Multistage VQ Based GMM For Text Independent Speaker identification System

Piyush Lotia, M.R. Khan

Abstract— **The use of Gaussian Mixture Models (GMM) are most common in speaker identification due to 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, in second phase of classification, decision tree rule are applied to separate out the similar speaker in same gender into two difference group. While in phase 3, 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 almost 20% reduce in time processing.**

Index Terms— MFCC, VQ, Cepstrum, LBG Algorithim.

I. Introduction

Automatic speaker recognition is the use of a machine to recognize a person from a spoken phrase. This makes it possible to use the speaker`s voice to verify their identity and can be used to control access to services such as voice dialing, banking by telephone, database access service, security control for confidential information area and remote access to computers. 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. The expected result is accepted or text independent, the user can utter any text during rejects the identity claim. When a system is said to be the input of test utterance. 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. Speech recognition is done [30] using audio visual features. Parameters of the mel-cepstral transformation are optimized in [29] for speech recognition. Das et.al [28] has introduced the scheme for speech processing in speaker verification. They have indicated that utterances should preferably be collected over a long period of time. Rosenberg et.al [12] has introduced new techniques for speaker verification.

Manuscript received May 29, 2011.

Piyush Lotia, Associate Professor, Department of Electronics and Telecommunication, SSCET-Bhilai (C.G.), India (e-mail: lotia_piyush@rediffmail.com).

Dr. M.R. Khan, Associate Professor, Department of Electronics and Telecommunication, GEC Raipur (C.G.), India (e-mail: mrkhan@cgdsteraipur.ac.in).

II. Literature Review

Vector Quantization (VQ)is a pattern classification technique applied to speech data to form a representative set of features. 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. Normalization and selection of speech segments technique is also used for voice recognition [33-35](1988,2000,2001). A second normalization is based on the range of matching scores of the supposed speaker's model against other speaker's models [36-37] (2000).

Neural network based speaker recognition is also an effective method, where supervised or unsupervised training is used [38-39](1994).The MNTN is found to perform better than full-search VQ classifiers [40-41](1990,1991).

III. PRE-PROCESSING AND FEATURE EXTRACTION

A. REEMPHASIS

Pre-emphasis is performed by passing the signal through a high pass filter. 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.

B. Silence Removal

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. The threshold used is a scaled function of the median of the envelope. The default threshold is one-fourth of the median of the envelope.

C. 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. The Hamming window is used as it is more selective than the rectangular window. Since the characteristic of the vocal tract is unique for each speaker, the vocal tract impulse response can be used to discriminate speakers. Therefore 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. In this feature extraction, we have used:

Sampling Frequency	11025Hz
No. of Coefficient Per Frame	12Hz
No. of Filters in Filter in Filter bank	29Hz
Length Of The Frame	256Hz
Frame Increment	128Hz

In 1968 by fant, used the Bark/ Mel scale for the frequency localization of the filters.

$$F_{mel} = 1000.(\log (1+f/1000))/ (\log 2)$$

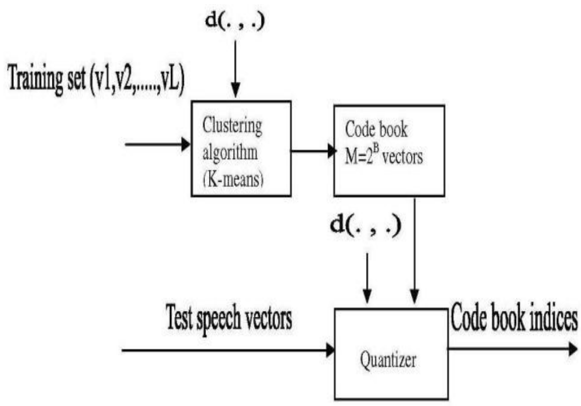
This algorithm transforms the speech signal which is the convolution between glottal pulse and the vocal tract impulse response into a sum of two components known as the Cepstrum that can be separated by band pass linear filters, if there is no frequency overlapping [19][20].

