

# Local Feature based Gender Independent Bangla ASR

Bulbul Ahamed

Senior Lecturer  
Northern University Bangladesh,  
Bangladesh

Khaled Mahmud

Lecturer Institute of Business  
Administration, University of Dhaka  
Bangladesh

B.K.M. Mizanur Rahman

Lecturer United International  
University Dhaka,  
Bangladesh

Foyzul Hassan

Senior S Q A Engineer  
Enosis Solutions  
Dhaka, Bangladesh

Rasel Ahmed

Lecturer Dhaka Residential  
Model College  
Bangladesh

Mohammad Nurul Huda

Associate Professor United  
International University Dhaka,  
Bangladesh

**Abstract—** This paper presents an automatic speech recognition (ASR) for Bangla (widely used as Bengali) by suppressing the speaker gender types based on local features extracted from an input speech. Speaker-specific characteristics play an important role on the performance of Bangla automatic speech recognition (ASR). Gender factor shows adverse effect in the classifier while recognizing a speech by an opposite gender, such as, training a classifier by male but testing is done by female or vice-versa. To obtain a robust ASR system in practice it is necessary to invent a system that incorporates gender independent effect for particular gender. In this paper, we have proposed a Gender-Independent technique for ASR that focused on a gender factor. The proposed method trains the classifier with the both types of gender, male and female, and evaluates the classifier for the male and female. For the experiments, we have designed a medium size Bangla (widely known as Bengali) speech corpus for both the male and female. The proposed system has showed a significant improvement of word correct rates, word accuracies and sentence correct rates in comparison with the method that suffers from gender effects using. Moreover, it provides the highest level recognition performance by taking a fewer mixture component in hidden Markov model (HMMs).

**Keywords-** Automatic speech recognition; Local features; gender factor; word correct rates; word accuracies; sentence correct rates; hidden Markov model.

## I. INTRODUCTION

Various methods were proposed to obtain robust automatic speech recognition (ASR) system; however, the ASR system that shows enough performance at any time and everywhere could not be realized now. One of the reasons is that the acoustic models (AMs) of an HMM-based classifier include many hidden factors such as speaker-specific characteristics that include gender types and speaking styles [1]-[3]. It is difficult to recognize speech affected by these factors, especially when an ASR system comprises only a classifier that made its training by a single type of gender.

One solution is to employ a acoustic model for both types of gender. Though the robustness of this acoustic model by utilizing the both gender specific characteristics is limited, but it resolves the gender effects more precisely.

On the other hand, only a very few works have been done in ASR for Bangla (can also be termed as Bengali) in spite of one of the largely spoken languages in the world. More than 220 million people speak in Bangla as their native language. It is ranked seventh based on the number of speakers [4]. A major difficulty to research in Bangla ASR is the lack of proper speech corpus. Some efforts are made to develop Bangla speech corpus to build a Bangla text to speech system [5]

However, this effort is a part of developing speech databases for Indian Languages, where Bangla is one of the parts and it is spoken in the eastern area of India (West Bengal and Kolkata as its capital). But most of the natives of Bangla (more than two thirds) reside in Bangladesh, where it is the official language. Although the written characters of Standard Bangla in both the countries are same, there are some sounds that are produced variably in different pronunciations of Standard Bangla, in addition to the myriad of phonological variations in non-standard dialects [6]. Therefore, there is a need to do research on the main stream of Bangla, which is spoken in Bangladesh, ASR. Some developments on Bangla speech processing or Bangla ASR can be found in [7]-[14]. For example, Bangla vowel characterization is done in [7]; isolated and continuous Bangla speech recognition on a small dataset using hidden Markov models (HMMs) is described in [8]. Again, Bangla digit recognition was found in [15]. Before us, there was no Bangla ASR system that incorporates gender specific characteristics, but our proposed method was based on Standard mel frequency cepstral coefficients (MFCCs) and consequently, it suffers from lower performance in the recognition stage [16].

