Code Separated Text Independent Speaker Identification System Using GMM

Piyush Lotia, Dr.M.R.Khan

[1] *Associate Professor, Department of Electronics and Telecommunication, SSCET Bhilai,India, Lotia_piyush@rediffmail.com*

[2] *Additional Director and Principal, Govt. Engineering College, Raipur, India mrkhan@cgdsteraipur.ac.in*

Abstract

In speaker Identification systems both parametric and nonparametric probability modeling is used. The Gaussian model is the basic parametric model that is used and this model is the basis of other sophisticated and it can be performed in a completely text independent situation. However, it sounds efficient to speaker identification application, but it results long time processing in practice. In this paper, we propose a decision function by using vector quantization (VQ) techniques to decrease the training model for GMM in order to reduce the processing time. In our proposed modeling, we take the superiority of VQ, which is simplicity computation to distinguish between male and female speaker. Then, GMM is applied into the subgroup of speaker to get the accuracy rates. Experimental result shows that our hybrid VQ/GMM method always yielded better improvements in accuracy and bring reduce in time processing. All the experiments have been done in both direct recording speech and mobile phone speech signals.

KeyWords: *MFCC, VQ, Cepstrum, LBG Algorithim.*

1. Introduction

Phonetics is part of the linguistic sciences. It is concerned with the sounds produced by the human vocal organs, and more specifically, the sounds which are used in human speech. One important aspect of phonetic research is the instrumental analysis of speech. This is often referred to as *experimental phonetics*, or *machine phonetics*.The instrumental analysis is performed using one or many of the available instruments. These include X-ray photography and film, air-flow tubes, electromyography (EMG), spectrographs,. The aim for most of these methods is to visualize the speech signal in some way, and to try and capture some aspects of the speech signal on paper or on a computer screen. Today the computer is the most readily available and used tool. We are making use of MATLAB software to process the speech. With the computer the analysis process is much simpler and usually faster than with other tools, however, it does not necessarily produce a result of higher quality

The process of automatically recognizing who is speaking by distinguishing qualities in the speaker's voice is called speaker recognition. For this purpose it is important to preserve the speaker specific information in the speech signal. This is in contrast to the speech recognition task, where the linguistic content of the speaker's voice signal is extracted. Speech recognition then corresponds to a classification problem. Only certain important features that are unique to individuals are extracted while other redundancies are discarded.

The use of personal features, unique to all human being to identify or to verify a person`s identity is a field that is being actively researched. The speaker recognition system consists of the following steps :Speaker identification; in this system, when a user inputs a test utterance, the system will identify which of the speaker made the utterance according to the speech patterns stored in the database. Speaker Verification [31] where the user inputs a test utterance together with his or her ID number, the system verifies a person`s identity claim by comparing the sample of their speech stored in the database to that of the claimed identity. In text independent, the user can utter any text during the identity claim. . In case of text dependent the restriction is that the text uttered during the test session should be identical to the one stored in the database. This is because the pattern extracted from the speech sample takes into account the word or text that is being uttered, therefore different text uttered from the same person will have different pattern.

2. Previous Works

Vector Quantization (VQ) modeling which is a nonparametric method is useful method for speaker identification. The algorithm is based on measurement of the similarity of distributions of features extracted from reference speech samples and from the sample to be attributed. The measure of feature distribution similarity employed is not based on any assumed form of the distributions involved. Quantization technique was proposed 1963 using block quantization of random variables [21]which was further refined with asymptotically quantization concept [22] in 1968.An algorithm was evolved for vector Quantization 1980[24], subsequently 1989 entropy constrained vector quantization was used [25]. Application of vector quantization in speech analysis is found in 2007.

The Gaussian model is the basic parametric model that is used in 1992 and this model is the basis of other sophisticated models. [31]. Using a GMM as the likelihood function, the background model is typically a large GMM trained to represent the speaker-independent distribution features [42] (2000).Finally the GMM-UBM classifier id proposed for process of decision and technique to optimize the size of GMM and UBM [43] (2008).

Both the technique was combined in 1979 as mathematical model [23].Sub band coding vector quantized was proposed [26], further 1996 vector quantization on image application is used [27]. In the year 2009 VQ and GMM are widely applied to the speaker verification, but both have some disadvantages [32]. To overcome those shortages, we introduce a new hybrid VQ decision/GMM model. Although in baseline form, the VQ-based solution is less accurate than the GMM, but it offers simplicity in computation. Therefore, we hope to make use of their merits via a hybrid VQ decision/GMM classifier. Before further discussing the proposed hybrid VQ/GMM model, reviews have been done on some recent works regarding the hybrid VQ/GMM model. There are many forms of GMM and other pattern classification techniques adaptation in the past and yet there are scantiness amount for VQ/GMM adaptation. In hybrid VQ/GMM, there are some researchers used VQ as optimization function to reduce the Expectation Maximization algorithm in order to improve the training speed. Besides, some researchers employed GMM as a post-processor after VQ cluster the speech signal into regions.

