

Phonetic Bengali Input Method for Computer and Mobile Devices

*Khan Md. Anwarus Salam*¹ *Setsuo Yamada*² *Tetsuro Nishino*¹

(1) The University of Electro-Communications, Tokyo, Japan.

(2) NTT Corporation, Tokyo, Japan.

salamkhan@uec.ac.jp, yamada.setsuo@lab.ntt.co.jp,
nishino@uec.ac.jp

ABSTRACT

Current mobile devices do not support Bangla (or Bengali) Input method. Due to this many Bangla language speakers have to write Bangla in mobile phone using English alphabets. During this time they used to write English foreign words using English spelling. This tendency also exists when writing in computer using phonetically input methods, which cause many typing mistakes. In this scenario, computer transliteration input method need to correct the foreign words written using English spelling. In this paper, we proposed a transliteration input method for Bangla language. For English foreign words, the system used International-Phonetic-Alphabet(IPA)-based transliteration method for Bangla language. Our proposed approach improved the quality of Bangla transliteration input method by 14 points.

KEYWORDS : Foreign Words, Bangla Transliteration, Bangla Input Method

1 Introduction

Bengali or Bangla is the official language of Bangladesh. Currently Bangladesh has 72.963 million mobile phone users. It is important to have Bengali input method for this huge number of Bengali language speakers. Although Bangladesh government declared standard keyboard layout for both computer and mobile device, currently there is no national standard for transliteration using English alphabets. Due to this there are many ambiguities in mapping 50 Bengali letters using 26 English letters. Different people have different assumptions on phonetic input system for Bengali language using English letters. These ambiguities effect the human communication, using mobile or emails, where people had no other choice except using English letters to write Bengali messages.

In this kind of scenario most people used to write English foreign words using English spelling. These ambiguities effect the human communication using SMS or email. In this kind of scenario most people used to write English foreign words using English spelling. To understand this kind of message needs a sophisticated phonetic input method for mobile devices. Bengali also needs a standard transliteration mechanism considering these issues. Such a transliteration scheme should be simple rule-base to minimize the computational resources.

In this paper, we propose a phonetic Bengali input method for computer and mobile devices. Our approach is a pattern-based transliteration mechanism. For handling foreign words, we used International-Phonetic-Alphabet(IPA)-based transliteration. Proposed system first tries to find if the word exists in English IPA diction-ary. If the word is not available in the English dictionary it uses the mechanism as proposed with Akkhor Bangla Software and a Bengali lexicon database to transliterate meaningful words. Our proposed approach improved the quality of Bangla transliteration input method by 14 points.

2 Related Works

There were several attempts in building Bengali transliteration systems. The first available free transliteration system from Bangladesh was Akkhor Bangla Software¹. Akkhor was first released on 2003 which became very popular among computer users.

Zaman et. al. (2006) presented a phonetics based transliteration system for English to Bangla which produces intermediate code strings that facilitate matching pronunciations of input and desired output. They have used table-driven direct mapping techniques between the English alphabet and the Bangla alphabet, and a phonetic lexicon-enabled mapping. However they did not consider about transliterating foreign words. Most of the foreign words cannot be mapped using their mechanism.

Rahman et. al. (2007) compared different phonetic input methods for Bengali. Following Akkhor, many other software started offering Bengali transliteration. But none of these works considered about transliterating foreign words using IPA based approach.

Amitava Das et. Al. (2010) proposed a comprehensive transliteration method for Indic languages including Bengali. They also reported IPA based approach improved the performance for Bengali language.

3 Transliteration Architecture

In this paper, we propose a transliteration input method for Bangla language with special handling of foreign words. For transliterating we considered only English foreign words and simple rule-based mechanism. For foreign words, we used International-Phonetic-Alphabet (IPA) based transliteration.

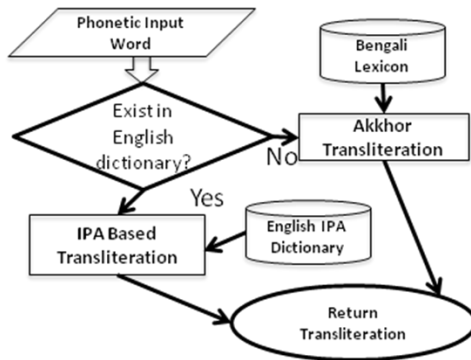


FIGURE 1. Proposed Transliteration Architecture

¹<http://www.akkhorbangla.com/>

Figure 1 shows the Bengali transliteration process in a flow chart. Proposed system first tries to find if the word exists in English IPA dictionary to detect foreign words. For these foreign words, it uses IPA based transliteration. If the word is not available in the English dictionary, it uses Akkhor transliteration mechanism. As Bengali language accepts many English foreign words, transliterating the English word into Bengali alphabet makes that a Bengali foreign word. In our assumption, when writing Bengali message people write English foreign words using English spelling. To identify such input words, the system first checks for a word for foreign (English) origin by looking up at the English IPA dictionary. If the word is not available in the English IPA dictionary, the system uses the transliteration mechanism as proposed with Akkhor Bangla Software and a Bengali lexicon database to transliterate Bengali words.

