiver along different paths can result in increase in signal strength or reduce it. This gives rise to fading that depends on the wavelength of transmitted signal , known as frequency-selective fading. When there is relative change They include: speech analysis, feature extraction, speaker modeling and speaker testing [2]. Feature extraction involves extracting speaker-specific features from the speech signal at reduced data rate. The extracted features are further combined using modeling techniques to generate speaker models. The speaker models are then tested using the features extracted from the test speech signal. The improvement in the performance can be achieved by employing new or improved techniques in one or more of these stages. A Universal Background Model (UBM) [7] is a model used in a biometric verification system to represent general, person independent feature characteristics to be compared against a model of person-specific feature characteristics when making an accept or reject decision. In a speaker verification system, the UBM is a speaker-independent Gaussian Mixture Model (GMM) trained with speech samples from a large set of speakers to represent general speech characteristics.

## III. FEATURE EXTRACTION

It is possible to extract number of features from audio samples, including both spectral and non-spectral features but in this project the speaker's voice is characterized exclusively through spectral features. The selected acoustic spectral feature set should reflect the unique characteristics of the speaker. For this purpose we use the magnitude component of the short-time Fourier transform (STFT) as a basis. Most of today's automatic speech recognition (ASR) systems are based on some type of Mel-frequency cepstral coefficients (MFCCs)[5-6], which have proven to be effective and robust under various conditions. The extraction of the best parametric representation of acoustic signals is an important task to produce a better recognition performance. MFCC is based on human hearing perceptions which cannot perceive frequencies over 1kHz. In other words, in MFCC is based on known
where
N = number of samples in each frame
The output after windowing is given in Equation 3

$$Y[k] = X[k] \,.\, W[k] \qquad (3)$$

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 6, June 2015**

Y[k] = Output signal
X[k] = input signal
W[k] = Hamming window function

**C.** Fast Fourier Transform

This process is to convert each frame of N samples from time domain into frequency domain. For each frame of signal (e.g., *N*=256) the Fast Fourier Transform (FFT) is performed to obtain its spectrum. This statement is supported by the Equation 4

$$X(w) = FFT[\ x[n]\ ] \qquad (4)$$

where
X(w) = output signal in frequency domain
x[n] = input signal after windowing function

D. Mel Filter Bank Processing

The frequencies range in FFT spectrum is very

A. Pre–emphasis

In this step processes signal is passed through a filter which emphasizes higher frequencies. This process will increase the energy of signal at higher frequency. The expression of pre-emphasis filter is given in Equation 1.

$$Y[n] = X[n] – a\ X[n-1] \qquad (1)$$

where
X[n] is the input speech signal
Y[n] is the pre-emphasized output signal
Let consider a = 0.95, which make 95% of any one sample is presumed to originate from previous sample.

B. Framing and Windowing

The process of segmenting the speech samples obtained from analog to digital conversion (ADC) into a small frame with the length within the range of 20 to 40 msec. The voice signal is divided into frames of N samples. Adjacent frames are being separated by M [3]. Typical values used are M = 100 and N= 256 (M < N).

Hamming window is used as window shape by considering the next block in feature extraction processing chain and integrates all the closest frequency lines. The Hamming window expression is given in Equation 2

$$w(n) = 0.54 – 0.46 \cos( 2\pi n/(N-1)) \quad 0< n < N-1 \qquad (2)$$

Mel Frequency Cepstrum Coefficient. The set of coefficient is called acoustic vectors. Therefore, each input utterance is transformed into a sequence of acoustic vector. The expression for DCT is given in Equation 6.

$$X_k = \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2}\right)k\right] \qquad k=0,1....N-1 \qquad (6)$$

F. Delta Energy and Delta Spectrum

The voice signal and the frames changes, such as the slope of a formant at its transitions. The energy in a frame for a signal X in a window from time sample $t_1$ to time sample $t_2$, is represented by the Equation 7

$$Energy = \Sigma\ X^2(t) \qquad (7)$$


DOI: 10.15662/ijareeie.2015.0406123

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 6, June 2015**

## III. GAUSSIAN MIXTURE MODELS (GMM)

A Gaussian Mixture Model (GMM)[4] is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system, such as vocal-tract related spectral features in a speaker recognition system. GMM parameters are estimated from training data using the iterative Expectation-Maximization (EM) algorithm or Maximum A Posteriori (MAP) estimation from a well-trained prior model.