3.1 Feature Extraction

The main objective of feature extraction is to extract characteristics from the speech signal that are unique to each individual which will be used to differentiate speakers. Since the characteristic of the vocal tract is unique for each speaker, the vocal tract impulse

response can be used to discriminate speakers. The first step of feature extraction is pre emphasis. The purpose of pre emphasis is to offset the attenuation due to physiological characteristics of the speech production system and also to enhance higher frequencies to improve the efficiency of the analysis as most of the speaker specific information lies within the higher frequencies.

The silence intervals are removed from the input speech based on an envelope threshold. The input signal is up-sampled, segmented to remove samples that fall below a threshold, and then re-sampled back to the original sampling rate, and filtered to smooth out the discontinuities where pauses in active speech occurred.

3.1.1 CEPSTRUM

The cepstrum can be seen as information about rate of change in the different spectrum bands. It was originally invented for characterizing the seismic echoes resulting from earthquakes and bomb explosions. It has also been used to analyze radar signal returns.The cepstrum is a representation used in homomorphic signal processing, to conver(such as a source and filter) combined by convolution into sums of the their cepstra, for linear separation. In particular, the power cepstrum is often used as a feature vector for representing the human voice and musical signals. For these applications, the spectrum is usually first transformed using the mel scale. The result is called the mel-frequency cepstrum or MFC (its coefficients are called mel-frequency cepstral coefficients or MFCCs). It is used for voice identification, pitch detection and much more. The cepstrum is useful in these applications because the low-frequency periodic excitation from the vocal cords and the formant filtering of the vocal tract, which convolve in the time domain and multiply in the frequency domain, are additive and in different regions in the frequency domain. In order to obtain the vocal tract impulse response from the speech signal, a deconvolution algorithm known as the **Mel Frequency Cepstrum Coefficient** is applied.

$$M = (\log_{10} 2)\,(\log_{10}(1 + f/1000)).$$

This algorithm transforms the speech signal which is the convolution between glottal pulse and the vocal tract impulse response into a sum of two components known as the Cepstrum that can be separated by band pass linear filters, if there is no frequency overlapping [19][20].

3.1.2 LPC

LPC is frequently used for transmitting spectral envelope information, and as such it has to be tolerant of transmission errors. Transmission of the filter coefficients directly is undesirable, since they are very sensitive to errors. LPC analyzes the speech signal by estimating the formants, removing their effects from the speech signal, and estimating the intensity and frequency of the remaining buzz. The process of removing the formants is called inverse filtering, and the remaining signal after the subtraction of the filtered modeled signal is called the residue. The numbers which describe the intensity and frequency of the buzz, the formants, and the residue signal, can be stored or transmitted somewhere else. LPC synthesizes the speech signal by reversing the process: use the buzz parameters and the residue to create a source signal, use the formants to create a filter (which represents the tube), and runs the source through the filter, resulting in speech .Because speech signals vary with time, this process is done on short chunks of the speech signal, which are called frames; generally 30 to 50 frames per second give intelligible speech.

4. Vector Quantization

Vector Quantization (VQ) modeling which is a nonparametric method is useful method for speaker identification. The algorithm is based on measurement of the similarity of distributions of features extracted from reference speech samples and from the sample to be attributed. The measure of feature distribution similarity employed is not based on any assumed form of the distributions involved.

Vector quantization is based on the principle of block coding. In automatic speaker recognition, vector quantization is used to cluster or group together feature vectors extracted from the speech sample according to their sound classes i.e. quasi periodic, noise like and impulse like sound. Hence each cluster or centroid represents a different class of speech signal. This enables a text independent speaker recognition system to be realized because the speech vectors are not clustered according to the spoken words but clustering is based on sound classes (VQ) is a data compression Technique, with several successful applications in speech and image coding or speech/speaker recognition [44].

The algorithm used for vector parameters quantization is LBG (Linde-Buzo-Gray) algorithm [39]. A codebook may be small in the beginning and may be gradually expanded to the final size. One method is to split an existing cluster in two smaller clusters and assign a codebook entry to each. The following steps describe this method of designing the codebook: - create an initial cluster consisting of the entire training set; this initial codebook contains a single centroid for the entire set; - split this cluster in two sub clusters, getting a codebook of twice the size; repeat this cluster-splitting process until the codebook reaches the desired size, ideally each cluster should be divided by a hyperlink normal to the direction of maximum distortion [40] .

