Noise Robust Speaker Identification using PCA based Genetic Algorithm

Md. Rabiul Islam
Assistant Professor
Department of Computer Science & Engineering
Rajshahi University of Engineering & Technology
Rajshahi-6204, Bangladesh.

Md. Fayzur Rahman
Professor
Department of Electrical & Electronic Engineering
Rajshahi University of Engineering & Technology
Rajshahi-6204, Bangladesh.

ABSTRACT
This paper emphasizes text dependent speaker identification system on Principal Component Analysis based Genetic Algorithm which deals with detecting a particular speaker from a known population under noisy environment. At first, the system prompts the user to get speech utterance. Noises are eliminated from the speech utterances by using wiener filtering technique. To extract the features from the speech, various types of feature extraction techniques such as RCC, LPCC, MFCC, ∆MFCC and ∆∆MFCC have been used. Principal Component Analysis has been used to reduce the dimensionality of the speech feature vector. To classify the speech utterances, Genetic Algorithm has been used. NOIZEOUS speech database has been used to measure the performance of this system under the condition of various SNRs. Experimental results show the superiority of the proposed close-set text dependent speaker identification system which can be used for security and access control purposes.

General Terms
Pattern Recognition, Soft Computing, Human Computer Interaction.

Keywords
Biometric Technology, Noise Robust Speaker Identification, Speech Feature Extraction, Principal Component Analysis, Genetic Algorithm.

1. INTRODUCTION
Biometrics are seen by many researchers as a solution to a lot of user identification and security problems now a days [1]. Speaker identification is one of the most important areas where biometric techniques can be used. There are various techniques to resolve the automatic speaker identification problem [2, 3, 4, 5, 6, 7, 8].

Most published works in the areas of speech recognition and speaker recognition focus on speech under the noiseless environments and few published works focus on speech under noisy conditions [9, 10, 11, 12]. In some research work, different talking styles were used to simulate the speech produced under real stressful talking conditions [13, 14, 15].

In this proposed system, Principal Component Analysis (PCA) based Genetic Algorithm(GA) with cepstral based features such as Real Cepstral Coefficients (RCC), Mel Frequency Cepstral Coefficients (MFCC), Delta Mel Frequency Cepstral Coefficients (∆MFCC), Delta Delta Mel Frequency Cepstral Coefficients (∆∆MFCC), Linear Prediction Coefficients (LPC) and Linear Prediction Cepstral Coefficients (LPCC) has been used to improve the performance of the text dependent speaker identification system under noisy environment. Results are compared according to different feature extraction techniques on the experimental and performance analysis section.
We ask that authors follow some simple guidelines. In essence, we ask you to make your paper look exactly like this document. The easiest way to do this is simply to down-load a template from [2], and replace the content with your own material.

2. PARADIGM OF THE PROPOSED SPEAKE IDENTIFICATION SYSTEM
The basic building blocks of speaker identification system are shown in the figure 1. Noises are eliminated from the speech utterances after acquisition of the speech. Then pre-emphasis filtering and silence part removal algorithm has been applied. Speech signal is segmented into some blocks, windowing technique is applied and features are extracted. Finally Genetic Algorithm has been used to classify the speech utterances.

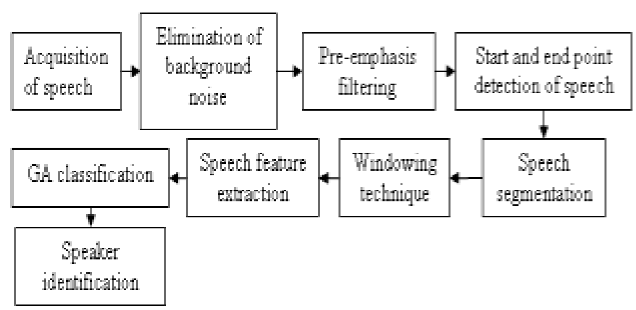

Figure 1: Block Diagram of the proposed automated speaker identification system

3. SPEECH SIGNAL PROCESSING FOR SPEAKER IDENTIFICATION
Sampling frequency of 11025 $H_{Z,}$ sampling resolution of 16-bits, mono recording channel and recorded file format = *.wav have been considered to capture the speech utterances. The speech preprocessing part has a vital role for the efficiency of learning. After acquisition of speech utterances, wiener filter has been used to remove the background noise from the original speech

utterances [16, 17, 18]. Speech end points detection and silence part removal algorithm has been used to detect the presence of speech and to remove pulse and silences in a background noise [19, 20, 21, 22, 23]. To detect word boundary, the frame energy is computed using the sort-term log energy equation [24],

$$E_i = 10\log \sum_{t=n_i}^{n_i+N-1} S^2(t) \qquad (1)$$

Pre-emphasis has been used to balance the spectrum of voiced sounds that have a steep roll-off in the high frequency region [25, 26, 27]. The transfer function of the FIR filter in the z-domain is [26],

$$H(Z) = 1 - \alpha . z^{-1}\ , 0 \le \alpha \le 1 \qquad (2)$$

Where α is the pre-emphasis parameter.

