Text Dependent Speaker Identification System using Discrete HMM in Noise

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
In this paper, an improved strategy for automated text dependent speaker identification system has been proposed in noisy environment. The identification process incorporates the Hidden Markov Model technique with cepstral based features. To remove the background noise from the source utterance, wiener filter has been used. Different speech pre-processing techniques such as start-end point detection algorithm, pre-emphasis filtering, frame blocking and windowing have been used to process the speech utterances. RCC, MFCC, ∆MFCC, ∆∆MFCC, LPC and LPCC have been used to extract the features. After parameterization of the speech, Discrete Hidden Markov Model has been used in the learning and identification purposes. Features are extracted by using different techniques to optimize the performance of the identification. The performance of this identification is almost different in each case. The highest speaker identification rate of 93[%] for noiseless environment and 69.27[%] for noisy environment have been achieved in the close set text dependent speaker identification system.

General Terms
Speaker Identification, Speech Feature Extraction.

Keywords
Noise Robust Speaker Identification, Discrete Hidden Markov Model, Speech Signal Processing, Speech Feature Extraction.

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, Discrete Hidden Markov Model with cepstral based features has been used to improve the performance of the text dependent speaker identification system under noisy environment. To extract the features from the speech, different types of feature extraction technique such as RCC, MFCC, ∆MFCC, ∆∆MFCC, LPC and LPCC have been used to achieve good result. Some of the tasks of this work have been simulated using Matlab based toolbox such as Signal processing Toolbox, Voicebox and HMM Toolbox.

2. PARADIGM OF THE SPEAKER IDENTIFICATION SYSTEM
The basic building blocks of speaker identification system are shown in the figure 1. The first step is the acquisition of speech from speakers. To remove the background noise from the original speech, wiener filter has been used. Then the start and end points of speech were detected. After detecting the start and end points, pre-emphasis filtering technique has been used. The speech signal is segmented into overlapping frames. After segmentation, windowing technique has been applied. Features were extracted from the segmented speech. The extracted features were then fed to the DHMM for learning and classification.

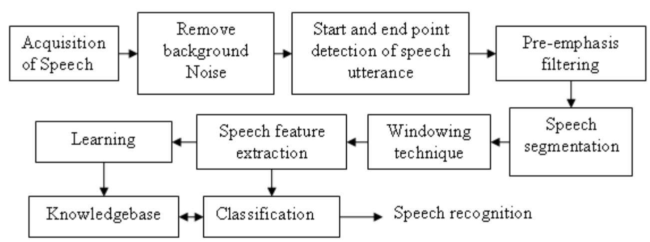

Fig 1: Block diagram of the proposed automated speaker identification system

3. SPEECH SIGNAL PROCESSING FOR SPEAKER IDENTIFICATION
To capture the speech signal, sampling frequency of 11025 H_Z, sampling resolution of 16-bits, mono recording channel and Recorded file format = *.wav have been considered. The speech preprocessing part has a vital role for the efficiency of learning. After acquisition of speech utterances, winner 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 frame [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 PARAMETERIZATION FOR SPEAKER IDENTIFICATION

This stage is very important in an ASIS because the quality of the speaker modeling and pattern matching strongly depends on the quality of the feature extraction methods. For the proposed ASIS, different types of speech feature extraction methods [34, 35, 36, 37, 38, 39] such as RCC, MFCC, ΔMFCC, ΔΔMFCC, LPC, LPCC have been applied.

5. TRAINING MODEL FOR TEXT DEPENDENT SPEAKER IDETIFICATION SYSTEM

Since DHMM can take only positive integer values as input, so it is required to transform the continuous valued features into discrete valued features. It has been performed by using vector quantization method. Vector quantization is a system for mapping a sequence of continuous or discrete vectors into a discrete codebook index.