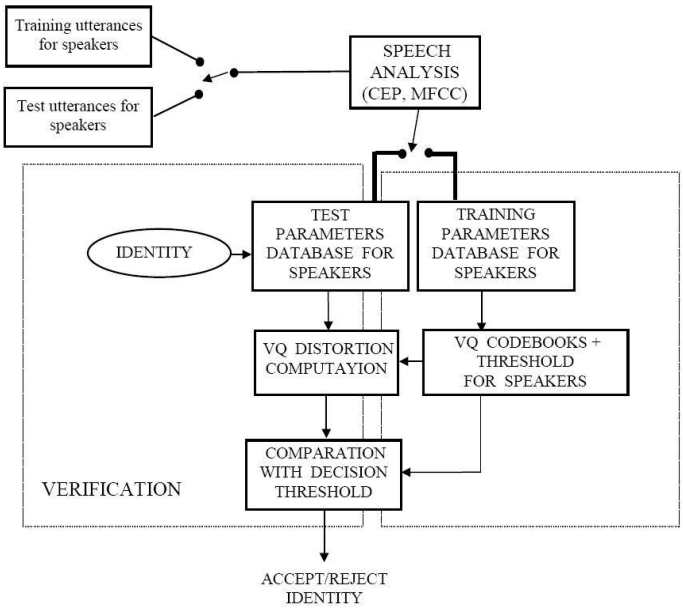

Figure 1: Block Diagram of Basic VQ Training and Classification Vector

In the testing or identification session, the Euclidean distance between the feature vector and codebook for each speaker is calculated and the speaker with the smallest average minimum distance is picked. The speaker models are constructed by clustering the feature

vectors in K separate clusters. Each cluster is then represented by a code vector, which is the centroid of the cluster. The resulting set of code vectors is stored in the speaker database. The matching of an unknown speaker is then performed by measuring the Euclidean distance between the feature vectors of the unknown speaker to the model of the known speakers in the database. The goal is to find the codebook that has the minimum distance measurement in order to identify the unknown speaker.

4.1. Training Model Based On Clustering Technique:

The way in which L training vectors can be clustered into a set of M code book vectors is by K-means clustering algorithm [17]. Classification procedure for arbitrary spectral analysis vectors that chooses the codebook vector is by computing Euclidean distance between each of the test vectors and M cluster centroid. The spectral distance measure for comparing features ***vi***and ***v j*** is as in (1).

$$d(v_i, v_j) = d_{ij} = \mathbf{0 when} v_1 = v_j \text{---- (1)}$$

If codebook vectors of an M-vector codebook are taken as $y_m, 1 \leq m \leq M$

and new spectral vector to be classified is denoted as v, then the index **** m*** of the best codebook entry is as in (2)

$$m^* = \arg\left(min\left(d(v, y_m)\right)\right) for\ 1 \leq m \leq M \text{---- (2)}$$

Clusters [12] are formed in such a way that they capture the characteristics of the training data distribution. It is observed that Euclidean distance is small for the most frequently occurring vectors and large for the least frequently occurring ones. Clustering is a method to reduce the number of feature vectors by using a codebook to represent centers of the feature vectors (Vector Quantization). The LBG (Linde, Buzo and Gray) algorithm [13,14] and the k-means algorithm are some of the most well known algorithms for Vector Quantization (VQ)[15]. The advantage of LBG lies in the generation of accurate codebooks with minimum distortion when a good quality initial codebook is used for LBG. However, due to the complexity, the computation cost is high [16].

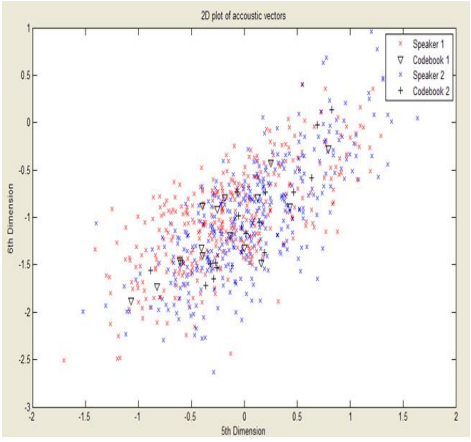

Figure 2: Acoustic Vectors of Two Speakers

5. Mixture Model

A statistical model is a set of mathematical equations which describe the behavior of an object of study in terms of random variables and their associated probability distributions. If the model has only one equation it is called a single-equation model, whereas if it has more than one equation, it is known as a multiple-equation model. In statistics a mixture model is a probabilistic model for density estimation using a mixture distribution. A mixture model can be regarded as a type of unsupervised learning or clustering. Mixture models should not be confused with models for compositional data, i.e., data whose components are constrained to sum to a constant value (1, 100%, etc.).