Frame blocking has been performed with an overlapping of 25% to 75% of the frame size. Typically a frame length of 10-30 milliseconds has been used. The purpose of the overlapping analysis is that each speech sound of the input sequence would be approximately centered at some frames [28, 29].

From different types of windowing techniques, Hamming window has been used for this system. The purpose of using windowing is to reduce the effect of the spectral artifacts that results from the framing process [30, 31, 32]. The hamming window can be defined as follows [33]:

$$w(n) = \begin{cases} 0.54 - 0.46\cos\frac{2\Pi n}{N}, & -(\frac{N-1}{2}) \le n \le (\frac{N-1}{2}) \\ 0, & \text{Otherwise} \end{cases} \qquad (3)$$

4. SPEECH FEATURE EXTRACTION AND DIMENSIONALITY REDUCTION OF THE SPEECH FEATURE VECTOR

RCC, LPCC, MFCC, ΔMFCC, ΔΔMFCC based various standard speech feature extraction techniques [34, 35, 36, 37] has been used to enhance the efficiency of the system because the quality of the system depends on the proper feature extracted values. A large dimension of speech features are extracted after applying the feature extraction values. To reduce the dimension of the feature vector, Principal Component Analysis method [38, 39, 40] has been used. After getting PCA values, vector normalization is used to normalize the features that will be further used in the speaker modeling.

5. SPEAKER MODELING

To identify the speaker, an unknown utterance is captured by the system. By applying preprocessing technique, features are extracted from the unknown speech. Then try to match with the existing all entire speaker utterance database. Finally the system identifies that speaker which has maximum similarity with the unknown speaker utterance. In the testing phase, for each unknown speaker to be recognized, the processing shown in figure 2 has been carried out.

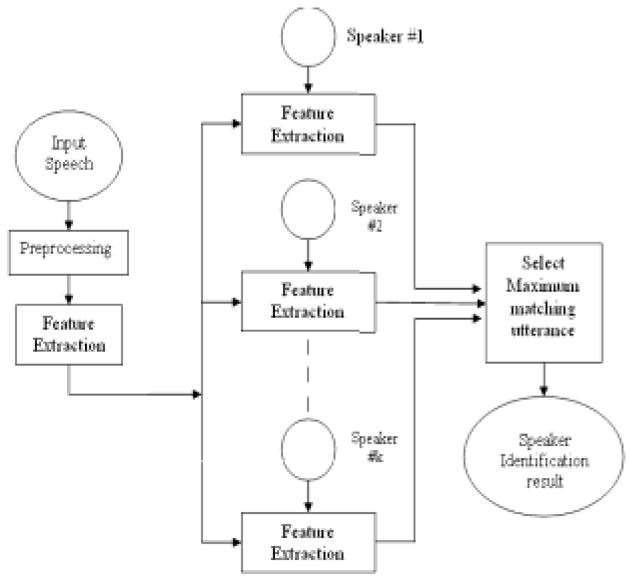

Figure 2: Speaker identification model

6. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

Experimental results and performance analysis has been analyzed in various dimensions. To select the optimum parameters values of the Genetic Algorithm such as crossover rate and number of generations, various experiment have been performed. Figure 3 and figure 4 show the results of the optimum parameters selection for GA. After finding out the optimum parameters, results of the close-set text dependent speaker identification system has been populated according to the NOIZEOUS speech database based on various speech feature extraction techniques which are shown the following sections.

6.1 Optimum Parameter Selection for GA

6.1.1 Experiment on the Crossover Rate of GA

The change of cross over rate for GA can enhance the performance of the system. In this experiment, crossover rate has been changed in various ways which are shown in figure 3. The highest speaker identification rate of (96%) was found at crossover rate 30.

