Bangla Phonetic Feature Table Construction for Automatic Speech Recognition

Foyzul Hassan, Mohammed Rokibul Alam Kotwal and Mohammad Nurul Huda
Department of Computer Science and Engineering
United International University
Dhaka, Bangladesh
foyzul.hassan@gmail.com, rokib_kotwal@yahoo.com, mnh@cse.uiu.ac.bd

***Abstract*—This This research constructs a phonetic feature (PF) table for all the phonemes pronounced in Bangla (widely known as Bengali) language where the whole study is divided into two parts. In the first part, a PF table is constructed, while the second part deals with Bangla automatic speech recognition (ASR) using PFs. For Bangla language, fifty three phonemes including both vowels and consonants are considered in which the phones, শ (/s/) and স (/s/), and, ণ (/n/) and ন (/n/) contain approximately same spectrum and hence, they share same PFs. In the PF table, twenty two PFs (Silence, Short Silence, Stop, …) are required for representing all the Bangla phonemes. On the other hand, the second part comprised of three stages: i) first stage deals with acoustic features, mel frequency cepstral coefficients (MFCCs) extraction, ii) second stage embeds PFs extraction procedure using a multilayer neural network (MLN) and iii) the final stage integrates a triphone-based hidden Markov model (HMM) for generating the output text strings by inputting log values of twenty two dimensional PFs. In the experiments on Bangla Newspaper Article Sentences, it is observed that the PF-based ASR system provides higher word correct rate, word accuracy and sentence correct rate in comparison with the standard MFCC-based method.**

Keywords— phonetic feature; mel frequency cepstral coefficient; multilayer neural network; automatic speech recognition; hidden Markov model

I. Introduction

There have been many literatures in automatic speech recognition (ASR) systems for almost all the major languages in the world. Unfortunately, only a very few works have been done in ASR for 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 seventh based on the number of speakers [1]. 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 [2]. 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 [3]. 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 [4]-[11], where various hidden Markov model (HMM)-based ASR systems have been developed. Most of these ASR systems make use of a preprocessed form, such as mel-frequency cepstral coefficients (MFCCs), of the speech signal, which encodes the time-frequency distribution of signal energy. However, these MFCC-based systems do not provide better recognition performance in real acoustic conditions (See Figure 1(a)). On the other hand, a system based on articulatory features (AFs) or phonetic features (PFs) exhibits higher recognition accuracy in practical conditions and models coarticulatory phenomena more naturally [12](See Figure 1(b)). From the Figures 1(a) and 1(b), it is shown that the PF-based system outputs few misclassifications. The main problem for the Bangla language is that PF table is yet to be constructed.

In this paper, we have designed a phonetic feature (PF) table for all the phonemes pronounced in Bangla language. The first part of the research deals with a PF table construction, while the second part constructs a Bangla ASR using PFs. In the PF table, twenty two PFs are required for representing all the Bangla phonemes. On the other hand, the second part comprised of three stages: i) first stage deals with acoustic features, mel frequency cepstral coefficients (MFCCs), extraction, ii) second stage embeds PFs extraction procedure using a multilayer neural network (MLN) and iii) the final stage integrates a triphone-based HMM for generating the output text strings by inputting log values of twenty two dimensional PFs.

The paper is organized as follows. Section II briefly describes an approximate phonetic scheme and speech corpus for Bangla and formation of words, and speech corpus for Bangla. Section III explains about Bangla PFs, while Section IV deals with Proposed ASR construction using Bangla PFs. Again, Section V gives experimental setup, results and discussion on Bangla continuous word recognition. Finally, Section VI draws some conclusions with future directions.

