*I.J. Information Engineering and Electronic Business*, 2019, 2, 1-9
Published Online March 2019 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijieeb.2019.02.01

# Text to Speech Synthesis for Bangla Language

**Khandaker Mamun Ahmed**
Department of Computer Science and Engineering, BRAC University, Dhaka, Bangladesh
Email: mamun.ahmed@bracu.ac.bd

**Prianka Mandal**
Department of Software Engineering, Daffodil International University, Dhaka, Bangladesh
Email: prianka.swe@diu.edu.bd

**B M Mainul Hossain**
Institute of Information Technology, University of Dhaka, Dhaka, Bangladesh
Email: mainul@iit.du.ac.bd

Received: 13 September 2018; Accepted: 14 December 2018; Published: 08 March 2019

***Abstract*—**Text-to-speech (TTS) synthesis is a rapidly growing field of research. Speech synthesis systems are applicable to several areas such as robotics, education and embedded systems. The implementation of such TTS system increases the correctness and efficiency of an application. Though Bangla is the seventh most spoken language all over the world, uses of TTS system in applications are difficult to find for Bangla language because of lacking simplicity and lightweightness in TTS systems. Therefore, in this paper, we propose a simple and lightweight TTS system for Bangla language. We converted Bangla text to Romanized text based on Bangla graphemes set and by developing a bunch of romanization rules. Besides, an xml-based data representation is developed as a feature of the system. It gives the flexibility to modify the data representation, parsing data and create speech based on one's own dialect. Our proposed system is very lightweight which takes less processing time and produces a good understandable speech.

***Index Terms*—**Synthesis, normalization, dialect, diphone, concatenation, tokenization, romanization.

## I. Introduction

Software systems have become an inevitable part of our daily life. Nowadays, the usage of software is tremendously increasing day by day. With the demand for different kinds of software systems, text-to-speech (TTS) synthesis system has come forward. There are hundreds of areas where TTS systems are very much important such as robotics, warning system, alarm system, email reading, human-computer interaction and especially for people with visual impairment and dyslexia. Considering the necessity of such systems, many popular technological organizations such as Mattel, SAM, Atari, Apple, Microsoft Windows, Amaga OS, Texas Instruments TI-99/4A offer speech synthesis as a built-in capability [23,24].

A TTS system converts natural language text into speech and then, a computer system able to read text aloud. A speech synthesizer converts written text to a phonemic representation and then converts the phonemic representation to waveforms which can be output as sound.

There are several ways to create synthesized speech. Among them, concatenative synthesis and formant synthesis are very popular. Concatenative synthesis is based on concatenating pre-recorded speech of phonemes, diphones, words or phrases. Concatenative synthesis produces the most natural sounding synthesized speech because of its use of pre-recorded data. Formant synthesis makes the timbre of a voice or instrument consistent over a wide range of frequencies and generates artificial, robotic sounding speech.

In this paper, we are using a concatenative synthesis technique to generate natural sounding speech. Bangla is one of the most important Indo-Iranian languages which is the seventh most popular language in the world and spoken by a population that now exceeds 250 million [16]. Bangla is the primary spoken language in Bangladesh and the second most spoken language in India [4]. Several researches were conducted in Bangla speech synthesis but these are not enough to build a complete TTS system. Sometimes a large lexicon is necessary to design a TTS system which needs long processing time [6]. Bangla language has 50 alphabets and the English language has 26 alphabets, which is almost half of Bangla alphabets. Taking this into concern, we translated Bangla text to English to reduce the processing time and to be able to use the existing English phone set to generate Bangla speech.

There is some text to speech synthesis engines available nowadays. Among them, *festival* is an open-source extremely flexible concatenative TTS engine which uses diphones or other units to generate synthesized speech [21,25]. It uses Bangla lexicon to produce Bangla speech [1,6]. Festival is a large system with slow compilation process and high runtime memory requirement [2,22]. Flite is another speech synthesizer which considers the size and performance on embedded platforms that reduces its flexibility [7]. Considering the performance issue and flexibility, another speech synthesizer FreeTTS is developed based on the two speech synthesizers. FreeTTS uses algorithms of Flite and the architecture of Festival. It is found that FreeTTS runs two to three times faster than Flite [7].

This paper presents a Bangla text-to-speech synthesis system which is flexible, needs small processing time and produces a good understandable speech. Besides, we developed an intermediate XML based data representation feature which will help users to create speech based on their own dialect. It reduces to know the technical details to synthesize speech. To the best of our knowledge, ours is the first work on synthesizing Bangla speech using English diphone set that reduces the processing time for synthesization.

The rest of the paper is organized as follows: section II presents several existing works regarding text to speech synthesis system. Section III presents the proposed approach of text-to-speech synthesis system. Section IV discusses the experimental results. Finally, conclusions of this work and suggestions for future work are summarized in section V.

## II. Background Study

Developing a text-to-speech synthesis system is a challenging task. There are many stages such as text normalization, text-to-phonemes conversion, prosodic emotional content detection, and speech synthesis are needed to accomplish to develop a complete TTS system.