In this paper, we have constructed a Gender-Independent (GI) ASR by utilizing the acoustic features [17], local features for suppressing the gender-factor up to a particular level. Here, the proposed technique trains the classifier with the both types of gender, male and female, and evaluates the classifier for the male and female. For the experiments, we have designed a medium size Bangla speech corpus for both the male and female. The proposed system has showed a significant improvement of word correct rates, word accuracies and sentence correct rates in comparison with the method that suffers from gender effects. Since the local features

incorporate frequency and time domain information, it shows significant improvement of recognition performance over the method based on MFCCs at fewer mixture components. Moreover, it requires a fewer mixture component in hidden Markov model (HMMs) and hence, computation time.

This paper is organized as follows. Sections II discusses Bangla phoneme schemes, Bangla speech corpus and triphone model. On the other hand, Section III and IV outline mel frequency cepstral coefficients (MFCCs) and Local features (LFs) extraction procedure, respectively and Section V explains the proposed GI-based technique. Section VI describes an experimental setup, and section VII analyzes experimental results. Finally, section VIII concludes the paper with some future remarks.

## II. BANGLA PHONEME SCHEMES,TRIPHONE DESIGN AND BANGLA SPEECH CORPUS

Bangla phonetic scheme and IPA (International Phonetic Alphabet) for Bangla were described in [16]. The paper [16] also showed characteristics of some Bangla words by using the spectrogra and triphone model based on HMM were also analyzed for Bangla words

At present, a real problem to do experiment on Bangla phoneme ASR is the lack of proper Bangla speech corpus. In fact, such a corpus is not available or at least not referenced in any of the existing literature. Therefore, we develop a medium size Bangla speech corpus, which is described below.

Hundred sentences from the Bengali newspaper “Prothom Alo” [18] are uttered by 30 male speakers of different regions of Bangladesh. These sentences (30x100) are used as male training corpus (D1). On the other hand, 3000 same sentences uttered by 30 female speakers are used as female training corpus (D2).

On the other hand, different 100 sentences from the same newspaper uttered by 10 different male speakers and by 10 different female speakers are used as male test corpus (D3) and female test corpus (D4), respectively. All of the speakers are Bangladeshi nationals and native speakers of Bangla. The age of the speakers ranges from 20 to 40 years. We have chosen the speakers from a wide area of Bangladesh: Dhaka (central region), Comilla – Noakhali (East region), Rajshahi (West region), Dinajpur – Rangpur (North-West region), Khulna (South-West region), Mymensingh and Sylhet (North-East region). Though all of them speak in standard Bangla, they are not free from their regional accent.

Recording was done in a quiet room located at United International University (UIU), Dhaka, Bangladesh. A desktop was used to record the voices using a head mounted close-talking microphone. We record the voice in a place, where ceiling fan and air conditioner were switched on and some low level street or corridor noise could be heard.

Jet Audio 7.1.1.3101 software was used to record the voices. The speech was sampled at 16 kHz and quantized to 16 bit stereo coding without any compression and no filter is used on the recorded voice.

## III. MFCC FEATURE EXTRACTOR

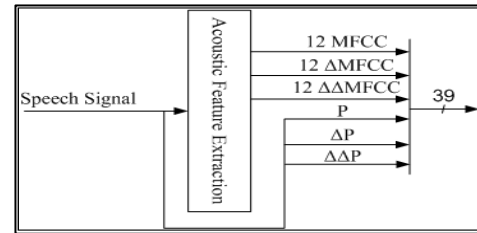


Figure 1. MFCC feature extraction

Conventional approach of ASR systems uses MFCC of 39 dimensions (12-MFCC, 12-ΔMFCC, 12-ΔΔMFCC, P, ΔP and ΔΔP, where P stands for raw energy of the input speech signal) and the procedure of MFCC feature extraction is shown in Fig.1. Here, hamming window of 25 ms is used for extracting the feature. The value of pre-emphasis factor is 0.97.