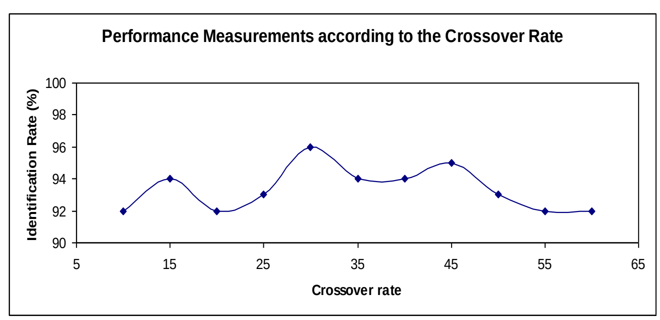

Figure 3: Speaker identification accuracy according to various crossover rates.

6.1.2 Experiment on the Number of Generations of GA

Different values of the number of generations have been tested to achieve the optimum number of generations for GA. Figure 4 shows the results of the accuracy measurement according to various numbers of generations. Finally the maximum identification rate of 98% was found at the number of generations 15.

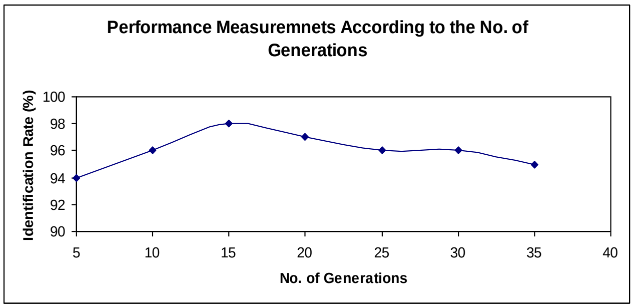

Figure 4: Identification rate according to the no. of generation at 15.

6.2 Performance Measurements of the Proposed System Based on GA

NOIZEOUS speech database [41, 42] has been used to measure the performance of the proposed speaker identification system. To measure the accuracy of the system, eight different types of environmental noises (i.e. Airport, Babble, Car, Exhibition, Restaurant, Street, Train and Train station) of NOIZEOUS database have been considered with four different SNRs such as 0db, 5db, 10db and 15db. The following tables show the experimental results of speaker identification rate at different types of noisy environments with various SNRs.

Table 1. Airport Noise Average Identification Rate (%) for NOIZEOUS Speech Corpus

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	RCC	LPCC
15dB	88.33	90.33	72.00	75.67	84.00
10dB	85.67	86.33	64.67	68.33	82.67
5dB	83.00	84.67	62.67	64.33	80.33
0dB	82.33	82.00	45.00	60.00	77.00
Average	84.83	85.83	61.09	67.08	81.00

Table 2. Babble Noise Average Identification Rate (%) for NOIZEOUS Speech Corpus

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	RCC	LPCC
15dB	90.00	92.33	70.33	75.00	88.00
10dB	87.67	88.00	62.33	72.67	83.33
5dB	83.67	82.67	60.00	72.67	80.00
0dB	77.33	80.67	50.00	57.33	65.67
Average	84.67	85.92	60.67	69.42	79.25

Table 3. Car Noise Average Identification Rate (%) for NOIZEOUS Speech Corpus

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	RCC	LPCC
15dB	90.67	92.67	70.00	72.67	83.00
10dB	86.00	87.33	60.33	62.33	75.33
5dB	79.67	80.67	54.00	62.00	70.33
0dB	76.33	77.33	57.67	58.33	65.00
Average	83.17	84.50	60.50	63.83	73.42

Table 4. Exhibition Hall Noise Average Identification Rate (%) for NOIZEOUS Speech Corpus

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	RCC	LPCC
15dB	89.00	91.00	67.67	78.00	86.67
10dB	87.33	87.67	65.00	76.67	82.33
5dB	78.33	80.00	56.67	67.00	75.00
0dB	82.00	85.33	53.33	61.00	68.33
Average	84.17	86.00	60.67	70.67	78.08

Table 5. Restaurant Noise Average Identification Rate (%) for NOIZEOUS Speech Corpus

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	RCC	LPCC
15dB	90.00	89.67	65.33	72.00	87.67
10dB	85.33	85.33	56.67	66.67	77.00
5dB	83.33	85.33	55.33	60.00	75.33
0dB	80.00	80.00	50.00	56.67	73.00
Average	84.67	85.08	56.83	63.84	78.25

Table 6. Street Noise Average Identification Rate (%) for NOIZEOUS Speech Corpus

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	RCC	LPCC
15dB	88.33	90.00	65.00	75.00	85.00
10dB	86.67	87.67	60.33	65.33	78.67
5dB	83.00	84.00	56.67	64.00	70.00
0dB	80.00	82.00	50.00	60.00	67.67
Average	84.50	85.92	58.00	66.08	75.34

