

Gender Effect Canonicalization for Bangla ASR

B.K.M. Mizanur
Rahman
Lecturer
United International
University,
Dhaka, Bangladesh

Bulbul Ahamed
Senior Lecturer
Northern University
Bangladesh, Dhaka,
Bangladesh

Md. Asfak-Ur-
Rahman
Technical Project
Manager
Grype Solutions
Bangladesh

Khaled Mahmud
Lecturer
Institute of Business
Administration
University of Dhaka

Mohammad Nurul
Huda
Associate Professor
United International
University,
Dhaka, Bangladesh

Abstract—This paper presents a Bangla (widely used as Bengali) automatic speech recognition system (ASR) by suppressing gender effects. Gender characteristic plays an important role on the performance of ASR. If there is a suppression process that represses the decrease of differences in acoustic-likelihood among categories resulted from gender factors, a robust ASR system can be realized. In the proposed method, we have designed a new ASR incorporating the Local Features (LFs) instead of standard mel frequency cepstral coefficients (MFCCs) as an acoustic feature for Bangla by suppressing the gender effects, which embeds three HMM-based classifiers for corresponding male, female and gender-independent (GI) characteristics. In the experiments on Bangla speech database prepared by us, the proposed system has achieved a significant improvement of word correct rates (WCRs), word accuracies (WAs) and sentence correct rates (SCRs) in comparison with the method that incorporates Standard MFCCs.

Keywords- Acoustic Model; AutomaticSpeech Recognition; Gender Effects Suppression; Hidden Markov Model.

I. INTRODUCTION

The current automatic speech recognition (ASR) system had been investigated for achieving the adequate performance at any time and everywhere; however, it could not be able for providing the highest level accuracies till to date. One of the reasons is that the acoustic models (AMs) of a hidden Markov model (HMM)-based classifier include many hidden factors such as speaker-specific characteristics that include gender types and speaking styles. It is difficult to recognize speech affected by these factors, especially when an ASR system contains only a single acoustic model. One solution is to employ multiple acoustic models, one model for each type of gender. By handling these gender effects appropriately the robustness of each acoustic model in an ASR can be extended to some limit.

A method of decoding in parallel with multiple HMMs corresponding to hidden factors has recently been proposed in [1], [2] for resolving these difficulties. Multi-path acoustic modeling, that represents hidden factors with several paths in the same AM instead of applying multiple HMMs, was also presented [3]. Unfortunately, only a very few works have been done for ASR in Bangla (can also be termed as Bengali), which is one of the largely spoken languages in the world. More than 220 million people speak in Bangla as their native language. It is ranked sixth 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].

Although some Bangla speech databases for the eastern area of India (West Bengal and Kolkata as its capital) were developed, but most of the natives of Bangla (more than two thirds) reside in Bangladesh, where it is the official language. Besides, 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 HMMs is described in [8].

Again, Bangla digit recognition was found in [15]. Since no work in Bangla was found for suppressing the gender factor, previously, we proposed a method [16] for that purpose by embedding multiple HMM-based classifiers. But this method of gender effect suppression did not incorporate any gender-independent (GI) classifier for resolving those male and female speakers whose voices have effect of opposite gender to some extent. To resolve this problem, we proposed another method for suppressing the gender factor more accurately by incorporating the GI classifier [17]; but this method could not be able to provide enough performance because of embedding standard mel frequency cepstral coefficients (MFCCs) as an input acoustic feature that did not include frequency domain information in its extraction process. Consequently, an exploitation of new feature is needed to obtain time and frequency domain information in its feature vector.

In this paper, we have designed a new ASR incorporating the Local Features (LFs) instead of standard mel frequency cepstral coefficients (MFCCs) as an acoustic feature for Bangla by suppressing the gender effects, which embeds three HMM-based classifiers for corresponding male, female and gender-independent (GI) characteristics. In the experiments on Bangla speech database prepared by us, the proposed system has achieved a significant improvement of word correct rates (WCRs), word accuracies (WAs) and sentence correct rates (SCRs) in

comparison with the method that incorporates Standard MFCCs.

This paper is organized as follows. Sections II discusses Bangla phoneme schemes, Bangla speech corpus and triphone model. On the other hand, Section III outlines MFCCs and LFs features extraction procedures and Section IV explain the gender effect suppression methods [16] and [17] including the proposed suppression technique by incorporating LFs. Section V describes the experimental setup and Section VI shows experimental results and provides a discussion. Finally, Section VII and VIII conclude the paper with some future remarks and references, respectively.

