Closed set and open set Speaker Identification using amplitude distribution of different Transforms

Conference Paper · January 2013

DOI: 10.1109/ICAdTE.2013.6524764

2 authors:

Hemant B Kekre
Narsee Monjee Institute of Management Studies

Vaishali Kulkarni
Narsee Monjee Institute of Management Studies

Closed Set and Open Set Speaker Identification using Amplitude Distribution of Different Transforms

H B Kekre, Vaishali Kulkarni

Abstract—**In this paper, closed set and open set Speaker Identification has been performed on two different databases. Feature extraction for the Identification has been done by using the amplitude distribution of four different Transforms i.e. DFT, DHT, DCT and DST. Two similarity measures i.e. Euclidean Distance (ED) and Manhattan Distance (MD) have been used for matching. The performance has been compared with respect to the best value of each Transform for following parameters: length of speech sample, similarity measure score, size of feature vector, FAR/FRR performance and data acquisition system. Amongst the transforms the best result is given by DFT at 99.06% for feature vector of size 32. Amongst similarity measures Manhattan distance outnumbers the Euclidean distance by 54 to 15, considering the results of all three lengths of speech. The best GAR is 90.65% with a threshold of 94.11% for DFT with MD as a similarity measure.**

Index Terms—**Euclidean Distance (ED), False Acceptance Rate (FAR), False Rejection Rate (FRR), Genuine Acceptance Rate (GAR), Manhattan Distance (MD),**

I. Introduction

SPEAKER identification uses voice as an unique characteristic to identify a person [1-3]. Speaker Identification task can also be classified into closed set and open set Speaker Identification [4, 6]. In the closed set problem, from N known speakers, the speaker whose reference template has the maximum degree of similarity with the template of input speech sample of unknown speaker is obtained. This unknown speaker is assumed to be one of the given set of speakers. Thus in closed set problem, system makes a forced decision by choosing the best matching speaker from the speaker database. In the open set speaker identification, matching reference template for an unknown speaker's speech sample may not exist. 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]. The general scheme for closed set Speaker Identification [7, 8] is shown in Fig. 1. Test and reference patterns (feature vectors) are extracted from speech utterances statically or dynamically. At the training stage, reference models are generated (or trained) from the reference patterns by various methods. A reference model (or template) is formed by obtaining the statistic parameters from the reference speech data. A test pattern is compared against the reference templates at the pattern matching stage. The comparison may be conducted by similarity measure using either distance or statistical parameters. After comparison, the test pattern is labeled to a speaker model at the decision stage. The labeling decision is generally based on the minimum risk criterion. The Open set Speaker Identification process is shown in Fig.1.3. In this process also, the feature vectors extracted from the test speaker are compared against the reference templates at the pattern matching stage. Matching is done based on the similarity measure. The decision to accept or reject a speaker depends on the threshold level.

Discrete Fourier Transform (DFT) is the most commonly used Transform for speaker recognition. It has been used to generate spectrograms [9 - 13], extract Mel Frequency Cepstral Coefficients [14 – 16] etc. Discrete Cosine Transform (DCT) has been used to generate MFCC coefficients. Apart from this, the DHT has been used in speaker recognition [17]. In this paper, four different transform techniques i.e. Discrete Fourier Transform (DFT), Discrete Hartley Transform (DHT), Discrete Cosine Transform (DCT) and Discrete Sine Transform (DST) have been used for feature extraction. Feature extraction technique has been explained in section II. For feature matching two similarity measures i.e. Euclidean distance (ED) and Manhattan Distance (MD) have been used. Feature matching has been explained in Section III for Open set as well as closed set Identification. Section IV describes the databases used for experimentation. Results have been discussed in Section V and conclusion in Section VI.

II. Feature Extraction

A. Amplitude Distribution Concept

The speech signal is a slowly varying quasi-stationary signal. The Transform when applied on a speech signal converts it from time domain to frequency domain. If the absolute magnitude of the DFT is plotted, the spectrum appears as shown in Fig. 3. As can be seen from this spectrum, the energy concentration is in the lower order coefficients. This spectrum is utilized for extracting the features for speaker identification.