Plenty of research works have already been proposed in Speech synthesis for different languages. Some early researchers tried to build machines to emulate human speech, long before the invention of electronic signal processing. In 1779 speech synthesis has come under the light when models of the human vocal tract were built that could produce the five vowel sounds (in International Phonetic Alphabet notation: [a], [e], [i], [o] and [u]) [8].

A pitch synchronous waveform processing technique for text-to-speech synthesis using diphones was presented in [19]. In this paper, several algorithms were reviewed in a common framework to improve the voice quality of a text-to-speech synthesis system. The framework was developed based on acoustical units concatenation technique [19,20]. A German text-to-speech synthesis system, MARY was proposed by Schröder, Marc, and Trouvain [17]. The systems main features are a modular design and an XML-based internal data representation. It allowed the user to access and modify the intermediate processing steps without having a technical understanding of the system. Though research in text to speech synthesization for western languages has reached in a good position but for Bangla language that is very few such as [1,10,11,14].

The work reported in F. Alam et al. developed a speech synthesizer for Bangla language [1,6]. This system is developed using diphone concatenation approach. It needs a lexicon with its pronunciation to produce speech. The lexicon contains ninety-three thousand entries [6]. The proposed system creates voice data for festival and additionally extends festival using its embedded scheme scripting interface to incorporate Bangla language support. It translates Bangla unicode text to ASCII according to Bangla phone set. However, there is no description of how the transliteration process works. Moreover, there is no description about letter-to-sound (LTS) rules developed for words that are absent in the lexicon.

Concatenative speech synthesis system based on Epoch Synchronous Non OverLap Add (ESNOLA) technique for Bangla text to speech synthesis is discussed in [10,11]. The ESNOLA algorithm is developed for concatenation, regeneration as well as for pitch and duration modification. Preprocessing module creates partnames database from the pre-recorded natural speech signals, text analysis module accepts input text and generates phoneme string and stress marker and synthesizer module generates speech through combining the slices of pre-recorded speech.

PDF text to speech conversion process is discussed in [9] where other tried to analysis sentiment from Bangla text [3]. PDF represents different types of data as objects such as text object, image object and multimedia object [9]. The pdf to unicode text conversion process extracts texts from pdf objects and unicode text to speech conversion process produces speech.

Every language has standard and non-standard words. To generate speech all the non-standard words should be converted to their correct pronounceable form. There are several ways to identify and normalize non-standard words. Some researchers have identified several semiotic classes like text normalization [12,13]. Regular expressions were written in .jflex format to recognize each semiotic class. And a set of rules were used for tokenization and verbalization. Another approach used decision tree and decision list for disambiguation [14]. Though some works have been done in this domain, but still there are some problems which need to be accomplished to get a good quality sound.

## III. Methodology

To synthesize speech from text, we proposed a text-to-speech synthesis system for Bangla language. The overall architecture of the proposed system is given in Figure 1 where we have normalized, tokenized, romanized and synthesized the input text.

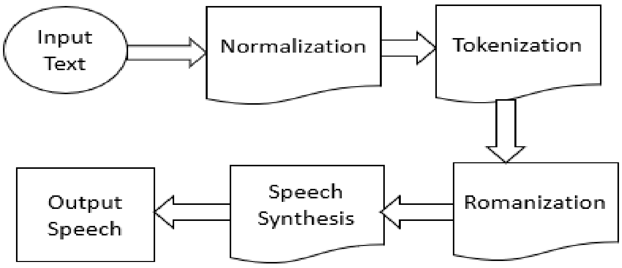


Fig.1. Architecture of Bangla text to speech synthesis system

## IV. Text Normalization

A text document contains not only full words but also various other language units such as numbers, dates, symbols, and currency. While speech can be synthesized from full words directly (subsection III), all the other language units must be first consistently expanded into full words before they get synthesized. The language unit conversion process which takes place internally is called text normalization. Table 1 contains the list of language units along with their expanded format. In the following subsections, the normalization process of some language units is discussed.

### A. Number normalization

Number is a mathematical notation to count, measure or label. Bangla numerals system has ten digits: ০, ১, ২, ৩, ৪, ৫, ৬, ৭, ৮, ৯ like Hindu–Arabic numeral system [15]. There are hundred numerals started from zero (0) to ninety-nine (99) (table 2). For numbers above 99 there are five main systems for naming numbers in Bangla (table 3).

If a text has only digits (০-৯) or digits separated by a comma (“,”), then it will be recognized as a number. After identifying the language unit as a digit, the following procedure converts a number to its pronounceable form.

1. Firstly, Bangla number is converted to the English number. This process works by replacing Bangla digits to the corresponding English digits. And, the relationship between Bangla and English digits is: ০->0, ১->1, ২->2, ৩->3, ৪->4, ৫->5, ৬->6, ৭->7, ৮->8, ৯->9.
2. After converting a number from Bangla to English, it is checked if the number is less than “১০০” (100). If the number is less than all number units, then the number’s corresponding pronounceable form is taken from Bangla numerals (table 2). But, if the number is greater or equal to a number unit (descending order), the number is divided by that unit and the quotient and remainder is calculated.
3. The calculated quotient and remainder are checked whether it is zero or not. If quotient or remainder is not equal to zero, it is passed again to process 2. And the units Bangla pronounceable form is added to the pronounceable text.