Finally in training phase, for each speaker *k*, an erogodic DHMM (Discrete HMM), θ_k has been built [40, 41, 42, 43]. The model parameters (A, B, θ) have been estimated to optimize the likelihood of the training set observation vector for the k^{th} speaker by using Baum-Welch algorithm. The Baum-Welch re-estimation formula has been considered as follows [44]:

$$\bar{\Pi}_i = \gamma_1(i) \qquad (4)$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)} \qquad (5)$$

$$\bar{b}_j(\vec{k}) = \frac{\sum_{t=1(s,t,\vec{o}_t=\vec{v}_k)}^{T}\gamma_t(j)}{\sum_{t=1}^{T}\gamma_t(j)} \qquad (6)$$

where, $\xi_t(i,j) = \dfrac{\alpha_t(i)a_{ij}b_j(\bar{o}_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)a_{ij}b_j(\bar{o}_{t+1})\beta_{t+1}(j)}$ and

$$\gamma_t(i) = \sum_{j=1}^{N}\xi_t(i,j)$$

6. TESTING MODEL FOR SPEAKER IDENTIFICATION

In the testing phase, for each unknown speaker to be recognized, the processing shown in figure 2 has been carried out. This procedure includes:

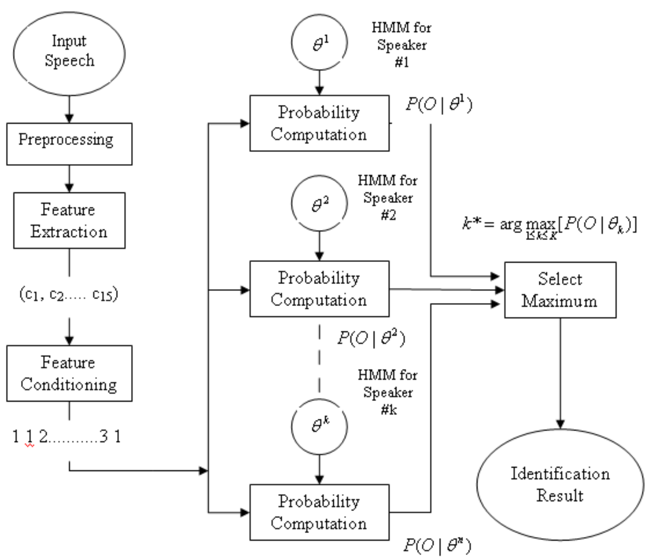

Fig 2: Block diagram of speaker DHMM recognizer

Measurement of the observation sequence $O = \{o_1, o_2, ... o_n\}$, via a feature analysis of the speech corresponding to a speaker.

Transformation the continuous values of *O* into integer values.

Calculation of model likelihoods for all possible models, $P(O\mid\theta_k), 1 \le k \le K$.

Declaration of the speaker as *k**speaker whose model likelihood is highest - that is,

$$k^* = \arg\max_{1 \le k \le K} [P(O \mid \theta_k)] \qquad (7)$$

In this proposed work the probability computation step has been performed using the Baum's Forward-Backward algorithm [44, 45].

7. PARAMETER SELECTION ON DHMM

There are some critical parameters (such as frame length, frame increment, number of cepstral coefficients, number of hidden states, pre-emphasizing parameter etc) that affect the performance of the developed system. A trade off is made to explore the optimal values of the above parameters and experiments are performed using those parameters. The optimal values of the above parameters are chosen and finally find out the results.

7.1 Experiment on Window Shift, N1

In this experiment hamming window has been used. The shifting effect of hamming window has been measured. By setting the window length, N_L = 15 ms, number of Mel-frequency Cepstral Coefficients excluding 0^{th} coefficients, N_{MC} =12, number of hidden states, N_H =5 and the pre-emphasizing parameter, α = 0.9, we have found the highest speaker identification rate of 85[%] at 65% window shift as shown in figure 3.

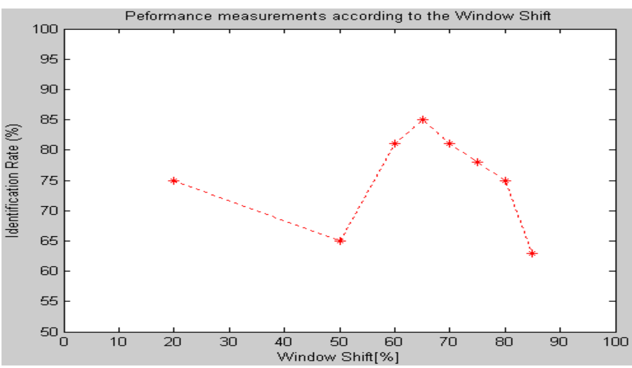

Fig 3: Performance measurement according to the window shift

7.2 Experiment on the Pre-emphasized parameter, α

The performance of the developed speaker identification system has been measured according to the pre-emphasized parameter α. We have set N_L = 15 ms, N_1 = 65%, N_{MC} =12 and N_H =5. We have studied the value of the parameter ranging from 0.7 to 0.99. We have found that the speaker identification performance was 86[%] at α = 0.95 which is shown in figure 4.