IV. VECTOR QUANTIZATION

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 [18]. 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.

D. Training Model Based On Clustering Technique

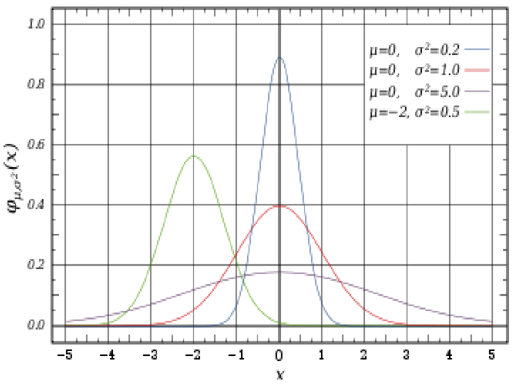
The way in which L training vectors can be clustered into a set of M code book vectors is by K-means clustering algorithm [17]. Block diagram for K-means clustering and classification is shown in fig. 1.



[Fig-1: Block Diagram of Basic VQ Training & Classification Vector]

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).

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



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)

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

Clusters [12] are formed in such a way that they capture the characteristics of the training data distribution. In testing phase, a function will computes the Euclidean distance between training data and testing data. The system will identify which calculation yields the lowest value and checks this value against aconstraint threshold. If the value is lower than the threshold, the system outputs an answer.It is observed that Euclidean distance is small for the most frequently occurring vectors and large for the least frequently occurring ones.

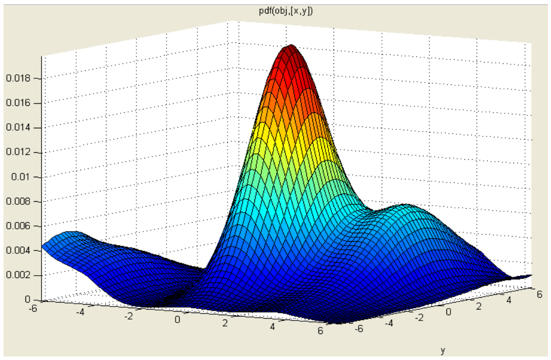
Clustering is a method to reduce the number of feature vectors by using a codebook to represent centers of the feature vectors (Vector Quantization). The LBG (Linde, Buzo and Gray) algorithm [13,14] and the k-means algorithm are some of the most well known algorithms for Vector Quantization (VQ)[15]. 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].

V. MIXTURE 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.).

A. Gaussian Mixture Model

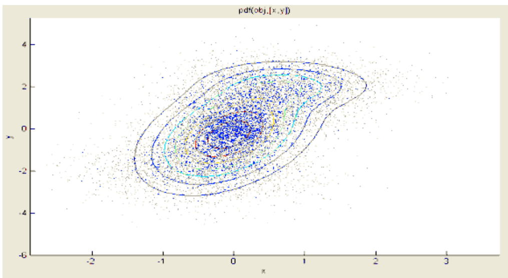
Gaussian Mixture Models (GMM) is among the most statistically mature methods for clustering (though they are also used intensively for density estimation). The concept of clustering includes that individual data points are generated by first choosing one of a set of multivariate Gaussians and then sampling from them...can be a well-defined computational operation. This optimization method is called Expectation Maximization (EM).



