

# **Comparative Analysis of Automatic Speaker Recognition using Kekre's Fast Codebook Generation Algorithm in Time and Transform Domain**

Dr. H. B. Kekre  
Senior Professor  
MPSTME, SVKM's NMIMS University  
Mumbai-400056.

Vaishali Kulkarni  
Ph.D Research Scholar,  
Associate Professor, MPSTME, SVKM's NMIMS University  
Mumbai-400056.

## **ABSTRACT**

In this paper, an approach based on Kekre's fast code book Generation (KFCG) Algorithm in the transform domain has been proposed. KFCG is used for feature extraction in both the training and testing phases. Three methods for codebook generation have been used. In the 1<sup>st</sup> method, codebooks are generated from the speech samples by using Discrete Fourier Transform (DFT). In the 2<sup>nd</sup> method, the codebooks are generated using Discrete Cosine Transform (DCT). In the 3<sup>rd</sup> method, the codebooks are generated using the Discrete Sine Transform (DST). For speaker identification, the codebook of the test sample is similarly generated and compared with the codebooks of the reference samples stored in the database. The results obtained for the above methods in the transform domain are compared with the results obtained in the time domain analysis. The results show that KFCG gives better results in transform domain than in time domain. Also the results improve as the vector dimension while generating the codebook is increased.

## **General Terms**

Speaker Recognition, Phone banking, Database services.

## **Keywords**

Vector Quantization (VQ), Code Vectors, Code Book, Discrete Fourier Transform (DFT), Discrete Sine Transform (DST), Discrete Cosine Transform (DCT).

## **1. INTRODUCTION**

Speaker Recognition [1] – [3] is the process of automatically recognizing who is speaking on the basis of individual information included in speech signals. It can be divided into Speaker Identification and Speaker Verification.[3] – [5] Speaker Identification determines which registered speaker provides a given utterance from amongst a set of known speakers (also known as closed set identification). Speaker Verification accepts or rejects the identity claim of a speaker (also known as open set identification).Speaker Recognition technology makes it possible to extract the identity of the person speaking. This technology has made it possible to use the speaker's voice to control access to restricted services, for example, for giving commands to computer, phone access to banking, database services, shopping or voice mail, and access to secure equipment.

Speaker identification task can be further classified into text-dependent or text-independent task [4], [5]. In the former case, the utterance presented to the system is known beforehand. In the latter case, no assumption about the text being spoken is made, but the system must model the general underlying properties of the speaker's vocal spectrum. In general, text-dependent systems are more reliable and accurate, since both the content and voice can be compared [3], [4].

Speaker Recognition systems have been developed for a wide range of applications like Phone banking, Database services, controlled access. [6] - [9]. Although many new techniques have been developed, widespread deployment of applications and services is still not possible. None of these systems gives accurate and reliable results.

Vector Quantization (VQ) maps a 'k' dimensional vector space to a finite set  $C = \{C_1, C_2, C_3 \dots C_N\}$ . The set C is called codebook consisting of 'N' number of codevectors and each code vector  $C_i = \{ci_1, ci_2, ci_3 \dots cik\}$  is of dimension k. The key to VQ is the good codebook. The method most commonly used to generate codebook is the Linde-Buzo-Gray (LBG) algorithm [10], [11] which is also called as Generalized Lloyd Algorithm (GLA).VQ [10] – [12], [20] is an efficient data compression technique and has been used in various applications involving VQ-based encoding [15] and VQ based recognition. VQ has been very popular in the field of speech recognition. [13] – [19]. We have proposed speaker identification using VQ by LBG algorithm [24] and KFCG algorithm in the time domain [25]. Since the results obtained by KFCG are far better than LBG, in this paper we propose speaker identification using VQ by KFCG algorithm in the transform domain. Also comparison of the results obtained by KFCG in the time and transform domain is shown.

In the next section we present Kekre's Fast Codebook Generation algorithm (KFCG). Section 3 consists of three approaches which have used for code book generation. Section 4 consists of results and conclusions in section 5.