For example, “১০০৩৩২” is a Bangla number and converted to English number 100332. 100332 is greater than number unit 100000. Therefore, 100332 is divided by 100000 and the remainder is 332 and quotient is 1. The quotient is not equal to zero and less than 100, therefore, the pronounceable form now is “এক লক্ষ” (one hundred thousand) (quotient + units pronounceable form). However, the remainder is greater than unit 100. Therefore, the remainder is divided by 100 and again the quotient 3 and remainder 32 is calculated. Now, both are less than 100, therefore the pronounceable form of the quotient is “তিন শত” (3 hundred) and the remainder is “বত্রিশ” (thirty-two). Finally, “১০০৩৩২” will be pronounced to “এক লক্ষ  তিনশত বত্রিশ ”.

### B. Date normalization

According to the national and official Calendar of Bangladesh, the date format is “দদ-মম-ববব‌ব” (dd-mm-yyyy). A text is identified as a date unit, if it contains a one to thirty-one, following a separator and a one- or two-digit number denoting a month ranged from one to twelve with the same separator and a two- or four-digit number denoting a year. People also use some other types of date formats like “Day number – month name – year” for example “২ জুলাই ২০১৭” (2 July 2017). These types of dates can be one- or two-digit number denoting a day ranged from identified if the text contains a one- or two-digit number, a separator, a text denoting the month and a four-digit number denoting the year sequentially.

Table 1. Language units with their expanded format

| Language unit | Non-standard format | Expanded format |
|---|---|---|
| Cardinal number | ১১৩০ | একহাজার একশত ত্রিশ |
| Fractional number | ১০৫.০২ | একশত পাঁচ দশমিক শূন্য শূন্য দুই |
| Ordinal number | ১ম, ২য় | প্রথম, দ্বিতীয় |
| Date | ০২/০৫/২০১৬ or<br>২ মে ২০১৬ | দুই মে দুই হাজার ষোল |
| Phone number | +৮৮০১৮১৩৩৭৫১৮২<br>or<br>+৮৮০-৯৬৩৯৫৩৭১৯০ | ধনাত্বক আট আট শূন্য এক আট এক তিন তিন সাত পাঁচ এক আট দুই or or ধনাত্বক আট আট শূন্য নয় ছয় তিন নয় পাঁচ পাঁচ তিন সাত এক নয় শূন্য |
| Range | ১০-১২ | দশ থেকে বার |
| Roman numerals | I, II, III, IV, V | প্রথম, দ্বিতীয়,  তৃতীয়, চতুর্থ,পঞ্চম.. |
| Time | ১২:৩০:১৫ | বারটা ত্রিশ মিনিট পনেরো পনেরো সেকেন্ড |
| Unit and measurement | °, ’, ”, % | ডিগ্রি, মিনিট, সেকেন্ড, শতাংশ |

There are four date component separators which are the followings.

1. "/" – stroke (slash)
2. "." – dots or full stops (periods)
3. "-" – hyphens or dashes
4. " " – spaces

After identifying the date, we have converted it to its pronounceable format. We have separated the date unit as day, month and year.

The day is expanded using the number normalization algorithm. Bangla calendar has twelve months in a year like English calendar system. If the month is a text like "জুলাই" (July), it remains as it is. If the month is a number, we replaced the number by the corresponding month text. For Bangla year there are two formats to pronounce a year. If the year is less than one thousand or not thousands such as 1YXX, 2YXX ... (here, x represents any digit and Y represents a digit not equal zero), it is pronounced by grouping.

Table 2. Bangla numerals

| | | | |
|---|---|---|---|
| ০ | শূন্‌য | ১১ | এগারো |
| ১ | এক | ১২ | বার |
| ২ | দুই | ১৩ | তের |
| ৩ | তিন | ১৪ | চৌদ্দ |
| ৪ | চার | ১৫ | পনের |
| ৫ | পাঁচ | ১৬ | ষোল |
| ৬ | ছয় | ১৭ | সতের |
| ৭ | সাত | ১৮ | আঠারো |
| ৮ | আট | ১৯ | উনিশ |
| ৯ | নয় | ২০ | বিশ |
| ১০ | দশ | ২১ | একুশ |
| ... | ... | ... | ... |
| ৯৪ | চুরানব্‌বই | ৯৭ | সাতানব্‌বই |
| ৯৫ | পঁচানব্‌বই | ৯৮ | আটানব্‌বই |
| ৯৬ | ছয়িানব্‌বই | ৯৯ | নরিানব্‌বই |

Table 3. Units for naming Bangla numbers

| Number notation | Power notation | English numbering unit | Bangla numbering unit |
|---|---|---|---|
| ১০০ | $10^2$ | One hundred | এক শত |
| ১০০০ | $10^3$ | One thousand | এক হাজার |
| ১০০০০০ | $10^5$ | Hundred thousand | এক লক্ষ |
| ১০০০০০০০ | $10^7$ | Ten million | এক কোটি |