A Gaussian mixture model is a weighted sum of M component Gaussian densities as given by the Equation 8wide and voice signal does not follow the linear scale. The bank of filters according to Mel scale as shown in Fig.2

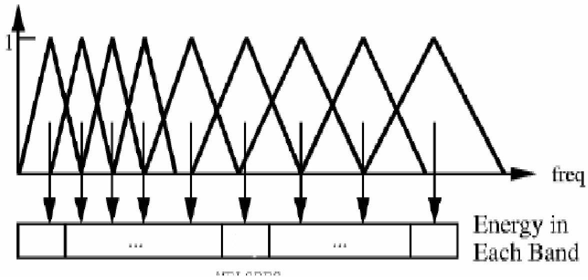


Fig. 2 MEL SCALE FILTER BANK

Fig.2 shows a set of triangular filters that are used to compute a weighted sum of filter spectral components so that the output of process approximates to a Mel scale. Each filter's magnitude frequency response is triangular in shape and equal to unity at the centre frequency and decrease linearly to zero at centre frequency of two adjacent filters. Then, each filter output is the sum of its filtered spectral components. After that the Equation 5 is used to compute the Mel for given frequency f in Hz.

F[Mel] = [ 2595 * log 10[1+f ] 700 ] (5)

E. Discrete Cosine Transform

This is the process to convert the log Mel spectrum into time domain using Discrete Cosine Transform (DCT). The result of the conversion is called be shared, or tied, among the Gaussian components, such as having a common covariance matrix for all components, The choice of model configuration (number of components, full or diagonal covariance matrices, and parameter tying) is often determined by the amount of data available for estimating the GMM parameters and how the GMM is used in a particular biometric application [4].

A. Maximum Likelihood Parameter Estimation

Given training vectors and a GMM configuration, we wish to estimate the parameters of the GMM, λ, which in some sense best matches the distribution of the training feature vectors. There are several techniques available for estimating the parameters of a GMM. By far the most popular and well-established method is maximum likelihood (ML) estimation. The aim of ML estimation is to find the model parameters which maximize the likelihood of the GMM given the training data. For a sequence of T training vectors $X = \{x_1, \ldots, x_T\}$, the GMM likelihood, can be written as in Equation 11.

$$P(X|\lambda) = \prod_{t=1}^{T} p(x_t|\lambda) \qquad (11)$$

This expression is a non-linear function of the parameters λ and direct maximization is not possible. However, ML parameter estimates can be obtained iteratively using a special case of the expectation-maximization (EM) algorithm.

B. Expectation-Maximization (EM) Algorithm

The basic idea of the EM[4] algorithm is, beginning with an initial model λ, to estimate a new model λ', such that $p(X|\lambda') \geq p(X|\lambda)$. The new model then becomes the

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 6, June 2015**

$p(x|\lambda) = \sum_{i=1}^{M} \omega_i \, g(x|\mu_i, \Sigma_i)$ (8)

where x is a D-dimensional continuous-valued data vector,

$_i$ are the mixture weights, and $g(x|\mu_i, {}_i)$ i = 1,2,............M, are the component Gaussian densities. Each component density is a D-variate Gaussian function of the form given in Equation 9

$g(x|\mu_i, \Sigma_i) = \frac{1}{2\pi^{D/2}|\Sigma_i|^{1/2}} \exp\{-1/2(x-\mu_i)' \Sigma_i^{-1}(x-\mu_i)\}$ (9)

where $\mu_i$ is the mean vector and $\Sigma_i$ is the covariance matrix. The mixture weights satisfy the constraint that $\sum_{i=1}^{M} \omega_i = 1$. The complete Gaussian mixture model is parameterized by the mean vectors, covariance matrices and mixture weights from all component densities. These parameters are collectively represented by the notation given in Equation 10

$\lambda = \{\omega_i, \mu_i, \Sigma_i\}$ i = 1,2,3.........M. (10)

The covariance matrices $\Sigma_i$, can be full rank or constrained to be diagonal. Additionally, parameters can

$P(i|x_t, \lambda) = \frac{\omega_i \, g(x_t|\mu_i, \Sigma_i)}{\sum_{k=1}^{M} \omega_k \, g(x_t|\mu_k, \Sigma_k)}$ (15)