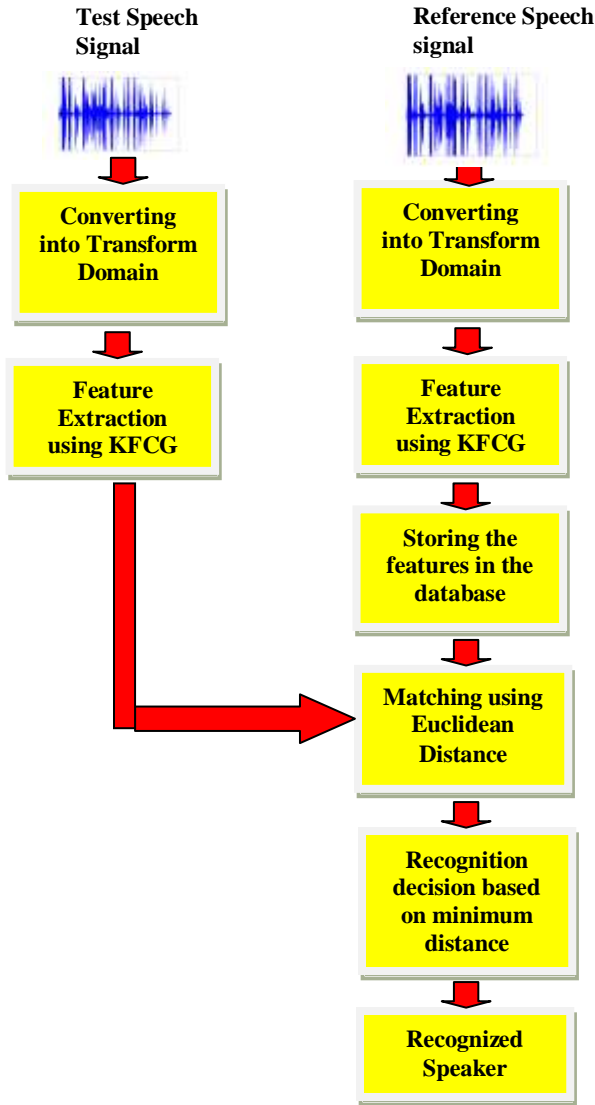


Figure 1 Speaker Recognition system

## 2. KEKRE'S FAST CODEBOOK GENERATION ALGORITHM (KFCG)

In this algorithm for generating the codebook the following procedure is used [20] – [23]:

- Initially we have only one cluster which is the entire set of training vectors. Design a 1-vector codebook; which is the centroid of the cluster.
- Split the cluster into two by comparing the first element of all the training vectors in the cluster with the first element of the centroid as follows:  
If  $v_{i,1} > c_{1,1}$  then  $v_{i,1}$  is grouped into C1 (cluster 1).  
Else  $v_{i,1}$  is grouped into C2 (cluster 2).

Where  $v$  is the training vector and  $c$  is the centroid.

- Find the centroids of C1 and C2 (this is 2-vector codebook). Now split C1 into two clusters by

comparing the second element of all the training vectors in C1 with the second element of its centroids explained in step 2 above. Similarly split C2 into two clusters by comparing the second element of all the training vectors in C2 with the second element of its centroid.

- Now four clusters are formed. Centroids of these four clusters are computed (this is 4-vector codebook). These four clusters are split further by comparing the third element of the training vectors in that cluster with the third element of its centroid as explained in step 2 above.
- The process is repeated until a codebook of size  $M$  is designed.

Figure 2 shows the generation of codevectors using the KFCG algorithm for 2 dimensional cases.