B. Feature Extraction Technique

As can be seen from Fig. 3, the second half of the DFT spectrum is a mirror image of the first half. So for feature extraction, only half the number of points in the DFT spectrum is sufficient, the remaining half is redundant. This concept was

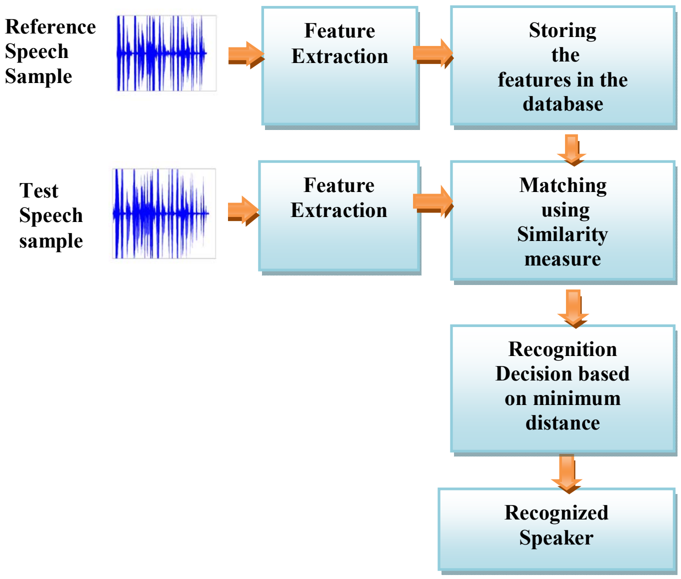


Fig.1 Closed Set Speaker Identification

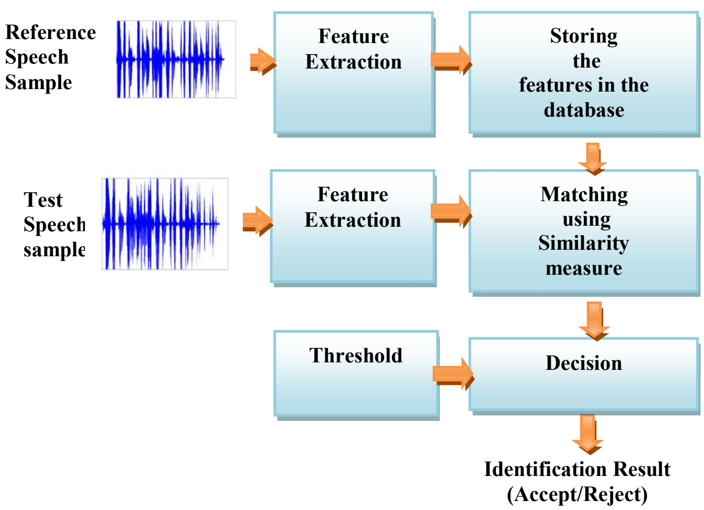


Fig.2 Open Set Speaker Identification

utilized and the magnitude spectrum was divided into groups and the sum of the magnitude for each group formed the element of the feature vector. The feature vectors for the database and the testing phase have been calculated as given by eq. (1), where n is the number of parts into which the magnitude spectrum is divided, m is the number of points of DFT in each of the n parts and Y_i is the coefficient of the magnitude spectrum. The k^{th} component of the feature vector F_k is as given by eq. (1).

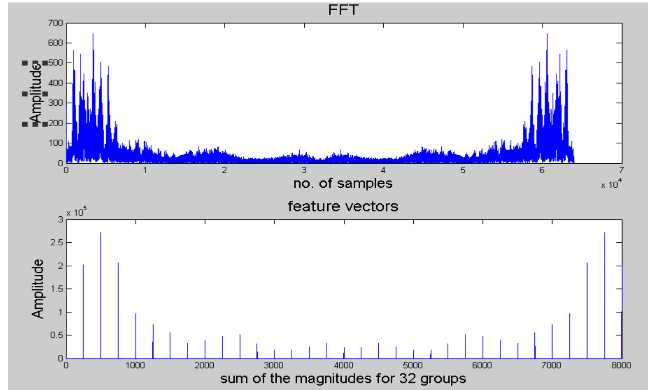


Fig.3 Feature Vectors for DFT by dividing the samples into 32 divisions

$$F_k = \sum_{i=m(k-1)+1}^{km} Y_i \qquad (1)$$

where k=1, 2, …, n.

This concept of amplitude distribution has been applied to other transforms i.e. DHT, DCT and DST also.