7.3 Experiment on the number of hidden states of DHMM, N_H

In the learning phase of DHMM, We have chosen the hidden states in the range from 5 to 20. We have set N_L = 15 ms, N_1 =65%, N_{MC} =12, and α = 0.95. The highest performance of 87[%] have been achieved at N_H =15 which is shown in figure 5.

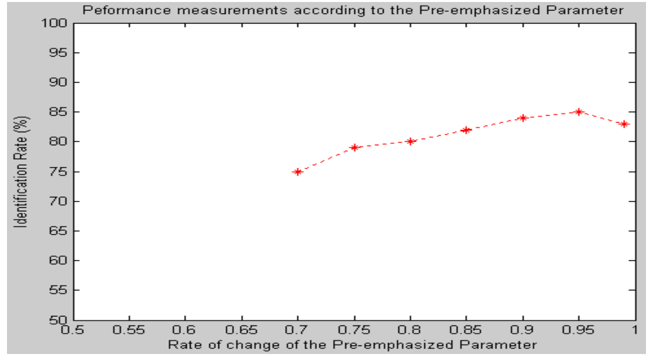

Fig 4: Speaker identification rate on the variation of pre-emphasis parameter

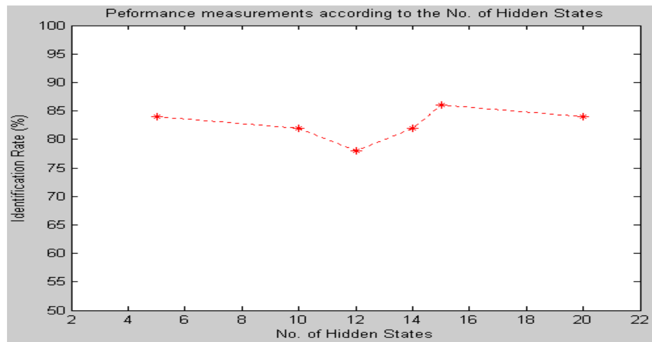

Fig 5: Results after setting up the hidden states of DHMM

7.4 Experiment of the Window Length, N_L

The performance of the identification system has also been investigated by varying the length of the window from 10 ms to 30 ms. By setting N_L = 15 ms, N_1 = 65%, N_{MC} = 12, N_H = 15 and α = 0.95, the highest performance has been achieved with MFCC based system to be 87[%] which is shown in the figure 6.

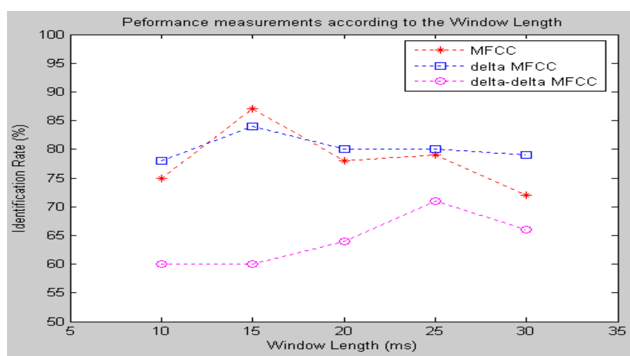

Fig 6: Effect of the window length on the identification rate

7.5 Experiment on the Number of Cepstral Coefficients, N_{MCL}

In this experiment, the numbers of cepstral coefficients were varied from 10 to 20. The highest speaker identification rate of 93[%] has been found at N_L = 15 ms, N_1 = 65%, α = 0.95 and N_{MC} =15 which is shown in figure 7.

