



Development of Phonetic Transcription System for Bangla to IPA

Abdullah Al Zubaer*¹, Iftikharul Islam*¹, Md. Manik Ahmed*², Rohani Amrin*³, Md. Mehedi Hasan
Naim*⁴,

*¹Lecturer, Computer Science and Engineering Department, Rabindra Maitree University, Kushtia,
Bangladesh, zubaerranacseiu@gmail.com

*¹Computer Science and Engineering Department, Islamic University, Kushtia, Bangladesh,
ifti_khar90@yahoo.com

*²Lecturer, Information and Communication Technology Department, Rabindra Maitree University,
Kushtia, Bangladesh, kiron.pust.ice@gmail.com

*³Lecturer, Information and Communication Technology Department, Rabindra Maitree University,
Kushtia, Bangladesh, rohani.amrin20@gmail.com

*⁴Lecturer, Computer Science and Engineering Department, Rabindra Maitree University, Kushtia,
Bangladesh, naim.cu11@gamil.com

Abstract: This project is dedicated to evolve a phonetic transcription system for Bangla to IPA. In producing a bona fide phonetic code the complex and often inconsistent rules of Bangla Word present note worthy challenges. For simplifying the problems a number of rules are collected from diverse sources. On the other hand, for producing proper phonetic code for Bangla language software has been developed using Java. To verify the system BDNC01 purpose was used. The noteworthy matter is that the system successful transcript 87% of the text which is taken for evaluation. On the contrary, it is expected that the effort can work out diverse problems in Bangla processing. The consequence of it can be used in multiple issues converting text to speech, Bangla dictionary, pronunciation, verifier, cross lingual application etc. Not only this but also in future this project will be used for proposing a phonetic encoder for Bangla. In fact, it will be used for taking into account the various context-sensitive rules. Moreover, it is wished that the effort will be able to solve various problems in Bangla processing.

I. Introduction

Between international phonetic Association and International Phonetic Alphabet the

acronym IPA can cite one. The International Phonetic Alphabet is the association's alphabet which is usually mentioned as the chart of the IPA [1]. This was published

periodically in chart form. Refining and updating of the chart have been taken place over the years so that the chart can accommodate the need to represent symbolically the sounder of the world's languages [2]. The IPA is used by lexicographers, foreign language students and teachers, linguists, speech-language pathologists, singers, actors, constructed language creators and translators [3][4]. The motives of the association and of its alphabet are to draw up and spread a uniform standard for phonetic writing which is commonly known as phonetic transcription. On the other hand the main object of the IPA design is to represent only those qualities of speech which are the part of oral language such as phones, phonemes, intonation and the separation of words and syllables [3]. Not only this but also for representing additional qualities of speech, such as tooth gnashing, lisping and sounder made with a cleft lip and cleft palate an extended set of symbols called the extensions to the international phonetic Alphabet may be used [5]. Moreover, IPA contains 107 letters, 52 diacritics and four prosodic marks that are are shown in the current IPA chart, posted below in this article and the website of the IPA [6][7].

History of IPA

In the late 19th century, with the evolution of the association and its announcement of erecting a phonetic system the story of the International Phonetic Alphabet (IPA) and the International phonetic Association (IPA) began. However, the phonetic system was used for interpreting the sounder of spoken languages. One motive of the International Phonetic Alphabet (IPA) was to deliver a

unique symbol for each particular sound in a language. That is why every sound or phoneme could be differentiated from one word to another [8][9]. The IPA (International phonetic Association) is not only the prominent organization but also the oldest representative organization for phoneticians. Previous statistics shows us that 2016 is marked as the 130th anniversary for the founding of the IPA, and 2018 is marked as the 130th anniversary for first launching of the International Phonetic Alphabet and the generation of the principles. The focus of the IPA is to encourage the scientific study of phonetics and the diverse pragmatic applications of that science. In furthermore of this object, the phonetic representation of all languages the International Phonetic Alphabet is delivering with a notational standard. In 2015, the latest version of the IPA Alphabet was published and IPA charts are reused annually [10].

II. Literature Review

There were several attempts in building Bengali transliteration systems. The first available free transliteration system from Bangladesh was Akkhor Bangla Software. 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 [11].

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 [12].

