

# Speaker Identification using Row Mean of Haar and Kekre's Transform on Spectrograms of Different Frame Sizes

Dr. H B Kekre

Senior Professor, Computer Dept.,  
MPSTME, NMIMS University,  
Mumbai, India.

Vaishali Kulkarni

Associate Professor, Electronics and Telecommunication,  
MPSTME, NMIMS University,  
Mumbai, India.

**Abstract**—In this paper, we propose Speaker Identification using two transforms, namely Haar Transform and Kekre's Transform. The speech signal spoken by a particular speaker is converted into a spectrogram by using 25% and 50% overlap between consecutive sample vectors. The two transforms are applied on the spectrogram. The row mean of the transformed matrix forms the feature vector, which is used in the training as well as matching phases. The results of both the transform techniques have been compared. Haar transform gives fairly good results with a maximum accuracy of 69% for both 25% as well as 50% overlap. Kekre's Transform shows much better performance, with a maximum accuracy of 85.7% for 25% overlap and 88.5% accuracy for 50% overlap.

**Keywords**—Speaker Identification; Spectrogram; Haar Transform; Kekre's Transform; Row Mean; Euclidean distance

## I. INTRODUCTION

Humans recognize the voice of someone familiar, and are able to match the speaker's name to his/her voice. This process is called Speaker Identification. Speaker Identification falls under the broad category of Speaker Recognition [1] – [3], which covers Identification as well as Verification. 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 Identification task can be further classified into text-dependent or text-independent task [4] – [6]. 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]. With a large number of applications like voice dialing, phone banking, teleshopping, database access services, information services, voice mail, security systems and remote access to computers etc., the automated systems need to perform as well or even better, than humans [7] – [10]. Work on Speaker Identification started as early as 1960. Since then many techniques such as filter banks [11], formant analysis [12], auto-correlation [13], instantaneous spectra covariance matrix [14], spectrum and fundamental frequency histograms [15], linear prediction

coefficients [16] and long term averaged spectra [17] for feature extraction have been implemented. Some of recent works on speaker identification depend on classical features including *cepstrum* with many variants [4], sub-band processing technique [18 - 21], Gaussian mixture models (GMM) [22], linear prediction coding [23, 24], wavelet transform [25 - 27] and neural networks [26 - 28]. A lot of work in this regard has been done. But still there is lack of understanding of the characteristics of the speech signal that can uniquely identify a speaker.

We have proposed speaker identification using power distribution in the frequency domain [29], [30]. We have also proposed speaker recognition using vector quantization in time domain by using LBG (Linde Buzo Gray), KFCG (Kekre's Fast Codebook Generation) and KMCG (Kekre's Median Codebook Generation) algorithms [31 – 33] and in transform domain using DFT (Discrete Fourier Transform), DCT (Discrete Cosine Transform) and DST (Discrete Sine Transform) [34].

The concept of row mean of the transform techniques has been used for content based image retrieval (CBIR) [35 – 38]. This technique also has been applied on speaker identification by first converting the speech signal into a spectrogram [39]. We have proposed Speaker Identification using row mean of DFT, DCT, DST and Walsh Transforms on the speech signal [40] – [41].

In this paper we have proposed a different approach by using the spectrograms. Row mean of Haar and Kekre's Transforms are taken on the spectrogram of the speech signal taken on different frame sizes. The generalized block diagram of the speaker identification system is shown in Fig. 1. As shown in Fig. 1, the reference signals in the database are first converted into their spectrograms and then the transforms are applied. The feature vectors are extracted and stored. The test signal to be identified is similarly processed and the feature vector is matched with the feature vectors stored in the database. The feature vector of the speaker in the database which gives the minimum Euclidean distance with the test signal is declared as the speaker identified. Section II describes the process of converting the speech signal into a spectrogram. The Haar and Kekre's transforms have been explained in section III. In Section IV, the feature vector

extraction is explained. Results are discussed in section V and conclusion ion section VI.