978-1-4799-3497-3/13/$31.00 ©2013 IEEE

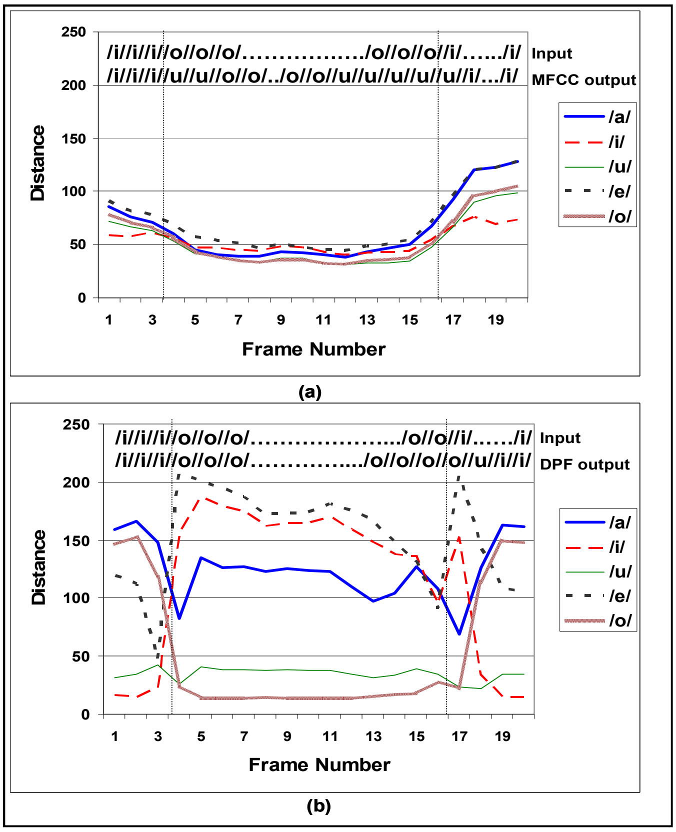

Fig. 1. Phoneme distances for utterance, /ioi/ using (a) MFCC-based system and (b) PF-based system.

II. PHONETIC SCHEME AND CORPUS FOR BANGLA

A. Bangla Phonemes

The phonetic inventory of Bangla consists of 14 vowels, including seven nasalized vowels, and 29 consonants. An approximate phonetic scheme in IPA is given in [13][14], where only the main 7 vowel sounds are shown, though there exists two more long counterpart of /i/ and /u/, denoted as /i:/ and /u:/, respectively. These two long vowels are seldom pronounced differently than their short counterparts in modern Bangla. There is controversy on the number of Bangla consonants.

B. Bangla Words

TABLE I. EXAMPLES OF SOME BANGLA WORDS WITH THEIR IPA

Bangla Word	English Pronunciation	IPA	Our Symbol
আমরা	AAMRA	/a m r a/	/aa m r ax/
আচরণ	AACHORON	/a tʃ r n/	/aa ch ow r aa n/
আবেদন	ABEDON	/a b æ d̪ n/	/ax b ae d aa n/

Table I lists some Bangla words with their written forms and the corresponding IPA. From the table, it is shown that the same 'আ' (/a/) has different pronunciation based on succeeding phonemes 'ম', 'চ' and 'ব'. These pronunciations are sometimes long or short. For long and short 'আ' we have used two different phonemes /aa/ and /ax/, respectively. Similarly, we have considered all variations of same phonemes and consequently, found total 51 phonemes excluding beginning and end silence (/sil/) and short pause (/sp/).

C. Bangla Speech Corpus

Hundred sentences from the Bengali newspaper "Prothom Alo" [15] are uttered by 30 male speakers of different regions of Bangladesh. These sentences (30x100) are used as training corpus (D1). On the other hand, different 100 sentences from the same newspaper uttered by 10 different male speakers are used as test corpus (D2). 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.

III. PROPOSED BANGLA PHONETIC FEATURES

A phoneme can easily be identified by its PFs [16][17]. Table II specified in succeeding page of this paper shows the proposed Bangla PFs for all the phonemes with their international phonetic alphabet (IPA) and Bangla orthographic transcription. Here, the fifty three Bangla phonemes and twenty two PFs for each phoneme are silence, short silence, stop, nasal, bilabial, fricative, liquid, lenis, vowel, front, central, back, unvoiced, long, short, dipthong, high, low, medium, round, unround and glottal, which shown in the table horizontally and vertically, respectively. Here, (Front, Back, Central) and (High, Low, Medium) represent tongue position in forward and backward, and upward and downward directions, respectively. Besides, plus (+) and minus (-) elements in the table represent whether corresponding element is present or absent, respectively.