5.1. Gaussian Mixture Model:

Gaussian mixture model based text-independent speaker verification has attracted the interests of many researchers in the past decade [41]. In many speaker verification applications, the accuracy and computational load are two major criteria for the selection of a proper system. Superior performance

of some GMM variants compared to the other known method in this area has promoted enormous new ideas to enhance the performance and/or to reduce the computational complexity of the system. A variant of Gaussian mixture models which uses universal background model (GMM-UBM) method for speaker verification has established high performance in several NIST evaluations and has become the dominant approach in text-independent speaker verification [42].

Different procedures have been reported in the literature to speed up the computation in a GMM-UBM based speaker verification system while maintaining the system error rate in an acceptable range [37], [38]. Shinoda and Lee proposed a hierarchical structure of model common to all speakers' GMMs and a multi-resolution GMM is used whose mean vectors are organized in a tree structure, with coarse-to-fine resolution when going down the tree [40. The resulting method is known as structural background model-structural Gaussian mixture model (SBM-SGMM). To compensate the performance degradation resulted from the employment of such lower complexity methods, the application of a post processing block, such as a neural network [39] or GMM Identifier [40] is recommended. Gaussian Mixture Models (GMM) is among the most statistically mature methods for clustering (though they are also used intensively for density estimation). The concept of clustering includes that individual data points are generated by first choosing one of a set of multivariate Gaussians and then sampling from them...can be a well-defined computational operation. This optimization method is called Expectation Maximization (EM).

5.1.1. Normal Distribution

In probability theory and statistics, the normal distribution or Gaussian distribution is a continuous probability distribution that describes data that clusters around a mean or average. The graph of the associated probability density function is bell-shaped, with a peak at the mean, and is known as the Gaussian function or bell curve.

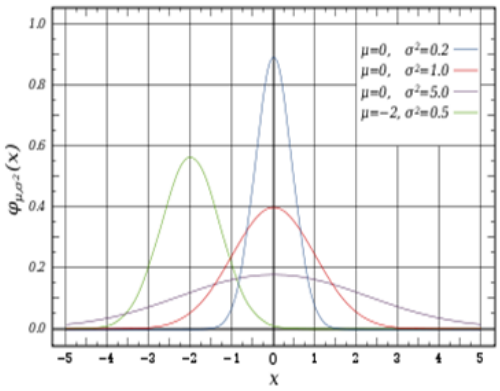

Figure 2: Gaussian Curves

The normal distribution can be used to describe, at least approximately, any variable that tends to cluster around the mean. The probability density function for a normal distribution is given by the formula

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \qquad \ldots\ldots (1)$$

Where μ is the mean, σ is the standard deviation (a measure of the "width" of the bell), and exp denotes the exponential function.

5.1.2 Expectation Maximization (EM)

The Expectation-maximization algorithm can be used to compute the parameters of a parametric mixture model distribution (the a_i's and θ_i's). It is an iterative algorithm with two steps: an expectation step and a maximization step.[5]

a) ***The expectation step***

With initial guesses for the parameters of our mixture model, "partial membership" of each data point in each constituent distribution is computed by calculating expectation values for the membership variables of each data point. That is, for each data point x_j and distribution Y_i, the membership value $y_{i,j}$ is:

$$y_{i,j} = \frac{a_i f_Y(x_j;\theta_i)}{f_X(x_j)}. \qquad \ldots\ldots (5)$$

b) ***The maximization step***

With expectation values in hand for group membership, plug-in estimates are recomputed for the distribution parameters. The mixing coefficients a_i are the means of the membership values over the N data points.

$$a_i = \frac{1}{N}\sum_{j=1}^{N} y_{i,j} \qquad \ldots\ldots (6)$$

The component model parameters θ_i are also calculated by expectation maximization using data points x_i that have been weighted using the membership values. For example, if θ is a mean

$$\mu_i = \frac{\sum_j y_{i,j} x_j}{\sum_j y_{i,j}}.$$

With new estimates for a_i and the θ_i's, the expectation step is repeated to **recompute** new membership values. The entire procedure is repeated until model parameters converge.

5.1.2 The GMM Assumption

There are K components (Gaussians)Each k is specified with three parameters: weight, mean, covariance matrix. The Total Density Function is given by :

$$f(x|\Theta) = \sum_{j=1}^{K} \alpha_j \frac{1}{\sqrt{(2\pi)^d \det(\Sigma_j)}} \exp\left(-\frac{(x-\mu_j)^T \Sigma_j^{-1} (x-\mu_j)}{2} \right)$$

$$\Theta = \{\alpha_j, \mu_j, \Sigma_j\}_{j=1}^{K}$$

$$\alpha = weight \qquad \alpha_j \geq 0 \quad \forall j \qquad \sum_{j=1}^{K} \alpha_j = 1$$