## II. SPECTROGRAM GENERATION

The first step in the speaker identification system is to convert the speech signal into a spectrogram. A spectrogram is a time-varying spectral representation [42] (forming an image) that shows how the spectral density of a signal varies with time. Spectrograms have been used for speaker Identification since a very long time [43 - 45]. Spectrograms are usually created in one of two ways: approximated as a filterbank that results from a series of bandpass filters (this was the only way before the advent of modern digital signal processing), or calculated from the time signal using the short-time Fourier transform (STFT). Creating a spectrogram using the STFT is usually a digital process. Digitally sampled data, in the time domain, is broken up into chunks, which usually overlap, and Fourier transformed to calculate the magnitude of the frequency spectrum for each chunk. Each chunk then corresponds to a vertical line in the image; a measurement of magnitude versus frequency for a specific moment in time. The spectrums or time plots are then "laid side by side" to form the image or a three-dimensional surface. This is done using the following steps:

1. The speech signal is first divided into frames, (of sizes 32, 64, 96, 128, 160, 192, 224, 256, 292 or 320) with an overlap of 25% or 50%.
2. These frames are arranged column wise to form a matrix. E.g. if the speech signal is a one dimensional signal of  $44096 \times 1$ . We divide this into frames of 256 samples each with an overlap of 25% between consecutive frames i.e. overlap of 64. These frames are then arranged column wise to form a matrix of dimension  $256 \times 229$ .
3. Discrete Fourier Transform (DFT) is applied to this matrix column wise.
4. The spectrogram is then plotted as the squared magnitude of this transform matrix.

Fig. 2 (a) shows a sample speech signal from the database. Fig.2 (b) shows the spectrogram plotted for the speech signal of fig. 2(a) with a frame size of 256 with a overlap of 25% between adjacent samples. Fig. 2 (c) shows the spectrogram plotted for the same speech signal with an overlap of 50% between adjacent samples. Fig. 3 shows the spectrograms generated for three different speakers for two iterations. Fig. 3a & Fig 3b are the spectrograms of the same speaker 1 for the same sentence for two iterations. Similarly Fig. 3c & Fig. 3d are for speaker 2 and Fig. 3e & Fig. 3f are for speaker 3. As can be seen from the spectrograms there is much similarity between two iterations of the same speaker whereas the spectrograms of different speakers vary. Looking at these features we decided to implement the row mean technique.

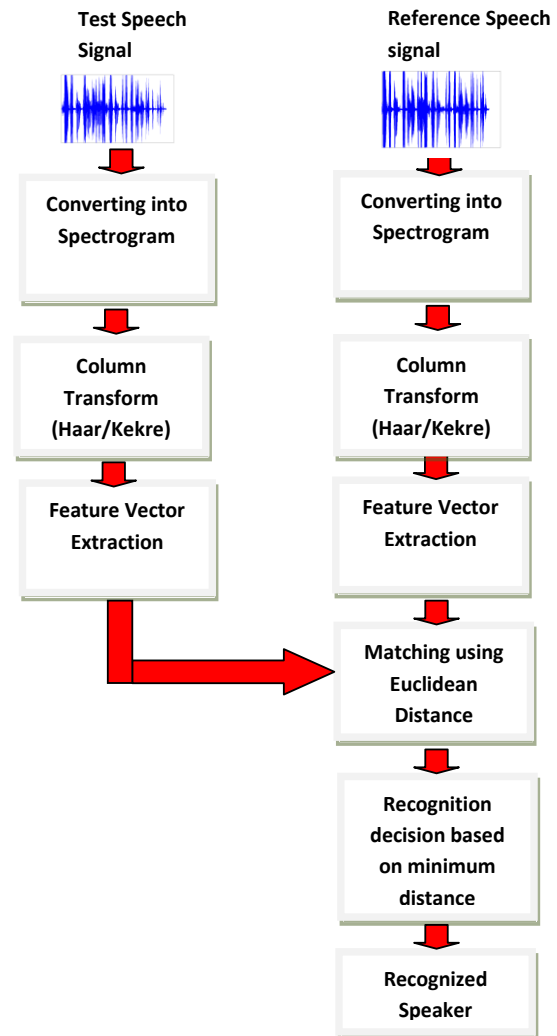
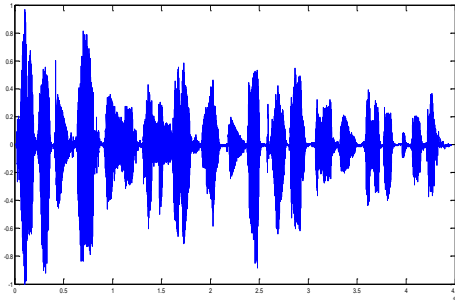