IV. PROPOSED ASR SYSTEM USING PFS

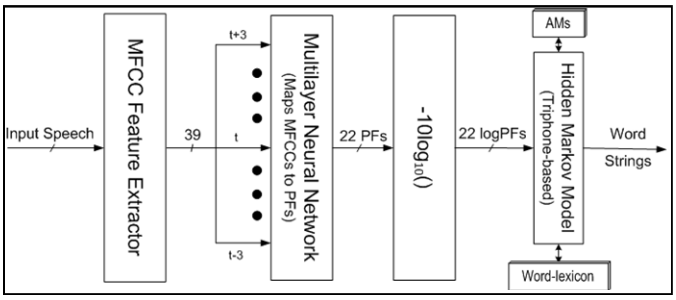

Fig. 2. Proposed PF-based ASR System.

We have implemented a PF-based ASR system with an input acoustic vector of MFCCs using an MLN which is shown in Figure 2. This system comprised of three stages: i) first stage deals with acoustic features, MFCCs extraction, ii) second stage embeds PFs extraction procedure using an MLN and iii) the final stage integrates a triphone-based HMM for generating the output text strings by inputting logarithmic values [17] of

TABLE II. PROPOSED BANGLA PHONETIC FEATURES

Bangla Orthographic Transcription	আ	চ	ছ		ও	র	ণ, ন	গ	উ	ল	আ (short)	ম	ব	এ্যা	দ	ঈ	ক	র (long)	ই (long)	ধ	ত	প	শ/স	ও	ষ	থ	ঙ	এ (long)
IPA	/a/	/tʃ/	/tʃʰ/	Short pause	/o/	/ʃ/	/n/	/g/	/µ/	/l/	/â/	/m/	/b/	/æ/	/d/	/i/	/k/	/ŕ/	/ē/	/d̪ʰ/	/t̪/	/p/	/s/	/o̤/	/s/	/tʰ/	/ŋ/	/e/
PFs/Phoneme	aa	ch	ey	sp	ow	r	n	g	uh	l	ax	m	b	ae	d	iy	k	er	ih	dh	t	p	sh	ao	s	th	ng	eh
Silence	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Short Silence	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Stop	-	-	-	-	-	-	-	+	-	-	-	-	+	-	+	-	+	-	-	+	+	+	-	-	-	-	-	-
Nasal	-	-	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
Bilabial	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
Fricative	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	+	+	-	-
Liquid	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Lenis	-	-	-	-	-	-	-	+	-	-	-	-	+	-	+	-	-	-	-	+	-	-	+	-	-	-	-	-
Vowel	+	-	+	-	+	-	-	-	-	-	+	-	-	-	-	+	-	+	+	-	-	-	-	+	-	-	-	+
Front	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	+	-	-	+	-	-	+	-	-	-	-	-	+
Central	+	-	-	-	-	+	+	-	-	+	-	-	-	-	+	-	-	+	-	+	+	-	+	+	+	+	-	+
Back	+	+	-	-	-	-	-	+	+	-	+	-	-	-	-	-	+	-	-	-	-	-	+	-	-	-	+	-
Unvoiced	-	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	+	+	-	+	+	-	-
Long	+	-	-	-	+	-	-	-	-	+	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
Short	+	-	+	-	-	-	-	-	+	-	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+
Dipthong	+	-	+	-	-	-	-	-	-	+	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+	-
High	+	-	-	-	-	-	-	-	-	-	+	-	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-
Low	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+
Medium	+	-	+	-	-	-	-	-	-	+	+	+	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	+
Round	+	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
UnRound	+	-	+	-	-	+	-	-	-	+	+	+	-	-	-	+	-	+	+	-	-	-	-	-	-	-	-	+
Glotal	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Bangla Orthographic Transcription	ভ	ঙ	য়		হ	ফ	ছ	আ (long)	আ	ঢ	ঘ	য	খ	অ	হ	ই	য	ঝ	জ	য় (long)	য়	ও (long)		ভ (long)	উ
IPA	/bʰ/	/ŋ/	/e̤/	/u/	/h/	/pʰ/	/tʃʰ/	/ā/	/a:/	/dʰ/	/gʰ/	/dʒ/	/kʰ/	/ɔ/	/h/	/i/	/dʒ/	/dʰʒ/	/dʒ/	/e̤//-	/e̤//-	/ō/	Silence	/bʰ/	/u/
PFs/Phoneme	v	nh	oy	uw	hh	f	chh	ay	ah	dha	gh	z	kh	aw	h	i	zh	jh	j	ky	y	w	sil	bh	u
Silence	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-
Short Silence	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Stop	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	+	+	+	-	-	-	-	+	-
Nasal	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Bilabial	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
Fricative	+	-	-	-	-	+	+	-	-	-	-	+	-	-	+	-	+	+	+	-	-	-	-	-	-
Liquid	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-
Lenis	+	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	+	-	-	-	-	-	-	-
Vowel	-	-	+	+	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
Front	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	+	-	-	-
Central	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
Back	-	-	-	+	+	-	+	-	-	-	+	-	+	-	-	-	-	+	-	+	+	-	-	-	+
Unvoiced	-	-	-	-	+	+	+	-	-	-	-	-	+	-	-	-	+	+	-	-	-	-	+	-	-
Long	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Short	-	-	+	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Dipthong	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
High	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+
Low	-	-	+	-	-	-	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-
Medium	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
Round	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	+
UnRound	-	-	-	-	+	-	-	+	+	-	-	-	-	+	-	-	-	-	-	-	+	-	-	-	-
Glotal	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-