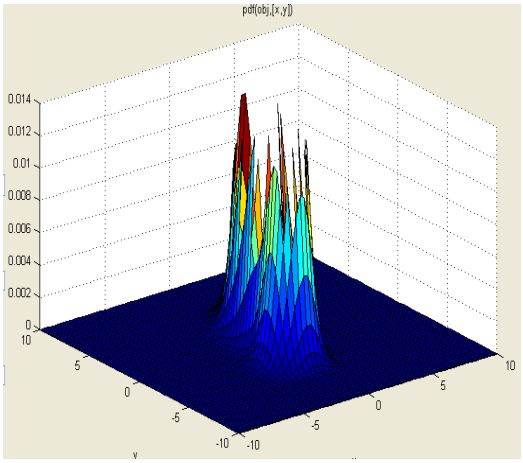

Figure 3: Probability Density Function Normal Distribution Curves

5.1.3. K-Mean Clustering:

Clustering algorithms are used to find groups of "similar" data points among the input patterns. K means clustering is an effective algorithm to extract a given number of clusters of patterns from a training set. Once done, the cluster locations can be used to classify data into distinct classes.

6. Code Separated GMM:

Here we use Quantization (VQ) modeling to speaker subgroups. The decision tree approach is applied to obtain distributed training for VQ model. GMM classification process is then employed on the initial result to achieve a final result. The efficiency of the model is evaluated by computational time and accuracy rate compared to GMM baseline models. Experimental result shows that the hybrid distributed VQ/GMM model yields better accuracy. Besides, it gives substantial reduction in processing time and is faster compared to GMM baseline models. The Decision Tree is one of the most popular classification algorithms in current use in data mining and machine learning. In speaker identification decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. It advantages are it provide robustness and it can perform well with large data a short time.

For VQ, the primary factor is the cookbook sizes [10], an experiment done by Kin Yu et al indicate that the optimum size is not dependent on the amount of training data. When a cookbook is generated, its only remains the centroid which can represent the whole cluster. The amount of data is significantly less, since the number of centroids is at least ten times smaller than the number of vectors in the original sample. This will reduce the amount of computations needed when comparing in later stages. In fact, VQ based solution is less accurate than the GMM. In our proposed hybrid modeling, we take the superiority of VQ, which is simplicity computation to distinguish between male and female speaker.

Besides, we combine the decision tree function and VQ classification techniques in order to fixed identification errors in huge database, this approach is used to separate out the very confusable speakers prior in the same gender group. Later on, we make use of GMM merits to identify the speaker identity in the smaller subgroup. The overall structure of our hybrid system is depicted in fig.2. After feature extraction process, the speech signal will transform to a feature vector form. For the phase 1 of the classification, VQ classifier clustering the speaker model into two subgroup which is subgroup I and subgroup II, In Next stage we use GMM within individual subgroup to find the desired speaker. This process aims to solve the similarity speaker problem

Speaker identification is the computing task of recognizing people's identity based on their voices. There are two main difficulties in this field. First is how to maintain the accuracy rate under large amount of training data. Second is how to reduce the processing time. Previous studies reported that Gaussian Mixture Model (GMM) for speaker identification appears to have many advantages. However, due to long processing time, this process does not always produce satisfying result in practice. Meanwhile, current mechanisms for hybrid production of speaker identification are directed more towards accuracy problems, not processing time optimization. This research focuses on constructing distributed data training on Vector Quantization and in order to make an improvement on the accuracy rate, we utilize dominance of GMM model to get the accuracy rates. GMM process will just applied in the particular subgroup to identify the speaker identity.

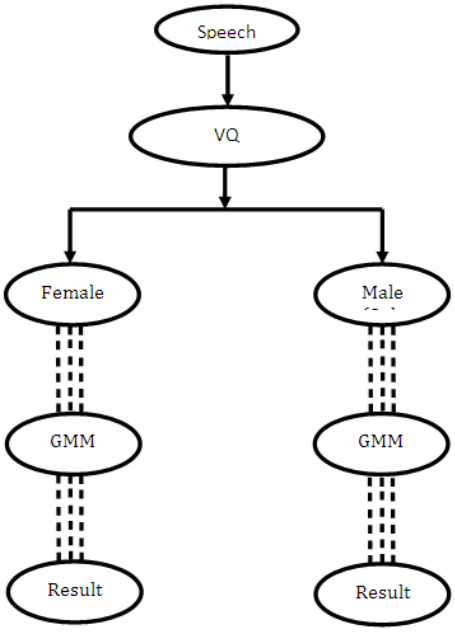

Figure 4. SIS Using VQ and GMM

While in n phase GMM classification engine will calculate log likelihood score for subgroup training speaker data and save it into a speaker model. While in testing phase, a comparison about training speaker and testing speaker will be done. GMM classification engine will make a decision followed by maximum posteriori probability. On account of the GMM model just need to train speaker data in the subgroup instead training all speaker data, the computation time will decrease.