Joyanta Basu ; Tulika Basu ; Mridusmita Mitra ; Shyamal Kr. Das Mandal (2009) presents Grapheme to Phoneme (G2P) conversion for Bangla based on orthographic rules. They stated that In Bangla G2P conversion sometimes depends not only on orthographic information but also on parts of speech (POS) information and semantics [13].

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 [14].

III. Main Objective

The object of this project is to develop a phonetic transcription System for Bangla to IPA (International Phonetic Alphabet). The aim of the Phonetic transcription from Bangla to International Phonetic Alphabet (IPA) is to provide a unique symbol for each distinctive sound in Bangla language that is, every sound, or phoneme, that serves to distinguish one word from another.

IV. Method and Materials

4.1 Corpus and Vocabulary

Corpus is an important resource for any language. We took Bangla corpus BDNC01.

BDNC01 Corpus: BDNC01 Corpus is a text corpus collected from web editions of several influential Bangla newspapers during 2005-2011. BdNC01 contains a large amount of Bangla text including more than 11 million word tokens [15].

4.2 IPA Equivalentents of Bangla Letters

In Bangla language every letter has its own sounds. There are two types of transcription [16].

I. Narrow transcription: captures as many aspects of a specific pronunciation as possible and ignores as few details as possible. Using the diacritics provided by the IPA, it's possible to make very subtle distinctions between sounds.

II. Broad transcription (or phonemic transcription): ignores as many details as possible, capturing only enough aspects of a pronunciation to show how that word differs from other words in the language.

The IPA corresponds to the Bangla letters, from the book 'Sadharan Bhasavijnan o Bangla Bhasa' by Dr. Rameswar Shaw, which are given below [16]:

Vowel letters	IPA	Vowel letters	IPA
অ	/ɔ/	ঋ	/ri/
আ	/a/	এ	/e/
ই	/i/	ঐ	/oi/
ঈ	/i/	ও	/o/
ঊ	/u/	ঔ	/ou/
ঋ	/u/	এ /অ্যা/	/æ/, /ɛ/

Consonant letters	IPA	Consonant letters	IPA
ক	/k/	প	/p/
খ	/k ^h /	ফ	/p ^h /
গ	/g/	ব	/b/
ঘ	/g ^h /	ভ	/b ^h /
ঙ	/ŋ/	ম	/m/
চ	/c/	য	/j/
ছ	/c ^h /	র	/r/
জ	/ʃ/	ল	/l/
ঝ	/ʃ ^h /	ব	/b/
ঞ	/n/	শ	/ʃ/
ট	/t/	ষ	/ʃ/
ঠ	/t ^h /	স	/s/
ড	/d/	হ	/h/
ঢ	/d ^h /	ং	/ŋ/
ণ	/n/	ঃ	/h/
ত	/t/	ঁ	/~/
থ	/t ^h /	ড়	/ɽ/
দ	/d/	ত	/ɽ ^h /
ধ	/d ^h /	শ্	/j/ /ë/
ন	/n/		

Mohammad Daniul Huq represents the IPA marks using for Bangla in the book 'Vashabiggner Kotha'. He stated that 'No matter how spellings are written in Bangla, 38 basic sounds are used in the language of speech.' these IPA marks are given below [17]:

স্বরধ্বনি :

১) অ = ɔ

২) আ = a

৩) ই = i

(ঐ=ই+ই= i+i)

৪) উ = u

(উ=উ+উ=u+u)

(ঋ its pronunciation is 'রী')

৫) এ = e (ঐ = অ+ই= ɔ+i)

৬) ও = o (ঔ=ও+উ= o+u)

৭) অ্যা = æ

৮) ও' = ɒ

ব্যঞ্জনধ্বনি :

৯) ক্ = k

১০) খ্ = k^h

১১) গ্ = g

১২) ঘ্ = g^h/g^h

১৩) ঙ্ (এবং 'ং') = ŋ

১৪) চ্ = c

১৫) ছ্ = c^h

১৬) জ্ = ʃ

১৭) ঝ্ = ʒ^h/ʒ^h

(ঞ pronounce as = ইঁ+অঁ)

১৮) ট্ = t

১৯) ঠ্ = t^h

২০) ড্ = d

২১) ঢ্ = d^h/d^h

(ণ্ its pronunciation is not exist in Bangla)

