Gaussian Mixture Model: A Modeling Technique for Speaker Recognition and its Component

Nilu Singh
SIST-DIT, Babasaheb Bhimrao Ambedkar University (Central University), Lucknow, UP, India

Alka Agrawal
Department of Computer Science, Khwaja Moinuddin Chishti Urdu, Arabi-Farsi University, Lucknow.

R. A. Khan
SIST-DIT, Babasaheb Bhimrao Ambedkar University (Central University), Lucknow, UP, India

ABSTRACT
This paper provides an overview of Gaussian Mixture Model (GMM) and its component of speech signal. During the earlier period it has been revealed that Gaussian Mixture Model is very much appropriate for voice modeling in speaker recognition system. For Speaker recognition, Gaussian mixture model is an essential appliance of statistical clustering. The task effortlessly performed by humans is not effortless for machine or computers such as voice recognition or face recognition so for this function speaker recognition technology makes available a solution, using this technology the computers/machines outperforms than humans.

General Terms
Speaker recognition, security, pattern matching etc.

Keywords
Speaker Recognition, GMM, Gaussian component.

1. INTRODUCTION
Gaussian Mixture Model (GMM) is a type of density model, this model containing a number of component functions and combining these component functions provides a multimodal density. Gaussian mixture models are use mostly among the statistical models for clustering. In terms of statistics a mixture model is a probabilistic representation that acquires all the data points are bring into existence from a mixture of a limited number of Gaussian distribution with unidentified parameters, for representing the existence of subgroup surrounded by an overall group i.e. it is an act of making formal of interaction involving variables in the form of mathematical equations. GMM attempt to characterize most part of the distribution of existing training data leave out any possible phonetic information which may be of significance. A well known definition of Gaussian mixture model is "it is a parametric probability density function represented the same as a weighted addition of Gaussian component densities and frequently used as a parametric representation of the probability distribution of continuous measurements/factors in a biometric system"[1].

Speaker recognition is a method of recognizing a human being as a result of using speaker specific information integrated in speech signal. Speaker recognition can be classified as speaker identification and speaker verification. In case of speaker identification the objective is to find out which tone of voice in an identified group of voices finest equalize the speaker, while in case of speaker verification the process for accepting or rejecting the speaker by the identity claims to be [2]. Speaker recognition technique also separated into text-dependent and text-independent recognition methods. In text dependent process the same wording is used for both training data and testing data while for text-independent process, the text is unconstrained for training and testing. For practical implementation of speaker recognition system, take the voice samples from speaker and these samples/sample is used to build up a model using statistical modeling technique GMM/ HMM. These models used for capturing the specific information of a speaker voice. The duration of the acquired data of speech signal or voice sample for text independent speaker recognition system is typically one (1) minute. The speaker model for each speaker is build by using Statistical Gaussian Mixture Model or Hidden Markov Model. To build speaker model, use statistical distributions of the spectral features which is extracted from the speech signal [3] [11].

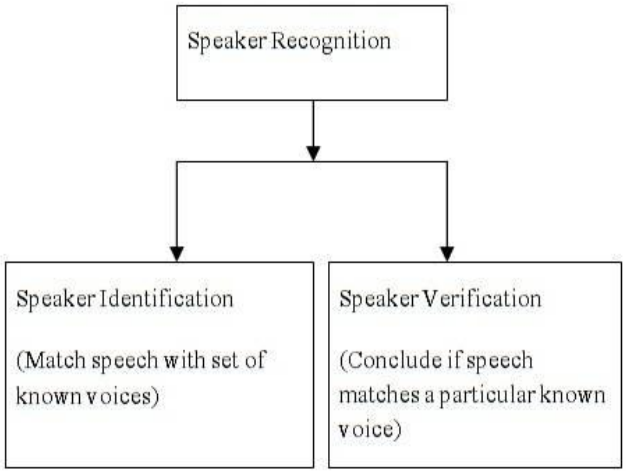

Fig 1: basic task of speaker recognition

Security has been thought for the basic of prevention for the peoples. But from last few decades, peoples have arisen interest increasing in security systems. The task of security systems to manage security in very efficient method hence these systems are very useful. Security systems also reduce the need of man power resources. Since human voice is unique due to their anatomical structure of vocalization hence it is possible to recognize people from their voice. Since this task can be estimated as automatic approach that's why it is also known as Automatic Speaker Recognition (ASR). To identify a people voice is the simplest way, voice based people recognition systems came in to the category of biometric systems. Biometric systems allow access control to the peoples [4].