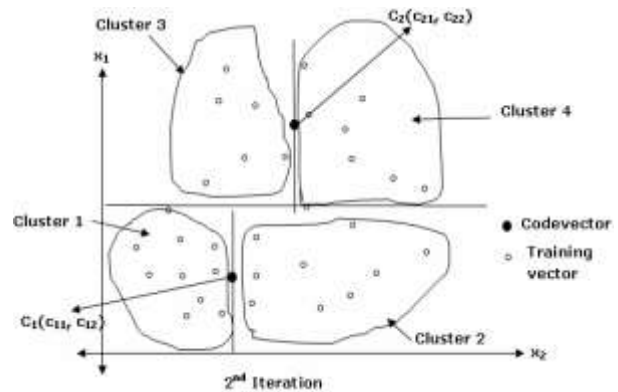


Figure 2 KFCG for 2 Dimensional cases

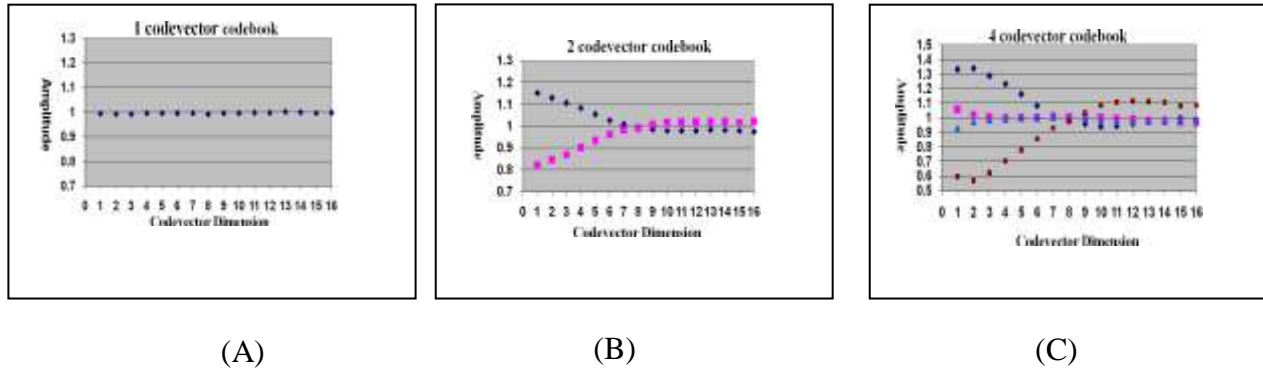
Figure 3 shows the process of obtaining four codevectors using the KFCG algorithm for 16 dimensional vectors. Figure 3(A) shows the one vector codebook which is the centroid of the entire training set. Figure 3(B) shows the two vector codebook obtained by splitting the training set by comparing with the first element of all the training vectors in the cluster with the first element of the centroid. Figure 3(C) shows the four vector codebook obtained similarly by splitting the two clusters.

## 3. CODE BOOK GENERATION APPROACH

The speech signal has amplitude range from -1 to +1. It is first converted into positive values by adding +1 to all the sample values. The samples are converted into 16, 32 or 64 dimensional vector space by considering an overlap of 4 between the samples of consecutive blocks. The codebooks for different size of code vectors are found using the following approaches:

### A. DFT

The DFT of the 16, 34 or 64 dimensional vectors is found and then code books for different size of code vectors are



**Figure 3:** Generation of four codevectors using KFCG algorithm

**Table 1** Comparison of time domain and transform analysis for 16 dimensional vectors using KFCG

No. of codevectors	Time Domain		DFT		DCT		DST	
	matches	%	matches	%	matches	%	matches	%
2	35	70	41	82	41	82	41	82
4	44	88	41	82	42	84	41	82
8	45	90	42	84	42	84	41	82
16	44	88	40	80	46	92	46	92
32	41	82	43	86	45	90	47	94

found using the KFCG algorithm discussed in the previous section.

**B. DCT**

The DCT of the 16, 34 or 64 dimensional vectors is found and then code books for different size of code vectors are found using the KFCG algorithm discussed in the previous section.

**C. DST**

The DST of the 16, 34 or 64 dimensional vectors is found and then code books for different size of code vectors are found using the KFCG algorithm discussed in the previous section.

**4. RESULTS**

**Basics of speech signal**

The speech samples used in this work are recorded using Sound Forge 4.5. The sampling frequency is 8000 Hz (8 bit, mono PCM samples). The samples are collected from different speakers. Samples are taken from each speaker in three sessions so that training model and testing data can be created. Twelve samples per speaker are taken. The samples recorded in one session are kept in database and the samples recorded in second session are used for testing.