4 IPA Based Transliteration

From English IPA dictionary the system can obtain the English words pronunciations in IPA format. Output for this step is the Bengali word transliterated from the IPA of the English word. In this step, we use following English-Bengali Transliteration map to transliterate the IPA into Bengali alphabet.

Mouth narrower vertically	[i:] ই / িঃ sleep /sli:p/	[ɪ] ই / িঃ slip /sɪp/	[ʊ] উ / ুঃ book /bʊk/	[u:] উ / ুঃ boot /bu:t/
	[e] এ / েঃ ten /ten/	[ə] আ / া after /a:ftə/	[ɜ:] আ / া bird /bɜ:d/	[ɔ:] ঝ bored /bɔ:d/
Mouth wider vertically	[æ] এষা / েয়া cat /kæt/	[ʌ] আ / া cup /kʌp/	[ɑ:] আ / া car /cɑ:r/	[ɒ] অ hot /hɒt/

Table 1. English-Bengali IPA chart for vowels

[ɪə] ইয়া / িয়া beer /bɪər/	[eɪ] এই / েই say /seɪ/	
[ʊə] উয়া / ুয়া fewer /fjuər/	[ɔɪ] অয়/য় boy /bɔɪ/	[əʊ] ও / ো no /nəʊ/
eə ঈয়া / ীয়া bear /beər/	[aɪ] াই / আই high /haɪ/	[aʊ] আউ / াউ cow /kaʊ/

Table 2. English-Bengali IPA chart for diphthongs

[p] প pan /pæn/	[b] ব ban /bæn/	[t] ট tan /tæn/	[d] ড day /deɪ/	[tʃ] চ chat /tʃæt/	[dʒ] জ judge /dʒʌdʒ/	[k] ক key /ki:/	[g] গ get /get/
[f] ফ fan /fæn/	[v] ভ van /væn/	[θ] থ thin /θɪn/	[ð] দ than /ðæn/	[s] স sip /sɪp/	[z] জ zip /zɪp/	[ʃ] শ ship /ʃɪp/	[ʒ] স vision /vɪʒən/
[m] ম might /maɪt/	[n] ন night /naɪt/	[ŋ] ঙ thing /θɪŋ/	[h] হ height /haɪt/	[l] ল light /laɪt/	[r] র right /raɪt/	[w] ঝ white /hwaɪt/	[j] য়ে yes /jes/

Table 3. English-Bengali IPA chart for consonants

Table 1, 2 and 3 shows our proposed English-Bengali IPA chart for vowels, diphthongs and consonants. Using rule-base we transliterate the English IPA into Bangla alphabets. The above IPA charts leaves out many IPA as we are considering about translating from English only. To translate from other language such as Japanese to Bangla we need to create Japanese specific IPA transliteration chart. Using the above English-Bangla IPA chart we produced transliteration from the English IPA dictionary. For examples: pan(pæn): প্যান; ban(bæn): ব্যান; might(maɪt): মাইট .

5 Akkhor Transliteration

বাংলা	অ	আ	ই	ঈ	উ	ঊ	ঋ	এ	ঐ	ও	ঔ
English	A	a/aa/	i/i	I/ee/	u/	U/ʊ	ri/	e/	oi/	o/	ou
হ		a		I	u		ri	e	oi	o	ou
বাংলা	ক	খ	গ	ঘ	ঙ	চ	ছ	জ	ঝ	ঞ	
English	k	kh	g	gh	Ng	ch	Ch	j	jh	Y	
বাংলা	ত	থ	দ	ধ	ন	ট	ঠ	ড	ঢ	ণ	
English	t	th	d	dh	n	T	Th	D	Dh	N	
বাংলা	প	ফ	ব	ভ	ম	য	র	ল	শ	ষ	
English	p	f/ph	b	bh/v	m	z	r	l	sh	S	
বাংলা	গ	ক্ষ	হ	ড়	ঢ়	য়	ৎ	ঃ	ড়		
English	S	k-S	h	R	rh	y	ng	:	~		
বাংলা	১	২	৩	৪	৫	৬	৭	৮	৯	০	
English	1	2	3	4	5	6	7	8	9	1	
বাংলা	কা	কে	কি	কু	কো	ক্র	ক্রে	ক্রি	ক্রু	ক্রু	
English	ka	ke	ki	ku	kO	kro	kre	kre	kru	krU	
বাংলা	কী	চী	মী	কু	মু	বু	ণু	নু	ক্য	ব্য	
English	kI	chI	mI	kU	mU	bU	NU	nU	k-z	b-z	