III. Feature Matching

A. Distance Measures

For this work two distance measures have been used and the performance of both has been compared. Manhattan distance (MD) is defined as the Minkowiski distance of the order 1 or 1-norm distance (where p=1). In n dimensions, the MD between two points A and B is given by eq. (2), where x_i (or y_i) is the coordinate of A (or B) in dimension i. Euclidean distance is defined as the Minkowiski distance of the order 2 or 2-norm distance (where p=2). In n dimensions, the ED between two points A and B is given by eq. (3), where x_i (or y_i) is the coordinate of A (or B) in dimension i. The results obtained by using these two distance measures for accuracy as well as performance measurement has been discussed.

$$d_{AB} = \sum_{i=1}^{n} | x_i - y_i | \qquad (2)$$

$$d_{AB} = \left(\sum_{i=1}^{n} | x_i - y_i |^2 \right)^{1/2} \qquad (3)$$

B. Decision Making

The final step in speaker recognition process is the decision making. The feature extraction and pattern matching are same for different speaker recognition tasks, but the decision depends on the task: closed set or open set. Let us denote generally a speaker model of speaker i by S_i, and let S = $\{S_1,\ldots\ldots,S_N\}$ be the speaker database of N known speakers. Without assuming a specific speaker model/classifier, let score(X, Si) be the match score between the unknown speaker's feature vectors X = $\{x_1,\ldots\ldots,x_T\}$ and the speaker model S_i. In the case of distance based classifiers, minimum distance corresponds to best match. In closed-set speaker identification task, the decision is simply the speaker index i that yields the minimum distance, where i is given by eq. (4).

$$i = \min_{i} dist(X, S_i) \quad \textbf{(4)}$$

where the minimum is taken over the speaker database S. In the open set identification task, the decision is given as given by eq. (5).

$$dist(X, S_i)\begin{cases} < \Theta_i, accept \\ \geq \Theta_i, reject \end{cases} \quad \textbf{(5)}$$

Where Θi is the threshold. The threshold can be set the same for all speakers, or it can be speaker-dependent. The threshold is determined so that a desired balance between the two types of errors False Acceptance Rate (FAR) and False Rejection Rate (FRR) is achieved [4, 18]. FRR and FAR can be defined as given by eq. (6) and eq. (7).

$$\text{FRR} = (\text{true claims rejected/total true claims}) \times 100 \quad \textbf{(6)}$$

$$\text{FAR} = (\text{imposter claims accepted/total imposter claims}) \times 100 \quad \textbf{(7)}$$

$$\text{GAR} = 100 - \text{FRR} \quad \textbf{(8)}$$

GAR given by eq. (8) is defined as the Genuine Acceptance Rate (GAR), in percentage. Thus FAR is the error with which an imposter is accepted and FRR is the error with which a genuine or true speaker is rejected. There is a trade-off between the two errors. When the decision threshold Θi is increased: FAR increases but FRR decreases, and vice versa. FAR and FRR are plotted against the decision threshold. The point of intersection of these two curves is defined as the Equal Error Rate (EER). The EER is the value for which the FAR and FRR are equal. The system performance can be given by Performance index (PI), which is defined as given by eq. (9).

$$\text{PI } (\%) = 100\text{-EER } (\%) \quad \textbf{(9)}$$

IV. Database Description

The work reported in this paper has been evaluated on two databases.

A. CSLU

CSLU, which is a text dependent Speaker Recognition database obtained from OGI. The CSLU Speaker Recognition Database consists of telephonically recorded speech spanning twelve sessions collected over a two year period. Table 1 shows the database description.

B. Local Database

Local Database, which has been prepared locally. 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 1 shows the database description. The samples have been collected from 107 different speakers. Five sessions of four different sentences of varying lengths have been recorded from each of the speakers. In all there are twenty samples per speaker.