## IV. LOCAL FEATURE EXTRACTOR

At the acoustic feature extraction stage, the input speech is first converted into LFs that represent a variation in spectrum along the time and frequency axes. Two LFs, which are shown in Fig. 2, are then extracted by applying three-point linear regression (LR) along the time (t) and frequency (f) axes on a time spectrum pattern (TS), respectively. Fig. 3 exhibits an example of LFs for an input utterance. After compressing these two LFs with 24 dimensions into LFs with 12 dimensions using discrete cosine transform (DCT), a 25-dimensional (12 Δt, 12 Δf, and ΔP, where P stands for the log power of a raw speech signal) feature vector called LF is extracted.

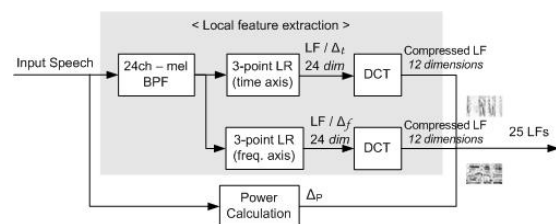


Fig.2 LFs extraction procedure.

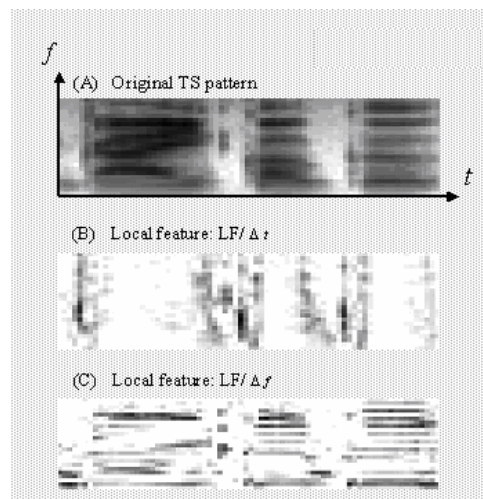


Fig. 3 Examples of LFs.

## V. PROPOSED LF-BASED GI ASR SYSTEM

Fig. 4 shows the proposed LF-based GI ASR system for Bangla Language. Here, an input speech is converted into LFs of 25 dimensions (12  $\Delta t$ , 12  $\Delta f$ , and  $\Delta P$ , where P stands for the log power of a raw speech signal) at the acoustic feature extraction stage, which is described in Section IV. Then, this extracted LFs (data set based on both male and female) of gender independent characteristics are used to train the GI classifier based on triphone HMM, while the Viterbi algorithm is used for evaluating the test data set for male and female.

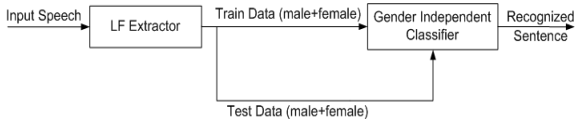


Fig. 4 The Proposed LF-based GI ASR System.

## VI. EXPERIMENTAL SETUP

The frame length and frame rate are set to 25 ms and 10 ms (frame shift between two consecutive frames), respectively, to obtain acoustic features (MFCCs) and local features (LFs) from an input speech. MFCCs and LFs comprised of 39 and 25 dimensional feature vectors, respectively.

For designing an accurate continuous word recognizer, word correct rate (WCR), word accuracy (WA) and sentence correct rate (SCR) for (D3+D4) data set are evaluated using an HMM-based classifier. The D1 (male) and D2 (female) data sets are used to design Bangla triphones HMMs with five states, three loops, and left-to-right models. Input features for the classifier are 39 dimensional MFCCs and 25 dimensional LFs. In the HMMs, the output probabilities are represented in the form of Gaussian mixtures, and diagonal matrices are used. The mixture components are set to 1, 2, 4 and 8.