II. BANGLA PHONEME SCHEME, SPEECH CORPUS AND TRIPHONE MODEL

Bangla phonetic scheme and triphone models design were presented in our papers, [16] and [17]. These papers show ed how the left and right contexts are used to design the triphone models.

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. ACOUSTIC FEATURE EXTRACTOR

3.1 MFCC Feature Extractor

Conventional approach of ASR systems uses MFCC of 39 dimensions (12-MFCC, 12- Δ MFCC, 12- $\Delta\Delta$ MFCC, P, Δ P and $\Delta\Delta$ P, 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.

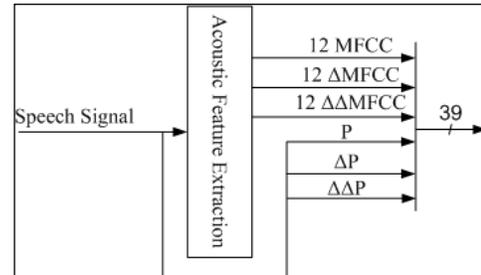


Fig. 1. MFCC feature extraction.

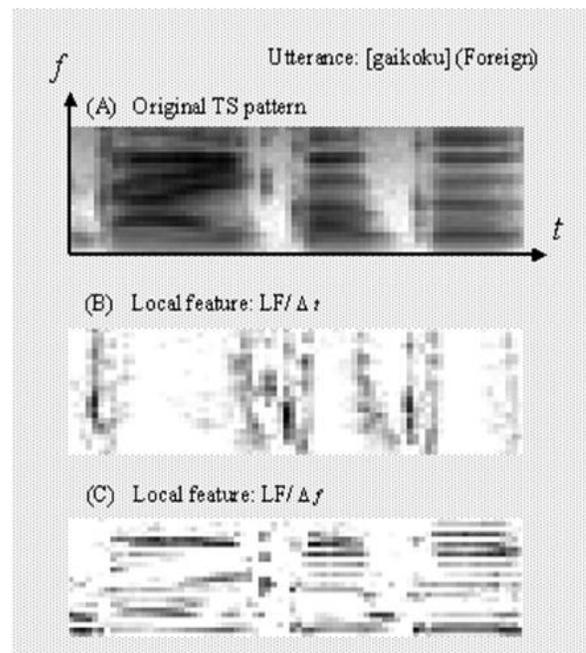


Fig. 2. Example of local features.

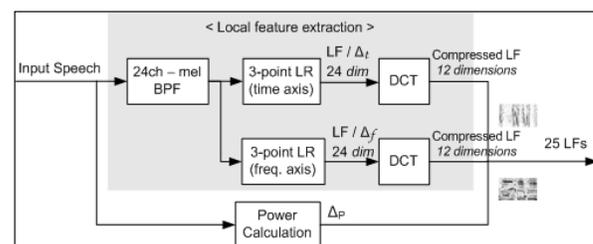


Fig. 3. Local feature extraction.

3.2 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 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. 2 exhibits an example of LFs for an input utterance, /gaikoku/. 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.3 shows the local feature extraction procedure.

IV. GENDER EFFECT SUPPRESSION METHODS

4.1 MFCC-Based Suppression method without GI classifier [16]

Fig. 4 shows the system diagram of the existing method [16], where MFCC features are extracted from the speech signal using the MFCC extractor described in Section 3.1 and then male and female HMM classifiers are trained using the D1 and D2 data sets, respectively. Here, triphone acoustic HMMs are designed and trained using D1 and D2 data sets. Output hypothesis is selected based on maximum output probabilities after comparing male and female hypotheses, and passed the best matches hypothesis to the output.

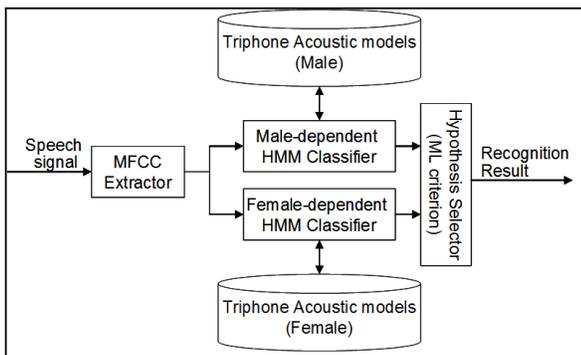


Fig. 4. MFCC-based gender effect suppression method without GI classifier [16].