২২) ত্ = t̪

২৩) থ্ = t̪^h

২৪) দ্ = d̪

২৫) ধ্ = d̪^h/d̪^h

২৬) ন্/ন্ = n

২৭) প্ = p

২৮) ফ্ = p^h

২৯) ব্ = b

৩০) ভ্ = b^h/b^h

৩১) ম্ = m

(Now the pronunciation of য is like জ)

৩২) র্ = r

৩৩) ল্ = l

(The pronunciation of অন্তস্থ-ব does not exists in Bangla)

৩৪) শ্ এবং ষ্ = ʃ/ʒ

৩৫) স্ = s

৩৬) হ্ = h (এবং বসির্গ [ঃ])

৩৭) ড়্ = r̥

৩৮) ঢ়্ = r̥^h/r̥^h

4.3 Letter Transcription

The Transcription from Bangla letter to IPA in this project we use the following:

	অ	আ	ই	ঈ	উ	ঊ	ঋ	এ	ঐ	ও	ঔ	
		ɒ	ɪ	ɪ	u	u	ɹ	e	ɔ	o	o	
	/ɔ/	/a/	/i/	/i/	/u/	/u/	/ri/	/e/	/oi/	/o/	/ou/	
ক	খ	গ	ঘ	ঙ	চ	ছ	জ	ঝ	ঞ	ট	ঠ	ড
/k/	/kʰ/	/g/	/gʰ/	/ŋ/	/tʃ/	/tʃʰ/	/dʒ/	/dʒʰ/	/ɛ̃/	/t/	/tʰ/	/d/
ত	ণ	ত	থ	দ	ধ	ন	প	ফ	ব	ভ	ম	য
/dʰ/	/n/	/t/	/tʰ/	/d/	/dʰ/	/n/	/p/	/pʰ/	/b/	/bʰ/	/m/	/dʒ/
		র	ল	শ	ষ	স	হ	ড়	ঢ়	য়		
		/r/	/l/	/ʃ/	/ʃ/	/ʃ/	/h/	/ɽ/	/ɽʰ/	/ẽ/		
				ং	ঃ							
				/t̃/	/ŋ/	/h/	/ʃ/					

4.4 Word Phonetic Transcription

In any word if it is not a consonant conjunct then every single letter of that word are pronounced individually, where consonant is typically pronounce as the pronunciation of the letter plus the inherent vowel অ ô /ɔ/ if it is not occurs in the last of the words. All single letters in a word is converted by Letter Transcription [21][22].

উপর = উ+প+র = /upɔr/

তখন = ত+খ+ন = /t̃ ɔkʰɔn/

সুবধি = স+ব+ধি+া = /ʃubid̃ʰa /

Here is a exception when the latter is ত /t̃ / appears at the end of the word, then typically it also pronounce with inherent vowel অ /ɔ/.

মোহিত = ম+ো+হ+ি+ত = /mohit̃ ɔ/

মূলত = ম+ূ+ল+ত = /mulɔt̃ ɔ/

But there are also some exceptional words.

কশাঘাত = ক+শ+া+ঘ+া+ত = /kɔʃagʰat̃ /

4.5 Consonants Conjuncts Transcription

Consonant clusters can be orthographically represented as a typographic ligature called

a consonant conjunct (যুক্তাক্ষর/যুক্তবর্ণ juktakkhôr/juktôbôrnô or more specifically যুক্তব্যঞ্জন). Typically, the first consonant in the conjunct is shown above and/or to the left of the following consonants. Many consonants appear in an abbreviated or compressed form when serving as part of a conjunct. Others simply take exceptional forms in conjuncts, bearing little or no resemblance to the base character [20].

There are some rules defined by linguists for consonant conjunct. The rules are given below [21][22]:

1. ক্ষ = ক্ + ষ :
 - a) If it appears at the first of the word then its pronunciation খ /kʰ/ ক্ষমা /kʰɔma/ .
 - b) If it appears at the middle or end of the word its pronunciation ক্ খ /kkʰ/ ভিক্ষুক /bʰikkʰuk/ .
2. হ্ম = হ্ + ম : Its pronunciation is ম্ভ /mbʰ/

Example : ব্রাহ্মণ
/brambʰɔn/.

such as পূর্বাহ্ণ
/purbannʰɔ/, চহ্ন /tʃinnʰɔ/.