Table 7. Train Noise Average Identification Rate (%) for NOIZEOUS Speech Corpus

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	RCC	LPCC
15dB	88.00	88.33	65.33	73.33	84.00
10dB	86.67	87.67	60.00	68.67	82.33
5dB	86.67	85.00	60.00	63.33	80.00
0dB	80.00	82.33	55.00	60.00	72.00
Average	85.34	85.83	60.08	66.33	79.58

Table 8. Train Station Noise Average Identification Rate (%) for NOIZEOUS Speech Corpus

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	RCC	LPCC
15dB	90.00	92.00	67.67	70.00	78.67
10dB	87.67	86.67	65.00	70.00	75.00
5dB	83.33	85.00	60.00	60.00	72.33
0dB	83.33	83.33	50.00	55.33	70.00
Average	86.08	86.75	60.67	63.83	74.00

Table 9 shows the overall average speaker identification rate for NOIZEOUS speech corpus. By comparing different feature extraction techniques, it was shown that ΔMFCC has higher performance (i.e. 85.73%) than any other methods. Figure 5 shows the performance comparison among different types of speech feature extraction techniques and it is clearly visible that ΔMFCC method dominated over all others though the performance between MFCC and ΔMFCC are nearly equal.

Table 9. Overall Average Speaker Identification Rate (%) for NOIZEOUS Speech Corpus

Method / Various Noises	MFCC	Δ MFCC	ΔΔ MFCC	RCC	LPCC
Airport Noise	84.83	85.83	61.09	67.08	81.00
Babble Noise	84.67	85.92	60.67	69.42	79.25
Car Noise	83.17	84.50	60.50	63.83	73.42
Exhibition Hall Noise	84.17	86.00	60.67	70.67	78.08
Restaurant Noise	84.67	85.08	56.83	63.84	78.25
Street Noise	84.50	85.92	58.00	66.08	75.34
Train Noise	85.34	85.83	60.08	66.33	79.58
Train Station Noise	86.08	86.75	60.67	63.83	74.00
Average Identification Rate (%)	84.68	85.73	59.81	66.39	77.37

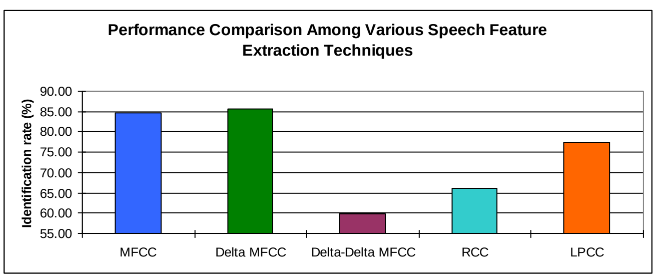

Figure 5: Identification rate according to various feature extraction technique.

7. CONCLUSIONS AND OBSERVATIONS

The parameters of genetic algorithm such as crossover rate and number of generations have a great impact on the identification performance of a GA based close set text dependent ASIS. The highest identification rate was 85.73% which has been achieved at ΔMFCC feature extraction technique. The system has some limitations such as when testing by the NOIZEOUS speech database, vocabulary was limited and the numbers of users were limited. The performance of this system can also be improved by improving the noise removing technique of the speech signal and by introducing the hybrid technique. By enhancing the speech independent speaker identification, increasing the number of user scan and identification of a male, female, child and adult can be the possible further research of this work.

8. REFERENCES

[1] Jain, R. Bole, S. Pankanti, *BIOMETRICS Personal Identification in Networked Society*, Kluwer Academic Press, Boston, 1999.

[2] Rabiner, L., and Juang, B.-H., *Fundamentals of Speech Recognition*, Prentice Hall, Englewood Cliffs, New Jersey, 1993.

[3] Jacobsen, J. D., "Probabilistic Speech Detection", *Informatics and Mathematical Modeling*, DTU, 2003.

[4] Jain, A., R.P.W.Duin, and J.Mao., "Statistical pattern recognition: a review", *IEEE Trans. on Pattern Analysis and Machine Intelligence 22 (2000)*, pp. 4–37, 2002.

[5] Davis, S., and Mermelstein, P., "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences", *IEEE* 74 *Transactions on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 28, no. 4, pp. 357-366, Aug. 1980.

[6] Sadaoki Furui, "50 Years of Progress in Speech and Speaker Recognition Research", *ECTI TRANSACTIONS ON COMPUTER AND INFORMATION TECHNOLOGY*, Vol.1, No.2, November 2005.