TABLE I
Database Description

Parameter	**CSLU**	**Local Database**
Language	English	English
No. of Speakers analyzed	77	107
Speech type	Read speech, telephonically recorded	Read speech, microphone recorded
Recording conditions	Home/Office	Home/Office
Sampling frequency	8000 Hz	8000 Hz
Resolution	16 bps	8 bps

V. Results

A. Closed Set Identification

The experiments were performed on three different lengths of speech samples. In the first set of experiment, for every sentence or phrase only the first 2.048 sec (16384 point transform @ 8 KHz) has been considered for feature extraction. The second set of experiments has been performed by considering 4.096 sec (32768 point transform @ 8 KHz) of the sentences. The speech signals which were shorter than 4.096 sec in length have been padded with zeros to make all the sentences of equal length. The third set of experiments has been performed by considering 8.192 sec (65536 point transform @ 8 KHz) of the sentences. The speech signals which were shorter than 8.192 sec in length have been padded with zeros to make all the sentences of equal length. The feature vectors have been calculated for the reference speech samples and stored in the database.

TABLE II
Best Results Of DFT Magnitude Spectrum

Sentence	2.046 sec				4.096 sec				8.192 sec			
	ED		MD		ED		MD		ED		MD	
	FV	A	FV	A	FV	A	FV	A	FV	A	FV	A
E1	64	91.58	96	**96.26**	64	**96.26**	56	**96.26**	64	96.26	56	**98.13**
E2	80	89.71	72	**93.45**	72	89.71	80	**93.45**	72	91.58	40	**93.45**
E3	88	86.91	56	**92.52**	80	94.39	72	**97.19**	80	97.19	56	**98.13**
E4	72	93.45	56	**96.26**	48	97.19	32	**99.06**	48	97.19	32	**98.13**
CSLU	72	**81.81**	88	**81.82**	80	**81.81**	88	80.51	80	**81.81**	88	80.51

FV=Feature vector size, A=Accuracy, ED=Euclidean Distance, MD=Manhattan Distance

TABLE III
BEST RESULTS OF DHT MAGNITUDE SPECTRUM

Sentence	2.046 sec				4.096 sec				8.192 sec			
	ED		MD		ED		MD		ED		MD	
	FV	A	FV	A	FV	A	FV	A	FV	A	FV	A
E1	64	91.58	64	**94.39**	64	95.32	56	**96.26**	64	96.26	56	**98.13**
E2	80	86.91	80	**92.52**	72	89.71	80	**93.45**	72	91.58	40	**93.45**
E3	64	87.85	40	**90.65**	80	94.39	72	**96.26**	80	97.19	56	**98.13**
E4	56	93.45	56	**96.26**	48	96.26	48	**98.13**	48	**97.19**	32	**97.19**
CSLU	72	**79.22**	48	77.92	72	**80.51**	88	**80.51**	80	**80.51**	88	**80.51**

FV=Feature vector size, A=Accuracy, ED=Euclidean Distance, MD=Manhattan Distance

TABLE IV
BEST RESULTS OF DCT MAGNITUDE SPECTRUM

Sentence	2.046 sec				4.096 sec				8.192 sec			
	ED		MD		ED		MD		ED		MD	
	FV	A	FV	A	FV	A	FV	A	FV	A	FV	A
E1	32	90.65	48	**92.52**	32	**94.39**	56	93.45	32	94.39	48	**95.32**
E2	32	85.98	40	**88.78**	40	86.91	40	**91.58**	40	87.85	40	**91.58**
E3	32	83.17	32	**84.11**	40	94.39	42	**95.32**	40	94.39	64	**98.13**
E4	40	89.71	32	**92.52**	48	95.32	48	**96.26**	40	**96.26**	40	**96.26**
CSLU	40	75.32	72	**79.22**	40	75.32	96	**76.62**	40	75.32	80	**76.62**

FV=Feature vector size, A=Accuracy, ED=Euclidean Distance, MD=Manhattan Distance

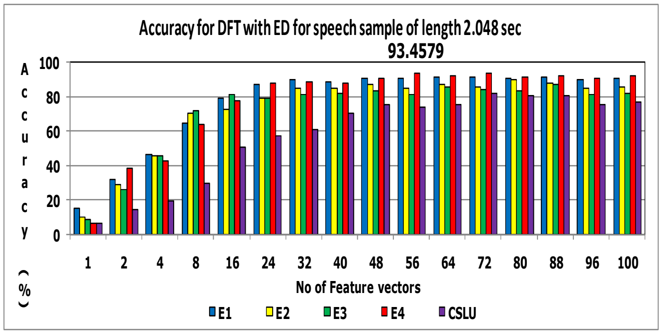
TABLE V
BEST RESULTS OF DST MAGNITUDE SPECTRUM