twenty two dimensional PFs. The thirty nine dimensional MFCCs extracted in the first stage are entered into the MLN with five layers including three hidden layers after combining a current frame x_t with the other two frames that are three points before and after the current frame (x_{t-3}, x_{t+3}) where the MLN generates twenty two PF values for each input frame of 39x3 features. The three hidden layers comprised of 400, 200 and 100 units, respectively. The MLN is trained using the standard back-propagation algorithm.

V. EXPERIMENTS

A. Setup

For evaluating word recognition performance, word correct rate (WCR), word accuracy (WA) and sentence correct rate (SCR) for D2 data set are evaluated using an HMM-based classifier. The D1 data set is used to design Bangla triphone HMMs with five states, three loops, and left-to-right models. Input features for the classifier are 39 dimensional MFCCs and log values of 22 dimensional PFs. The mixture components are set to 1, 2, 4 and 8.

For evaluating the performance of standard MFCC-based method including the proposed method, we have designed the following experiments:

(a) MFCC:dim-39 [Baseline]

(b) PF:dim-22 [Proposed]

In our experiments the range of output is from 0 to 1, where the non-linear function is a sigmoid, (1/(1+exp(-x))) for the hidden and output layers of MLN. For evaluating PF correct rate we have considered 0.20 as threshold to obtain better segmentation. Here, 0.20 is considered as threshold by observing the experimental results.

B. Result Analysis and Discussion

Segmentation for silence, short silence, stop, nasal, bilabial, fricative, liquid, lenis, vowel, front, central, back, unvoiced, long, short, dipthong, high, low, medium, round, unround and glottal PFs are depicted in Figures 3 and 4 for ideal and real cases for utterance, /prothom/. From both the figures, it is observed that segments of nasal, liquid, vowel and front are more precise (follows ideal line) in Figure 3, and unvoiced, long, dipthong, high, low, medium, unround and glottal exhibit better segments with respect to ideal segmentation in Figure 4. Again, Figure 5 shows correct rates for each of the PFs using the test utterances in D2 data set, where PF correct rates for the corresponding PFs are 97.83%, 52.88%, 75.15%, 75.88%, 64.30%, 84.68%, 49.20%, 84.67%%, 95.72%, 87.83%, 88.22%, 78.42%, 93.79%, 87.49%, 86.65%, 82.97%, 77.82%, 70.75%, 92.62%, 86.15%, 89.00%, and 100.00%, respectively.