The algorithms are implemented using MATLAB 7.7.0. The feature vectors of all the reference speech samples are stored in the database in the training phase. In the matching phase, the test sample that is to be identified is taken and similarly processed as in the training phase to form the feature vector. The stored feature vector which gives the minimum Euclidean distance with the input sample feature vector is declared as the speaker identified.

Table 1 shows the comparison of time domain and transform domain analysis on a sample set of 50 speakers for text dependent identification using KFCG algorithm. The shaded cells in each row show the maximum accuracy for the particular number of codevectors. The results show that for 16 dimensional vectors DFT does not show any improvement over time domain. In fact, time domain gives better results. For DCT, the accuracy improves as the codevectors size is increased. For 16 codevectors the accuracy is 92%. For DST, the results are better than DCT and the accuracy is 94% for 32 codevectors.

We have also analyzed the performance of DFT, DCT and DST for 32 and 64 dimensional vectors. Figure 4 shows the performance comparison 32 dimensional vectors for the three transforms. As seen from the graph the performance of DFT has improved for 32 dimensional vectors as compared to 16 dimensional vectors shown in Table 1. The maximum accuracy seen here is 92% for a codevector size of 16. For DCT, the results are still better, with the maximum accuracy of 94% for a codevector size of 8. DST gives the best results and the accuracy is as high as 96% for a codevector size of 16. Overall we can say that the performance has improved by increasing the vector dimension from 16 to 32 for generating the codebooks. Figure 5 shows the performance comparison 64 dimensional vectors for the three transforms. As seen from the graph the performance of DFT has improved for 64 dimensional vectors as compared to 32 dimensional vectors shown in Figure 4. The maximum accuracy seen here is as high as 98% for a codevector size of 16. For DCT, the maximum accuracy of 96% for a codevector size of 4. For DST gives the maximum accuracy 96% for a codevector size of 2 only. 64 dimensional vectors gives the best results for all the three transforms.

## 5. CONCLUSION

Very simple techniques based on the lossy compression using vector quantization have been introduced. KFCG is simple and faster technique for codebook generation. As can be seen from the results, KFCG gives better results in the transform domain than in the time domain. Also the results improve as the vector dimension while generating the codebook is increased. The maximum accuracy obtained is as high as 98% for a codevector size of 16 using DFT. The present study is still ongoing, which may involve different techniques to find the feature vectors and their comparison.