Table 4. Date normalization

| Regular text | Identified date | Normalized date |
|---|---|---|
| ১ জুলাই ২০১৬ তারিখে নয়জন নয়জন হামলাকারী ঢাকার হলি হলি আর্টিসান বেকারিতে গুলিবর্ষণ করে। | ১ জুলাই ২০১৬ | এক জুলাই দুই হাজার হাজার ষ৹োল |
| ২/৩/২০১৭ তারখিে বাংলাদশেে ক্‌রকিটে দল শ্‌রীলঙ্‌কা সফর করে। | ২/৩/২০১৭ | দুই মার্‌চ দুই হাজার বার |

The last two digits make a group and the rest digits make another group and the word "শো" (sho) is added to the last group. For example, 1971 will be pronounced as "উনিশশো একাত্তর" (unissho ekattor). And, if the year is like 10XX or 20XX, it is converted using the number conversion algorithm. For instance, 2012 will be pronounced as "দুই হাজার বার" (two thousand twelve).

## C. Currency normalization

The Bangladeshi taka (Bangla: `টাকা') is the currency of People's Republic of Bangladesh and its sign is `৳'. There are two ways to represents currency or amount of money. Firstly, the currency sign `৳' which comes before a number like `৳১০০'. Secondly, when the word "টাকা" (taka) comes after a number like "১০০ টাকা" (100 taka). The currency unit needs to be normalized to the corresponding pronounceable form for both of these situations. The correct pronounceable form of `৳১০০' or "১০০ টাকা" is "এক শত টাকা".

To recognize currency from the text, we have created two currency recognition formats like:

- "৳ - Space - N" or "৳ - N"
- "N - Space – "টাকা"

Here, N refers to a number. The number may have a comma (",") to separate special units ("শত, হাজার, লক্ষ, কোটি").

We have used the same algorithm to normalize currency which is used to normalize number. After recognizing a text as a currency unit, we separated the word "টাকা" or the currency sign `৳' and get the number. Then, the number normalization algorithm generates pra onounceable form of that number.

Finally, the word "টাকা" is added after the pronounceable text.

## D. Phone number normalization

A telephone number is a sequence of digits assigned to a fixed-line telephone subscriber station connected to a telephone line or to a wireless electronic telephony device such as a radio telephone or a mobile telephone or to other devices for data transmission via the public switched telephone network (PSTN) or other private networks. The subscriber phone number in Bangladesh is a unique 11-digit long number. The country calling code for Bangladesh is +880.

The typical format for a mobile phone number is: "+880-1X-NNNN-NNNN" and typical format for a telephone number is: "+880-96XX-NNNNNN".

For mobile and telephone number, +880 is the a country code, X is operator code and N is subscriber number.

When dialing a Bangladesh number from inside Bangladesh, the format is:

- 0 – operator code (X) – subscriber number (N) or
- 96 – operator code (X) – subscriber number (N)

If a text has +880 following 1 or 96 and an eight-digit number, it will be recognized as a phone number. When dialing inside Bangladesh, the country code is not necessary. In that situation, if a text has 1 or 96 following an eight-digit number, it will be identified as a phone number.

Phone number is actually a sequence of digits. After identifying a text as phone number, the digits of that number are replaced with their corresponding pronounceable form.

## V. TOKENIZATION

Tokenization is the process of demarcating and possibly classifying sections of a string of input characters. In tokenization, a given character sequence is chopped into pieces called tokens. A token is an instance of a sequence of characters in some particular document that is grouped together as a useful semantic unit for processing.

Table 5. Bangla alphabets romanization

| Grapheme category | Bangla grapheme | | Romanized form |
|---|---|---|---|
| Vowel | Vowel | Vowel mark | - |
| | অ | - | o |
| | আ | ◌া | a |
| | ই | ি◌ | i |
| | ঈ | ◌ী | i |
| | উ | ◌ু | u |
| | … | … | … |
| Consonant (ব্যঞ্জনবর্ণ) | ক | | K |
| | খ | | Kh |
| | গ | | g |
| | ঘ | | gh |
| | ঙ | | ng |
| | … | | … |
| Consonant conjuncts (যুক্তবর্ণ) | ক্ক | | kk |
| | ন্ট | | nt |
| | দ্ধ | | dh |
| | ক্ষ | | kkho |
| | চ্ছ | | cch |
| | … | | … |

Tokens are identified based on the specific rules of the lexer. Some methods used to identify tokens include: regular expressions, specific sequences of characters termed a flag, specific separating characters called delimiters and explicit definition by a dictionary. Special characters, including punctuation characters, are commonly used by lexers to identify tokens because of their natural use in writing. Like English, Hindi and other South Asian language, Bangla language also uses whitespaces to tokenize a sequence of characters into individual tokens.

In this paper, the punctuation characters are used to tokenize sentences and then the sentences are further tokenized to words by a whitespace character.

## VI. ROMANIZATION

Romanization is the representation of a script in Latin script. Bangla is a segmental writing system and its graphemes represent the phonemes. Bangla Script has 11 vowel graphemes and 39 consonant graphemes and more than two hundred consonant conjunctions. We have Romanized Bangla script according to Bangla grapheme set. Table 5 shows Bangla graphemes sets for vowel, consonant and consonant conjuncts with their corresponding Romanized from and table 6 shows the romanization process.