For evaluating the performance of different methods including the proposed method, we have designed the following experiments:

### Experiment-I [Exp-I]

- (a) MFCC (Train: 3000 male, Test: 1000 male + 1000 female).
- (b) LF(Train: 3000 male, Test: 1000 male + 1000 female).

### Experiment-II [Exp-II]

- (c) MFCC (3000 female, Test: 1000 male + 1000 female).
- (d) LF(3000 female, Test: 1000 male + 1000 female).

### Experiment-III [Exp-III]

- (e) MFCC (Train: 3000 male + 3000 female, Test: 1000 male + 1000 female).
- (f) LF(Train: 3000 male + 3000 female, Test: 1000 male + 1000 female) [Proposed].

## VII. EXPERIMENTAL RESULT AND ANALYSIS

Figure 5 shows sentence correct rates for MFCC and LF-based ASR using the mixture component one, where total numbers of input sentences were 2000. From the figure it is shown that, LF-based ASR provides higher sentence correct

rate over all the experiments evaluated by MFCC-based ASR.

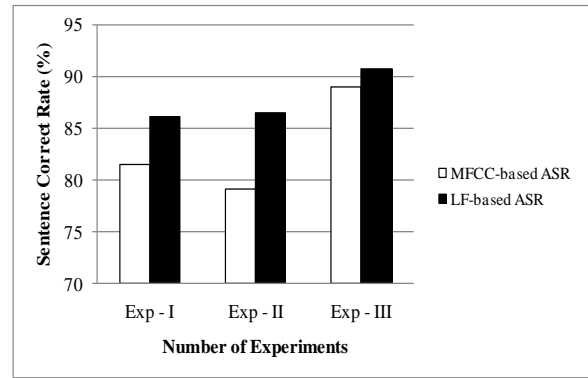


Fig.5 Sentence Correct Rates for MFCC and LF-based ASR.

It is noted that, MFCC-based method provides 81.45%, 79.05% and 88.90% SCRs for the experiments I, II and III, respectively, while corresponding experiments of LF-based method generates 86.10%, 86.45% and 90.65%, respectively. On the other hand, experiment III, which is done by gender independent condition, provides significant improvement of SCR over the experiments I and II that are gender dependent. For an example, GI LF-based method (experiment III (f)) shows 90.65% SCR that is significant improvement in comparison with the values, 86.10% and 86.45% which are provided by experiments I(b) and II(d). The reason for the better results exhibited by the LF-based method is the incorporation of frequency and time domain information in the input features, where the MFCC-based method only includes time domain features. Moreover, the GI LF-based method (experiment III(f)) gives better result over experiments I(b) and II(d) because training of HMM-based classifier for GI LF-based method embeds both male and female voices.

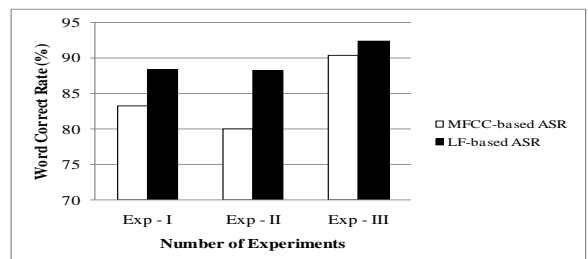


Fig. 6 Word Correct Rates for MFCC and LF-based ASR.

WCRs for the experiments I, II and III using the MFCC-based and LF-based methods are shown in Figure 6 using the mixture component one. From the experiments, it is exhibited that the LF-based methods provides higher word correct rates than the MFCC-based methods. Maximum improvement is shown in the Exp-II. On the other hand, the highest correctness is provided by the LF-based method for Exp-III, where gender-independent training was performed.

On the other hand, Figure 7 depicts the WAs for the experiments I, II and III using the MFCC-based and LF-based methods for the mixture component one. From the experiments, it is exhibited that the LF-based methods provides higher word accuracies than the MFCC-based methods. Exp-II exhibits the maximum improvement. Moreover, the highest level accuracy is generated by the LF-based method for Exp-III, where training was done by incorporating the male and female data sets.