C. Speaker Identification

The identification system is a straight forward maximum likelihood classifier. For a group of speakers S={1,2,…S} represented by models $\lambda_1, \lambda_2, \ldots \lambda_S$, the objective is to find the speaker model which has the maximum posterior probability for the input feature vector sequence X ={$x_1, x_2, \ldots x_T$}.The minimum error Baye's rule for this problem is

$\hat{S} = \arg\max_{0 \le s \le S} Pr(\lambda_s/X)$

$= \arg\max_{0 \le s \le S} \cdot \frac{p(X/\lambda_s)}{p(X)} Pr(\lambda_s)$ (16)

Assuming equal prior probabilities between $Pr(\lambda_s)$ and p(X) are constant for all speakers and can be ignored in the maximum. So the decision rule becomes

$\hat{S} = \arg\max_{0 \le s < S} \cdot \sum_{t=1}^{T} \log p(x_t|\lambda_s)$ (17)

A block diagram of speaker identification system is shown in Fig.3.

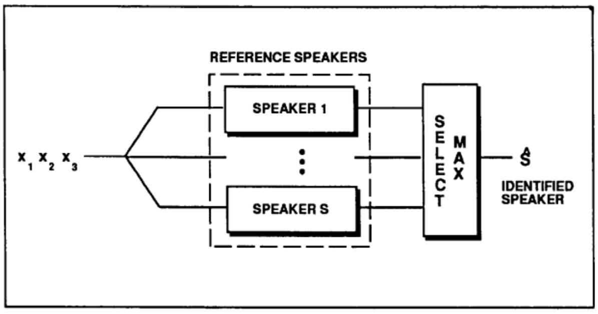


Fig. 3 SPEAKER IDENTIFICATION SYSTEM

initial model for the next iteration and the process is repeated until some convergence threshold is reached. The initial model is typically derived by using some form of binary VQ estimation. On each EM iteration, the following re-estimation formulas are used which guarantee a monotonic increase in the model's likelihood value,

Mixture Weights

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2015

$$\omega_i = \frac{1}{T}\sum_{t=1}^{T} P(i|x_t, \lambda) \quad (12)$$

Mean

$$\mu_i = \frac{\sum_{t=1}^{T} P(i|x_t, \lambda)\, x_t}{\sum_{t=1}^{T} P(i|x_t, \lambda)} \quad (13)$$

Covariance

$$\Sigma_i = \frac{\sum_{t=1}^{T} P(i|x_t, \lambda)(x_t-\mu_i)'(x_t-\mu_i)}{\sum_{t=1}^{T} P(i|x_t, \lambda)} \quad (14)$$

The a posteriori probability for component i is given by Equation 15

Applying Baye's rule and discarding the constant prior probabilities, the likelihood ratio in the log domain becomes

$$\Lambda(x) = \log p(X|\lambda_c) - \log p(X|\lambda_{\hat{c}}) \quad (19)$$

The term $p(X|\lambda_c)$ is the likelihood of utterances given it is from the claimed speaker and $p(X|\lambda_{\hat{c}})$ is the likelihood of utterances given it is not from the claimed speaker The ratio is compared with the threshold θ and the claimed speaker is accepted if $\Lambda(x) > \theta$ and rejected if $\Lambda(x) < \theta$. The block diagram of speaker verification system is shown in Fig.4.

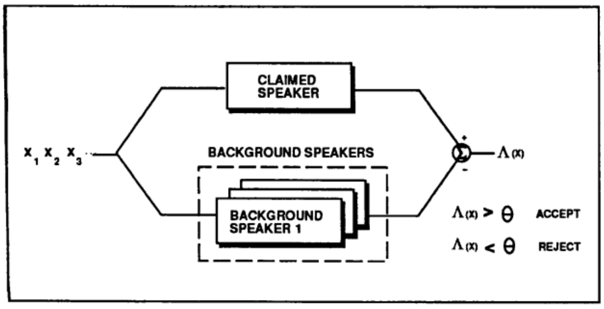


Fig.4 SPEAKER VERIFICATION SYSTEM

## IV. PROPOSED SYSTEM

A. Phase Information

The short-term spectrum S(ω,t) for the ith frame of a signal is obtained by the DFT of an input speech signal S(ω,t) sequence is given in Equation 20.

$$S(\omega,t) = X(\omega,t) + jY(\omega,t)$$

$$= \sqrt{(X^2(\omega,t) + Y^2(\omega,t))} \times e^{j\,\theta(\omega,t)} \quad (20)$$

For conventional MFCCs, the power spectrum $X^2(\omega,t) + Y^2(\omega,t)$ is used, but the phase information θ(ω,t) is ignored. In this project, phase θ(ω,t) is also extracted as one of the feature parameter set for speaker recognition

C. Speaker Verification

It requires only a binary decision but the verification task is more difficult than the identification task. The system must decide whether the input voice came from the claimed speaker, with a well defined model, or not from the claimed speaker, which is ill defined. For a given utterance X and a claimed identity the choice is between $H_0$ and $H_1$.