Figure 1. Speaker Identification System

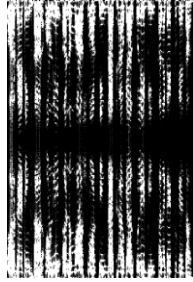
Fig. 2 (a) shows a sample speech signal from the database. Fig.2 (b) shows the spectrogram plotted for the speech signal of fig. 2(a) with a frame size of 256 with a overlap of 25% between adjacent samples. Fig. 2 (c) shows the spectrogram plotted for the same speech signal with an overlap of 50% between adjacent samples.

Fig. 3 shows the spectrograms generated for three different speakers for two iterations. Fig. 3a & Fig 3b are the spectrograms of the same speaker 1 for the same sentence for two iterations.

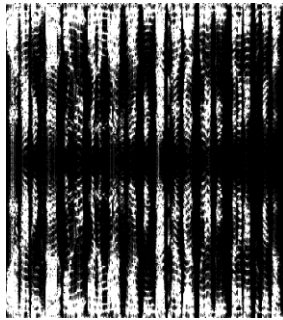
Similarly Fig. 3c & Fig. 3d are for speaker 2 and Fig. 3e & Fig. 3f are for speaker 3. As can be seen from the spectrograms there is much similarity between two iterations of the same speaker whereas the spectrograms of different speakers vary. Looking at these features we decided to implement the row mean technique.



a. Sample Speech signal



b. Spectrogram of frame size 256 with a 25% overlap



c. Spectrogram of frame size 256 with a 50% overlap

Figure 2. Speech Signal and its Spectrogram

### III. TRANSFORM TECHNIQUES

For the present work on Speaker Identification, we have used two transform techniques, namely Haar transform and Kekre's Transform. The two transforms have been explained in this section.

#### A. Haar Transform

This sequence was proposed in 1909 by Alfréd Haar [46]. Haar used these functions to give an example of a countable orthonormal system for the space of square-integrable functions on the real line [47, 48]. The Haar transform is derived from the Haar matrix. The Haar transform is separable and can be expressed in matrix form as:

$$[F] = [H] [f] [H]^T \quad (1)$$

Where  $f$  is an  $N \times N$  image,  $H$  is an  $N \times N$  Haar transform matrix and  $F$  is the resulting  $N \times N$  transformed image. The transformation  $H$  contains the Haar basis function  $h_k(t)$  which are defined over the continuous closed interval  $t \in [0, 1]$ .

The Haar basis functions are

- When  $k=0$ , the Haar function is defined as a constant

$$h_0(t) = 1/\sqrt{N} \quad (2)$$

- When  $k>0$ , the Haar function is defined by

$$h_k(t) = \frac{1}{\sqrt{N}} \begin{cases} 2^{p/2} & (q-1)/2^p \leq t < (q-0.5)/2 \\ -2^{p/2} & (q-0.5)/2^p \leq t < q/2^p \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Where  $0 \leq p < \log_2 N$  and  $1 \leq q \leq 2^p$

The  $N$  Haar functions can be sampled at  $t = (2n+1) \Delta$ , where  $\Delta = T/(2N)$  and  $n = 0, 1, 2, 3, \dots, N-1$  to form an  $N \times N$  matrix for discrete Haar transform. For example, when  $N=4$ , we have

$$H_4 = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{bmatrix} \quad (4)$$