7. Experimental Setup

In this section, we describe the experiments carried out in order to test the different recognizers as stated as above and make a comparison result with our hybrid technique. through a set of preliminary experiments. We performed our evaluation on the our own database of different speech signals of different durations of 20 seconds,40 seconds . In our database we have collected different speech signals over land-line phone, mobile phones and directly via mike.

In very first set we kept training and testing speech sample identical. We performed all the Experiments for direct recording. Both training and testing speech samples are taken from direct recording database.

First method evaluated uses GMM as pattern classification techniques. Table 1 shows the effect of increasing the speakers on performance of the GMM speaker identification system. The next method evaluated uses hybrid VQ decision/GMM as pattern classification techniques. Table 1 show the effect of increasing the speakers on performance of the hybrid VQ/GMM speaker identification system. Accuracy starts off high and slowly declines.

In next set the same procedure is followed for the mobile phone speech. It is found that the accuracy is decreased but here also and improved result is obtained when Hybrid method is used.

In last set of experiment the speech samples for training is taken from direct recording and for test phase mobile phone speech sample is used. In this case there is a little improvement in the performance of hybrid method. It can be also be observed that even hybrid VQ decision/GMM speaker identification accuracy rate has decrease when the training data increase.

8. Results

The results of the all the three different sets of experiments (direct, mobile, and mixed mode) are shown. For all these Experiments were performed with 20 speakers. It is found that execution time is less in VQ/GMM than baseline GMM. Thus, our implementation can categorize as more amplified version for classification techniques in speaker identification system.. The results indicate that with the hybrid modeling, the performance of the speaker identification system is improved. Moreover, the speed of verification is significantly increased because number of features is reduced which consequently decrease the complexity of our identification system.

All the result is summarized in table 1 in a compact manner. Some of the results are shown in bar Charts**.**

RESULTS		Direct Recording		Mobile Recording		Mixed Mode	
SPEAKERS	Rec. Time in sec.	GMM	VQ & GMM	GMM	VQ & GMM	GMM	VQ & GMM
5	20	76	83	58	61	44	42
	40	78	89	54	60	41	37
10	20	71	79	56	59	43	43
	40	73	81	56	57	40	39
15	20	68	74	51	54	37	39
	40	72	83	49	52	38	38
20	20	67	74	50	53	35	34
	40	69	77	48	49	33	35

Table 1. Accuracy (%) for All Thee Sets of Experiments for 20 and 40 sec Training Speech.

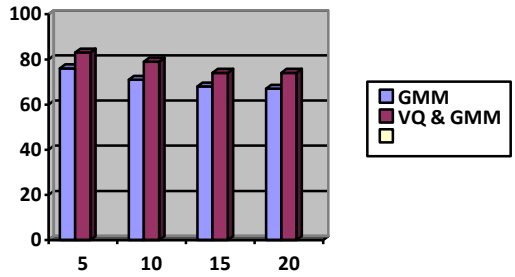

Chart 1. For Direct Recording

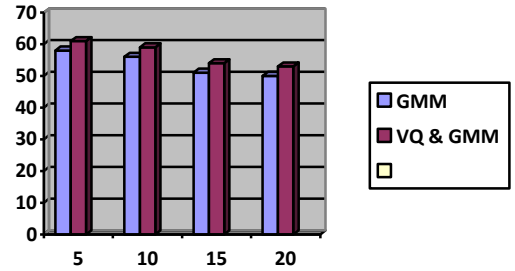

Chart 2. For Mobile Phone Recoding

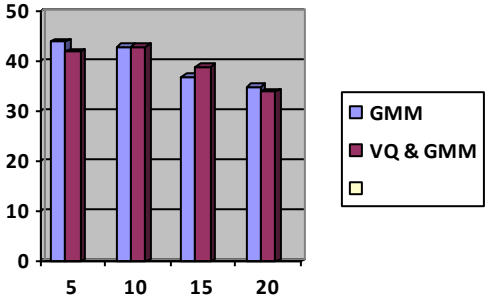

Chart 3. For Mixed Mode

9. Conclusion

In conclusion, this research successfully improves the computational time and accuracy of the text-independent speaker identification system. Future work will be concentrating on investigation of the effectiveness of hybrid VQ decision/GMM for more robust speaker recognition. Investigation on a better adaptation function also will be done to ensure that the hybrid classifier get the better accuracy. VQ and GMM both has their advantages and disadvantages, both of their merits can used to recover their disadvantages of each other. We are intended to improve the computation, the approximation quality and the accuracy of the speaker identification system by the proposed method.