[Fig-3: Probability Density Function Normal Distribution Curves]

B. 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. 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. 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. [11]



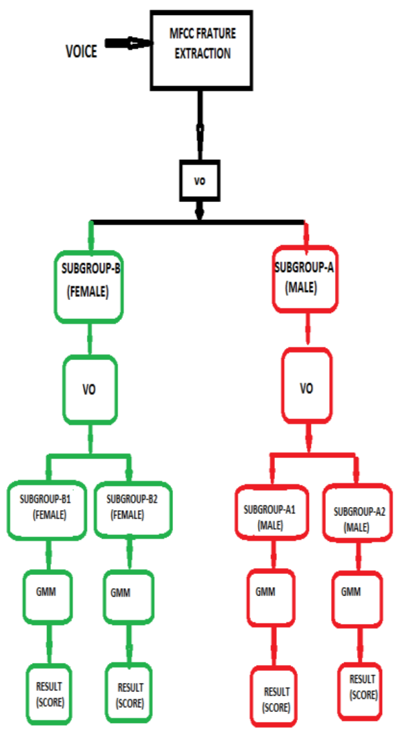
[Fig-4: Mixture obtained by K-Means Clustering and contour plot]

VI. VECTOR QUANTIZATION BASED GAUSSIAN MIXTURE MODELING

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 in 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 novel 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 MFCC 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 A and subgroup B. In phase 2 classification, we use the decision tree function to separate out the speaker model that gain the similar score into 4 difference group which are subgroup A1, A2, B1 and B2. This process aims to solve the similarity speaker problem in order to make an improvement on the accuracy rate when the 3 classification, 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.



[Fig-5. Speaker identification system based on vector quantization decision function for Gaussian mixture modeling]

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.

VII. EXPERIMENTAL SETUP AND RESULTS

The focus of this paper is to compare the performance of the baseline GMM system and the proposed VQ/GMM described earlier in Fig 5. 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. Experiments are conducted on a clean condition. In orders to get a fair comparison between 2 types of classifier, for each of them we have properly selected the same datasets and done some pre-processing for enhanced the feature data 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 and 60 seconds. In our database we have collected different speech signals over land-line phone, mobile phones and directly via mic.

A. Gaussian Mixture Model Baseline System Evaluation

The first method evaluated uses GMM as pattern classification techniques. The first set of experiments; we use our database to check the result when the number of speakers is increased from 5 to 25 Fig. 6 shows the effect of increasing the speakers on performance of the GMM speaker identification system. Accuracy starts off highly 82.1% as would be expected, and slowly declines to approximately 73.2%.As can be observed, GMM speaker verification accuracy rate has decrease when the training data increase; this is due to the complexity of the computation. Besides, it ignores knowledge of the underlying phonetic content of the speech; therefore it does not take advantage of all available information.

B. Hybrid Vector Quantization/Gaussian Mixture Model System Evaluation

The next method evaluated uses hybrid VQ decision/GMM as pattern classification techniques. This is the new hybrid pattern classification as we proposed for speaker identification system. Here, we classified speaker by three phase of classification which the first phase we distinguish the male and female speakers using VQ decision approach and in the second phase of classification, decision tree rule are applied to separate out the similar speaker in same gender into two difference group. While in phase 3, GMM is applied into the sub group of speaker. Fig. 6 shows the effect of increasing the speakers on performance of the hybrid VQ/GMM speaker identification system. Accuracy starts off highly 89.4%, and slowly declines to approximately 77.5%.

As can be observed, even hybrid VQ decision/GMM speaker identification accuracy rate has decrease when the training data increase, but it still obtain the better result if compare with baseline GMM. Besides, it seems more stable to handle the large data set.

International Journal of Soft Computing and Engineering (IJSCE)
ISSN: 2231-2307, Volume-1, Issue-2, May 2011

VII. RESULT

The result of time processing for (5, 10, 15, 20, and 25) speakers by using VQ, baseline GMM and hybrid VQ/GMM shows in table in fig.6. By our observation the 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. Obviously, a significant improvement compared to the baseline system is reported, a reduction identification times up to 20% is reached. 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 over 50% which consequently decrease the complexity of our identification system.

In this set up two mixtures are used. If it is increased, the acquisition time is reduced. Also if the speakers increase accuracy reduces. To solve this problem, the recording time is increased.