For  $N=8$

$$H_8 = \frac{1}{\sqrt{8}} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} \\ 2 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & -2 \end{bmatrix} \begin{matrix} \varphi_0(t) \\ \psi_0(t) \\ \psi_{1,0}(t) \\ \psi_{1,1}(t) \\ \psi_{2,0}(t) \\ \psi_{2,1}(t) \\ \psi_{2,2}(t) \\ \psi_{2,3}(t) \end{matrix} \quad (5)$$

#### B. Kekre's Transform

Kekre Transform matrix [49, 51] can be of any size  $N \times N$ , which need not have to be in powers of 2 (as is the case with most of other transforms including Haar Transform). All upper diagonal and diagonal values of Kekre's transform matrix are one, while the lower diagonal part except the values just below diagonal are zero. Generalized  $N \times N$  Kekre Transform Matrix can be given as in (6).

The formula for generating the term  $K_{xy}$  of Kekre's transform matrix is given by (7).

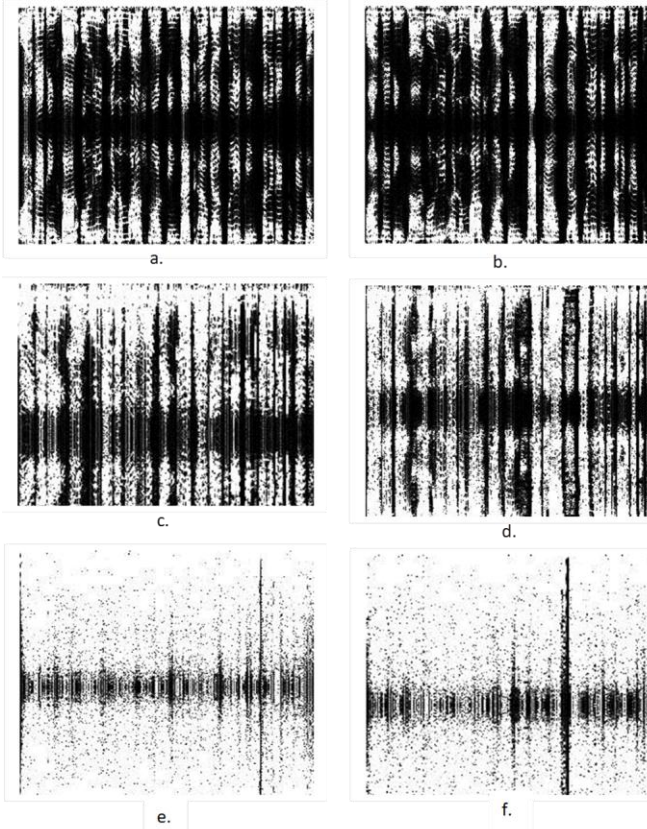


Figure 3. Spectrograms for three different speakers. a & b for speaker 1, c & d for speaker 2 and e & f for speaker 3 for the text, "All great things are only a number of small things that have carefully been collected together".

Fig. 4 shows eight waveforms generated using (3) for  $N=8$ . Writing this in matrix form we get  $8 \times 8$  Haar matrix.

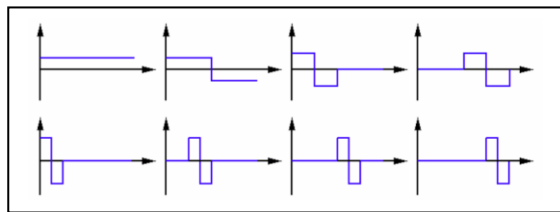


Figure 4. Eight waveforms generated using (3) for  $N=8$

$$K_{N \times N} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ -N+1 & 1 & 1 & \dots & 1 & 1 \\ 0 & -N+2 & 1 & \dots & 1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 1 \\ 0 & 0 & 0 & \dots & -N+(N-1) & 1 \end{bmatrix} \quad (6)$$

$$K_{xy} = \begin{cases} 1 & ; x \leq y \\ -N + (x-1) & ; x = y + 1 \\ 0 & ; x > y + 1 \end{cases} \quad (7)$$

Kekre's Transform has been used for image retrieval [50].