Sentence	2.046 sec				4.096 sec				8.192 sec			
	ED		MD		ED		MD		ED		MD	
	FV	A	FV	A	FV	A	FV	A	FV	A	FV	A
E1	32	91.58	48	**93.45**	32	**95.52**	32	92.52	32	95.52	48	**96.26**
E2	40	85.04	40	**88.78**	56	88.78	40	**89.71**	40	87.85	40	**89.71**
E3	32	82.24	56	**85.98**	40	93.45	40	**95.32**	32	94.39	64	**98.13**
E4	32	90.65	32	**92.52**	40	**95.32**	32	**95.32**	40	**96.26**	32	**96.26**
CSLU	56	77.92	72	**79.22**	40	**75.32**	64	**75.32**	56	**79.22**	96	77.92

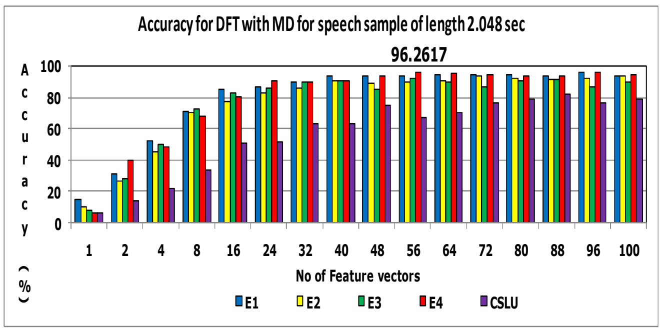
FV=Feature vector size, A=Accuracy, ED=Euclidean Distance, MD=Manhattan Distance

For testing, the test speech sample has been similarly processed and feature vector has been computed. The similarity measure Euclidean distance (ED) or Manhattan Distance (MD) between the database feature vectors and test feature vector has been calculated. The speaker whose reference feature vector gives the minimum distance with the test feature vector has been declared as the speaker recognized. The results have been computed on the four sentences (E1, E2, E3 and E4) of the local database and one phrase of the CSLU database. These results have been obtained by increasing the size of the feature vector for the different sentences. Fig. 4 shows the results obtained for the speech signals of the local database and the CSLU database by considering only the first 2.048 sec. The results have been computed on the four sentences (E1, E2, E3 and E4 where the length of the sentences is in the ascending order as follows: E2<E1<E4<E3.) of the local database and one phrase of the CSLU database. These results have been obtained by increasing the size of the feature vector for the different sentences. The best result for ED is 93.45% which is obtained for E4 for a feature vector of size 56. The best result for MD is

96.26% which is obtained for E4 for a feature vector of size 56. Similar experiments were conducted for speech signal of length 4.096 sec and 8.192 sec also.



(a)



(b)

Fig. 4 Accuracy of DFT for speech signal of length 2.048 sec. a) with ED as similarity measure b) with MD as similarity measure

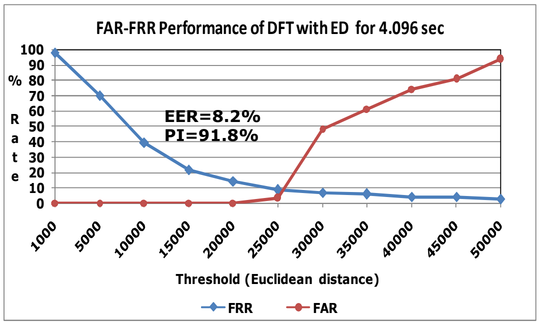
The comparison of the best performance of DFT for all the three sets of experiments for ED and MD is shown in Table II. As can be seen from Table II, the maximum accuracy is 99.06% for a feature vector of size 32 with MD as similarity measure. The same set of experiments was performed on the two databases using DHT, DCT and DST for feature extraction. Table III shows the results obtained for DHT magnitude spectrum. As can be seen from Table III, the maximum accuracy is 98.13% for a feature vector of size 48 with MD as similarity measure. Table IV shows the results obtained for DCT magnitude spectrum. As can be seen from Table IV, the maximum accuracy is 96.26% for a feature vector of size 40 with ED and MD as similarity measure. Table V shows the results obtained for DCT magnitude spectrum. As can be seen from Table V, the maximum accuracy is 96.26% for a feature vector of size 32 with MD as similarity measure.