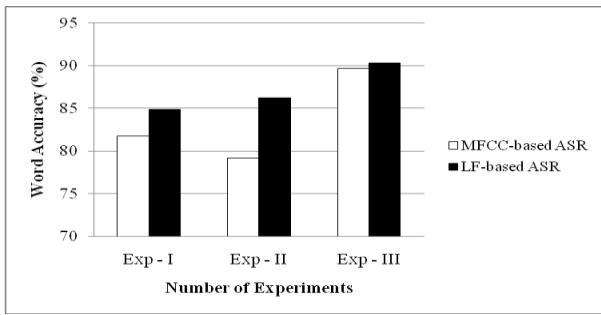


Fig. 7 Word Accuracy for MFCC and LF-based ASR.

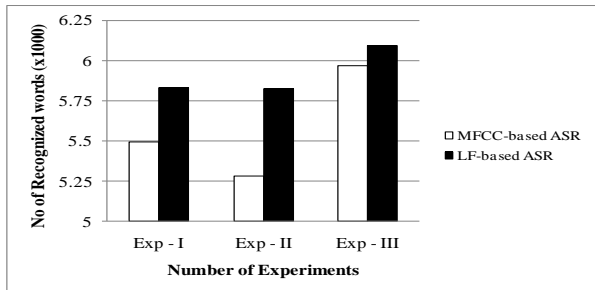


Fig. 8 No. of correctly recognized words for MFCC and LF-based ASR.

Again, the number of correctly recognized words out of 6600 input words is shown in Figure 8. From the figure, it is observed that the LF-based method in Exp-III recognizes the highest number of input words. Besides, the highest improvement by the LF-based method over the method based on MFCC is shown in Exp-II.

Tables 1 and 2 show the speech recognition Performance for the Exp-I where MFCC and LF-based methods for the mixture components 1, 2, 4 and 8 are investigated. Here, training and testing are done by using the D1 (male) and D3+D4 (male and female) speech corpora, respectively. For all the mixture components in the Table 1, LF-based method shows higher word correct rate, word accuracy and sentence correct rate in comparison with the method that incorporated MFCCs as input feature. It may be mentioned that the mixture component one provides the highest level performance among the entire mixture component investigated.

From the Table 2, it is exhibited that the methods incorporating LFs show the higher number of sentence recognition and the highest number at mixture component one compared to the counterpart.

Tables 3 and 4 generate the performance of similar pattern for the female dependent training in Exp-II. On the other hand, Tables 5 and 6 exhibit the gender independent performance where training and testing are done in the gender independent environment in Exp-III. It is claimed from the Tables 1, 3 and 5 that the LF-based method provides the higher recognition performance for all the mixture components.

Besides, the proposed LF-based method in Exp-III provides the higher performance among the three experiments and outputs the highest recognition performance for all the investigated mixture components. Among the experimented mixture components, the best result is achieved in component one.

On the other hand, from the Tables 2, 4 and 6 it is observed that the proposed method recognized the highest number of sentences.

Table I: Speech Recognition Performance for Exp-I using MFCC and LF-based methods using the mixture components 1, 2, 4 and 8. Training and testing are done by using the D1 and (D3+D4) speech corpora, respectively.

Mixture Components	Methods	Recognition Performance (%)		
		Word Correct Rate	Word Accuracy	Sentence Correct Rate
1	MFCC-based	83.20	81.71	81.45
	LF-based	88.32	84.85	86.10
2	MFCC-based	82.71	81.26	81.25
	LF-based	87.82	84.39	85.95
4	MFCC-based	78.02	77.26	77.20
	LF-based	86.79	83.24	84.65
8	MFCC-based	68.05	67.59	67.25
	LF-based	86.85	83.65	84.80

Table II: Word Recognition Performance for Exp-I using MFCC and LF-based methods using the mixture components 1, 2, 4 and 8. Training and testing are done by using the D1 and (D3+D4) speech corpora, respectively.