Also Kekre's Wavelet transform ( $N^2 \times N^2$ ) has been developed using ( $N \times N$ ) Kekre's Transform Matrix [51].

#### IV. FEATURE VECTOR EXTRACTION

The procedure for feature vector extraction is given below:

1. Column Transform (Haar or Kekre's Transform) is applied on the spectrogram of the speech signal.
2. The mean of the absolute values of the rows of the transform matrix is then calculated.
3. These row means form a column vector ( $M \times 1$ ) where  $M$  is the number of rows in the transform matrix.
4. This column vector forms the feature vector for the speech sample.
5. The feature vectors for all the speech samples are calculated for different values of  $n$  and stored in the database. Fig. 5 shows the Feature Vector generation technique.

Figure 5. Feature Vector Generation from Spectrogram

#### V. RESULTS

##### A. Database Description

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).

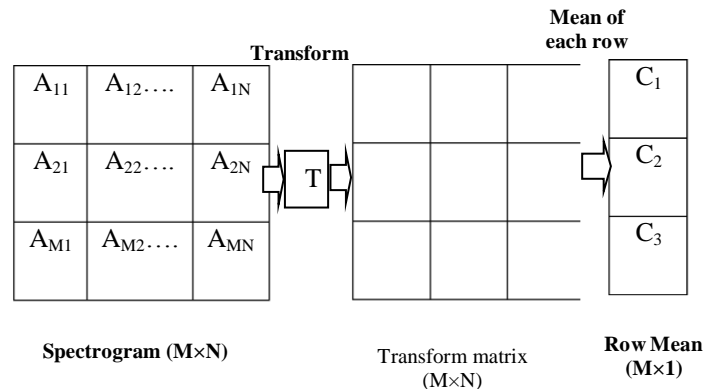


Table I shows the database description. Five iterations of four different sentences of varying lengths are recorded from each of the speakers. Twenty samples per speaker are taken. For text dependent identification, four iterations of a particular sentence are kept in the database and the remaining one iteration is used for testing.

TABLE I. DATABASE DESCRIPTION

Parameter	Sample characteristics
Language	English
No. of Speakers	42
Speech type	Read speech
Recording conditions	Normal. (A silent room)
Sampling frequency	8000 Hz
Resolution	8 bps

### B. Experimental Results

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.

### C. Accuracy of Identification

The accuracy of the identification system is calculated as given by (8).

$$\text{Accuracy (\%)} = \frac{\text{number of matches}}{\text{number of samples tested}} \times 100 \quad (8)$$

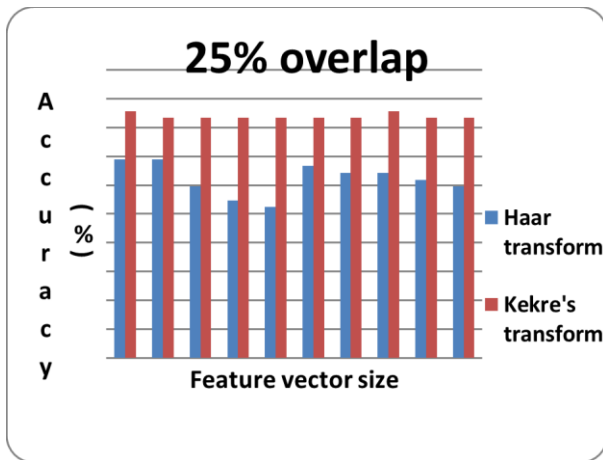


Figure 6. Performance comparison of Haar and Kekre's Transform for a overlap of 25%