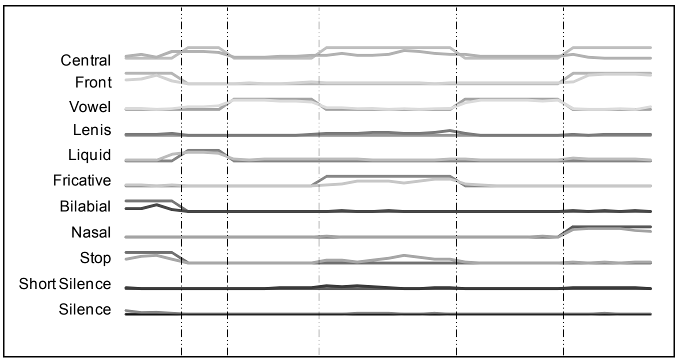

Fig. 3. Segmentation for silence, short silence, stop, nasal, bilabial, fricative, liquid, lenis, vowel, front and central PFs using the utterance /prothom/.

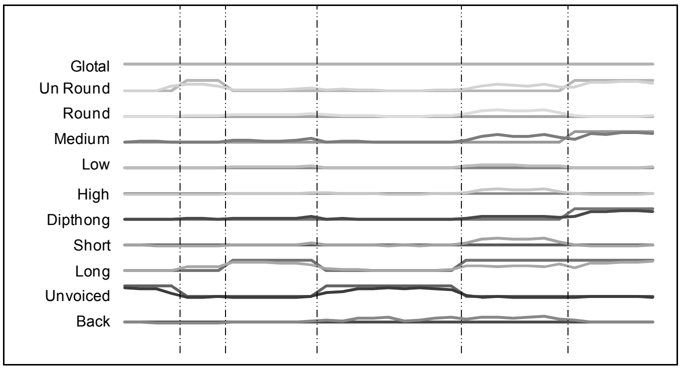

Fig. 4. Segmentation for back, unvoiced, long, short, dipthong, high, low, medium, round, unround and glottal PFs using the utterance /prothom/.

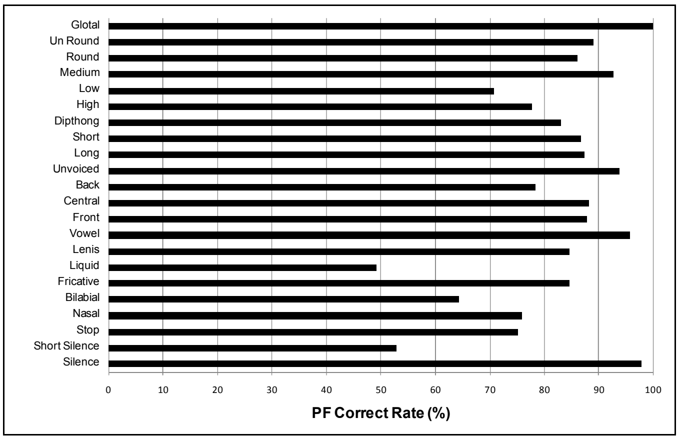

Fig. 5. Correct rates (%) for silence, short silence, stop, nasal, bilabial, fricative, liquid, lenis, vowel, front, central, back, unvoiced, long, short, dipthong, high, low, medium, round, unround and glottal PFs using the test utterances in D2 data set.