$H_0$: X is from the claimed speaker.

H1: X is not from the claimed speaker.

The general approach used in the speaker verification system is to apply a likelihood ratio test to an input utterance to determine if the claimed identity is accepted or rejected. For a utterance $X = \{x_1, x_2, \ldots x_T\}$ and a claimed speaker identity with corresponding model $\lambda_c$, the likelihood ratio is

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2015

$$\frac{\Pr(X \text{ is from the claimed speaker})}{\Pr(X \text{ is not from the claimed speaker})} = \frac{\Pr(\lambda_c|X)}{\Pr(\lambda_{\bar{c}}|X)} \quad (18)$$

sampling sequence for a cyclic function, where sample per period $L_w = f_s/f = 2\pi/\omega f_s$ on radian frequency ω and $f_s$ sampling frequency. The phase of $s_1,s_2,s_3,\ldots s_{Lw}$ and the phase of $s_1,s_2,s_3,\ldots\ldots s_{Lw}$, $s_{Lw+1}$ are different from each other [9]. The difference of phase on the radian frequency ω is $2\pi/L_w$. To overcome the influence of the phase response with respect to frame position, phases with the basis radian frequency $\omega_b$ for all frames is converted to a constant, and the phase with the other frequency is estimated relative to this. Without loss of generality, setting the phase with the basis radian frequency $\theta(\omega_b,t)$ to 0, the expression is given in Equation 21.

$$S'(\omega_b,t) = \sqrt{(X^2(\omega_b,t) + Y^2(\omega_b,t))} \times e^{j\,\theta(\omega b,t)} \times e^{j\,(-\theta(\omega b,t))} \quad (21)$$

The difference between the un-normalized wrapped phase $\theta(\omega_b,t)$ with basis frequency $\omega_b$ in Equation 20 and the normalized wrapped phase in Equation 21 is $-\theta(\omega_b,t)$. The spectrum on frequency ω is normalized as

$$S'(\omega,t) = \sqrt{(X^2(\omega,t) + Y^2(\omega,t))} \times e^{j\theta(\omega b,t)} \times e^{j\omega/\omega b(-\theta(\omega b,t))}$$

$$= \sqrt{(X^2(\omega,t) + Y^2(\omega,t))} \times e^{j\breve{o}(\omega,t)}$$

$$= X(\omega,t) + j\acute{Y}(\omega,t) \quad (22)$$

and the phase information is normalized as

$$\breve{o}(\omega,t) = \theta(\omega,t) + \omega/\omega_b(-\theta(\omega_b,t)) \quad (23)$$

where $\breve{o}(\omega,t)$ it is referred to the proposed original phase or the original normalized phase.

B. Bhattacharyya-Based GMM Distance Measure

The Bhattacharyya kernel between two probability distributions is given in Equation 24

$$\Lambda_{Bhatt}(p_a||p_b) = \int_{Rn}\sqrt{p_a(x)}\,\sqrt{p_b(x)}\,dx \quad (24)$$

The Bhattacharyya distance of the two probability distributions is given in Equation 25

$$\Psi_{Bhatt}(p_a||p_b) = -\ln(\Lambda_{Bhatt}(p_a||p_b)) \quad (25)$$

[8]. The GMMs used in this paper are insensitive to the temporal aspects of speech, and do not capture the dependence of features extracted from each frame. Phase information of the same person with the same voice extracted from different frames may be $\theta(\omega,t)$ and $2\pi + \theta(\omega,t)$. They express different phase values and the different speaker characteristics using phase-based GMMs. So phase value is constrained to $[-\pi,\pi]$.