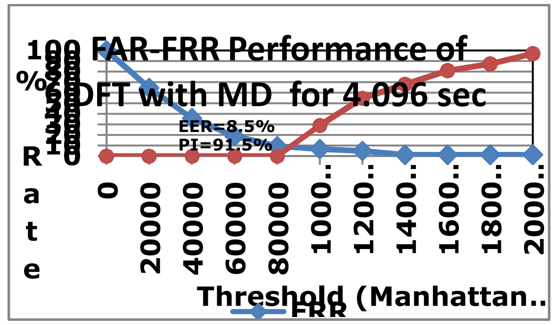
B. Open Set Identification

Open set Identification has been done on one sentence from the local database i.e. E4 for which imposter speech signals were collected. There are 31 imposter speakers. For each transform technique, open set identification has been done for the speech signal length which gives maximum accuracy for the closed set identification. For DFT, FRR and FAR has been calculated for E4 for 4.096 sec for a feature vector of size 48 for ED and 32 for MD by varying the threshold. Fig. 4.5 (a) above shows the % rate for FAR and FRR with ED for varying threshold. The EER is 8.2% and the PI is 91.8%. Fig. 4.5 (b) above shows the % rate for FAR and FRR with MD for varying threshold. The EER is 8.5% and the PI is 91.5%. For DHT, FRR and FAR has been calculated for E4 for 8.192 sec for ED with the feature vector size of 48 and for 4.096 sec for MD with the feature vector size of 48 by varying the threshold. Fig. 6 (a) above shows the % rate for FAR and FRR with ED for varying threshold. The EER is 10.1% and the PI is 89.9%. Fig. 6 (b) above shows the % rate for FAR and FRR with MD for varying threshold. The EER is 11.7% and the PI is 88.3%. For DCT, FRR and FAR has been calculated for E4 for 8.192 sec for a feature vector of size 40 by varying the threshold. Fig. 7 (a) above shows the % rate for FAR and FRR with ED for varying threshold. The EER is 12.3% and the PI is 87.7%. Fig. 7 (b) above shows the % rate for FAR and FRR with MD for varying threshold. The EER is 11.7% and the PI is 88.3%. For DST, FRR and FAR has been calculated for E4 for 8.192 sec with ED and MD for a feature vector of size 40 and 32 respectively, by varying the threshold. Fig. 8 (a) shows the % rate for FAR and FRR with ED as the threshold. The EER is 11.7% and the PI is 88.3%. Fig. 8 (b) shows the % rate for FAR and FRR with MD as the threshold. The EER is 11.9% and the PI is 88.1%.

In the problem of Speaker Identification, the parameter EER does not play any important role. However, the threshold value, which gives the margin of operation for 0% FAR, is important. Hence for comparing performance of different transforms for threshold parameter, the ratio of maximum permissible threshold at 0% FAR and at crossover point (EER) of FAR and FRR has been considered. The conflict for same ratio is resolved by considering GAR at a point where FAR is 0%. Table VI gives the comparative threshold performance of different transforms. The performance has been compared with respect to the best value of each Transform for following parameters: length of speech sample, similarity measure score, size of feature vector, FAR/FRR performance and data acquisition system.

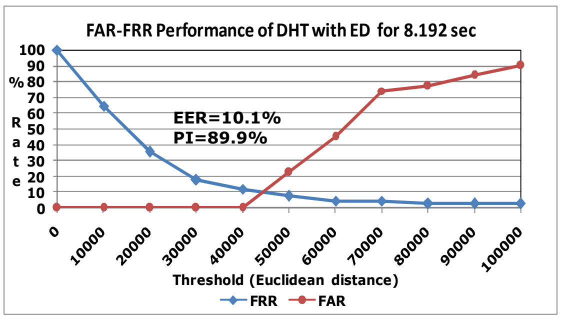
From Table VI, Accuracy depends on the length of speech samples. The best results for each transform are obtained for the longest speech sample of 8.192 sec, except for DFT. Amongst the transforms the best result is given by DFT at 99.06% for feature vector of size 32. Amongst similarity measures Manhattan distance outnumbers the Euclidean distance by 54 to 15, considering the results of all three lengths of speech. Instrumentation has a major role in speaker