4.2 MFCC-Based Suppression method incorporating GI classifier [17]

The diagram of the method [17] is depicted in Fig. 5. Here, the extracted MFCC features from the input speech signal are inserted into the male, female and GI HMM-based classifiers. The male, female and GI HMM-based classifiers are trained using the D1, D2 and (D1+D2) data sets. Here, output hypothesis is selected based on maximum output probabilities after comparing male, female and gender independent hypotheses, and passed the best matches hypothesis to the output.

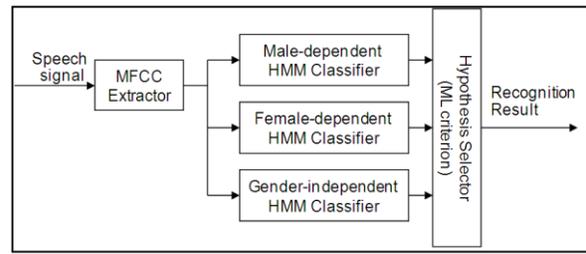


Fig. 5. MFCC-based gender effect suppression method incorporating GI classifier [17].

4.3 LF-based Proposed Suppression Method incorporating GI classifier

The diagram of the LF-based method is depicted in Fig. 6. Here, the extracted LFs [19] from the input speech signal are inserted into the male, female and GI HMM-based classifiers. The male, female and GI HMM-based classifiers are trained using the D1, D2 and (D1+D2) data sets. Here, output hypothesis is selected based on maximum output probabilities after comparing male, female and gender independent hypotheses, and passed the best matches hypothesis to the output.

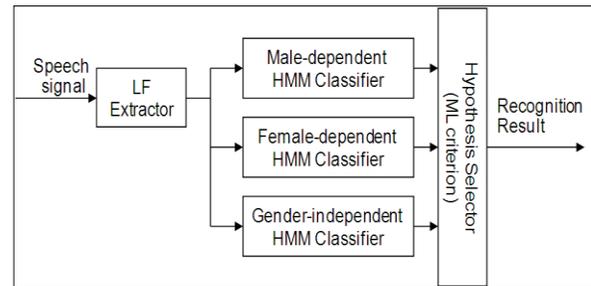


Fig. 6. Proposed LF-based suppression method incorporating GI classifier.

V. 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 LFs) from an input speech. MFCC and comprised of 39 (12-MFCC, 12- Δ MFCC, 12- $\Delta\Delta$ MFCC, P, Δ P and $\Delta\Delta$ P, where P stands for raw energy of the input speech signal) and 25 (12 delta coefficients along time axis, 12 delta coefficients along frequency axis, and delta coefficient of log power of a raw speech signal) dimensions, 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 and D2 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 the methods, [16] and [17], we have designed the following experiments:

- (a) MFCC (Train: 3000 male, Test: 1000 male+1000 female).
- (b) MFCC (Train: 3000 female, Test: 1000 male+1000 female).
- (c) MFCC (Train: 3000 male + 3000 female, Test: 1000 male +1000 female).
- (d) MFCC (Train: 3000 male, Train: 3000 female, Test: 1000 male + 1000 female) [16].
- (e) MFCC (Train: 3000 male, Train: 3000 female, Train: 3000 male + 3000 female, Test: 1000 male + 1000 female).

New experiments given below are designed for mixture component one using the LFs for evaluating the performance:

- (a) LF (Train: 3000 male, Test: 1000 male+1000 female).
- (b) LF(Train: 3000 female, Test: 1000 male+1000 female).
- (c) LF(Train: 3000 male + 3000 female, Test: 1000 male +1000 female).
- (d) **Proposed**, LF(Train: 3000 male, Train: 3000 female, Train: 3000 male + 3000 female, Test: 1000 male + 1000 female).

VI. EXPERIMENTAL RESULTS AND ANALYSIS

Fig. 7 shows the comparison of word correct rates among all the MFCC-based investigated methods, (a), (b), (c), (d) and (e). Among all the mixture components investigated, the method, (e) shows higher performance in comparison with the other method evaluated. It is noted that the method, (e) exhibits its best performance (92.17%) at mixture component two.

