Likelihood Ratio Based Score Fusion for Audio-Visual Speaker Identification in Challenging Environment

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
It is well known to enhance the performance of noise robust speaker identification using visual speech information with audio utterances. This paper presents an approach to evaluate the performance of a noise robust audio-visual speaker identification system using likelihood ratio based score fusion in challenging environment. Though the traditional HMM based audio-visual speaker identification system is very sensitive to the speech parameter variation, the proposed likelihood ratio based score fusion method is found to be stance and performs well for improving the robustness and naturalness of human-computer-interaction. In this paper, we investigate the proposed audio-visual speaker identification system in typical office environments conditions. To do this, we investigated two approaches that utilize speech utterance with visual features to improve speaker identification performance in acoustically and visually challenging environment: one seeks to eliminate the noise from the acoustic and visual features by using speech and facial image pre-processing techniques. The other task combines speech and facial features that have been used by the multiple Discrete Hidden Markov Model classifiers with likelihood ratio based score fusion. It is shown that the proposed system can improve a significant amount of performance for audio-visual speaker identification in challenging official environment conditions.

General Terms
Speaker Identification, Human Computer Interaction, Biometrics.

Keywords
Audio-Visual Speaker Identification, Cepstral Base Features, Feature Fusion, Decision Fusion, Likelihood Ratio Based Score Fusion, Discrete Hidden Markov Model.

1. INTRODUCTION
Human speaker identification is bimodal in nature [1, 2]. Visual speech information can plays a vital role for the improvement of natural and robust human-computer interaction [3, 4, 5, 6, 7]. 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 [8, 9, 10, 11]. Indeed, various important human-computer components, such as speaker identification, verification [12, 13], localization [14], speech event detection [15], speech signal separation [16], coding [17], video indexing and retrieval [18], and text-to-speech [19, 20], have been shown to benefit from the visual channel [21].

In this paper, log likelihood ratio based score fusion for audio-visual speaker identification system has been proposed at official environmental conditions. Discrete Hidden Markov Model with cepstral based feature such as RCC, MFCC, ΔMFCC, ΔΔMFCC, LPC and LPCC has been used to improve the performance of this proposed system. VALID audio-visual database has been used to measure the performance which has been shown in the experimental results and performance analysis section in this paper. Section 2 shows the audio-visual system components, section 3 elaborates the audio identification and visual identification process has been focused on section 4.

2. AUDIO-VISUAL SPEAKER IDENTIFICATION COMPONENTS
The block diagram for the proposed log likelihood ratio based audio-visual speaker identification system is shown in figure 1. At first speech utterance and facial image are captured, pre-processing techniques are applied, features are extracted and HMM classification are applied for both audio and visual features. Finally audio and visual reliability are measured from the audio and visual classification output and audio-visual decision fusion are performed to get the final speaker identification result.

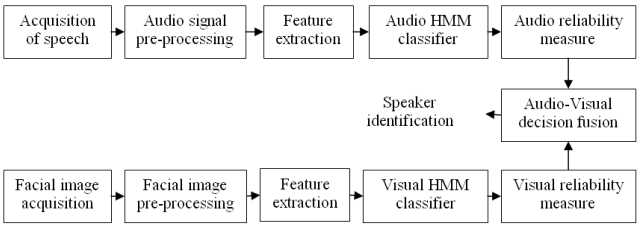

Figure 1: Paradigm of the log likelihood ratio based audio-visual speaker identification.

3. AUDIO 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 [22, 23, 24]. 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 [25, 26, 27, 28, 29]. To detect word boundary, the frame energy is computed using the sort-term log energy equation [24],

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

(1)

Pre-emphasis has been used to balance the spectrum of voiced sounds that have a steep roll-off in the high frequency region [30, 31, 32]. The transfer function of the FIR filter in the z-domain is [33],

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

(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 [34, 35].

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 [36, 37, 38]. The hamming window can be defined as follows [39]:

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

(3)

To extract the features from the speech utterances, various types of standard speech feature extraction techniques [40, 41, 42, 43] such as RCC, MFCC, ΔMFCC, ΔΔMFCC, LPC, LPCC have been applied. Principal Component Analysis method has been used to reduce the dimensionality of the speech feature vector. Finally, HMM learning and classification and algorithms [44, 45, 46] has been applied to classify the speakers.