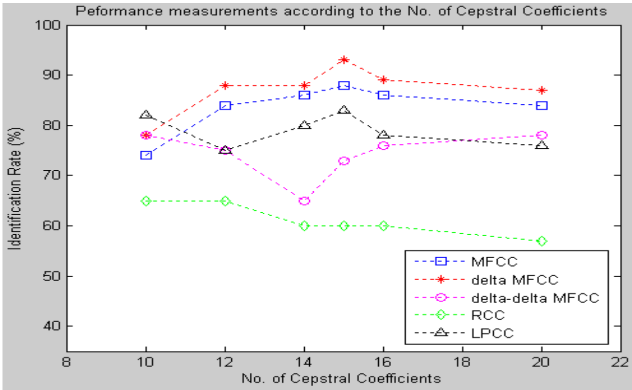

Fig 7 Speaker identification accuracy according to the number of cepstral coefficients

In figure 7, it is found that in HMM the highest speaker identification rate was 93% which was achieved for ΔMFCC per frame.

8. PERFORMANCE MEASUREMENT OF THE TEXT-DEPENDENT SPEAKER IDENTIFICATION SYSTEM

NOIZEOUS speech database [46, 47] 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	85.00	83.33	63.33	63.33	66.67
10dB	82.00	75.00	53.33	56.67	60.00
5dB	63.33	70.00	43.33	53.33	53.33
0dB	53.33	63.33	43.33	50.00	46.67
Average	70.92	72.92	50.83	55.83	56.67

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

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	RCC	LPCC
15dB	70.00	82.00	53.33	53.33	66.67
10dB	66.67	76.67	43.33	46.67	60.00
5dB	53.33	63.33	36.67	46.67	60.00
0dB	53.33	53.33	36.67	43.33	53.33
Average	60.83	68.83	42.50	47.50	60.00

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

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	RCC	LPCC
15dB	66.67	83.33	53.33	63.33	66.67
10dB	63.33	73.67	43.33	53.33	60.00
5dB	53.33	63.33	43.33	53.33	60.00
0dB	53.33	53.33	36.67	43.33	50.00
Average	59.17	68.42	44.17	53.33	59.17

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

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	RCC	LPCC
15dB	80.00	86.67	66.67	70.00	77.67
10dB	73.33	73.33	53.33	66.67	66.67
5dB	66.67	70.00	66.67	66.67	63.33
0dB	63.33	66.67	43.33	53.33	60.00
Average	70.83	74.17	57.50	64.17	66.92

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

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	RCC	LPCC
15dB	75.00	83.33	43.33	73.33	73.33
10dB	70.00	70.00	43.33	66.67	63.33
5dB	63.33	66.67	40.00	53.33	63.33
0dB	50.00	53.33	36.67	53.33	53.33
Average	64.58	68.33	40.83	61.67	63.33

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

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	RCC	LPCC
15dB	73.33	80.00	53.33	66.67	73.33
10dB	66.67	70.00	46.67	53.33	63.33
5dB	63.33	66.67	43.33	66.67	63.33
0dB	53.33	63.33	36.67	53.33	53.33
Average	64.17	70.00	45.00	60.00	63.33

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

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	RCC	LPCC
15dB	80.00	86.67	53.33	63.33	75.00
10dB	70.00	75.00	43.33	60.00	66.67
5dB	56.67	66.67	43.33	53.33	53.33
0dB	56.67	63.33	36.67	56.67	53.33
Average	65.84	72.92	44.17	58.33	62.08

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

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	RCC	LPCC
15dB	76.67	86.67	43.33	60.00	66.67
10dB	66.67	66.67	43.33	56.67	63.33
5dB	53.33	56.67	36.67	46.67	53.33
0dB	50.00	53.33	36.67	43.33	50.00
Average	61.67	65.84	40.00	51.67	58.33

Table 9 shows the overall average speaker identification rate for NOIZEOUS speech corpus. From the table it is easy to compare the performance among MFCC, ΔMFCC, ΔΔMFCC, RCC and LPCC methods for DHMM based text-dependent speaker identification system. It is shown that ΔMFCC has greater performance (i.e. 69.27%) than any other methods such as MFCC, ΔΔMFCC, RCC and LPCC.