Table 6. Bangla word romanization process

| Bangla word | Bangla syllable | Corresponding English syllable |
|---|---|---|
| আমার | আ + ম + া + র | a + m + a + r |
| দশে | দ + ে + শ | d + e + s |
| বাংলাদশে | ব + া + ং + ল + া + দ া + দ + ে +শ | b + a + ng + l + a + d + e + sh |
| রহমি | র + হ + ি + ম | ro + h + i + m |
| ক◌োকলি | ক + ◌ো + ক + ি + ল ল | k + o + k + i + l |

Based on the romanization process, each token is romanized to Latin scripts. We have designed romanization rules based on vowel and consonant combinations. Some of the rules are described below:

1. The vowels are romanized directly according to its corresponding romanized form (table 5).
2. If a consonant is in the last position of a word, it is replaced according to its romanized form (table 5). For example, in the word "বকুল" (bokul), "ল" is a consonant which is in the last position, so 'l' replaces the letter "ল".
3. If a consonant is not in the last position of a word and if there is no vowel after it, 'o' is added along with the consonant. For example, "রহমি" is romanized as "rohim". Here, 'r' is for "র" and 'o' is added after it to make the pronunciation correct.
4. If the character "◌্" is found in a word, the letter before and after it is taken into consideration to make a consonant conjunct and then the conjunct is looked in the Bangla alphabets romanization table (table 5). If the consonant conjunct found, its

corresponding romanized form replaces it. If it is not found “্” is escaped.

5. If the character “ঃ” is found in the middle of a word, the consonant after “ঃ” is placed in its position. For example, in the word “দুঃখ” is pronounced as “দুখখ” and its romanization form is “Dukkho”.

## VII. Speech Synthesis

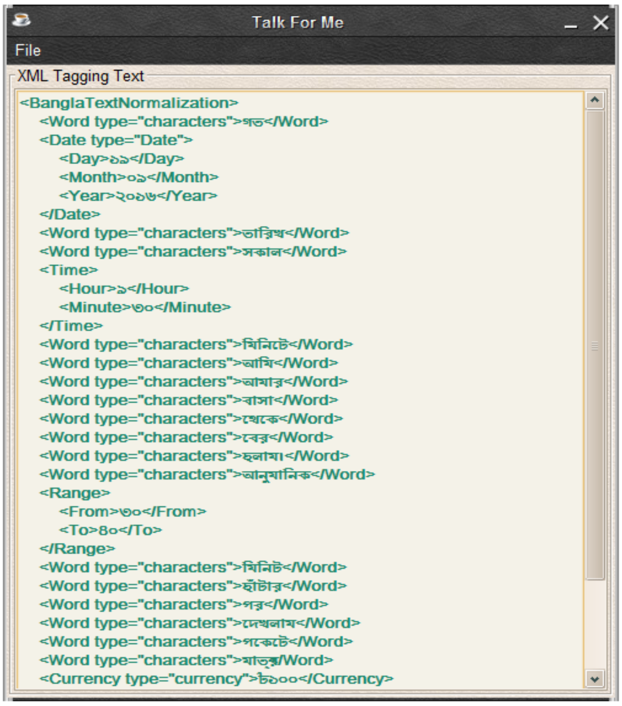
Synthesized speech is the ultimate production of a TTS system. The text is converted to phonemes based on the phonemes database. Phoneme is the fundamental unit of sound in a language. Then the prosody analysis analyzes the prosody of the phonemes, words, and sentences to determine the appropriate prosody. The prosody and phonemes information are used to produce audio waveforms of the sentences.

In this paper, to synthesize speech, we have used MBROLA voice [27]. It is a 16 kilohertz (kHz) male voice. After romanization, the text is converted to an interface named FreeTTSSpeakable. The source text that need to be spoken is first converted to it. Then the FreeTTSSpeakable is sent to voice interface which is the central processing point of speech synthesis. It takes FreeTTSSpeakable as input, converts the it into a series of utterances using the MBROLA voice and generates audio output.

## VIII. Xml Data Representation

Along with speech synthesis system, we have developed an xml-based data representation system depending on the language units. Each language unit is given a specific tag. After tokenization (subsection II), all the tokens are given their corresponding language units tags such as date, time, word and currency. We have used regular expression for the purpose of langauge unit identification. This is an intermediate data representation where users can modify the data representation without knowing the technical details of the system and can generate speech based on their own dialect from the xml.