Mixture Components	Methods	Sentence recognition performance (out of 2000)		Word recognition performance (out of 6600)			
		Correctly recognized Sentence, H	Substitution, S	Correctly recognized Words, H	Deletion, D	Substitution, S	Insertion, I
1	MFCC-based	1629	371	5491	240	869	98
	LF-based	1722	278	5829	54	717	229
2	MFCC-based	1625	375	5459	264	877	96
	LF-based	1719	281	5796	57	747	226
4	MFCC-based	1544	456	5149	419	1032	50
	LF-based	1693	307	5728	67	805	234
8	MFCC-based	1345	655	4491	734	1375	30
	LF-based	1696	304	5732	71	797	211

Table III: Speech Recognition Performance for Exp-II using MFCC and LF-based methods using the mixture components 1, 2, 4 and 8. Training and testing are done by using the D2 and (D3+D4) speech corpora, respectively.

Mixture Components	Methods	Recognition Performance (%)		
		Word Correct Rate	Word Accuracy	Sentence Correct Rate
1	MFCC-based	79.94	79.14	79.05
	LF-based	88.20	86.23	86.45
2	MFCC-based	83.45	82.62	82.35
	LF-based	85.48	83.38	83.65
4	MFCC-based	80.33	79.65	79.20
	LF-based	84.05	82.09	82.35
8	MFCC-based	71.11	70.70	70.30
	LF-based	80.20	77.91	78.50

Table IV: Word Recognition Performance for Exp-II using MFCC and LF-based methods using the mixture components 1, 2, 4 and 8. Training and testing are done by using the D2 and (D3+D4) speech corpora, respectively.

Mixture Components	Methods	Sentence recognition performance (out of 2000)		Word recognition performance (out of 6600)			
		Correctly recognized Sentence, H	Substitution, S	Correctly recognized Words, H	Deletion, D	Substitution, S	Insertion, I
1	MFCC-based	1581	419	5276	322	1002	53
	LF-based	1729	271	5821	94	685	130
2	MFCC-based	1647	353	5508	239	853	55
	LF-based	1673	327	5642	145	813	139
4	MFCC-based	1584	416	5302	380	918	45
	LF-based	1647	353	5547	165	888	129
8	MFCC-based	1406	594	4693	661	1246	27
	LF-based	1570	430	5293	203	1104	151

Table V: Speech Recognition Performance for Exp-III using MFCC and LF-based methods using the mixture components 1, 2, 4 and 8. Training and testing are done by using the (D1+D2) and (D3+D4) speech corpora, respectively.

Mixture Components	Methods	Recognition Performance (%)		
		Word Correct Rate	Word Accuracy	Sentence Correct Rate
1	MFCC-based	90.36	89.67	88.90
	LF-based	92.27	90.30	90.65
2	MFCC-based	89.59	88.76	87.95
	LF-based	91.09	88.53	89.25
4	MFCC-based	91.23	90.53	89.85
	LF-based	91.50	89.18	89.50
8	MFCC-based	91.26	90.73	90.15
	LF-based	91.03	88.88	89.30

Table VI: Word Recognition Performance for Exp-III using MFCC and LF-based methods using the mixture components 1, 2, 4 and 8. Training and testing are done by using the (D1+D2) and (D3+D4) speech corpora, respectively.

Mixture Components	Methods	Sentence recognition performance (out of 2000)		Word recognition performance (out of 6600)			
		Correctly recognized Sentence, H	Substitution, S	Correctly recognized Words, H	Deletion, D	Substitution, S	Insertion, I
1	MFCC-based	1778	222	5964	123	513	46
	LF-based	1813	187	6090	40	470	130
2	MFCC-based	1759	241	5913	120	567	55
	LF-based	1785	215	6012	44	544	169
4	MFCC-based	1797	203	6021	100	479	46
	LF-based	1790	210	6039	37	524	153
8	MFCC-based	1803	197	6023	127	450	35
	LF-based	1786	214	6008	43	549	142

### VIII. CONCLUSION

This paper has proposed a gender independent automatic speech recognition technique for Bangla language by inputting local features. The following information concludes the paper.