2. RELEVANT LITERATURE OF SPEAKER RECOGNITION

Before the development of Speaker Recognition machine researchers must thought that can humans recognize one another only by voice or some other related features such as voice or perceptual cues. Human voice is not only the concatenation of sounds although it is a mixture of different sounds often times by way of no separate limitations among transitions. Automatic speaker recognition system is usually used in speech signal processing to categorize experimental utterances by the speaker's detection. Such type of personal belongings frequently requires high processing throughput. Speaker recognition technology is use to recognizing the people automatically as a result of using particular information of speaker integrated in speech signal. There are some well known applications of this technology in many fields such as identity of people through voice, voice dialing, telephone banking, database access services, accessing remote computers, and security control for confidential information [5] [11].

Speaker recognition can be categories as speaker identification and speaker verification, identification and verification can be done by using user voice or speaker voice by their speech signal. Speech signal contains the required information for speaker recognition such as what are saying i.e. speech message, who are speaking i.e. speaker identity, mind set i.e. emotions, tone identification i.e. identify the location, language identification i.e. Hindi , English or some native language or any local language. Speech signal also hold the information about channel & environment of recording of voice [3].

As discussed in [4] [6] using the spectrum of the speech signal we can be able to obtained most of the parameters which used for recognition process. Speech signal spectral moderate information regarding the vocal tract as well as the excitation source in the glottis by resources of the formants and the fundamental frequency. For speaker recognition technique, the parameters obtained by spectral typically the equivalent as the ones used in speech recognition technique.

3. GAUSSIAN MIXTURE MODEL FOR SPEAKER (GMM): A STATISTICAL MODELING TECHNIQUE

There are a lot of the discipline dealing with the art or science of applying scientific knowledge to practical problems applications where should be predictable from data. As discussed in [5] GMM is an essential purpose of statistical clustering to speaker recognition, at the time of training features vectors a number of prototypes are generated in the direction of characterize the feature space as a mixture of Gaussian distribution. Gaussian mixtures particularly when mixture of distributions have been used comprehensively as models in such type of problems where data know how to be viewed the same as arising from two or more populations [6]. In a GMM based system setup either each speaker is modeled by a particular, individual GMM or each speaker by several GMMs. GMM are frequently used in biometric systems mostly in speaker recognition systems because it have the potential of representing a huge class of taster distributions. There is one influential property of GMM that it have the ability to shape smooth estimation to arbitrarily shaped densities, standard Gaussian model represents feature distributions by means of a mean vector and a covariance matrix and a vector quantizer. The basic purpose of GMM in biometric system for representing feature distributions and also is provoked through the instinctive conception that the individual component densities could model a number of fundamental set of concealed classes [1] [12].