Besides, users can parse data from the xml data tree (Fig.2). For example, if a user wants to get all the date, by parsing the xml s/he can get all the date.



```
<BanglaTextNormalization>
    <Word type="characters">গত</Word>
    <Date type="Date">
        <Day>১১</Day>
        <Month>০১</Month>
        <Year>২০১৬</Year>
    </Date>
    <Word type="characters">তারিখ</Word>
    <Word type="characters">সকাল</Word>
    <Time>
        <Hour>৯</Hour>
        <Minute>৩০</Minute>
    </Time>
    <Word type="characters">মিনিটে</Word>
    <Word type="characters">আমি</Word>
    <Word type="characters">আমার</Word>
    <Word type="characters">বাসা</Word>
    <Word type="characters">থেকে</Word>
    <Word type="characters">বের</Word>
    <Word type="characters">হলাম।</Word>
    <Word type="characters">আনুমানিক</Word>
    <Range>
        <From>৩০</From>
        <To>৪০</To>
    </Range>
    <Word type="characters">মিনিট</Word>
    <Word type="characters">হাঁটার</Word>
    <Word type="characters">পর</Word>
    <Word type="characters">দেখলাম</Word>
    <Word type="characters">পকেটে</Word>
    <Word type="characters">মাত্র/Word>
    <Currency type="currency">৳১০০</Currency>
```

Fig.2. XML data presentation

## IX. Experimental Evaluation

To produce output speech, input text is taken from different sources. The sources of input texts are daily newspaper, poem and short stories. We considered the most popular daily newspaper prothom-alo, the famous poem of Bangla literature “কানা বগীর ছা” and Bangla short story “Chuti” (“ছুটি”) [27], written by the Nobel laureate Rabindranath Tagore. In table 7, the input text along with its corresponding romanized form is shown.

To get the result, we have selected two groups of people. One is graduate students and the other group of people is senior citizens. We let them to hear the produced speech and write down the words they have heard. Moreover, result shows that graduate students are more attentive and understand more words than senior citizens. The result shows that graduate students understand clearly 68% of the produced speech where senior citizens understand 60% of the produced speech (Table 8). Senior citizens understand less of the produced speech, because of their physiological aging and changes in cognitive ability [26]. Moreover, result also varies based on the sources where the best result is found for poem and the least result is found for short stories. The average accuracy for newspaper, poem and short stories are 67.37, 71.87 and 64.52 for graduate students, and 59.6, 62.5 and 59.6 for senior citizens.