10. Summary

The principal contributions of this experiment are presented a series of evaluation and comparison performance. From the findings of the experiment, the proposed model - hybrid distributed VQ/GMM has been proven to be a powerful tool for text-independent speaker identification system. It has successfully achieved the goal of this research which is solving the time consuming issue for GMM model. Although hybrid distributed VQ/GMM that applied in this study has performed well for several comparisons in experiment, it retains some constraints.

Acknowledgement

We express appreciation to all reviewers for their helpful criticisms and suggestions to our manuscript.

References

[1] Campbell, J.P., "Speaker Recognition: A Tutorial", Proc. of the IEEE, vol. 85, no. 9, 1997, pp. 1437-1462.
[2] Sakoe, H.and Chiba, S., "Dynamic programming algorithm optimization for spoken word recognition", Acoustics, Speech, and Signal Processing, IEEE Transactions on Volume 26, Issue 1, Feb 1978, Page 43 - 49.
[3] Vlasta Radová and Zdenek Svenda, "Speaker Identification Based on Vector Quantization", Proceedings of the Second International Workshop on Text, Speech and Dialogue, Vol. 1692, 1999, Pages: 341 -344.
[4] Lawrence R. Rabiner., "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", Proceedings of the IEEE,77 (2), 1989, p. 257–286.
[5] Reynolds, D. A. and Rose, R. C. "Robust text-independent speaker identification using Gaussian mixture speaker models. IEEE Trans.Speech Audio Process. 3, 1995, pp 72–83.

[6] Solera, U.R., Martín-Iglesias, D., Gallardo-Antolín, A., Peláez-Moreno,C. and Díaz-de-María, F, "Robust ASR using Support Vector Machines",Speech Communication, Volume 49 Issue 4, 2007.
[7] J. Pelecanos, S. Myers, S. Sridharan and V. Chandran, "Vector Quantization Based Gaussian Modeling for Speaker Verification", 15thInternational Conference on Pattern Recognition, Volume 3, 2000, p.3298.
[8] Qiguang Lin, Ea-Ee Jan, ChiWei Che, Dong-Suk Yuk and Flanagan, J,"Selective use of the speech spectrum and a VQGMM method for speaker identification", Fourth International Conference on Spoken Language, Vol 4, 1996, Pg:2415 - 2418.
[9] Davis, S. B. and Mermelstein, P., "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences", IEEE Trans. on Acoustic, Speech and Signal Processing, ASSP-28, 1980, No. 4.
[10] Yu, K., Mason, J., Oglesby, J., "Speaker recognition using hidden Markov models, dynamic time warping and vector quantization" Vision, Image and Signal Processing, IEE Proceedings, Oct 1995.
[11] Vijendra Raj Apsingekar and Phillip L. De Leon; Speaker Model Clustering for Efficient Speaker Identification in Large Population Applications; IEEE transactions on Speech and Audio Signal Processing; Vol. 17, No. 4; May 2009.
[12] Aaron. E. Rosenberg, "New techniques for automatic speaker verification", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol.ASSP-23, No.2, pp.169-176, April 1975
[13] Y. Linde, A. Buzo, and R.M. Gray,. "An algorithm for vector quantizer design,". *IEEE Trans. Communications, vol. COM-28(1), pp. 84-96, Jan. 1980.*
[14] R. Gray. "Vector quantization,". *IEEE Acoust., Speech, Signal Process. Mag., vol. 1, pp. 4-29, Apr. 1984.*
[15] F.K. Soong, A.E. Rosenberg, L.R. Rabiner, and B.H. Juang,. "A Vector quantization approach to speaker recognition,". *in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., vol. 10, Detroit, Michingon, Apr. 1985, pp. 387-90.*
[16]. Sookpotharom Supot, Reruang Sutat, Airphaiboon Surapan, and Sangworasil Manas. Medical Image Compression Using Vector Quantization and Fuzzy C-Means. [Online] http://www.kmitl.ac.th/biolab/member/sutath/final_paper_iscit02.pdf.
[17] Rabiner.L.& Juang B.H., "Fundamentals of speech recognition", Prentice Hall, NJ 1993.
[18] Wai C. Chu, "Speech Coding Algorithm", Wiley Interscience, New York, NY, 2003.
[19] C. Becchetti and Lucio Prina Ricotti, "Speech Recognition", John Wiley and Sons, England, 1999.
[20] E. Karpov, "Real Time Speaker Identification," Master`s thesis, Department of Computer Science, University of Joensuu, 2003
[21]J.Y. Hwang , P.M. Schuelthess , " Block quantization of coreleted Gaussian Random Variables," IEEE Tran. Commu Vol COM 11, pp 289-296 Sep 1963.
[22] H. Gish, J.N. Pierce, " Asympoticaaly efficient Quantizing ," IEEE Tran. Infor Tehory Vol IT-14, pp, 676-683, Sept 1968.
[23] AGeirsoi , "Asympoticaaly optimal block Quantization ," IEEE Tran. Infor Tehory Vol IT-25, pp, 378-380, July 1979.
[24] Y. Londe, A. Buzo, R.M. Gray," An Algoii. For vector quantizer design," IEEE Tran. Commun. Vol Comm 28, ppi, 84-95, Jan 1980.
[25] P.A. Chou, T. Lookabaugh, and R. M. Gray, "Entropy Constrained Vector Quantization ," IEEE Tran Acustic speec signal processing. Vol 37, pp, 31-42, Jan 1989.
[26] N. M. Akrout, C. Diab, R. Prost, and R. Goutte,"Code Word Orientation an improved subband vector quantization scheme for image coding," Opt. Engg. VOl. 33 No. 7, pp 2294-2398 July 1994.
[27] Poonam Bansal, Amita Dev, Shail Bala Jain, " Automatic speaker identification using vector Quantization, " Asian Journal of Information Technology pp938-942, 2007.
[28] S.R.Das, W.S. Mohn, "A scheme for speech processing in automatic speaker verification", IEEE Transactions on Audio And Electroacoustics, Vol.AU-19, pp.32-43, March 1971.
[29] Chulhee Lee, Donghoon Hyun, Euisun Choi, Jinwook Go and Chungyong Lee, "Optimizing feature extraction for speech recognition", IEEE Transactions on Speech and Audio Processing, Vol.11, No.1, January 2009.
[30] Guorong Xuan, Wei Zhang and Peiqi Chai; EM Algorithms Of GaussianMixture ModelandHidden Markov Model; IEEE Transactions; 2001.
[31] D.A.Reynols, R.C.Rose "A Gaussian mixture modeling approach to text independent speaker recognitions system" in Proc. Int. Conf. Signal Processing Apll. Tech. Nov. 1992 pp. 967-973.
[32]A.Revathi and Y.Venkataramani,"Text independent speaker identification/verification using multiple features", Proceedings of IEEE International Conference on Computer Science and Information Engineering, April 2009, Los Angeles, USA.
[33] J. Koolwaaij and L. Boves, "Local normalization and delayed decision making in speaker detection and tracking," *Digital Signal Processing*, vol. 10, no. 1–3, pp. 113–132, 2000.