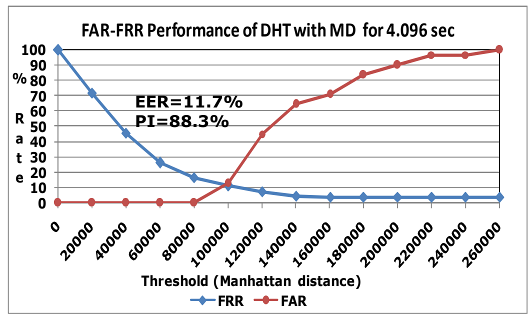




(a) (b)

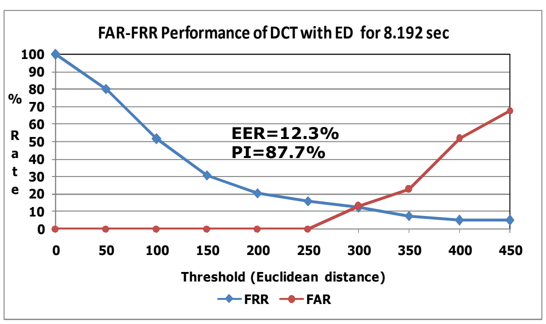
Fig.5 FAR-FRR of DFT for E4 with ED and MD for varying threshold.

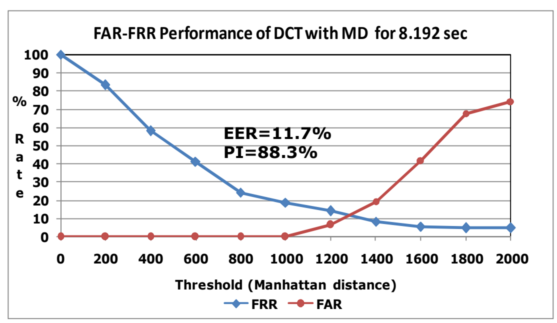




(a) (b)

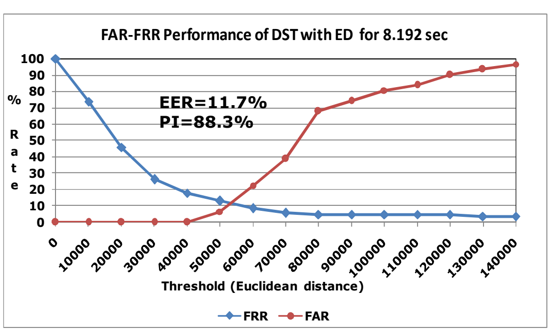
Fig.6 FAR-FRR of DHT for E4 with ED and MD for varying threshold.

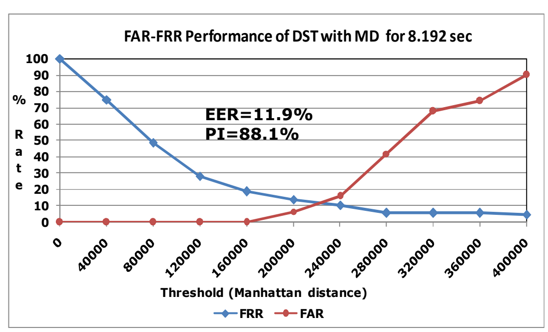




(a) (b)

Fig.7 FAR-FRR of DCT for E4 for 8.192 sec for a feature vector of size 40 with ED and MD for varying threshold.





(a) (b)

Fig.8 FAR-FRR of DST for 8.192 sec for E4 with ED and MD for varying threshold.

TABLE VI
COMPARATIVE PERFORMANCE OF THE SINUSOIDAL TRANSFORMS

Sr. No.	Parameter	Transform							
		DFT		DHT		DCT		DST	
	Length of the speech sample	**(Max. % Accuracy/Feature Vector size)**							
1.	2.048 sec	**96.26 FV=56**		**96.26 FV=56**		92.52 FV=32		93.48 FV=48	
2.	4.096 sec	**99.06 FV=32**		98.13 FV=48		96.26 FV=48		95.32 FV=32	
3.	8.192 sec	**98.13 FV=32**		98.13 FV=56		98.13 FV=64		98.13 FV=64	
	FAR/FRR performance	**(%)**							
4.	GAR for 0% FAR with ED	85.98		**87.85**		84.11		82.24	
5.	% Threshold with ED	80		**90.90**		83.33		74.07	
	GAR for 0% FAR with MD	**90.65**		83.17		81.3		81.3	
	% Threshold with MD	**94.11**		80		76.92		72.72	
	Similarity measure score for Length of speech sample	**ED**	**MD**	**ED**	**MD**	**ED**	**MD**	**ED**	**MD**
6.	(2.048 sec)	1	**5**	1	**4**	0	**5**	0	**5**
7.	(4.096 sec)	2	**4**	1	**5**	1	**4**	3	**4**
8.	(8.192 sec)	1	**4**	2	**5**	1	**5**	2	**4**
9.	Best results of Telephonically recorded speech	**81.82% FV=88**		80.51% FV=72		79.22% FV=72		79.22% FV=56	