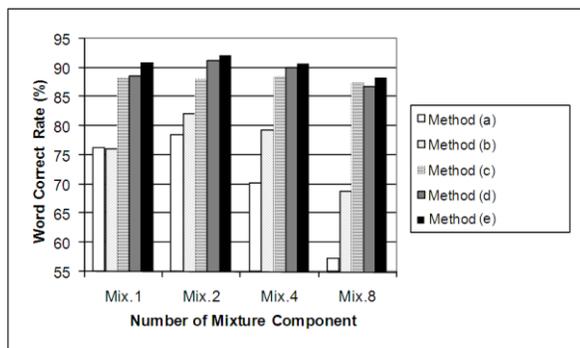


Fig. 7. Word correct rate comparison among MFCC-based investigated methods.

Word accuracies for different investigated methods based on MFCCs are depicted in Fig. 8. From the figure, it is observed that the highest level performance (91.64%) at mixture component two is found by the method, (e) compared to the other methods investigated. Here, the performance of the methods, (a), (b), (c), (d) and (e) at mixture component two

are 77.58%, 81.47%, 87.39%, 90.78% and 91.64%, respectively.

It is shown from the Fig. 9 that sentence correct rates for the MFCC-based investigated methods, (a), (b), (c), (d) and (e) are 77.20%, 81.45%, 86.60%, 90.45% and 91.30%, respectively, where the method, (e) provides its best performance. The methods, (a), (b) and (c) give less performance in comparison with the method (d) because (d) incorporates both HMM-based classifiers for male and female. Again, the method, (e) incorporates GI HMM-based classifier over the method (d), which increases sentence correct rate significantly. Since the maximum output probability is generated by the MFCC based method, (e) after comparing the probabilities among male, female and GI classifiers, the suppression method, (e) shows its superiority.

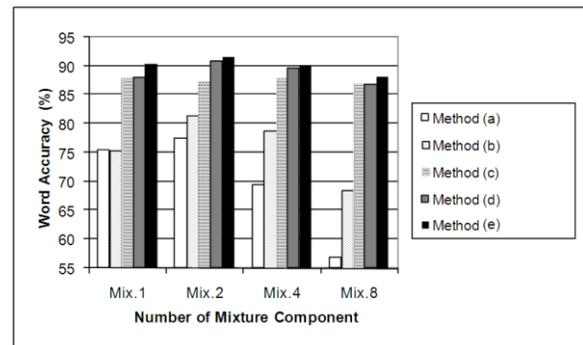


Fig. 8. Word accuracy comparison among MFCC-based investigated methods.

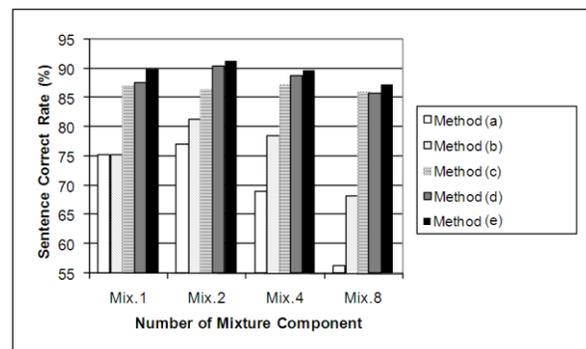


Fig. 9. Sentence correct rate comparison among MFCC-based investigated methods.

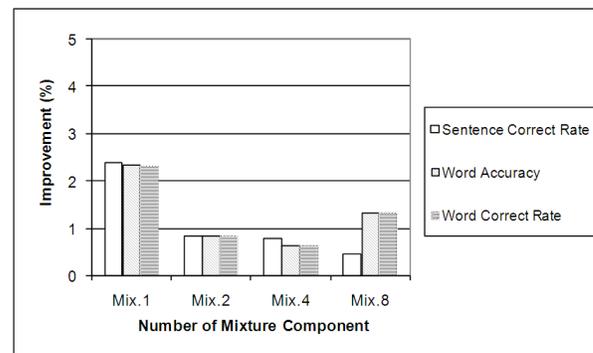


Fig. 10. Effect of Gender-Independent classifier in the MFCC-based method (e) over the method (d).

Improvements by the GI classifier in the MFCC-based method (e) over the method (d) that does not incorporate GI classifier is shown in Fig. 10. From the figure, it is observed that the method, (e) shows its highest level improvement at mixture components one, where the improvement of sentence correct rate, word accuracy and word correct rate are 2.4%, 2.36% and 2.32%, respectively.