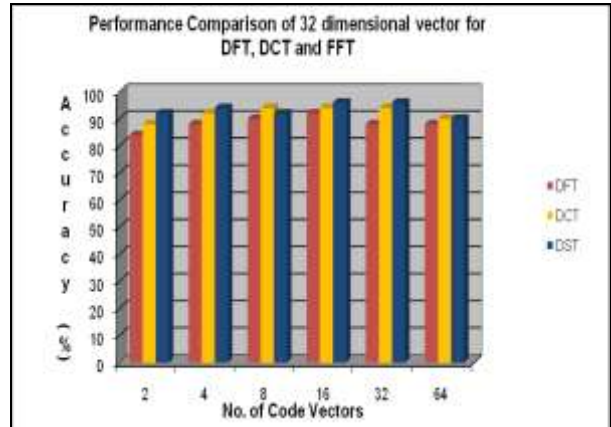


Figure 4 Performance Comparison of 32 dimensional vectors for DFT, DCT and FFT

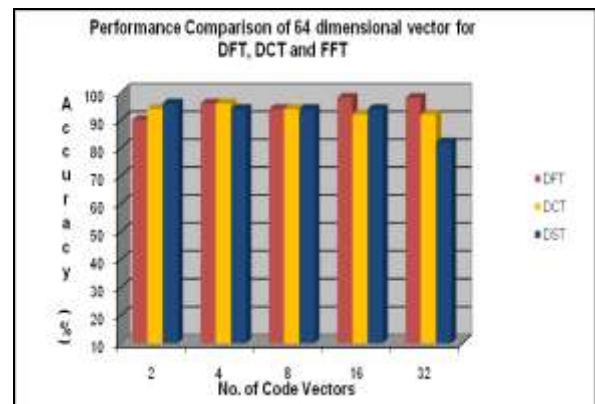


Figure 5 Performance Comparison of 64 dimensional vectors for DFT, DCT and FFT

## 6. REFERENCES

- [1] Lawrence Rabiner, Biing-Hwang Juang and B.Yegnanarayana, "Fundamental of Speech Recognition", Prentice-Hall, Englewood Cliffs, 2009.
- [2] S Furui, "50 years of progress in speech and speaker recognition research", ECTI Transactions on Computer and Information Technology, Vol. 1, No.2, November 2005.
- [3] D. A. Reynolds, "An overview of automatic speaker recognition technology", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP'02)*, 2002, pp. IV-4072–IV-4075.
- [4] Joseph P. Campbell, Jr., Senior Member, IEEE, "Speaker Recognition: A Tutorial", *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1437-1462, September 1997.
- [5] F. Bimbot, J.-F. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-García, D.Petrovska-Delacrétaz, and D. A. Reynolds, "A tutorial on text-independent speaker verification," *EURASIP J. Appl. Signal Process.*, vol. 2004, no. 1, pp. 430–451, 2004.

- [6] D. A. Reynolds, "Experimental evaluation of features for robust speaker identification," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 4, pp. 639–643, Oct. 1994.
- [7] Tomi Kinnunen, Evgeny Karpov, and Pasi Fr'anti, "Realtime Speaker Identification", ICSLP2004.
- [8] Marco Grimaldi and Fred Cummins, "Speaker Identification using Instantaneous Frequencies", *IEEE Transactions on Audio, Speech, and Language Processing*, vol., 16, no. 6, August 2008.
- [9] Zhong-Xuan, Yuan & Bo-Ling, Xu & Chong-Zhi, Yu. (1999). "Binary Quantization of Feature Vectors for Robust Text-Independent Speaker Identification" in *IEEE Transactions on Speech and Audio Processing*, Vol. 7, No. 1, January 1999. IEEE, New York, NY, U.S.A.
- [10] R. M. Gray.: 'Vector quantization', *IEEE ASSP Marg.*, pp. 4-29, Apr. 1984.
- [11] Y. Linde, A. Buzo, and R. M. Gray.: 'An algorithm for vector quantizer design,' *IEEE Trans. Commun.*, vol. COM-28, no. 1, pp. 84-95, 1980.
- [12] A. Gersho, R.M. Gray.: '*Vector Quantization and Signal Compression*', Kluwer Academic Publishers, Boston, MA, 1991.
- [13] F. K. Soong, et. al., "A vector quantization approach to speaker recognition", *At & T Technical Journal*, 66, pp. 14-26, 1987.
- [14] A. E. Rosenberg and F. K. Soong, "Evaluation of a vector quantization talker recognition system in text independent and text dependent models", *Computer Speech and Language* 22, pp. 143-157, 1987.
- [15] Jeng-Shyang Pan, Zhe-Ming Lu, and Sheng-He Sun.: 'An Efficient Encoding Algorithm for Vector Quantization Based on Subvector Technique', *IEEE Transactions on image processing*, vol 12 No. 3 March 2003.
- [16] F. Soong, E. Rosenberg, B. Juang, and L. Rabiner, "A Vector Quantization Approach to Speaker Recognition", *AT&T Technical Journal*, vol. 66, March/April 1987, pp. 1426.
- [17] Md. Rashidul Hasan, Mustafa Jamil, *Md. Golam Rabbani Md. Saifur Rahman* , "Speaker Identification using Mel Frequency Cepstral Coefficients", 3rd International Conference on Electrical & Computer Engineering ICECE held at Dhaka, Bangladesh , 28-30 December 2004.
- [18] Poonam Bansal, Amrita Dev, Shail Bala Jain, "Automatic Speaker Identification using Vector Quantization", *Asian Journal of Information Technology* 6 (9): 938-942, 2007.
- [19] Jyoti Singhai, "Automatic Speaker Recognition :An Approach using DWT based Feature Extraction and Vector Quantization", *IETE Technical Review*, vol. 24, No 5, pp 395-402, September-October 2007
- [20] H. B. Kekre, Tanuja K. Sarode, "Speech Data Compression using Vector Quantization", *WASET International Journal of Computer and Information Science and Engineering (IJCISE)*, Fall 2008, Volume 2, Number 4, pp.: 251-254, 2008. <http://www.waset.org/ijcise>.
- [21] H. B. Kekre, Tanuja K. Sarode, "New Fast Improved Codebook Generation Algorithm for Color Images using Vector Quantization," *International Journal of Engineering and Technology*, vol.1, No.1, pp. 67-77, September 2008.
- [22] H. B. Kekre, Tanuja K. Sarode, "Fast Codebook Generation Algorithm for Color Images using Vector Quantization," *International Journal of Computer Science and Information Technology*, Vol. 1, No. 1, pp: 7-12, Jan 2009.
- [23] H. B. Kekre, Tanuja K. Sarode, "An Efficient Fast Algorithm to Generate Codebook for Vector Quantization," *First International Conference on Emerging Trends in Engineering and Technology, ICETET-2008*, held at Raisoni College of Engineering, Nagpur, India, 16-18 July 2008, Available at online IEEE Xplore.
- [24] H B Kekre, Vaishali Kulkarni, "Speaker Identification by using Vector Quantization", *International Journal of Engineering Science and Technology*, May 2010 edition.
- [25] H B Kekre, Vaishali Kulkarni, "Performance Comparison of Speaker Recognition using Vector Quantization by LBG and KFCG", *International Journal of Computer Applications*, vol. 3, July 2010.
- [26] H.B. Kekre, Archana Athawale, Tanuja K. Sarode, Kalpana Sagvekar, "Comparative Performance of Information Hiding in Vector Quantized Codebooks using LBG, KPE, KMCG and KFCG", *International Journal of Computer Science and Information Security*, 2010 Vol: 8 Issue: 2, pp 89-95.
- [27] H B Kekre, Archana Athawale, Tanuja Sarode and Kalpana Sagvekar, "Increased Capacity of Information Hiding using Mixed Codebooks of Vector Quantization Algorithms: LBG, KPE and KMCG, *International Journal of Advances in Computational Sciences and Technology*, Volume 3 Number 2 (2010) pp. 245–256.