- i) The methods based on local features provide a higher speech recognition performance than the method that incorporates the standard MFCCs for all the experimented mixture components.
- ii) For the LF-based methods, the mixture component one generates the highest level performance.
- iii) The proposed LF-based gender independent method has showed the significant improvement of word correct rate, word accuracy and sentence correct rate in comparison with the methods that are experimented in gender dependent environments.

In future, the authors would like to incorporate neural network based systems in gender independent for evaluating the performance.

### REFERENCES

- [1] S. Matsuda, T. Jitsuhiro, K. Markov and S. Nakamura, "Speech Recognition system Robust to Noise and Speaking Styles," Proc. ICSLP'04, Vol.IV, pp.2817-2820, Oct. 2004.
- [2] M. A. Hasnat, J. Mowla, and Mumit Khan, "Isolated and Continuous Bangla Speech Recognition: Implementation Performance and application perspective, " in Proc. International Symposium on Natural Language Processing (SNLP), Hanoi, Vietnam, December 2007.

- [3] R. Karim, M. S. Rahman, and M. Z Iqbal, "Recognition of spoken letters in Bangla," in Proc. 5th International Conference on Computer and Information Technology (ICCIT02), Dhaka, Bangladesh, 2002.
- [4] [http://en.wikipedia.org/wiki/List\\_of\\_languages\\_by\\_total\\_speakers](http://en.wikipedia.org/wiki/List_of_languages_by_total_speakers), Last accessed April 11, 2009.
- [5] S. P. Kishore, A. W. Black, R. Kumar, and Rajeew Sangal, "Experiments with unit selection speech databases for Indian languages," Carnegie Mellon University.
- [6] S. A. Hossain, M. L. Rahman, and F. Ahmed, "Bangla vowel characterization based on analysis by synthesis," Proc. WASET, vol. 20, pp. 327-330, April 2007.
- [7] A. K. M. M. Houque, "Bengali segmented speech recognition system," Undergraduate thesis, BRAC University, Bangladesh, May 2006.
- [8] K. Roy, D. Das, and M. G. Ali, "Development of the speech recognition system using artificial neural network," in Proc. 5th International Conference on Computer and Information Technology (ICCIT02), Dhaka, Bangladesh, 2003.
- [9] M. R. Hassan, B. Nath, and M. A. Bhuiyan, "Bengali phoneme recognition: a new approach," in Proc. 6th International Conference on Computer and Information Technology (ICCIT03), Dhaka, Bangladesh, 2003.
- [10] K. J. Rahman, M. A. Hossain, D. Das, T. Islam, and M. G. Ali, "Continuous bangla speech recognition system," in Proc. 6th International Conference on Computer and Information Technology (ICCIT03), Dhaka, Bangladesh, 2003.
- [11] S. A. Hossain, M. L. Rahman, F. Ahmed, and M. Dewan, "Bangla speech synthesis, analysis, and recognition: an overview," in Proc. NCCPB, Dhaka, 2004.
- [12] S. Young, et al, The HTK Book (for HTK Version. 3.3), Cambridge University Engineering Department, 2005. <http://htk.eng.cam.ac.uk/prot-doc/kitkbook.pdf>.
- [13] [http://en.wikipedia.org/wiki/Bengali\\_script](http://en.wikipedia.org/wiki/Bengali_script), Last accessed April 11, 2009.
- [14] C. Masica, *The Indo-Aryan Languages*, Cambridge University Press, 1991.
- [15] Ghulam Muhammad, Yousef A. Alotaibi, and Mohammad Nurul Huda, "Automatic Speech Recognition for Bangla Digits," ICCIT'09, Dhaka, Bangladesh, December 2009.
- [16] Foyzul Hasan, Rokibul Alam Kotwal, Saiful Alam Khan and Mohammad Nurul Huda, "Gender Independent Bangla Automatic Speech Recognition," IEEE/IAPR International Conference on Informatics, Electronics and Vision (ICIEV) 2012, May 2012, Dhaka, Bangladesh.
- [17] T. Nitta, "Feature extraction for speech recognition based on orthogonal acoustic-feature planes and LDA," Proc. ICASSP'99, pp.421-424, 1999.
- [18] Daily Prothom Alo. Online: [www.prothom-alo.com](http://www.prothom-alo.com).