[7] Lockwood, P., Boudy, J., and Blanchet, M., "Non-linear spectral subtraction (NSS) and hidden Markov models for robust speech recognition in car noise environments", *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, pp. 265-268, Mar. 1992.

[8] Matsui, T., and Furui, S., "Comparison of text-independent speaker recognition methods using VQ-distortion and discrete/ continuous HMMs", *IEEE Transactions on Speech Audio Process*, no. 2, pp. 456-459, 1994.

[9] Reynolds, D.A., "Experimental evaluation of features for robust speaker identification", *IEEE Transactions on SAP*, Vol. 2, pp. 639-643, 1994.

[10] Sharma, S., Ellis, D., Kajarekar, S., Jain, P. & Hermansky, H., "Feature extraction using non-linear transformation for robust speech recognition on the Aurora database", *Proc. ICASSP2000*, 2000.

[11] Wu, D., Morris, A.C. & Koreman, J., "MLP Internal Representation as Disciminant Features for Improved Speaker Recognition", *Proc. NOLISP2005*, Barcelona, Spain, pp. 25-33, 2005.

[12] Konig, Y., Heck, L., Weintraub, M. & Sonmez, K., "Nonlinear discriminant feature extraction for robust text-independent speaker recognition", *Proc. RLA2C, ESCA workshop on Speaker Recognition and its Commercial and Forensic Applications*, pp. 72-75, 1998.

[13] Ismail Shahin, "Improving Speaker Identification Performance Under the Shouted Talking Condition Using the Second-Order Hidden Markov Models", *EURASIP Journal on Applied Signal Processing 2005:4*, pp. 482–486, ,2005, Hindawi Publishing Corporation.

[14] S. E. Bou-Ghazale and J. H. L. Hansen, "A comparative study of traditional and newly proposed features for recognition of speech under stress", *IEEE Trans. Speech, and Audio Processing*, vol. 8, no. 4, pp. 429–442, 2000.

[15] G. Zhou, J. H. L. Hansen, and J. F. Kaiser, "Nonlinear feature based classification of speech under stress", *IEEE*

Trans. Speech, and Audio Processing, vol. 9, no. 3, pp. 201–216, 2001.

[16] Simon Doclo and Marc Moonen, "On the Output SNR of the Speech-Distortion Weighted Multichannel Wiener Filter", IEEE SIGNAL PROCESSING LETTERS, vol. 12, no. 12, 2005.

[17] Wiener, N., *Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications*, Wiely, Newyork, 1949.

[18] Wiener, N., Paley, R. E. A. C., "Fourier Transforms in the Complex Domains", American Mathematical Society, Providence, RI, 1934.

[19] Koji Kitayama, Masataka Goto, Katunobu Itou and Tetsunori Kobayashi, "Speech Starter: Noise-Robust Endpoint Detection by Using Filled Pauses", *Eurospeech 2003*, Geneva, pp. 1237-1240, 2003.

[20] S. E. Bou-Ghazale and K. Assaleh, "A robust endpoint detection of speech for noisy environments with application to automatic speech recognition", *in Proc. ICASSP2002*, vol. 4, pp. 3808–3811, 2002.

[21] Martin, D. Charlet, and L. Mauuary, "Robust speech / non-speech detection using LDA applied to MFCC", *in Proc. ICASSP2001*, vol. 1, pp. 237–240, 2001.

[22] Richard. O. Duda, Peter E. Hart, David G. Strok, *Pattern Classification*, A Wiley-interscience publication, John Wiley & Sons, Inc, Second Edition, 2001.

[23] Sarma, V., Venugopal, D., "Studies on pattern recognition approach to voiced-unvoiced-silence classification", *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '78.* , Volume: 3, pp. 1-4, Apr 1978.

[24] Qi Li. Jinsong Zheng, Augustine Tsai, Qiru Zhou, "Robust Endpoint Detection and Energy Normalization for Real-Time Speech and Speaker Recognition", *IEEE Transaction on speech and Audion Processing*, Vol.10, No.3, March, 2002.

[25] Harrington, J., and Cassidy, S., *Techniques in Speech Acoustics*. Kluwer Academic Publishers, Dordrecht, 1999.

[26] Makhoul, J., "Linear prediction: a tutorial review", *Proceedings of the IEEE 64*, 4 (1975), pp. 561–580, 1975.