Table 7. Input Bangla text with corresponding romanized text

| **Bangla text** | **Corresponding Romanized text** |
|---|---|
| সাদা ও কালো। যদি বলি এই দুটি আলাদা কোনো রং নয়, চমকে উঠতে পারেন অনেকেই। বিজ্ঞান বলছে সাদা ও কালো এই দুটি একক কোনো রং রং নয়। সাদা রং তৈরি করে সূর্যরশ্মি। সূর্যের আলো যখন প্রিজমের মধ্য দিয়ে যায় , তখন লাল-সবুজ-নীল—এই তিন রং দেখা যায়। তা অনেকে অনেকে আগেই প্রমাণ করেছেন বিজ্ঞানী স্যার আইজাক নিউটন। আবার আবার সাতরঙা গোলাকার শক্ত কোনো কাগজ বা বোর্ড জোরে ঘুরতে থাকলে দেখা যায় সব রংই উধাও - সাদা একটা কিছু চোখে পড়ছে। বিজ্ঞানীরা বলেন আলোর অনুপস্থিতি হলো কালো। আবার **শিল্পীদের** কাছে সব রং মিশিয়ে তৈরি হয় কালো , সাদা তো ক্যানভাস। বলা হয় এই এই সাদা ও কালো একক রং নয়। দুটোই অনেক রঙের সমন্বয়। তাই এই এই দুটি রঙের প্রভাব হয়তো অনেক বেশি। বিশেষ বিশেষ উপলক্ষ , বিশেষ বিশেষ আবেগ বা অনুভূতির প্রকাশ করা যায় সাদা-কালো দিয়ে। পশ্চিমে পশ্চিমে নানা অনুষ্ঠান , দিন কিংবা আয়োজনে ড্রেসকোড , কালারকোড কালারকোড থাকলেও আমাদের দেশে নির্দিষ্টভাবে সে রকম কিছু নেই। তবে উপলক্ষ , দিনের আবহ বুঝে আমরাও কিন্তু পোশাকের রং ঠিক করি। করি। জাতীয় শোক দিবস কিংবা একুশে ফেব্রুয়ারির কোনো আয়োজনে আয়োজনে গেলে সাদা-কালোর বাইরে খুব একটা আমরা যাই না। একুশের একুশের প্রথম প্রহরে বা ভোরের প্রভাতফেরিতে সাদা-কালো পোশাক , খালি পা—কাউকে বলে দিতে হয় না। যে বাড়িতে শোকের ঘটনা ঘটছে , ঘটছে , সেখানেও দেখা যায় স্বজন হারানো মানুষেরা শোকের প্রাথমিক ধাক্কা সামলে নিজের পরনের পোশাকটি বদলে সাদা বা কালো পোশাক পরে পরে শামিল হচ্ছেন শেষযাত্রায়। আসলে পরিবেশ-পরিস্থিতি , আবেগ-অনুভূতি বলে দিচ্ছে ওই সময়ে কোন রং থাকবে পরনে। | sada o kalo. jodi boli ai duti alada kono rong noy , chomoke uthote paren onekei. biggan boloche sada o kalo ai duti akok kono rong noy.,sada rong toiri kore surjroshi. Surjer alo jokhon prijomer moddho diye jay , tokhon lal sobuj nil ai tin rong dekha jay.,ta onek agei prman korechen biggani sjar aijak niuton. kono kagoj ba bord jore ghurote thakole dekha jay sob rongoi udhao , sada akota kichu chokhe poroche. bigganira bolen alor onuposthiti holo kalo. abar iosolpider kache sob rong misiye toiri hoy kalo , sada to kanovas. bola hoy ai sada o kalo akok rong noy . dutoi onek ronger somonnoy. tai ai duti ronger prvab hoyoto onek besi. bises bises upolokkho , bises abeg ba onuvutir prkas kora jay sada kalo diye. poschime nana onusthan , din kingoba ayojone dresokod , kalarokod thakoleo amader dese nirdistvabe se rokom kichu nei .,tobe upolokkho , diner aboh bujhe amorao kintu posaker rong thik kori.,jatiy sok dibos kingoba akuse februyarir kono ayojone gele sada kalor baire khub akota amora jaina .,akuser prthom prhore ba vorer prvatoferite sada kalo posak , khali pa kauke bole dite hoy na.,je barite soker ghotona ghoteche , sekhaneo dekha jay sojon harano manusera soker prathomik dhakka samole nijer poroner posakoti bodole sada ba kalo peaosak pore samil hocchen sesojatray., asole poribes poristhiti , abeg onuvuti bole dicche oi somoye kon rong thakobe porone . |
| ঐ দেখা যায় তাল গাছ , ঐ আমাদের গা , ওই খানেতে বাস করে কানা বগি বগি ছা। ও বগি তুই খাস কি , পান্তা ভাত চাস কি , একটা যদি পাস অমনি ধরে গাপুস গুপুস খাস। | oi dekha jay tal gach , oi amader ga , oi khanete bas kore kana bogir cha . o bogi tui khas ki , panta vat chas ki , akota jodi pas omoni dhore gapus gupus khas . |
| বালকদিগের সর্দার ফটিক চক্রবর্তীর মাথায় চট করিয়া একটা নূতন ভাবোদয় হইল। নদীর ধারে একটা প্রকাণ্ড শালকাষ্ঠ মাস্তুলে রূপান্তরিত হইবার প্রতীক্ষায় পড়িয়া ছিল। স্থির হইল, সেটা সকলে মিলিয়া গড়াইয়া লইয়া যাইবে। স্থির হইল , **সেটা** সকলে মিলিয়া গড়াইয়া লইয়া যাইবে। যে ব্যক্তির কাঠ , আবশ্যক কালে তাহার যে কতখানি বিস্ময় বিস্ময় বিরক্তি এবং অসুবিধা বোধ হইবে , তাহাই উপলব্ধি করিয়া বালকেরা এ প্রস্তাবে সম্পূর্ণ অনুমোদন করিল। কোমর বাধিয়া সকলেই যখন মনোযোগের সহিত কাযে প্রবৃও হইবার উপক্রম করিতেছে এমন সময়ে ফটিকের কনিষ্ঠ মাখনলাল গম্ভীরভাবে সেই গুড়ির উপরে গিয়া বসিল বসিল। ছেলেরা তাহার এইরুপ উদার ঔদাসীন্য দেখিয়া কিছু বিমর্ষ হইয়া হইয়া গেল। | balokodiger sordar fotik chokrbortir mathay chot koriya akota nuton vabodoy hoilo . nodir dhare akota prkando salokastho mastule rupantrit hoibar prtikkhoay poriya chil . sthir hoilo, seta sokole miliya goraiya loiya jaibe. je bektir kath , abosok kale tahar je kotokhani bishoy birokti abong osubidha bodh hoibe , tahai upolobdhi koriya balokera a prstabe sompurn onumodon koril . komor badhiya<br>sokolei jokhon monojoger sohit karje prbritt hoibar upokrm koriteche amon somoye fotiker konistho makhonolal gomvirovabe sei gurir upore giya bosil . chelera tahar airup udar<br>oudasinj dekhiya kichu bimors hoiya gelo. |

Table 8. Result evaluation

| Testers category | Testers | Correctly written words from total words | | | Percentage (%) | | |
|---|---|---|---|---|---|---|---|
| | | **Newspaper - 220 words** | **Poem - 32 words** | **Literature - 85 words** | **Newspaper** | **Poem** | **Literature** |
| Graduate student | Student 1 | 147 | 23 | 53 | 66.8 | 71.8 | 62.3 |
| | Student 2 | 155 | 25 | 57 | 70.5 | 78.1 | 67.1 |
| | Student 3 | 141 | 21 | 51 | 64.1 | 65.6 | 60 |
| | Student 4 | 137 | 20 | 55 | 62.3 | 62.6 | 65.9 |
| | Student 5 | 154 | 24 | 55 | 70 | 75 | 64.7 |
| | Student 6 | 155 | 25 | 57 | 70.5 | 78.1 | 67.1 |
| Senior citizen | Citizen 1 | 127 | 22 | 48 | 57.7 | 68.8 | 56.5 |
| | Citizen 2 | 133 | 21 | 55 | 60.4 | 65.6 | 64.7 |
| | Citizen 3 | 122 | 19 | 45 | 55.5 | 59.3 | 52.9 |
| | Citizen 4 | 129 | 20 | 51 | 58.6 | 62.5 | 60.0 |
| | Citizen 5 | 127 | 23 | 50 | 57.7 | 71.9 | 58.8 |
| | Citizen 6 | 130 | 15 | 55 | 67.7 | 46.9 | 64.7 |

## X. Discussion

The produced speech of the system is understandable but has lacking in its naturalness. In this system, we have used MBROLA diphone database, because developing a diphone database for Bangla language itself is a lengthy process. We have produced Bangla speech using existing English diphone database which reduces lots of task to synthesize speech. English language has 26 alphabets and Bangla language has 49 alphabets.