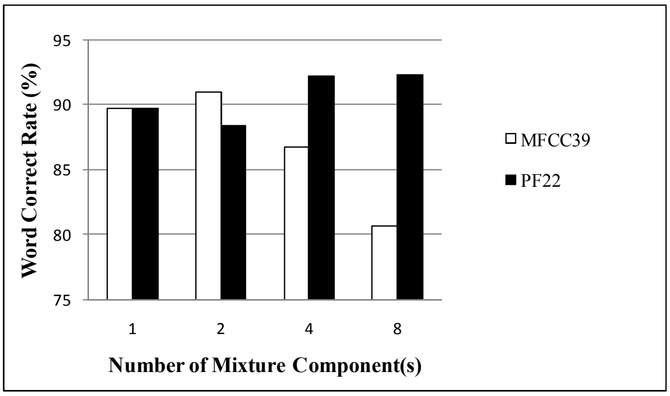

Fig. 6. Word Correct Rates for MFCCs and Proposed Method.

Figure 6 shows the comparison of word correct rates among all the investigated methods, standard MFCC-based method and proposed method. Among all the mixture components except two, the proposed method shows higher correctness in comparison with baseline. It is observed from the figure that the proposed method exhibits its best performance (92.25%) at mixture component eight. Besides, the mixture components, four and eight in the proposed method exhibit almost the same performance. Therefore, further investigation for higher correctness in higher mixture component is not required.

Word accuracies for the different investigated mixture components in standard MFCC-based and proposed methods are depicted in Figure 7. In mixture components one, two, four and eight, the proposed method provides 89.45%, 88.02%, 91.43% and 91.64% accuracies respectively, whereas 89.03%, 90.33%, 86.17% and 80.43% are observed in baseline method for the corresponding mixture components respectively.

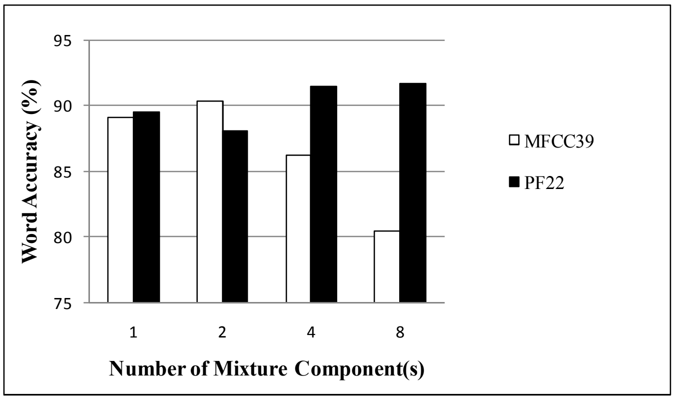

Fig. 7. Word Accuracies for MFCCs and Proposed Method.

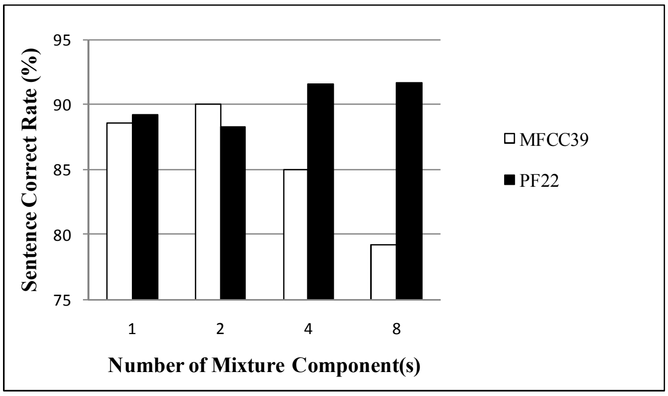

Fig. 8. Sentence Correct Rates for MFCCs and Proposed Method.

Sentence correct rate which is shown in Figure 8 gives an idea about the performance of ASR systems investigated. For the experimented mixture components, there are 89.20%, 88.20%, 91.50% and 91.60% SCRs are found in the proposed method respectively, while baseline system generates 88.60%, 90.00%, 85.00% and 79.20% for the same experimental conditions.