4. VISUAL IDENTIFICATION

The first step in image pre-processing is image acquisition. To do so, an imaging sensor along with signal digitization capability has been used so that captured image can be converted to digital form directly. After acquisition of face image, Stams [47] Active Appearance Model (ASM) has been used to detect the facial features. Then the binary image has been taken. The Region Of Interest (ROI) has been chosen according to the ROI selection algorithm [48, 49]. Lastly the background noise has been eliminated [50] and finally appearance based facial feature has been found. The procedure of the facial image pre-processing parts is shown in figure 2.

To reduce the dimensionality of the facial feature vector, PCA and HMM training and testing algorithm has been used to classify the facial images.

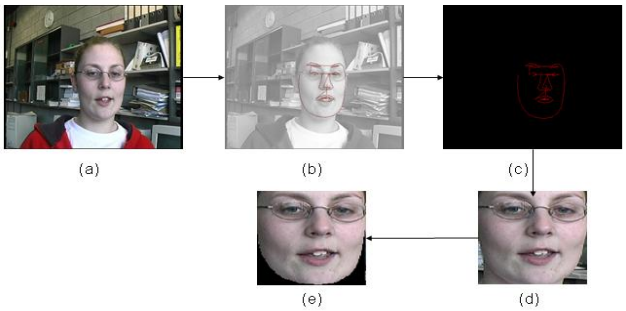

Figure 2: Facial image pre-processing for the proposed system (a) Original image (b) Output taken from Stams Active Appearance Model (c) Facial edges are extracted (d) Region Of Interest (ROI) selection with background noise (e) Appearance based facial features.

5. AUDIO-VISUAL LIKELIHOOD RATIO BASED SCORE FUSION

After the acoustic and visual sub-systems perform identification separately, their outputs are combined by a weighted sum rule to produce the final decision. Sensor level fusion and feature level fusion can be used before matching and after matching match score level, rank level and decision level fusion can be introduced. In this work, match score level used to combine the audio and visual identification outputs. For a given audio-visual speaker test datum of O_A and O_V, the identification utterance C^* is given by [51],

$$C^* = \arg\max_i \{\gamma \log P(O_A / \lambda_A^i) + (1-\gamma)\log P(O_V / \lambda_V^i)\} \quad (4)$$

Where λ_A^i and λ_V^i are the acoustic and the visual HMMs for the i^{th} utterance class respectively and $\log P(O_A / \lambda_A^i)$ and $\log P(O_V / \lambda_V^i)$ are there log likelihood against the i^{th} class.

Among various types of score fusion techniques, baseline reliability ratio-based integration has been used to combine the audio and visual identification results. The reliability of each modality can be measured from the outputs of the corresponding HMMs. When the acoustic speech is not corrupted by any noise, there are large differences between the acoustic HMMs output otherwise the differences become small. The reliability of each modality can be calculated by the most appropriate and best in performance [52],

$$S_m = \frac{1}{N-1}\sum_{i=1}^{N}(\max_j \log P(O/\lambda^j) - \log P(O/\lambda^i)) \quad (5)$$

Which means the average difference between the maximum log-likelihood and the other ones and *N* is the number of classes being considered to measure the reliability of each modality, $m \in \{A,V\}$.

Then the integration weight of audio reliability measure γ_A can be calculated by [53]

$$\gamma_A = \frac{S_A}{S_A + S_V}$$

(6)

Where S_A and S_V are the reliability measure of the outputs of the acoustic and visual HMMs respectively.

The integration weight of visual modality measure can be found as,

$$\gamma_V = (1 - \gamma_A)$$

(7)

6. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

There are some critical parameters such as frame length, frame increment for the speech pre-processing and number of cepstral coefficients, number of hidden states, pre-emphasizing parameter etc for HMM 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 which are shown in the following subsections.

6.1 Optimum Parameter Selection for Speech Pre-preprocessing

6.1.1 Experiment on the Window Shift, N_1

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.