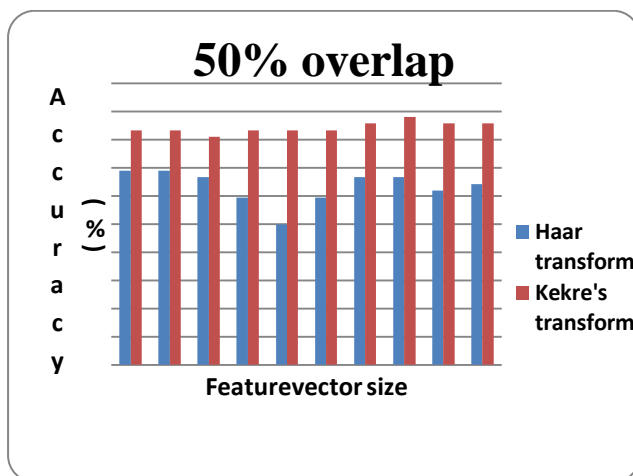


Figure 7. Performance comparison of Haar and Kekre's Transform for a overlap of 50%

Fig. 6 shows the results obtained by using the two transforms for an overlap of 25% between the adjacent frames while creating the spectrograms of the speech signals. As can

be seen from the graphs, the Haar transform gives an average performance of around 60%. The maximum accuracy is obtained for a feature vector size of 32 and 64 (69%) and minimum for a feature vector size of 160 (51%). As the feature vector size is increased further, the accuracy drops and there is no improvement.

For Kekre's transform, the average accuracy is around 83.33%, with a maximum accuracy of around 85% for a feature vector size of 256 and minimum accuracy for feature vector size of 64 to 160 (82%). Fig.7 shows the results obtained by using the two transforms for an overlap of 50% between the adjacent frames while creating the spectrograms of the speech signals. Here also Haar transform gives an average accuracy of around 60%. The behavior of haar transform for both the cases is much similar.

For Kekre's transform, the average accuracy is slightly more i.e. 85.7%. The maximum accuracy is 88.5% for a feature vector size of 256. Overall Kekre's transform gives much better results as compared to Haar transform.

## VI. CONCLUSION

In this paper we have compared the performance of Haar and Kekre's transforms for speaker identification for two different cases (25% and 50% overlap). Haar transform gives an average accuracy of around 60% for both the cases. Accuracy does not increase as the feature vector size is increased from 64 onwards.

Kekre's transform gives an accuracy of more than 80% for both the cases. The maximum accuracy obtained for Kekre's transform 88.5% for a feature vector size of 256. The present study is ongoing and we are analyzing the performance on other transforms.

## 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.Petrovsk-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] S. Furui. Recent advances in speaker recognition. AVBPA97, pp 237--251, 1997.
- [7] 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.
- [8] Tomi Kinnunen, Evgeny Karpov, and Pasi Franti, "Realtime Speaker Identification", ICSLP2004.
- [9] Marco Grimaldi and Fred Cummins, "Speaker Identification using Instantaneous Frequencies", IEEE Transactions on Audio, Speech, and Language Processing, vol., 16, no. 6, August 2008.