For speaker recognition using phase information[2], the phases extracted from two different windows of a same sentence from same people should be as small as possible. Thus, it is necessary to normalize the phase response with respect to the frame position. A basic method for eliminating the influence of the phase response with respect to frame position is explained as follows. Let $s_1,s_2,s_3,\ldots\ldots\ldots\ldots s_{Lw}$ be the

$$\psi_{Bhatt}'(p_a||p_b) = \sum_{i=1}^{M}(\Psi_{Bhatt}(p_a||p_b))$$

$$= \frac{1}{8}\sum_{i=1}^{M}\left((m_i^{(b)} - m_i^{(a)})^T\,[\Sigma_i^{(a)} + \Sigma_i^{(b)}\,/2]^{-1}\right.$$

$$\left.+ (m_i^{(b)} - m_i^{(a)})\right) + \frac{1}{2}\sum_{i=1}^{M}\{\ln \frac{|(\Sigma_i^{(a)} + \Sigma_i^{(b)})/2|}{\sqrt{(|\Sigma_i^{(a)}|\,|\Sigma_i^{(b)}|)}}$$

$$- \frac{1}{2}\sum_{i=1}^{M}(\ln(w_i^{(a)} w_i^{(b)})) \quad (26)$$

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2015

## V. RESULTS AND ANALYSIS

### A. Experimental Setup

The evaluation is conducted on four male and four female. We carried out both gender-independent and gender-dependent speaker verification separately.

For feature extraction, an 10-dimensional Mel Frequency cepstral coefficient (MFCC) vector is computed from pre-emphasized speech every 8 ms using a 16 ms Hamming window. An energy-based speech detector is applied to discard silence and noise frames [10].

The GMM parameters are initialized, mean from k-means clustering and covariance matrix is initialized with identity matrix. The weights of the clusters are given equal probabilities.

### B. Performance Evaluation

The test speech was fist processed by the front end analysis to produce the sequence of feature vector $\{x_1, x_2 \ldots x_T\}$. To evaluate different test utterance length, the sequence was divided into overlapping segments of T feature vectors. The first two segments from the sequence is given as

$$\{\underbrace{x_1, x_2, x_3, \ldots x_T}_{\text{Segment 1}}, x_{T+1}, x_{T+2,\ldots\ldots}\}$$

$$\{x_1, x_2, x_3, \ldots x_T, x_{T+1}, x_{T+2,\ldots\ldots}\}$$

Segment 2

A test segment length of 5 seconds corresponds to T = 500 feature vectors at a 16ms frame rate. Each segment of where $p_a$ and $p_b$ denote the probabilistic models $GMM_a$ and $GMM_b$, respectively.

Since $GMM_a$ and $GMM_b$ are adapted from the same UBM, what is more concerned is the difference between each pair of the Gaussian components rather than the difference of the entire mixture model [12]. Therefore, we use the equation $\sum_{i=1}^{M}( \Psi_{Bhatt}(p_a||p_b))$. Hence, we introduce the Bhattacharyya based GMM-distance measure between the two Gaussian mixture models as in Equation 26.

### C.Results

TABLE 1 GMM IDENTIFICATION PERFORMANCE FOR DIFFERENT MODEL ORDER AND TEST LENGTH

| Classifier | Model order | Test Length | | |
|---|---|---|---|---|
| | | 5 sec | 10 sec | 20 sec |
| Maximum likelihood detector | 4 | 79.8% | 85.6% | 88.3% |
| | 8 | 87.3% | 90.5% | 91.4% |
| Bhattacharya based distance | 4 | 81.8% | 90.1% | 92.3% |
| | 8 | 89.2% | 91.7% | 93.1% |

The table shows the identification performance between maximum likeli-hood classifier and Bhattacharya based distance measure. In the above results the identification performance is insensitive to model order if we increase the

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2015

test length. The GMM with Bhattacharya distance give high identification rate than the Maximum Likelihood detector. The above experiment is conducted with the speech signal of length 60 seconds, by increasing the amount of training signal the performance is increasing.

## VI. CONCLUSION

This paper provided a performance evaluation of the model parameters used in a text independent speaker recognition system. From the above simulation result it is clear that the accuracy of the identification process can be influenced by certain factors such as the MFCC technique should be applied for feature extraction. The GMM with Bhattacharya distance gives higher performance. In a GMM based text-independent speaker identification system on increasing the amount of training data increases the identification rate. Experimental result shows that increasing the mixture components of the speaker model improves the performance, limited by amount of training T vectors was treated as separate test utterance [11]. The identified speaker of each segment was compared to each actual speaker of the test utterance and the number of segments which were correctly identified was tabulated in Table 1.

$$\% \text{ of correct identification} = \frac{\#\text{correctly identified segments}}{\text{total number of segments}}$$

The evaluation is repeated for different values of T to evaluate performance with respect to test utterances.
data. Future work depends on combining phase information in the MFCC feature extraction process and evaluating performance with GMM–Bhattacharya distance measure.