[34] P. Delacourt and C. J. Wellekens, "DISTBIC: A speaker based segmentation for audio data indexing," *Speech Communication*, vol. 32, no. 1-2, pp. 111–126, 2000.
[35] R. Auckenthaler,M. Carey, and H. Lloyd-Thomas, "Score normalization for text-independent speaker verification system, *"Digital Signal Processing*, vol. 10, no. 1, 2000.
[36] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, no. 1, pp. 19–41, 2000.
[37] Peter Varchol, Duston Levicky, "Optimization of GMM for text Independent Speaker Verification System" IEEE transaction on signal processing 978-1-4244-2088-9/08/2008/IEEE.
[38] Rabiner, L.R., Juang, B.H. *"*Fundamentals of speech recognition*"*, Prentice-Hall International , Inc. (1993).
[39] Lupu E., Toderean G., "Speaker verification using single section vector quantization", International Symposium on Signals Circuits and System SCS'97 iasi, 2-3 Oct. 1997.
[40] Furui, S. Digital speech processing, synthesis and recognition, Marcel Dekker Publications, 1989.
[41] D. A. Reynolds and R. C. Rose, "Robust textindependent speaker identification using Gaussian mixture speaker models," *IEEE Trans. on Speech Audio Processing*, vol. 3, no. 1, pp. 72-83, Jan. 1995.
[42] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, no. 1-3, pp. 19-41, Jan. 2000.
[43] J. Mclaughlin, D. A. Reynolds, and T. Gleason, "A study computation speed-ups of the GMM-UBM speaker recognition system," in *Proc. Eurospeech'99*, pp. 1215-1218, 1999.
[44] K. Shinoda and C. H. Lee, "A structural Bayes approach to speaker adaptation," *IEEE Trans. Speech Audio Processing*, vol. 9, no. 3, pp. 276 287, May 2001.
[45] B. Xiang and T. Berger, "Efficient text-independent speaker verification with structural Gaussian mixture models and neural network," *IEEE Trans. Speech Audio Processing*, vol. 11, no. 5, pp. 447-456, Sept. 2003.
[46] R. Saeidi, H. R. Sadegh Mohammadi, M. Khalaj Amirhosseini, "An efficient GMM classification post-processing method for structural Gaussian mixture model based speaker verification," *to be Presented at ICASSP'06*, Toulouse, France, May 2006.