[27] Picone, J., "Signal modeling techniques in speech recognition", *Proceedings of the IEEE 81*, 9 (1993), pp. 1215–1247, 1993.

[28] Clsudio Becchetti and Lucio Prina Ricotti, *Speech Recognition Theory and C++ Implementation*, John Wiley & Sons. Ltd., pp.124-136, 1999.

[29] L.P. Cordella, P. Foggia, C. Sansone, M. Vento., "A Real-Time Text-Independent Speaker Identification System", *Proceedings of 12th International Conference on Image Analysis and Processing*, IEEE Computer Society Press, Mantova, Italy, pp. 632 - 637 , September , 2003.

[30] J. R. Deller, J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*. Macmillan, 1993.

[31] F. Owens., *Signal Processing Of Speech*, Macmillan New electronics. Macmillan, 1993.

[32] F. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform", *Proceedings of the IEEE 66*, vol.1 (1978), pp.51-84, 1978.

[33] J. Proakis and D. Manolakis, *Digital Signal Processing, Principles, Algorithms and Aplications*, Second edition, Macmillan Publishing Company, New York, 1992.

[34] A.V. Oppenheim, and R.W. Schafer, *Digital Signal Processing*, Prentice Hall, Englewood Cliffs, 1975.

[35] Svetoslav Marinov., "Text Dependent and Text Independent Speaker Verification Systems. Technology and Applications", *Overview article*, 2003. http://www.speech.kth.se/~rolf/gslt_papers/SvetoslavMarinov.pdf

[36] Brett Richard Wildermoth. "Text-Independent Speaker Recognition Using Source Based Features", *Master of Philosophy Thesis*, 2001. http://www4.gu.edu.au:8080/adt-root/uploads/approved/adt-QGU20040831.115646/public/01Front.pdf

[37] Tomi Kinnunen. "Spectral Features for Automatic Text-Independent Speaker Recognition.", *Licentiate's Thesis,* 2003. http://www.cs.joensuu.fi/pages/pums/public_results/2004_PhLic_Kinnunen_Tomi.pdf

[38] K. I. Diamantaras and S. Y. Kung, *Principal Component Neural Networks: Theory and Applications*, John Wiley & Sons,Inc., 1996.

[39] M.A. Turk and A.P. Pentland, "Face Recognition Using Eigenfaces", *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 586-591, 1991.

[40] Omar Daoud, Abdel-Rahman Al-Qawasmi and Khaled daqrouq, "Modified PCA Speaker Identification Based System Using Wavelet Transform and Neural Networks", *International Journal of Recent Trends in Engineering,* Vol 2, No. 5, November 2009.

[41] Hu, Y., and Loizou, P., "Subjective comparison of speech enhancement algorithms", *Proceedings of ICASSP-2006*, I, pp. 153-156, Toulouse, France, 2006.

[42] Hu, Y., and Loizou, P., "Evaluation of objective measures for speech enhancement", *Proceedings of INTERSPEECH-2006*, Philadelphia, PA, 2006.

Authors Biographies

Md. Rabiul Islam was born in Rajshahi, Bangladesh, on December 26, 1981. He received his B.Sc. degree in Computer Science & Engineering and M.Sc. degrees in Electrical & Electronic Engineering in 2004, 2008, respectively from the Rajshahi University of Engineering & Technology, Bangladesh. From 2005 to 2008, he was a Lecturer in the Department of Computer Science & Engineering at Rajshahi University of Engineering & Technology. Since 2008, he has been an Assistant Professor in the Computer Science & Engineering Department, University of Rajshahi University of Engineering & Technology, Bangladesh. His research interests include bio-informatics, human-computer interaction, speaker identification and authentication under the neutral and noisy environments.

Md. Fayzur Rahman was born in 1960 in Thakurgaon, Bangladesh. He received the B. Sc. Engineering degree in Electrical & Electronic Engineering from Rajshahi Engineering College, Bangladesh in 1984 and M. Tech degree in Industrial Electronics from S. J. College of Engineering, Mysore, India in 1992. He received the Ph. D. degree in energy and environment electromagnetic from Yeungnam University, South Korea, in 2000. Following his graduation he joined again in his previous job in BIT Rajshahi. He is a Professor in Electrical & Electronic Engineering in Rajshahi University of Engineering & Technology (RUET). He is currently engaged in education in the area of Electronics & Machine Control and Digital signal processing. He is a member of the Institution of Engineer's (IEB), Bangladesh, Korean Institute of Illuminating and Installation Engineers (KIIEE), and Korean Institute of Electrical Engineers (KIEE), Korea.