## REFERENCES


1. "A Gaussian Mixture Model Spectral representation for Speech Recognition", Matthew Nicholas Stuttle, Cambridge University Engineering Department, July 2003.
2. Subhashini V., Ponnusamy S., Muthamizhchelvan C., "Growth and characterization of novel organic optical crystal: Anilinium d-tartrate (ADT)", Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy, ISSN : 1386-1425, 87() (2012) pp.265-272.
3. "Speaker Identification and Verification by Combining MFCC and Phase Information", Nakagawa Seiichi, Wang Longbiao, Ohtsuka Shinji, IEEE transactions on audio, speech, and language processing, vol. 20, NO. 4, May2012.
4. Thomas J., Ragavi B.S., Raneesha P.K., Ahmed N.A., Cynthia S., Manoharan D., Manoharan R., "Hallermann-Streiff syndrome", Indian Journal of Dermatology, ISSN : 0019-5154, 58(5) (2013) pp.383-384.
5. "GMM-SVM Kernel with a Bhattacharyya-based Distance for Speaker Recognition", Chang Huai You, Kong Aik Lee, Haizhou Li, IEEE transactions on audio, speech, and language processing, vol. 18, NO. 6, Aug 2010.
6. Subha Palaneeswari M., Ganesh M., Karthikeyan T., Manjula Devi A.J., Mythili S.V., "Hepcidin-minireview", Journal of Clinical and Diagnostic Research, ISSN : 0973 - 709X, 7(8) (2013) pp.1767-1771.
7. "Robust text-independent speaker identification using Gaussian mixture speaker models", Reynolds, D. A. and Rose, R. C., IEEE Trans.Speech Audio Process. 3, 1995, pp 72–83.
8. "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques", Muda Lindasalwa, Begam Mumtaj and I. Elamvazuthi, Journal Of Computing, Volume 2, Issue 3, March 2010, Issn 2151-9617.
9. Laljee R.P., Muddaiah S., Salagundi B., Cariappa P.M., Indra A.S., Sanjay V., Ramanathan A., "Interferon stimulated gene - ISG15 is a potential diagnostic biomarker in oral squamous cell carcinomas", Asian Pacific Journal of Cancer Prevention, ISSN : 1513-7368, 14(2) (2013) pp.1147-1150.
10. "Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains," J. L. Gauvain and C. H. Lee, IEEE Trans. Speech Audio Process., vol. 2, no. 2, pp. 291–298, 1994.
11. "An efficient scoring algorithm for Gaussian mixture model based speaker identification", B. L. Pellom and J. H. L. Hansen, *IEEE Signal Process. Lett.*, vol. 5, no. 11, pp. 281–284, Nov. 1998.
12. Kumar S., Das M.P., Jeyanthi Rebecca L., Sharmila S., "Isolation and identification of LDPE degrading fungi from municipal solid waste", Journal of Chemical and Pharmaceutical Research, ISSN : 0975 – 7384 5(3) (2013) pp.78-81.
13. [13]B Karthik, S Rajeswari, Design and Analysis of a Transceiver on a Chip for Novel IR-UWB Pulses, Middle-East Journal of Scientific Research 19 (6), PP 817-820, 2014.
14. Shriram, Revati; Sundhararajan, M; Daimiwal, Nivedita; , Application of High & Low Brightness LEDs to Human Tissue to Capture Photoplethysmogram at a Finger TipRed, V-620,PP 700.
15. Muralibabu, K; Sundhararajan, M; , PAPR performance improvement in OFDM system using DCT based on adjacent symbol groupingTrans Tech Publ, Applied Mechanics and Materials, V-550,PP 204-209, 2014.
16. sivaperumal, s; sundhararajan, m; , advance feature extraction of mri brain image and detection using local segmentation method with watershed.
17. MURALIBABU, K; SUNDHARARAJAN, M; , PAPR reduction using DCT based sub carrier grouping with Companding Technique in OFDM system.
18. Kanniga, E; Srikanth, SMK; Sundhararajan, M; , Optimization Solution of Equal Dimension Boxes in Container Loading Problem using a Permutation Block AlgorithmIndian Journal of Science and Technology, V-7, I-S5,PP 22-26, 2014