### AUTHORS PROFILE

**Bulbul Ahamed** was born in Munshiganj, Bangladesh in 1982. He obtained his B. Sc. in Computer Science and Engineering and MBA (major in MIS & Marketing) from Northern University Bangladesh. Now he is pursuing his M.Sc. in Computer Science and Engineering at United International University, Bangladesh. He is now working as Senior Lecturer in Northern University Bangladesh. His research interests include Speech Recognition, Artificial Intelligence, Neural Network and Business. He has published his articles in different journals of Pakistan, Dubai and Bangladesh.

**B.K.M. Mizanur Rahman** was born in Jhenaidah, Bangladesh in 1972. He completed his B.Sc. in Electrical and Electronic Engineering Degree from BUET, Dhaka, Bangladesh. He is a student of Masters in Computer Science and Engineering at United International University, Dhaka, Bangladesh. He is now working as a Lecturer in the Department of Electrical and Electronic Engineering of the same university. His research interests include Speech Recognition, Digital Signal Processing and Renewable Energy.

**Rasel Ahmed** was born in Shariatpur, Bangladesh in 1983. He completed his Bachelor in Computer Science from National University, Bangladesh. Now he is pursuing his M.Sc. in Computer Science and Engineering at United International University, Bangladesh. He is now working as Lecturer in the Dhaka Residential Model College.

His research interests include Speech Recognition, Artificial Intelligence, and Information Technology.

**Khaled Mahmud** was born in 1984 at Pabna, Bangladesh. He was graduated from Bangladesh University of Engineering and Technology (BUET) in Computer Science and Engineering. He had his MBA (Marketing) from Institute of Business Administration, University of Dhaka. He was awarded gold medals both in his secondary and higher secondary school level for excellent academic performance. He is a Fulbright Scholar, now pursuing his MBA at Bentley University, Massachusetts, USA. He previously worked as Assistant Manager in Standard Chartered Bank. He has research interest business, technology, e-learning, e-governance, human resource management and social issues. He has his articles published in journals and conferences of USA, Canada, Australia, United Arab Emirates, Malaysia, Thailand, South Korea, India and Bangladesh.

**Foyzul Hassan** was born in Khulna, Bangladesh in 1985. He completed his B.Sc. in Computer Science and Engineering Degree from

Military Institute of Science and Technology (MIST), Dhaka, Bangladesh in 2006. He has participated several national and ACM Regional Programming Contests. He is currently doing M. Sc. in CSE in United International University, Dhaka, Bangladesh and also has been working as a Senior Software Quality Assurance Engineer at Enosis Solutions, Dhaka, Bangladesh. His research interests include Speech Recognition, Robotics and Software Engineering.

**Mohammad Nurul Huda** was born in Lakshmipur, Bangladesh in 1973. He received his B. Sc. and M. Sc. in Computer Science and Engineering degrees from Bangladesh University of Engineering & Technology (BUET), Dhaka in 1997 and 2004, respectively. He also completed his Ph. D from the Department of Electronics and Information Engineering, Toyohashi University of Technology, Aichi, Japan. Now, he is working as an Associate Professor in United International University, Dhaka, Bangladesh. His research fields include Phonetics, Automatic Speech Recognition, Neural Networks, Artificial Intelligence and Algorithms. He is a member of International Speech Communication Association (ISCA).