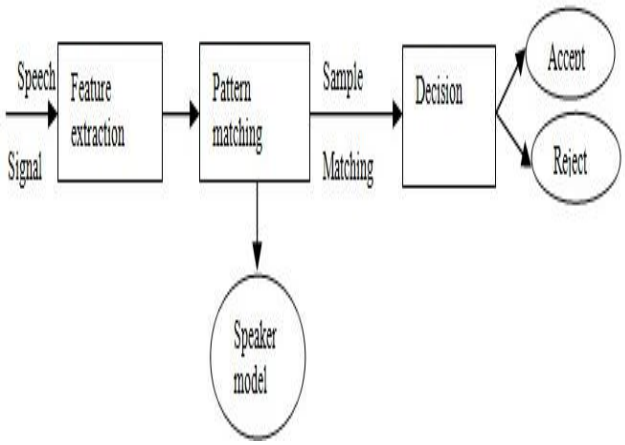

Fig 2: A view of Speaker Recognition System

As discussed in [5] usually GMM is a statistical clustering technique, its algorithm be able to relate as a prototype based algorithm. In this prototypes are bring into existence from the training feature vectors through representing the feature space like a mixture distributions and each image/prototype comprises a model parameter set as well as mean vector, covariance matrix and mixture weight. GMM Parameters is trained inside an unsupervised taxonomy using the expectation maximization (EM) algorithm, this algorithm make available an iterative aspect of maximum likelihood estimation technique. Investigation of experiments has shown that the training samples of data cover an enough assortment of the speaker speech signal. For speaker recognition system the most useful modeling technique is GMM because it proficient for accomplishing prominent identification accuracy for short length utterance from the speech signal. As discussed in [6][8] for signal processing the quality of being of practical use of the GMM depends on two factors, primary is whether or not the estimate is satisfactorily powerful to signify a large class of density functions, which is especially noticeable for that cases which are encountered in engineering applications. And secondary using parameter estimation scheme, if such an estimation be able to be obtained in a logical way which permit the user to calculate the best possible values of the mixture parameters from a predetermined set of data samples.

The arbitrary PDF (probability density function) can be estimated by the linear combination of Gaussian density unimodel [4] [8] [11]. As discussed in [7] about Gaussian mixture density that It is a weighted sum of M component densities and can be stated as-

$$P(x \mid \lambda) = \sum_{i=1}^{M} p_i b_i(x) \qquad (1)$$

Where-

x is D dimensional vector.

P_i is the weight of component.

$b_i(x)$ = component densities.

Further it can be written as-

$$bi(x) = 1/((2\pi)\ D/2\ |\sum i|\ 1/2) * \exp\{-1/2\ (x - \mu i)\ ^T\ (\sum i)\ ^(-1)\ ]\ (x - \mu i)\}$$

Where -

μi = mean vector

∑ i= covariance matrix

$$\sum_{i=1}^{m} p\ i = 1 \qquad (3)$$

i.e. Mixture weights must satisfy constraint.

So we can say that Gaussian mixture density is parameterized by the mixture weights, mean vectors and covariance matrices.

These parameters are represented by:

$\lambda = \{p_i\ \mu_i\ \sum i\}$ i = 1,2……M.

are the Gaussian component and mixture weight.

Therefore one model for each speaker and is represented by λ. or in other words every speaker is recognizing by λ model which is acquire from GMM analysis [4] [7] [11]. *GMM* is a significant application of biometric systems and widely used for speaker recognition system. The popularity of GMM because of their potential of representing a large class of samples distribution [1].

4. GAUSSIAN COMPONENTS

The task of the speaker recognition done by using individual mixture components i.e. Gaussian components [8] [9]. For speaker recognition features are obtained from the speech signal. The fundamental information of speaker discrimination can be characterized by Gaussians. For the expected GMM parameters i.e. covariance and Gaussian component weight is associated to the location of formant, magnitude of speech signal and bandwidth of speech signal [8]. As discussed in [9] for good quality system performance at least 8 to 16 Gaussian components are mandatory where voice/speech is consider as noiseless. The GMM created by using these components through diagonal covariance matrix. And when build multi conditional robust systems the minimum number of essential Gaussian components involving 64 and 128. Since likelihood function computationally inexpensive this is the advantage of using GMM. GMM is collected from a finite mixture of Gaussian components [8]. Since Gaussian components have the potential to characterize the discriminative information of speaker hence it is used widely for speaker recognition [8] [10].