- [10] 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.
- [11] S. Pruzansky, "Pattern-matching procedure for automatic talker recognition", J.A.S.A., 35, pp. 354-358, 1963.
- [12] G.R.Doddington, "A method of speaker verification", J.A.S.A., 49,139 (A), 1971.
- [13] P.D. Bricker, et. al., "Statistical techniques for talker identification", B.S.T.J., 50, pp. 1427-1454, 1971.
- [14] K.P.Li, et. al., "Experimental studies in speaker verification using a adaptive system", J.A.S.A., 40, pp. 966-978, 1966.
- [15] B. Beek, et. al., "Automatic speaker recognition system", Rome Air Development Center Report, 1971.
- [16] M.R.Sambur, Speaker recognition and verification using linear prediction analysis, Ph. D. Dissert., M.I.T., 1972.
- [17] S. Furui, et. al., "Talker recognition by long time averaged speech spectrum", Electronics and Communications in Japan, 55-A. pp. 54-61, 1972.
- [18] Besacier, L., J.F. Bonnastre and C. Fredouille, 2000. Localization and Selection of Speaker-Specific Information with Statistical Modeling. Speech Communications. 31: 89-106.
- [19] Besacier, L. and J.F. Bonnastre, 2000. Subband Architecture for Automatic Speaker Recognition. Signal Processing. 80: 1245-1259.
- [20] Dampier, R.I. and J.E. Higgins, 2003. Improving Speaker Identification in Noise by Subband Processing and Decision Fusion. Pattern Recognition Letters. 24: 2167-2173.
- [21] Sivakumaran, P., A.M. Ariyaeinia and M.J. Loomes, 2003. Subband Based Text-dependent Speaker Verification. Speech Communications. 41: 485-509.
- [22] Reynolds, D.A, T.F. Quatieri and R.B. Dunn., 2000. Speaker Verification Using Adapted Gaussian Mixture Models. Digital Signal Processing. pp. 19-4.
- [23] Bassam A. Mustafa, B. Y. Thanoon and S.D. Al- Shamaa., 2005. A Database System for Speaker Identificatoin. Proceedings of The 2<sup>nd</sup> International Conference on Information Technology. Al-Zaytoonah University of Jordan. May 2005.
- [24] Mohd Saleem, A. M., K. Mustafa and I. Ahmad., 2005. Spoken Word of German Digits Uttered by Native and non Native Speakers. Proceedings of The 2<sup>nd</sup> International Conference on Information Technology. Al-Zaytoonah University of Jordan. May 2005.
- [25] Prina Ricotti, L., 2005. Multitapring and Wavelet Variant of MFCC in Speech Recognition. IEE Proceedings on Vis. Image Signal Process., pp: 29- 35.
- [26] Dokur, Z. and T. Olmz., 2003. Classification of Respiratory Sounds By using An Artificial Neural Networks. International Journal of Pattern Recognition and artificial Intelligence. 4: 567-580.
- [27] Abduladheem A., M.A. Alwan, and A.A. Jassim, 2005. Hybrid Wavelet-Network Neural/FFT Nural Phoneme Recognition. Proceedings of The 2<sup>nd</sup> International Conference on Information Technology. Al-Zaytoonah University of Jordan, May 2005.
- [28] Farrell, K.R., R.J. Mammone and K.T. Assalah., 1994. Speaker Recognition Using Neural. Networks and Conventional Classifiers. IEEE Trans. on Speech and Audio Proc. pp: 194-205.
- [29] Dr. H B Kekre, Vaishali Kulkarni, "Speaker Identification using Power Distribution in Frequency Spectrum", Technopath, Journal of Science, Engineering & Technology Management, Vol. 02, No.1, January 2010.
- [30] Dr. H B Kekre, Vaishali Kulkarni, "Speaker Identification by using Power Distribution in Frequency Spectrum", ThinkQuest - 2010 International Conference on Contours of Computing Technology", BGIT, Mumbai, 13th -14th March 2010.
- [31] H B Kekre, Vaishali Kulkarni, "Speaker Identification by using Vector Quantization", International Journal of Engineering Science and Technology, May 2010.
- [32] 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.
- [33] H B Kekre, Vaishali Kulkarni, "Performance Comparison of Automatic Speaker Recognition using Vector Quantization by LBG KFCG and KFCG", International Journal of Computer Science and Security, Vol: 4 Issue: 5, 2010.
- [34] H B Kekre, Vaishali Kulkarni, "Comparative Analysis of Automatic Speaker Recognition using Kekre's Fast Codebook Generation Algorithm in Time Domain and Transform Domain", International Journal of Computer Applications, Volume 7 No.1. September 2010.
- [35] Dr. H.B.Kekre, Sudeep D. Thepade, Akshay Maloo "Performance Comparison of Image Retrieval using Row Mean of Transformed Column Image", International Journal on Computer Science and Engineering Vol. 02, No. 05, 2010, 1908-1912