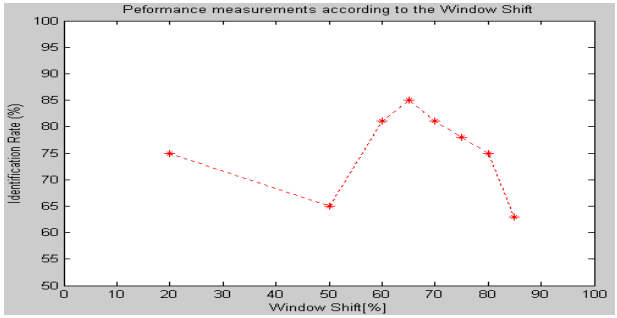

Figure 3: Performance measurement according to the window shift.

6.1.2 Experiment on the Pre-emphasize 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.

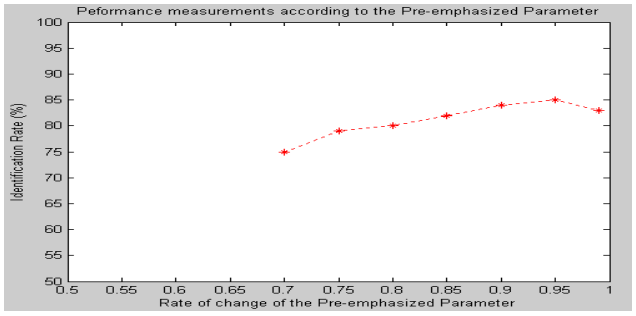

Figure 4: Speaker identification rate on the variation of pre-emphasis parameter.

6.2 Optimum Parameter Selection for HMM

6.2.1 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.

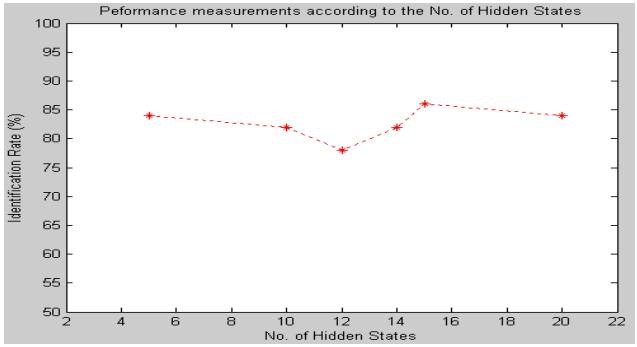

Figure 5: Results after setting up the hidden states of DHMM.

6.2.2 Experiment on 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_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.

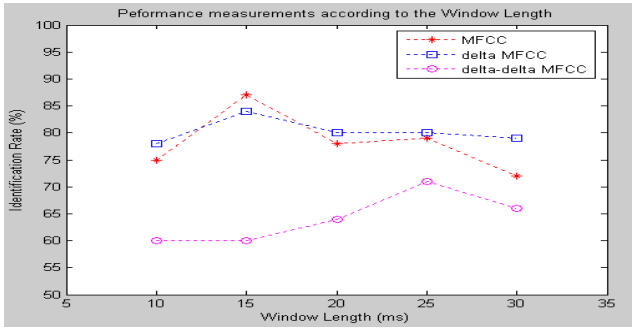

Figure 6: Effect of the window length on the identification rate.

6.2.3 Experiment on the Number of Cepstral Coefficients, N_{MC}

In this experiment, the number of cepstral coefficients varies from 10 to 20. The highest speaker identification rate 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.

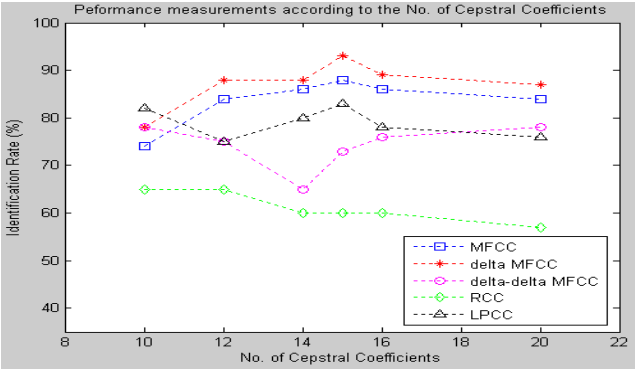

Figure 7: Speaker identification accuracy according to the number of cepstral coefficients.

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

6.3 Accuracies of Speaker Identification under Various SNRs

VALID audio-visual database [54] has been used to measure the performance of the proposed speaker identification system. Artificial white Gaussian noise was added to the original clean speech utterances to simulate various SNR levels. The models were trained at clean speech utterances and tested under SNR level ranging from 0dB to 30dB at 5dB intervals. Figure 8 shows the results of the performance of the proposed system under various SNR levels.