#### **Author Biographies**

**Dr. H. B. Kekre** has received B.E. (Hons.) in Telecomm. Engg. from Jabalpur University in 1958, M.Tech (Industrial Electronics) from IIT Bombay in 1960, M.S.Engg. (Electrical Engg.) from University of Ottawa in 1965 and Ph.D. (System Identification) from IIT Bombay in 1970. He has worked Over 35 years as Faculty of Electrical Engineering and then HOD Computer Science and Engg. at IIT Bombay. For last 13 years worked as a Professor in Department of Computer Engg. at Thadomal Shahani Engineering College, Mumbai. He is currently Senior Professor working with Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS University, Vile Parle(w), Mumbai, INDIA. He ha guided 17 Ph.D.s, 150 M.E./M.Tech Projects and several B.E./B.Tech Projects. His areas of interest are Digital Signal processing, Image Processing and Computer Networks. He has more than 300 papers in National / International Conferences / Journals to his credit. Recently ten students working under his guidance have received best paper awards. Two of his students have been awarded Ph. D. of NMIMS University. Currently he is guiding ten Ph.D. students. He is member of ISTE and IETE.

**Vaishali Kulkarni** has received B.E in Electronics Engg. from Mumbai University in 1997, M.Tech (Electronics and Telecom) from Mumbai University in 2006. Presently she is pursuing Ph. D from NMIMS University. She has a teaching experience of more than 8 years. She is Associate Professor in telecom Department in MPSTME, NMIMS University. Her areas of interest include Speech processing: Speech and Speaker Recognition. She has 7 papers in National / International Conferences / Journals to her credit.