3. জ্ঞ = জ্ + ঞ :

a) If it appears at the first of the word then its pronunciation

গ্ৰ্ণ /gɽ/ জ্ঞান /gɽãn/ .

b) If it appears at the middle or end of the word its pronunciation গ্ৰ্ণ
/ggɽ/ অজ্ঞান
/ɔggɽãn/ .

4. ঞ :

a) If it conjuncts with consonants its pronunciation ন /n/. such as পঞ /pɔntʃɔ/,
ব্জ্ঞ /bɛndʒɔn/.

b) There are some word which doesn't change its original pronunciation such as মঞ /miɛã/ .

5. হ্ন = হ + ল :

a) If it appears at the first of the word then its pronunciation ল /l/ such as হ্লাদ /lhɑd/.

b) If it appears at the middle or end of the word its pronunciation ল্ল /ll/ such as আহ্লাদ /allhɑd/.

6. হ্ণ or হ্ন = হ + ন : Its pronunciation is ন্হ /nnʰ/

7. হ্য় = হ্ + য : It doubles the য /dʒ/ and its pronunciation as জ্য় /dʒdʒʰɔ/ such as উহ্য়
/udʒdʒʰɔ/, ঐতহ্য়
/oitidʒdʒʰɔ/.

8. হ্র = হ্ + র : Its pronunciation is র্হ /rʰ/ such as হ্রদ
/rʰɔd/, হ্রতি /rʰitɔ/.

9. য-ফলা :

a) If it conjuncts with first of the word its pronunciation is অ্যা /ɛ/ such as ব্যক্ত
/bɛktɔ/, ব্যতীত /bɛtitɔ/ .

b) If it conjuncts with Consonants Conjuncts then it remain silent, such as সন্ধ্যা /ʃɔndʰa/, সন্ধ্যাসী
/ʃɔnnɑʃi/.

c) If it appear in the middle or last of the words, then it doubles the previous consonant, such as ধন্য
/dʰɔnnɔ/, কল্যাণ /kollan/.

10. য-ফলা :

a) If it conjuncts with first of the word or conjuncts with Consonants Conjuncts then it remain silent, such as স্মরণ /ʃmɽn/.

b) If it appear in the middle or last of the words, then it doubles the previous

consonant, such as আত্ম /attɔ/, বস্মরণ /bissɔrɔn/.

c) But in the middle or last of the words if it conjunct with গ, ঙ, ট, ণ, ন, ম, ল then it don't change its pronunciation, such as যুগ্ম /dzugmɔ/, বাঙময় /banɔmɔɛ/, কুট্মল /kutɔml/, মৃগ্ময় /mrinmɔɛ/, উন্মাদ /unmad/, সম্মান /ɔmmɔn/, বাল্মীকি /balmiki/.

d) There are some exception word which doesn't change its pronunciation such as কুষ্মাণ্ড /kufmandɔ/, স্মৃতি /smɪtɪ/, শূচস্মৃতি /ʃutʃismɪtɪ/, সুস্মৃতি /ʃusmɪtɪ/, কাশ্মীর /kɔʃmir/, আয়ুষ্মতী /æyusmɔtɪ/.

11. ল-ফলা :

a) If it conjuncts with first of the word or conjuncts with Consonants Conjuncts then its pronunciation doesn't change such as ক্লান্ত /klantɔ/.

b) If it appear in the middle or last of the words, then it doubles the previous consonant, such as অশ্লীল /ɔʃʃlil/.

12. র-ফলা :

a) If it conjuncts with first of the word or conjuncts with Consonants Conjuncts then its pronunciation doesn't change প্রকাশ /prɔkɔʃ/, কেন্দ্র /kendrɔ/.

b) If it appear in the middle or last of the words, then it doubles the previous consonant, such as নম্র /nɔmmrɔ/, বদ্রিহ /bidrɔhɔ/.

13. ব-ফলা :

a) If it conjuncts with first of the word or conjuncts with Consonants Conjuncts then it remain silent, such as স্বদশে /ʃɔdɔʃ/.

b) If it conjunct with উদ্, গ, ব, ম then it doesn't change its pronunciation such as উদ্বগে /udɔbeg/, উদ্বোধন /udɔbɔdʰɔn/, দগ্বিজয় /digɔbidʒɔɛ/, তিব্বত /tɪbbɔt/, সম্বর্ধনা /ɔmbɔrdʰɔna/.

c) If it appear in the middle or last of the words, then it doubles the previous consonant, such as বশ্বাস /bissɔʃɔʃ/, রাজত্ব /radʒɔttɔ/.