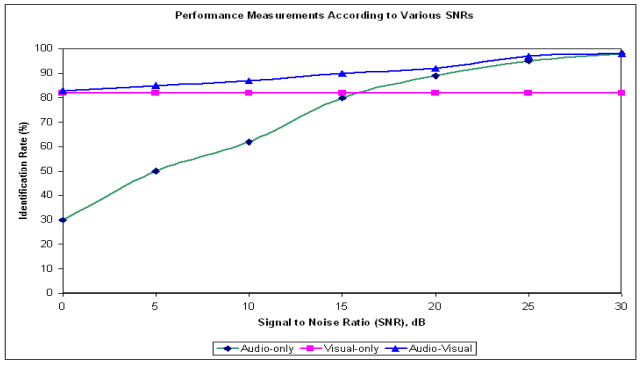

Figure 8: Accuracies (%) of speaker identification under different SNRs on VALID database.

From figure 8, it has been seen that when the noise level is low, the acoustic modality performs better than the visual one and, thus, the audio-visual recognition performance should be at least as good as that of the acoustic speech recognition. When the noise level is high and the visual recognition performance is better than the acoustic one, the integrated recognition performance should be at least the same to or better than the performance of the visual-only recognition.

7. CONCLUSIONS AND OBSERVATIONS

The experimental results show the versatility of the Audio-visual speaker identification system. This paper also investigates the correlations between audio and visual features. Experiment on the VALID database shows that the proposed strategy achieves the best accuracies of speaker identification at all levels of acoustic signal-to-noise ratio, ranging from 0dB to 30dB. The identification rate of this system revels that this proposed system can be used in various security and access control purposes. The performance of the system can be improved by using efficient speech and signal pre-processing techniques. Finally the performance of this proposed system can be populated according to the largest audio-visual speech database.

8. REFERENCES

[1] D. G. Stork andM. E. Hennecke, Eds., *Speechreading by Humans and Machines*. Berlin, Germany: Springer, 1996.

[2] R. Campbell, B. Dodd, and D. Burnham, Eds., *Hearing by Eye II*. Hove, United Kingdom: Psychology Press Ltd. Publishers, 1998.

[3] S. Dupont and J. Luettin, "Audio-visual speech modeling for continuous speech recognition*," IEEE Trans. Multimedia,* 2(3):141–151, 2000.

[4] G. Potamianos, J. Luettin, and C. Neti, "Hierarchical discriminant features for audio-visual LVCSR," *Proc. Int. Conf. Acoust., Speech, Signal Process.*, pp. 165–168, 2001.

[5] G. Potamianos and C. Neti, "Automatic speechreading of impaired speech," *Proc. Conf. Audio-Visual Speech Process.*, pp. 177–182, 2001.

[6] F.J. Huang and T. Chen, "Consideration of Lombard effect for speechreading," *Proc. Works. Multimedia Signal Process.*, pp. 613–618, 2001.

[7] G. Potamianos, C. Neti, G. Gravier, A. Garg, and A.W. Senior, "Recent advances in the automatic recognition of audio-visual speech," *To Appear: Proc. IEEE*, 2003.

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

[9] 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.

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

[11] 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*, 1998, 72-75.

[12] C. C. Chibelushi, F. Deravi, and J. S. D. Mason, "A review of speech-based bimodal recognition," *IEEE Trans. Multimedia*, vol. 4, pp. 23–37, Mar. 2002.

[13] X. Zhang, C. C. Broun, R. M. Mersereau, and M. Clements, "Automatic speechreading with applications to human-computer interfaces," *EURASIP J. Appl. Signal Processing*, vol. 2002, pp. 1228–1247, Nov. 2002.

[14] D. N. Zotkin, R. Duraiswami, and L. S. Davis, "Joint audio-visual tracking using particle filters," *EURASIP J. Appl. Signal Processing*, vol. 2002, pp. 1154–1164, Nov. 2002.

[15] P. De Cuetos, C. Neti, and A. Senior, "Audio-visual intent to speak detection for human computer interaction," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Istanbul, Turkey, June 5–9, 2000, pp. 1325–1328.