recognition. It is seen that microphone recording gives far better performance than telephonic recording consistently for all the transforms. Their best performances for accuracy are 99.06% for DFT for microphone speech whereas the best for telephonic recording is 81.81%. The best GAR is 90.65% with a threshold of 94.11% for DFT with MD as a similarity measure.

VI. CONCLUSION

In this paper, a comparative analysis of Speaker Identification, using four transform techniques has been done. The transforms have been applied on three different lengths i.e. 2.048 sec, 4.096 sec and 8.192 sec respectively. Two similarity measures namely ED and MD have been using for feature matching. The best performance of every Transform is for a speech signal of length 8.192 sec, except for DFT where the maximum accuracy has been obtained for 4.096 sec. DFT gives maximum accuracy of 99.06%. To conclude DFT seems to give best performance compared to other transforms with respect to all performance parameters, followed by DHT with a close margin.

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, et al., “A tutorial on text-independent speaker verification,” EURASIP J. Appl. Signal Process., no. 1, pp. 430–451, 2004.

[6] D. O’Shaughnessy, “Speech communications - Man and Machine”, New York, IEEE Press, 2nd Ed., pp. 199, pp. 437-458, 2000.

[7] Dr. H B Kekre, Vaishali Kulkarni, “Performance Comparison of Automatic Speaker Recognition using Vector Quantization by LBG KFCG and KMCG”, International Journal of Computer Science and Security, CSC Journals, Vol: 4 Issue: 6, 2010.

[8] Dr. 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.

[9] S. Pruzansky, “Pattern-matching procedure for automatic talker recognition,” J.A.S.A., 35, pp. 354-358, 1963.

[10] W. Endress, et. al., “Voice spectrograms as a function of age, Voice Disguise and Voice Imitation”, J.A.S.A., 49, 6(2), pp. 1842-1848, 1971.

[11] Tridibesh Dutta, “Text dependent speaker identification based on spectrograms”, Proceedings of Image and vision computing, pp. 238-243, New Zealand 2007.

[12] Azzam Sleit, Sami Serhan, and Loai Nemir, “A histogram based speaker identification technique”, International Conference on ICADIWT, pp. 384-388, May 2008.

[13] 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.

[14] Molau, S, Pitz, M, Schluter, R, and Ney, H., Computing Mel-frequency coefficients on Power Spectrum, Proceedings of IEEE ICASSP-2001, 1: 73-76 (2001).

[15] S. Khan, et al., “Speaker recognition using MFCC”, IJCSES (International Journal of Computer Science and Engineering System) 2(1): 2008.

[16] Mohd Rasheedur Hassan, et al., “Speaker identification using MFCC coefficients”, 3rd international conference on electrical and computer engineering (ICECE), (2004).

[17] Gedo C. N., “Automatic speaker recognition system using the discrete Hartley transform and an artificial neural network”, Twenty-fifth Asilomar IEEE Conference on Signals, Systems and Computers, 1991

[18] Micki Krause and Harold F. Tipton, “Handbook of Information Security Management”, Auerbach Publications, CRC Press, ISBN: 0849399475.

(1) AUTHORS PROFILE

Dr. H. B. Kekre has received B.E. (Hons.) in Telecomm. Engineering, 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 450 papers in National / International Conferences / Journals to his credit. Recently twelve students working under his guidance have received best paper awards. Recently seven research scholars have received Ph. D. degree from NMIMS University Currently he is guiding six Ph.D. students. He is member of ISTE and IETE.

Vaishali Kulkarni. Author 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 12 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 21 papers in National / International Conferences / Journals to her credit.