No. of speakers	Time (in sec)	VQ	GMM	VQ based GMM
5	20	62	75.4	83.2
	40	63.5	77.2	87.8
	60	65.2	82.1	89.4
10	20	60.2	73.2	81.8
	40	61.4	73.2	85.2
	60	62.5	79.5	86.5
15	20	58.1	71.5	77.1
	40	54.2	72.9	80.2
	60	59.2	77.1	81.6
20	20	53.8	68.5	74.7
	40	54.2	70.5	78.4
	60	55.8	75.2	79.8
25	20	51.2	67.1	71.8
	40	52.4	69.7	75.8
	60	53.7	73.2	77.5

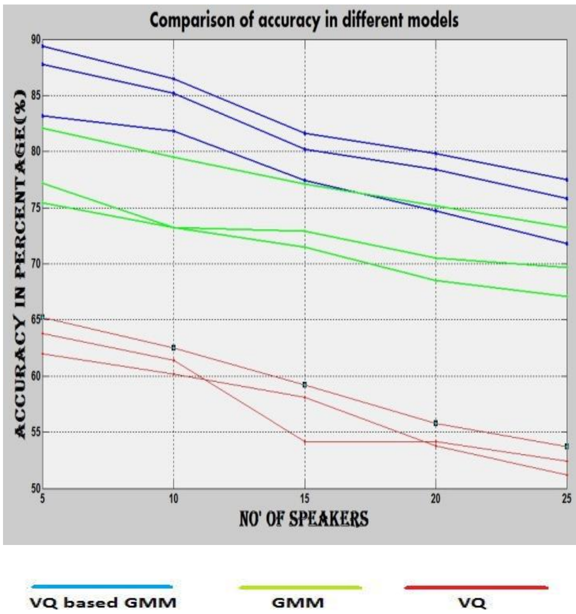


Fig-6: Accuracy (in %) of different modeling techniques for different number of speakers for different times (in sec)

IX. CONCLUSION

In this paper, a new, hybrid, robust and simplicity computation method of pattern classification technique for speaker identification system is proposed. From the experiments, we observe that one good way of applying hybrid method between VQ, decision tree and GMM because of their difference ways to classified data. 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. 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 rate.

ACKNOWLEDGMENT:

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 lanagan, 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 continuouslyspoken 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

[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] P. Li and J. E. Porter, *"Normalizations and selection of speech segments for* speaker recognition scoring, "in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '88)*, vol. 1, pp. 595–598, New York, NY, USA, April 1988.

[34] Y. GU and T. Thomas, "A text-independent speaker verification system using support vector machines classifier," in Proc. European Conference on Speech Communication and Technology(Euro speech '01), pp. 1765–1769, Aalborg, Denmark, September 2001. speech recognition", IEEE Transactions on Speech and Audio Processing, Vol.11, No.1, January 2009.

[35]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.

[36] 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.

[37] R. Auckenthaler,M. Carey, and H. Lloyd-Thomas, "Score normalization for text-independent speaker verification system, *"Digital Signal Processing*, vol. 10, no. 1, 2000.

[38]Lippman, R.P. "Neural Classifiers Useful for Speech Recognition", IEEEMagazine on Acoustic. Speech and Signal Processing, vo1.4, no.2, pp. 4-22.

[39] K. R. Farrell, R.Mammone, and K. Assaleh, "Speaker recognition using neural networks and conventional classifiers," *IEEE Trans. Speech, and Audio Processing*, vol. 2, no. 1, pp. 194–205,1994.

[40]M. D. Richard and R. P. Lippmann, "Neural network classifiers estimate Bayesian a posteriori probabilities," *Neural Computation*, vol. 3, no. 4, pp. 461–483, 1991.

[41] Niles, L.T. and Silverman,querque, "Combining Hidden Markov Model and Neural Classifiers", April, 1990 **H.F.,** PIOC. ICASSP-1990, pp.417-420.

[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, pp. 19–41, 2000.

[43] 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.