[16] D. Sodoyer, J.-L. Schwartz, L. Girin, J. Klinkisch, and C. Jutten, "Separation of audio-visual speech sources: A new approach exploiting the audio-visual coherence of speech stimuli," *EURASIP J. Appl. Signal Processing*, vol. 2002, pp. 1165–1173, Nov. 2002.

[17] E. Foucher, L. Girin, and G. Feng, "Audiovisual speech coder: Using vector quantization to exploit the audio/video correlation," in *Proc. Conf. Audio-Visual Speech Processing*, Terrigal, Australia, Dec. 4–6, 1998, pp. 67–71.

[18] J. Huang, Z. Liu, Y. Wang, Y. Chen, and E. Wong, "Integration of multimodal features for video scene classification based on HMM," in *Proc. Works. Multimedia Signal Processing*, Copenhagen, Denmark, Sept. 13–15, 1999, pp. 53–58.

[19] M. M. Cohen and D. W. Massaro, "What can visual speech synthesis tell visual speech recognition?," in *Proc. Asilomar Conf. Signals, Systems, Computers*, Pacific Grove, CA, 1994.

[20] E. Cosatto and H. P. Graf, "Photo-realistic talking-heads from image samples," *IEEE Trans. Multimedia*, vol. 2, pp. 152–163, Sept. 2000.

[21] *Gerasimos Potamianos, Chalapathy Neti,* and *Sabine Deligne, "*Joint Audio-Visual Speech Processing for Recognition and Enhancement," *Auditory-Visual Speech Processing Tutorial and Research Workshop (AVSP),* pp. 95-104, St. Jorioz, France, September 2003.

[22] 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, DECEMBER 2005.

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

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

[25] 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.

[26] 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, 2002, pp. 3808–3811.

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

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

[29] 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, Apr 1978, Pages: 1-4.

[30] 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.

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

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

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

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

[35] 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.

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

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

[38] 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.

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

[40] D. Kewley-Port and Y. Zheng, "Auditory models of formant frequency discrimination for isolated vowels", *Journal of the Acostical Society of America*, 103(3):1654–1666, 1998.

[41] D. O'Shaughnessy, *Speech Communication - Human and Machine*, Addison Wesley, 1987.

[42] E. Zwicker., "Subdivision of the audible frequency band into critical bands (frequenzgruppen)", *Journal of the Acoustical Society of America*, 33:248–260, 1961.

[43] 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*, 28:357–366, Aug 1980.

[44] 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, April 1993.

[45] 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*, 41, 1970, pp. 164-171.

[46] 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.

[47] Stephen Milborrow and Fred Nicolls, "Locating Facial Features with an Extended Active Shape Model," available at http://www.milbo.org/stasm-files/locating-facial-features-with-an-extended-asm.pdf.

[48] R. Herpers, G. Verghese, K. Derpains and R. McCready, "Detection and tracking of face in real environments," *IEEE Int. Workshop on Recognition, Analysis and Tracking of Face and Gesture in Real- Time Systems*, Corfu, Greece, pp. 96-104, 1999.

[49] J. Daugman, "Face detection: a survey," *Comput. Vis. Imag. Underst*, 83, 3, pp. 236- 274, 2001.

[50] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing.* Addison-Wesley, 2002.

[51] A. Rogozan, P.S. Sathidevi, "Static and dynamic features for improved HMM based visual speech recognition," 1st *International Conference on Intelligent Human Computer Interaction*, 9Allahabad, India, 20090, pp. 184-194.

[52] J. S. Lee, C. H. Park, "Adaptive Decision Fusion for Audio-visual speech Recognition", *Speech Recognition, Technologies and Applications*, ed. F. Mihelic, J. Zibert, (Vienna, Australia, 2008), pp. 550, 2008.

[53] A. Adjoudant, C. Benoit, "On the integration of auditory and visual parameters in an HMM-based ASR," *Speechreading by Humans and Machines: Models, Systems, and Speech Recognition, Technologies and Applications*, ed. D.G. Strok and M. E. Hennecke, (Springer, Berlin, Germany, 1996), pp. 461-472.

[54] N. A. Fox, B. A. O'Mullane and R. B. Reilly, "The Realistic Multi-modal VALID database and Visual Speaker Identification Comparison Experiments," *Proc. of the 5th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA-2005),* New York, 2005.

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.