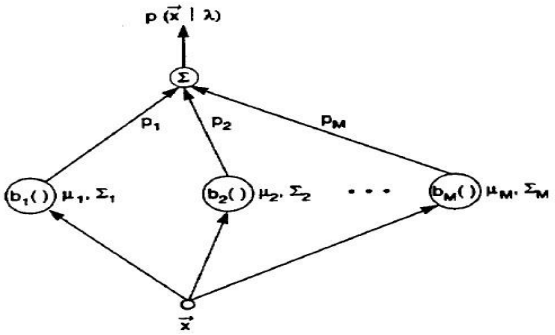

Fig 3: representation of M component of Gaussian mixture density

5. CONCLUSION

The objective of this paper is to describe the use of mixture models as a technique to provide efficient and accurate solutions to problems of essential engineering significant. In this study we reviewed about the GMM for speaker recognition technology. The studies shows that since mixture model analysis yields a large number of theorems, methods, applications and test procedures, there is much related theoretical work as well as research on Gaussian mixture applications which has been not there, here we give description only in terms of automatic speaker recognition technology.

6. ACKNOWLEDGEMENT

This work is sponsored by the CST-UP, Lucknow, India, under CST/D-413.

7. REFERENCES

[1] Reynolds, Douglas. "Gaussian Mixture Models." MIT Lincoln Laboratory, 244 Wood St., Lexington, MA 02140, USA: 1-5. Print.

[2] Neiberg, Daniel . "Text Independent Speaker Verification Using Adapted Gaussian Mixture Models." Centre for Speech Technology (CTT) Department of Speech, Music and Hearing KTH, Stockholm, Sweden: 1-47. Print.

[3] Yegnanarayana, B. , K. Sharat Reddy, and S. P. Kishore. "SOURCE AND SYSTEM FEATURES FOR SPEAKER RECOGNITION USING AANN MODELS." Speech and Vision Laboratory Department of Computer Science and Engineering Indian Institute of Technology Madras: 1-4. Print

[4] Alfredo, Maesa, Fabio Garzia, Michele Scarpiniti, and Roberto Cusani. "Text Independent Automatic Speaker Recognition System Using Mel-Frequency Cepstrum Coefficient and Gaussian Mixture Models." Journal of Information Security, 2012, : 335-340. Print.

[5] Tran, Dat , and Michael Wagner. "Fuzzy Gaussian Mixture Models for Speaker Recognition." Human-Computer Communication Laboratory School of Computing, University of Canberra, ACT 2601, Australia: 1-8. Print.

[6] Plataniotis, K.N. , and D. Hatzinakos. "Gaussian Mixtures and their Applications to Signal Processing." Gaussian Mixtures and their Applications to Signal Processing. Toronto: Department of Electrical and Computer Engineering University of Toronto , Ontario, M5S 3G4, Canada, 2000. . Print.

[7] Kamarauskas, J. . "Speaker Recognition using Gaussian Mixture Models." ELECTRONICS AND ELECTRICAL ENGINEERING 2008. No. 5(85): 29-32. Print.

[8] "Robust Features for Automatic Text-Independent Speaker Recognition Usin g Gaussian Mixture Model." International Journal of Soft Computing and Engineering (IJSCE) Volume-1, November 2011: 330-335. Print.

[9] Q. D'Almeida, Frederico , Francisco A. O. Nascimento, Pedro A. Berger, and Lúcio M. da Silva. "Automatic Speaker Recognition with Multi-

resolution Gaussian Mixture Models (MR-GMMs)." The International Journal of FORENSIC COMPUTER SCIENCE: 9-21. Print.

[10] Reynolds, Douglas A, and Richard C Rose. "Robust Text Independent Speaker Identification Using gaussian Mixture Speaker Model." IEEE transactions on speech and audio processing Volume 3: 72-83. Print.

[11] A. Reynolds, Douglas . "Automatic Speaker Recognition Using Gaussian Mixture Speaker Models." THE LINCOLN LABORATORY JOURNAL VOlUME 8, NUMBER 2,1995 : 173-192. Print.

[12] CAMPBELL, JOSEPH P.. "Speaker Recognition: A Tutorial." PROCEEDINGS OF THE IEEE VOL. 85, : 1437-1462. Print.