14. ঃ (বস্বির্গ) :

a) If it appears at the end of the word then its pronunciation হ্ /h/, such as উঃ /uh/.

b) If it appear in the middle or last of the words, then it doubles the previous consonant, such as দুঃসময় /d̪uʃʃɔmɔɐ̃/.

c) Sometimes it Silent in spellings like অন্তঃনগর /ɔnt̪ɔɔɔgr̪/.

d) Also used as abbreviation, like ডাঃ = ডাক্তার /dakt̪ar/.

4.6 Digits and Numerals

When a word is a number first it should be converted into its Bangla pronunciation. When writing large numbers with many digits, commas are used as delimiters to group digits, indicating the thousand (হাজার hazar), the hundred thousand or lakh (লাখ lakh or লক্ষ lôkkhô), and the ten million or hundred lakh or crore (কোটি koti) units. In other words, leftwards from the decimal separator, the first grouping consists of

three digits, and the subsequent groupings always consist of two digits [23]. Such as ২০১৮ = দুই হাজার আঠারো /d̪ui hadʒar at̪harɔ/.

4.7 Punctuation Marks

Bengali punctuation Mark's are down stroke, commas, semicolon, colon, quotation mark's etc. Among these punctuation marks except down stroke others have been adopted from western scripts and their usages are similar. In bangla down stroke which is equivalent of a full stop in western scripts has the same usages in both bengali and western scripts.

4.8 Others Languages

For any other language or any other character it will remain same as the Document.

V. System design

5.1 Transcription Architecture

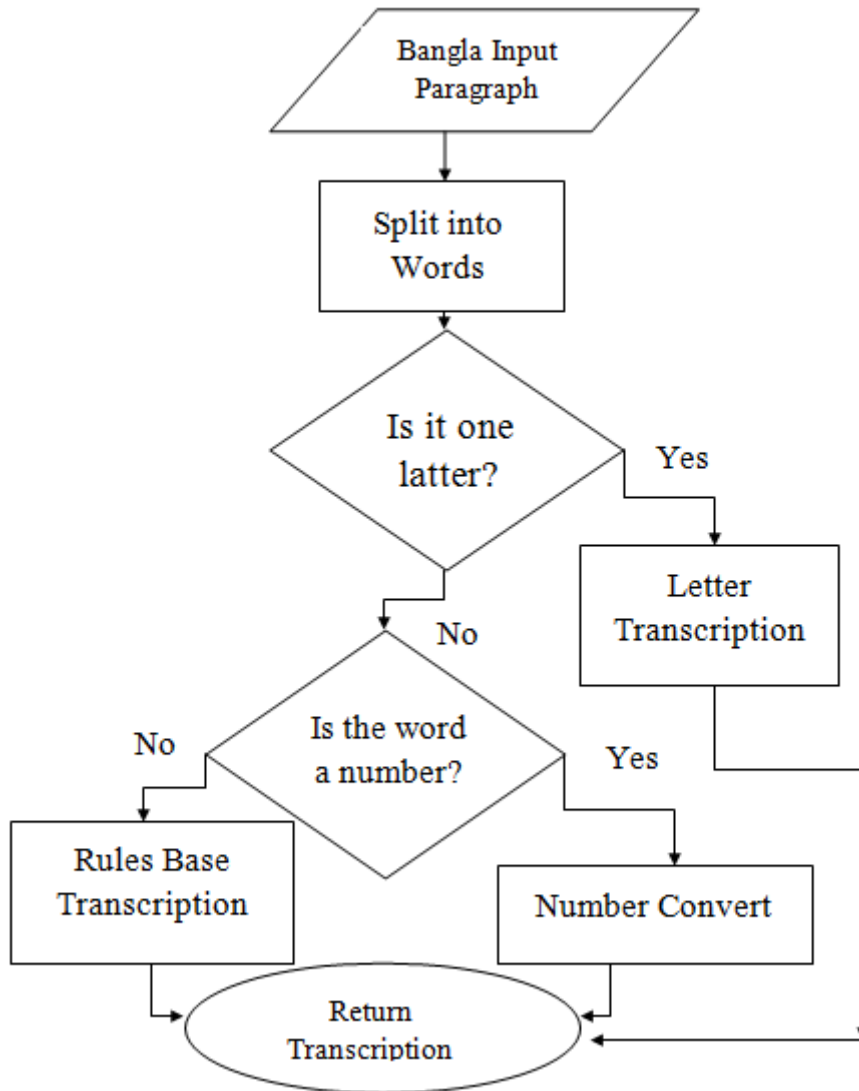


Figure above shows the Bengali Transcription process in a flow chart.