Word Correct Rates (WCRs), Word Accuracies (WAs) and Sentence Correct Rates (SCRs) for LF and MFCC based methods, (a), (b), (c) and (d)/(e) using (D3+D4) data set are shown in Table I. Here, methods (d) and (e) represent LF-based and MFCC-based suppression methods with gender independent (GI) classifiers, respectively. From the experiment #1, where the HMM-based classifier is trained with D1 data set and evaluated with (D3+D4) data set, a tremendous improvement of WCRs, WAs and SCRs are exhibited by the LF-based ASR that incorporates gender effect canonicalization module in the ASR process. Similarly, the same pattern of performance is also achieved in the experiment #2, where the HMM-based classifier is trained with D2 data set and evaluated with (D3+D4) data set. This trend explicates the LFs, which embeds time and frequency domain information in its extraction procedure, as excellent feature for the Bangla automatic speech recognition system. Again, the experiment #3, which shows the GI ASR for Bangla language based on LF and MFCC features, is trained with (D1+D2) and evaluated with (D3+D4) data sets and provides 3.83%, 2.60% and 3.70% improvements by the LF-based method in comparison with the MFCC-based counterpart. Finally, the methods in the experiment #4 imply two ASR systems for GI Bangla ASR by integrating gender factor canonicalization process in its architecture and shows the highest level performance for WCRs, WAs and SCRs compare to the corresponding the methods in the experiments #1, #2 and #3. Since these methods of the experiment #4 always maximize the output probabilities obtained from the three classifiers: male, female and GI, it shows its maximum level of performance. Besides, the LF-based methods improves the WCRs, WAs and SCRs by 3.94%, 3.43% and 3.90%, respectively than the method that inputs MFCC features in the HMM-based classifier in the gender effect suppression process.

Table I. Word Correct Rates (WCRs), Word Accuracies (WA) and Sentence Correct Rates (SCRs) for LF and MFCC based methods, (a), (b), (c) and (d)/(e) using (D3+D4) data set for mixture component one. Here, methods (d) and (e) are LF-based and MFCC-based suppression methods with gender independent (GI) classifiers, respectively.

Experiments	Methods	WCR(%)	WA(%)	SCR(%)
#1	LF-based (a)	88.63	85.53	86.60
	MFCC-based (a)	76.38	75.41	75.30
#2	LF-based (b)	86.29	83.65	84.50
	MFCC-based (b)	76.16	75.50	75.45
#3	LF-based (c)	92.22	90.43	90.85
	MFCC-based (c)	88.39	87.83	87.15
#4	LF-based (d)	94.94	93.86	93.95
	MFCC-based (e)	91.00	90.43	90.05

On the other hand, Table II exhibits sentence recognition performance for LF and MFCC based methods, (a), (b), (c) and (d)/(e) using (D3+D4) data set, where the methods (d) and (e) represent LF-based and MFCC-based suppression methods with gender independent (GI) classifiers, respectively. Here, experiments #1, #2, #3 and #4 use same corpora for training and evaluation that we explained earlier. From all the experiments, it is evident that the LF-based method reduces the number of incorrectly recognized sentences with respect to its counterpart. The experiment #4 shows the highest number of correctly recognized sentences than the corresponding methods in the other experiments, #1, #2 and #3 investigated. For an example, in the LF-based and MFCC-based methods of experiment #4, the numbers of correctly recognized sentences are 1879 and 1801, respectively that are the highest numerical figures among the corresponding methods of all the experiments. It is noted that two significant phenomenon contributed more for obtaining the best experimental results by the LF-based method of experiment #4: i) both time and frequency domain information and ii) selection of maximum probability among the three output probabilities calculated by the male, female and GI HMM-based classifiers.

Table II. Sentence recognition performance for LF and MFCC based methods, (a), (b), (c) and (d)/(e) using (D3+D4) data set for mixture component one. Here, methods (d) and (e) are LF-based and MFCC-based suppression methods with gender independent (GI) classifiers, respectively

Experiments	Methods	Sentence Recognition Performance (Total 2000 Sentences)	
		Correctly Recognized Sentences, H	Incorrectly Recognized Sentences, S
#1	LF-based (a)	1732	268
	MFCC-based (a)	1506	494
#2	LF-based (b)	1690	310
	MFCC-based (b)	1509	491
#3	LF-based (c)	1817	183
	MFCC-based (c)	1743	257
#4	LF-based (d)	1879	121
	MFCC-based (e)	1801	199

Moreover, word recognition performance for LF and MFCC based methods, (a), (b), (c) and (d)/(e) using (D3+D4) data set are summarized in the Table III. In the table, methods (d) and (e) represent LF-based and MFCC-based suppression methods with gender independent (GI) classifiers, respectively. The same speech corpora for training and evaluation are used for the experiments #1, #2, #3 and #4 which is already described in the earlier. It is observed from all the experiments that the LF-based method increases the number of correctly recognized words in comparison with its counterpart. The highest number of correctly recognized words shown by the experiment #4 with respect to their corresponding methods in the other experiments, #1, #2 and #3 are evident from the table. It can be mentioned as an example that the LF-based and MFCC-based methods of experiment #4 provide the highest numbers of correctly recognized words, which are 6247 and 5988, respectively, dictates the respective numerical figures obtained for the corresponding methods of all the other experiments. The reason for obtaining the best experimental results by the LF-based method of experiment #4 is also illustrated earlier.