- [36] Dr.H.B.Kekre,Sudeep Thepade "Edge Texture Based CBIR using Row Mean of Transformed Column Gradient Image", International Journal of Computer Applications (0975 – 8887) Volume 7– No.10, October 2010
- [37] Dr. H.B.Kekre, Sudeep D. Thepade, Akshay Maloo "Eigenvectors of Covariance Matrix using Row Mean and Column Mean Sequences for Face Recognition", International Journal of Biometrics and Bioinformatics (IJBB), Volume (4): Issue (2)
- [38] Dr. H.B.Kekre, Sudeep Thepade, Archana Athawale, "Grayscale Image Retrieval using DCT on Row mean, Column mean and Combination", Journal of Sci., Engg. & Tech. Mgt. Vol 2 (1), January 2010
- [39] Dr. H. B. Kekre, Dr. T. K. Sarode, Shachi J. Natu, Prachi J. Natu "Performance Comparison of Speaker Identification Using DCT, Walsh, Haar on Full and Row Mean of Spectrogram", International Journal of Computer Applications (0975 – 8887) Volume 5– No.6, August 2010
- [40] Dr. H B Kekre, Vaishali Kulkarni "Comparative Analysis of Speaker Identification using row mean of DFT, DCT, DST and Walsh Transforms", International Journal of Computer Science and Information Security, Vol. 9, No.1, January 2011.
- [41] Dr. H B Kekre, Vaishali Kulkarni, Sunil Venkatraman, Anshu Priya, Sujatha Narashiman, "Speaker Identification using Row Mean of DCT and Walsh Hadamard Transform", International Journal on Computer Science and Engineering, Vol. 3, No.1, March 2011.
- [42] S. Haykin, editor, Advances in Spectrum Analysis and Array Processing, vol.1, Prentice-Hall, 1991.
- [43] Tridibesh Dutta, "Text dependent speaker identification based on spectrograms", Proceedings of Image and vision computing, pp. 238-243, New Zealand 2007
- [44] Azzam Sleit, Sami Serhan, and Loai Nemir, "A histogram based speaker identification technique", International Conference on ICADIWT, pp. 384-388, May 2008.
- [45] Tridibesh Dutta and Gopal K. Basak, "Text dependent speaker identification using similar patterns in spectrograms", PRIP'2007 Proceedings, Volume 1, pp. 87-92, Minsk, 2007.
- [46] Haar, Alfred, "Zur Theorie der orthogonalen Funktionensysteme". (German), Mathematische Annalen, volume 69, No. 3, 1910, pp. 331-371.
- [47] Minh N. Do, Martin Vetterli, "Wavelet-Based Texture Retrieval Using Generalized Gaussian Density and Kullback-Leibler Distance", IEEE Transactions On Image Processing, Volume 11, Number 2, pp.146-158, February 2002.
- [48] Charles K. Chui, "An Introduction to Wavelets", Academic Press, 1992, San Diego, ISBN 0585470901.
- [49] Dr. Kekre H. B. and Thepade Sudeep, "Image Retrieval using Non-Involuntional Orthogonal Kekre's Transform", International Journal of MultiDisciplinary Research And Advnces in Engineering,IJMRAE, Vol.1, No.1,November 2009,pp189-203.
- [50] Dr. H.B.Kekre, Sudeep D. Thepade, Archana Athawale, Anant Shah, Prathamesh Verlekar and Suraj Shirke, "Kekre Transform over Row Mean, Column Mean and Both Using Image Tiling for Image Retrieval", International Journal of Computer and Electrical Engineering, Vol.2, No.6, December, 2010 1793-8163.
- [51] Dr. H.B.Kekre, Archana Athawale, Deepali Sadavarti, "Algorithm to Generate Kekre's Wavelet Transform from Kekre's Transform", International Journal of Engineering Science and Technology Vol. 2(5), 2010, 756-767.

#### AUTHORS PROFILE



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 has 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 350 papers in National / International Conferences / Journals to his credit. Recently fifteen students working under his guidance have received best paper awards. Recently five research scholars have received Ph. D. degree from NMIMS University Currently he is guiding seven 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.E (Electronics and Telecom) from Mumbai University in 2006. Presently she is pursuing Ph. D from NMIMS University. She has a teaching experience of around 10 years. She is Associate Professor in telecom Department in MPSTME, NMIMS University. Her areas of interest include networking, Signal processing, Speech processing: Speech and Speaker Recognition. She has 15 papers in National / International Conferences / Journals to her credit.