For Phonetic Transcription from Bangla word to IPA we considered only simple rule-based mechanism. Proposed system first take a paragraph then split it into words. Then if the word is only one letter then transcript it from Letter Transcription. Then find if the word is a number then transfer the number into Number Convert function which returns the equivalent IPA form of pronunciation. If it is not a number then it will follow the Rules Base Transcription.

5.2 Number Conversion

We map ০ = শূন্য , ১ = এক up to ৯৯ = নরানব্বই [24][25]. Then find its corresponding IPA, Transcription and put into a TreeMap as

```

ipaNumber.put("০", "ʃunnɔ");
ipaNumber.put("১", "ek");
ipaNumber.put("২", "d̪ui");
ipaNumber.put("৩", "t̪in");

```

...
ipaNumber.put("৯৯", "niranɔbbɔi");
First set a string to null. For a number we cut from the last position.

Number Convert Function

1. First we cut two digit and check as a key into map then find its value and add it to string.
2. Then cut one digit and find value + শত(/ʃɒt/) and add it to string as prefix.
3. Then cut two digit and find value + হাজার (hadʒar) and add it to string as prefix.
4. Then cut two digit and find value + লক্ষ (lɒkkʰɔ) and add it to string as prefix.
5. Then cut two digit and find value + কোটি (koʈi) and add it to string as prefix.
6. The process from 2-5 will continue until digit is zero..

VI. Result and Discussion

6.1 Result & Discussion

This software can't transcript 100% because of we use only rule base transcription. In Bangla there are many words which do not follow any rules. We use only one pronunciation for one letter there are some letter that have different pronunciation such as: অ pronunciation will be /ɔ/ or /o/. এ pronunciation will be /e/ or /ɛ/. স pronunciation will be /s/, /ʃ/ or /ʒ/ but we use only /ʒ/. If consonant letter appears first or middle of the word then the letter typically are pronounced with the inherent vowel অ ô /ɔ/. But in Bangla there are many words that do not follow this rules such as করত /kɔrɒt/ but actual transcription is /kɔrɒtɛ/. If consonant letter appears at the end of the word typically are pronounced without the inherent vowel অ ô /ɔ/. But in Bangla there are many words that do not follow this rules such as অথচ /ɔtʰɔtʃ/ but actual transcription is /ɔtʰɔtʃɔ/. But in a

word if end of the letter is ত then it is pronounced with the inherent vowel অ ô /ɔ/. But in Bangla there are many words that does not follow this rules such as ভাত /bʰatɔ/ but actual transcription is /bʰatɛ/. মফলা: There are some exception word which doesn't change its pronunciation such as কুম্ভাণ্ড /kuʃʃandɔ/ but actual transcription is /kuʃʃmandɔ/, স্মৃতি /ʃritɔ/ but actual transcription is /ʃmitɔ/. ঃ (বসির্গ): Sometimes it Silent in spellings like অন্তঃনগর /ɔntɔnnɔgr/ but actual transcription is /ɔntɔnɔgr/. When used as abbreviation, like ডাঃ = ডাক্তার /daktar/.

6.2 Accuracy

We took 5,000 words frequency from Bangla corpus BdNC01 [18]. Generate the IPA of the words and then check manually that is the IPA correct or not. We observed that the System can successfully transcript to IPA 87% of the text taken for evaluation. So this software accuracy rating is 87%.

6.3 Applications

This Bangla to IPA converting software can be applicable in the following fields:

- Bangla to IPA translator
- Text to Speech
- Bangla dictionary
- Pronunciation Verifier
- Cross-lingual application

VII. Future Development

We will store the word collection which can't convert by the Rules Base Transcription. We will add some new rules or modification if required. In future we will develop a Text to Speech System.