Table III. Word recognition performance for LF and MFCC based methods, (a), (b), (c) and (d)/(e) using (D3+D4) data set for mixture component one. Here, methods (d) and (e) represent LF-based and MFCC-based suppression methods with gender independent (GI) classifiers, respectively.

Experiments	Methods	Word Recognition Performance (Total 6580 words)			
		Correctly Recognized Words, H	Deletion, D	Substitution, S	Insertion, I
#1	LF-based (a)	5832	58	690	204
	MFCC-based (a)	5026	438	1116	64
#2	LF-based (b)	5678	89	813	174
	MFCC-based (b)	5011	448	1121	43
#3	LF-based (c)	6068	55	457	118
	MFCC-based (c)	5816	152	612	37
#4	LF-based (d)	6247	33	300	71
	MFCC-based (e)	5988	140	452	38

VII. CONCLUSION

This paper has proposed an automatic speech recognition technique based on LFs for Bangla language by suppressing the gender effect incorporating HMM-based classifiers for male, female and Gender-Independent characteristics. The following information concludes the paper.

- i) The MFCC-based method incorporating GI classifier provides the higher performance than the method that does not incorporate GI classifier. The MFCC-based method incorporating GI exhibits its superiority at all the mixture components investigated.
- ii) The incorporation of GI HMM classifier improves the word correct rates, word accuracies and sentence correct rates significantly.
- iii) The proposed LF-based method shows a significant improvement of word correct rates, word accuracies and sentence correct rates for mixture component one.

In future, the authors would like to evaluate performance by incorporating neural network based systems.

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] T. Shinozaki and S. Furui, "Spontaneous Speech Recognition Using a Massively Parallel Decoder," Proc. ICSLP'04, Vol.III, pp.2817-2820, Oct. 2004.
- [3] A. Lee, Y. Mera, K. Shikano and H. Saruwatari, "Selective multi-path acoustic model based on database likelihoods," Proc. ICSLP'02, Vol.IV, pp.2661-2664, Sep. 2002.
- [4] http://en.wikipedia.org/wiki/List_of_languages_by_total_speakers, Last accessed July04, 2012.
- [5] S. P. Kishore, A. W. Black, R. Kumar, and Rajeev 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/ktkbook.pdf>.
- [13] http://en.wikipedia.org/wiki/Bengali_script, Last accessed September 28, 2011.
- [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] Mohammed Rokibul Alam Kotwal, Foyzul Hasan and Mohammad Nurul Huda, "Gender effects suppression in Bangla ASR by designing multiple HMM based classifiers," CICN 2011, Gwalior, India.
- [17] Foyzul Hasan, Mohammed Rokibul Alam Kotwal and Mohammad Nurul Huda, "Bangla ASR design by suppressing gender factor with gender-independent and gender-based HMM classifiers," WICT 2011, December 2011, Mumbai, India.
- [18] Prothom Alo. Online: www.prothom-alo.com.
- [19] T. Nitta, "Feature extraction for speech recognition based on orthogonal acoustic-feature planes and LDA," Proc. ICASSP'99, pp.421-424, 1999.

AUTHORS PROFILE

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 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.

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 and conferences of USA, Pakistan, United Arab Emirates and Bangladesh.

Md.Asfak-Ur-Rahman was born in Kishoreganj, Bangladesh in 1985. He obtained his B. Sc. in Computer Science and Engineering BRAC University, Bangladesh. Now he is pursuing his M.Sc. in Computer Science and Engineering at United International University, Bangladesh. He is now working as Technical Project Manager in Grype Solutions, Bangladesh. His research interests include Speech Recognition, Artificial Intelligence and web programming.

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 published his articles in

journals and conferences of USA, Canada, Australia, United Arab Emirates, Malaysia, Thailand, South Korea, India and Bangladesh.

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).