Table 4. Akkhor phonetic mapping for Bengali alphabets

Akkhor Bangla Software first implemented Bangla phonetic input method for computers. As a result this phonetic mapping become very popular among Bangladeshi computer users. However initially Akkhor did not consider about using Bangla Lexicon database. In this research we used Akkhor phonetic mapping with Bangla Lexicon database. Table 4 shows the phonetic mapping for Bengali alphabets.

Because of the ambiguities in mapping 50 Bengali letters using 26 English letters, any fixed Bengali Phonetic mapping is debatable. As a result different people have different assumptions on phonetic input system for Bengali language using English letters. To overcome this problem we used Bengali Lexicon which includes the IPA for each Bengali words. The system produces the Bengali transliteration by ranking the words using IPA string edit distance.

5.1 IPA String Edit Distance

IPA string edit distance assign a score to a sequence of acoustic observations X for a hypothesized string of phones P, but rely on different distributions for P (Droppo et. al. 2010). From these IPA string edit distance scores, we choose the highest scored transliteration and show the word candidates in descending order.

6 Experiment

Based on the above method we implemented the phonetic Bengali input method for computer and mobile devices. We evaluated the produced transliteration for our test-set including 200 words. This test-set mainly covered the foreign words from different domains. We have taken these words by averaging across multiple users' behaviour.

Table 5 shows our evaluation result, where we compared our proposed phonetic input method with the Akkhor Baseline. We manually checked the produced transliteration quality and assigned each test-set entry as wrong (W), correct (C) or neutral (N). Neutral refers to such words which can be correct in different context and it depends on the user intention. For example, ban can refer to both Bengali word বান and বান. Therefore, we assign such ambiguous words as neutral or N. In our experiment, our proposed approach improved the quality of Bangla transliteration input method by 14 points. The proposed phonetic input method could correct transliterate 68% of the test-set words.

Transliteration Quality	Akkhor Baseline	Proposed Phonetic Input Method
Correct (C)	54%	68%
Wrong (W)	39%	22%
Neutral (N)	7%	10%
Total	100%	100%

Table 5. Comparison between Akkhor transliteration and proposed phonetic input method

For example, Table 6 shows 3 sample transliterations produced by Akkhor Baseline and our proposed Phonetic Input Method. As we can see for first two examples Akkhor produced wrong (W) transliteration and our proposed phonetic input method produced correct (C) transliteration.

In third case, both are marked as neutral (N), because both words is correct in different context and it depends on the user intention.

#	Input	Akkhor Baseline	Proposed Phonetic Input Method
1.	School	স্চুল (W)	স্কুল (C)
2.	University	উনিভেরসিতয় (W)	ইউনিভার্সিটি (C)
3.	ban	বান (N)	ব্যান (N)

Table 6. Sample Transliterations produced by Akkhor baseline and proposed method

7 Conclusion

We proposed a phonetic Bengali input method which is useful for computer and mobile devices by transliteration mechanism. Our proposed solution is effective especially for mobile devices due to low computational resources. For English foreign words, the system used International-Phonetic-Alphabet(IPA)-based transliteration method for Bangla language. Our proposed approach improved the quality of Bangla transliteration input method by 14 points. In future, this method can be expanded to consider about handling foreign words in other Indian languages.

References

- Naushad UzZaman, Arnab Zaheen and Mumit Khan. A comprehensive roman (English)-to-Bengali transliteration scheme. Proceedings of International Conference on Computer Processing on Bangla. Dhaka, Bangladesh, 2006.
- M. Ziaur Rahman and Foyzur Rahman. On the Context of Popular Input Methods and Analysis of Phonetic Schemes for Bengali Users., In Proceedings of the 10th International Conference on Computer & Information Technology, Jahangirnagar, Bangladesh, 2007.
- Jasha Droppo and Alex Acero. Context Dependent Phonetic String Edit Distance for Automatic Speech Recognition, IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), Dallas, TX. 2010.
- Amitava Das, Tanik Saikh, Tapabrata Mondal, Asif Ekbal, Sivaji Bandyopadhyay. English to Indian Languages Machine Transliteration System at NEWS 2010. Proceedings of the 2010 Named Entities Workshop, ACL 2010.