Table III exhibits word recognition performance with respect to correctly recognized words (H), deletion (D), substitution (S) and insertion (I), respectively for the experimented mixture components in both the investigated ASR systems using the 3290 input words. For proposed and baseline methods, the H, D, S and I are 3035, 52, 203 and 20 respectively; and 2654, 202, 434 and 8, respectively for the investigated mixture component eight. Here, proposed method inserted more words than baseline. On the other hand, sentence recognition information for the investigated mixture components is provided in Table IV using 1000 input spoken sentences.

TABLE III. WORD INFORMATION FOR INVESTIGATED METHODS WHERE H, D, S AND I REPRESENT CORRECT WORDS, DELETION, SUBSTITUTION AND INSERTION OUT OF 3290 RESPECTIVELY

		Mixture Components			
		Mix 1	Mix 2	Mix 4	Mix 8
MFCC 39	H	2950	2992	2851	2654
	D	91	75	114	202
	S	249	223	325	434
	I	21	20	16	8
AF 22	H	2952	2908	3033	3035
	D	77	106	58	52
	S	261	276	199	203
	I	9	12	25	20

TABLE IV. SENTENCE INFORMATION FOR INVESTIGATED METHODS WHERE H, AND S REPRESENTS CORRECTLY AND INCORRECTLY RECOGNIZED SENTENCES RESPECTIVELY OUT OF 1000

		Mixture Components			
		Mix 1	Mix 2	Mix 4	Mix 8
MFCC 39	H	886	900	850	792
	S	1144	100	150	208
PF 22	H	892	882	915	916
	S	108	118	85	84

VI. CONCLUSION

This paper has constructed a phonetic feature table for Bangla automatic speech recognition. In the first part of the research twenty two phonetic features are considered for Bangla spoken language and the second part of the research designs an ASR system using the PFs considered here. The following conclusions are given:

(i) Segmentation for each of the PFs follows ideal boundaries for an input spoken sentence.

(ii) Correct rates for most of the PFs are above 80%.

(iii) Word correct rate, word accuracy and sentence correct rate for the proposed method using all the investigated mixture components except two are better in comparison with the standard MFCC-based method.

In near future, the author would like to evaluate PFs using recurrent neural network (RNN), which accommodates longer context window in its architecture. Moreover, the authors evaluate the experiments for gender independent environments.

REFERENCES

[1] http://en.wikipedia.org/wiki/List_of_languages_by_total_speakers, Last accessed April 11, 2009.

[2] S. P. Kishore, A. W. Black, R. Kumar, and Rajeev Sangal, "Experiments with unit selection speech databases for Indian languages," Carnegie Mellon University.

[3] http://en.wikipedia.org/wiki/Bengali_phonology, Last accessed April 11, 2009.

[4] 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.

[5] 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.

[6] 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.

[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, 2002.

[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 bangle 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] K. Kirchhoff, et. al, "Combining acoustic and articulatory feature information for robust speech recognition," Speech Commun.,vol.37, pp.303-319, 2002.

[13] C. Masica, The Indo-Aryan Languages, Cambridge University Press, 1991.

[14] Ghulam Muhammad, Yousef A. Alotaibi and Mohammad Nurul Huda, “Automatic Speech Recognition for Bangla Digits,” ICCIT’09, Dhaka, Bangladesh, December 2009.

[15] Daily Prothom Alo. Online: www.prothom-alo.com

[16] S. King and P. Taylor, "Detection of Phonological Features in Continuous Speech using Neural Networks," Computer Speech and Language 14(4), pp. 333-345, 2000.

[17] S. King, et. al, “Speech recognition via phonetically features syllables,” Proc ICSLP’98, Sydney, Australia, 1998.

[18] T. Fukuda and T. Nitta, “Noise-robust ASR by Using Distinctive Phonetic Features Approximated with Logarithmic Normal Distribution of HMM,” Proc. Eurospeech 2003, Vol.III, pp.2185-2188, Sep. 2003.