References

- [1] The introduction of IPA see <http://www.oxfordbibliographies.com/view/document/obo-9780199772810/obo-9780199772810-0022.xml>.
- [2] “The Handbook of the International Phonetic Association”, Cambridge University Press, pp. 194-197, 1999.
- [3] MacMahon, Michael K. C, ISBN 0-19-507993-0, pp. 821–846, 1996, “Phonetic Notation”, In P. T. Daniels and W. Bright (eds.). “The World's Writing Systems”, New York: Oxford University Press.
- [4] Wall, Joan, ISBN 1-877761-50-8, 1989. “International Phonetic Alphabet for Singers: A Manual for English and Foreign Language Diction”, Pst.
- [5] “IPA: Alphabet”. Langsci.ucl.ac.uk. Archived from the original on 10 October 2012. Retrieved 20 November 2012.
- [6] “Full IPA Chart”, Retrieved 24 April 2017, International Phonetic Association.
- [7] Amalia Arvaniti, ISSN: 0025-1003 “Journal of the International Phonetic Association”, University of Kent, UK.
- [8] International Phonetic Association, 19 (2), pp. 67-80, 1989 “Report on the 1989 Kiel convention”, Journal of the International Phonetic Association.
- [9] “A guide to the use of the International Phonetic Alphabet” Handbook of the International Phonetic Association: Cambridge: Cambridge University Press, (1999).
- [10] IPA (International Phonetic Association) Home <https://www.internationalphoneticassociation.org/>.
- [11] Naushad UzZaman, Arnab Zaheen and Mumit Khan (2006), a comprehensive roman (English)-to-Bengali transliteration scheme. Proceedings of International Conference on Computer Processing on Bangla. Dhaka, Bangladesh.
- [12] M. Ziaur Rahman and Foyzur Rahman (2007) ,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.
- [13] Joyanta Basu , Tulika Basu , Mridusmita Mitra, Shyamal Kr. Das Mandal, (2009) , At International Conference on Speech Database and Assessments, Urumqi, China, IEEE.
- [14] Amitava Das, Tanik Saikh, Tapabrata Mondal, Asif Ekbal, Sivaji Bandyopadhyay (2010) English to Indian Languages Machine Transliteration System at NEWS 2010.

[15] Dr. Md. Farukuzzaman Khan, Afraza Ferdousi, Dr. M. Abdus Sobhan, ISSN: 2321-9653, Volume 5 Issue XI, November 2017. "Creation and Analysis of a New Bangla TeCorpus BDNC01", International Journal for Research in Applied Science & Engineering Technology (IJRASET) .

[16] Dr. Rameswar Shaw, (1996), " সাধারণ ভাষাবিজ্ঞান ও বাংলা ভাষা " , কলকাতা : পুস্তক বণিগি. 3rd Ed. ISBN: 8185471126, pp 192-211.

[17] Mohammad Daniul Huq, ISBN: 9844103053, pp 42-44, 2007, " ভাষাবিজ্ঞানের কথা ", মাওলা ব্রাদার্স. 2nd Ed.

[18] Sameer ud Dowla Khan (2010) "Bengali (Bangladeshi Standard)", Journal of the International Phonetic Association, 40 (2), doi: 10.1017/S0025100310000071, pp. 221–225.

[19] Wiktionary: Bengali transliteration [https://en.wiktionary.org/wiki/Wiktionary: Bengali _transliteration](https://en.wiktionary.org/wiki/Wiktionary:_Bengali_transliteration).

[20] 'Consonant conjunct' part from Bengali alphabet [https://en.wikipedia.org/wiki/Bengali _alphabet](https://en.wikipedia.org/wiki/Bengali_alphabet).

[21] Noren Biswas, (2003), "বাংলা একাডেমী বাঙলা উচ্চারণ অভিধান", বাংলা একাডেমি. 2nd Ed. ISBN: 9840756478.

[22] Hayat Mahmud Enhanced 3rd Ed, ISBN: 984446045-X, 2003. "বাংলা লেখার নহিমকানুন" প্রতীক প্রকাশনা সংস্থা.

[23] 'Digits and numerals' part from Bengali https://en.wikipedia.org/wiki/Bengali_alphabet

[24] Bengali numbers <https://www.omniglot.com/language/numbers/bengali.htm>.

[25] BengaliNumbers–Counting from 1 to100<https://www.2indya.com/2016/06/02/bengali-number-counting-from-1-to-100/>.

[26] Java SE Downloads <https://www.oracle.com/technetwork/java/javase-downloads/index.html>.

[27] Eclipse Packages <https://www.eclipse.org/downloads/packages/>.