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

Method / Various Noises	MFCC	Δ MFCC	ΔΔ MFCC	RCC	LPCC
Airport Noise	70.92	72.92	50.83	55.83	56.67
Babble Noise	60.83	68.83	42.50	47.50	60.00
Car Noise	59.17	68.42	44.17	53.33	59.17
Exhibition Hall Noise	70.83	74.17	57.5	64.17	66.92
Restaurant Noise	64.58	68.33	40.83	61.67	63.33
Street Noise	64.17	70.00	45.00	60.00	63.33
Train Noise	65.84	72.92	44.17	58.33	62.08
Train Station Noise	61.67	65.84	40.00	51.67	58.33
Average Identification Rate (%)	64.07	69.27	42.50	57.92	61.77

9. CONCLUSIONS AND OBSERVATIONS

The critical parameters such as frame length, frame increment, number of cepstral coefficients, number of hidden states and the emphasizing parameter have a great impact on the identification performance of a DHMM based close set text dependent ASIS. To find out the best performance of this system, the optimal values of the above parameters where the highest speaker identification rate was 93[%] have been selected effectively. Therefore the highest identification rate of 93[%] has been achieved at ΔMFCC in noiseless environment. Five experiments have been performed for this purpose. According to the NOIZEOUS speech database, 69.27[%] accuracy has been achieved under eight different types of noisy environment. Since the highest speaker identification rate was 93[%] in clean environment and 69.27[%] in noisy environment, this can satisfy the practical demand. The performance of this system can also be enhanced by the improvement of speech signal processing part and by using the hybrid system. Open set text independent speaker identification system with noisy speech can be the further work of this system.

10. 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, Vol. 22, pp.* 4–37, 2000..

[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, Nov. 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 Discriminant 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,* Vol. 4, pp. 482–486, 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, Dec. 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*, 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* , Vol. 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*, Vol. 64, No. 4, pp. 561–580, 1975.

[27] Picone, J., “Signal modeling techniques in speech recognition”, *Proceedings of the IEEE*, Vol. 81, No. 9, pp. 1215–1247, 1993.

[28] Clsudio Beccchetti 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 , Sep. 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, 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] D. Kewley-Port and Y. Zheng, “Auditory models of formant frequency discrimination for isolated vowels”, *Journal of the Acostical Society of America*, Vol. 103, Issue 3, pp. 1654–1666, 1998.

[35] D. O’Shaughnessy, Speech Communication - Human and Machine, Addison Wesley, 1987.

[36] E. Zwicker., “Subdivision of the audible frequency band into critical bands (frequenzgruppen)”, *Journal of the Acoustical Society of America*, Vol. 33, pp. 248–260, 1961.

[37] S. Davis and P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences”, *IEEE Transactions on Acoustics Speech and Signal Processing*, Vol. 28, pp. 357–366, Aug. 1980.

[38] S. Furui., “Speaker independent isolated word recognition using dynamic features of the speech spectrum”, *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 3, pp. 52–59, Feb. 1986.

[39] S. Furui, “Speaker-Dependent-Feature Extraction, Recognition and Processing Techniques”, *Speech Communication*, Vol. 10, pp. 505-520, 1991.

[40] Huang X.D**.**, Ariki Y**.**, Jack M.A., *Hidden Markov Models for Speech Recognition***,** Edinburgh. university Press, 1990.

[41] M**.** Hwang**, X.** Huang**,** **"**Shared-Distribution Hidden**.** Markov Models for Speech Recognition**",** *IEEE. Trans. on. Speech and Audio Processing*, vol. 1, No. 4, pp. 414-420, Apr. 1993.

[42] L.E. Baum**,** T**.** Petrie**,** G. Soules, and N. Weiss, “A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains", *The Annals of Mathematical Statistics*, Vol. 41, pp. 164-171, 1970.

[43] R.J.Elliott, L. Aggoun, and J.B. Moore, “Hidden Markov Models: Estimation and Control”, *Applications of Mathematics: Stochastic Modeling and Applied Probability,* Vol. 29, Springer, Berlin, 1997.

[44] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition”, *Proceedings of the IEEE*, Vol. 77, No. 2, pp. 257–286, 1989.

[45] P. A. Devijver, **"**Baum's forward-backward algorithm revisited**",** *Pattern Recognition Letter*, Vol. 3, pp. 369-373, 1985.

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

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