The English diphone database is small considering Bangla diphone set and it reduces the memory requirement of the system.Therefore, we can say that the computation time will be short considering the small diphone set in English language. The result of our proposed system is satisfactory as we have achieved this result with a very lightweight system without creating any new Bangla lexicon or a diphone database.

## XI. Conclusion and Future Work

TTS system has become popular due to its use in various sectors. But Speech synthesis technique for Bangla language is not very satisfactory. This paper introduces a new lightweight TTS system for Bangla language which uses existing English diphone sets to generate Bangla speech. The proposed system produces good quality understandable speech in small processing time. However, the sound quality is not very natural. To generate good quality natural speech, our future focus is to explore other state of the art techniques.

## Acknowledgement

This research has been supported by The University Grant Commission, Bangladesh under the Dhaka University Teachers Research Grant No-Regi/Admn-3/58763, dated on 21.03.2017.

## References

[1] Firoj Alam, Promila Kanti Nath, Mumit Khan (2007 'Text to speech for Bangla language using festival', BRAC University

[2] Mukherjee, Sankar and Mandal, Shyamal Kumar Das (2012) 'A Bengali speech synthesizer on Android OS', Association for Computational Linguistics, Proceedings of the 1st Workshop on Speech and Multimodal Interaction in Assistive Environments, pp.43–46

[3] Hasan, KM Azharul and Islam, Md Sajidul and Mashrur-E-Elahi, GM and Izhar, Mohammad Navid2013 'Sentiment Recognition from Bangla Text', Technical Challenges and Design Issues in Bangla Language Processing, pp.315

[4] Languages of India http://censusindia.gov.in/Census_Data_2001/Census_Data_Online/Language/Statement1.h (Accessed 8 August 2017)

[5] K. M. A. Hasan and M. Hozaifa and S. Dutta and R. Z. Rabbi, A framework for Bangla text to speech synthesis, pp.60-64. doi:10.1109/ICCITechn.2014.6997307

[6] Firoj Alam, Promila Kanti Nath, Mumit Khan (2011) 'Bangla text to speech using festival',Conference on human language technology for development, pp.154-161

[7] Walker, Willie and Lamere, Paul and Kwok, Philip (2002) 'FreeTTS: a performance case study', Sun Microsystems Inc.

[8] History and Development of Speech Synthesis, Helsinki University of Technology, http://research.spa.aalto.fi/publications/theses/lemmetty_mst/chap2.html (Accessed 11 September 2018)

[9] Islam, Md Rafiqul and Saha, Ram Shanker and Hossain, Ashif Rubayat (2009 'Automatic reading from Bangla PDF document using rule-based concatenative synthesis', IEEE, pp.521–525

[10] DasMandal, Shyamal Kr and Pal, Barnali (2002 'Bengali text to speech synthesis system a novel approach for crossing literacy barrier', CSIYITPA (E)

[11] Mandal, Shyamal Kr Das and Datta, Asoke Kumar (2007) 'Epoch synchronous non-overlap-add (ESNOLA) method-based concatenative speech synthesis system for Bangla', SSW, pp.351–355

[12] Text Normalization, http://developer.ivona.com/en/ttsresources/text_normalization/text_normalization_en.html, (Accessed 25 July 2017)

[13] Alam, Firoj and Habib, SM and Khan, Mumit (2008 'Text normalization system for Bangla', BRAC University

[14] Panchapagesan, K and Talukdar, Partha Pratim and Krishna, N Sridhar and Bali, Kalika and Ramakrishnan, AG (2004 'Hindi text normalization', Fifth International Conference on Knowledge Based Computer Systems (KBCS), Citeseer, pp.19–22

[15] David Eugene Smith and Louis Charles Karpinski 'The HinduArabic Numerals', Fifth International Conference on Knowledge Based Computer Systems (KBCS), http://www.gutenberg.org/ebooks/22599

[16] Bengali at Ethnologue (18th ed., 2015), http://www.ethnologue.com/18/language/ben, (Accessed 23 October 2017)

[17] Schröder, Marc and Trouvain, Jürgen (2003) 'The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching', International Journal of Speech Technology, vol. 6, No. 4, pp.365–377, issn.1572-8110, https://doi.org/10.1023/A:1025708916924

[18] Eric Moulines and Francis Charpentier (1990) 'Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones', Speech Communication, vol. 9, No. 5, pp.453 - 467, issn.0167-6393, http://www.sciencedirect.com/science/article/pii/016763939090021Z

[19] Charpentier and E. Moulines (1988) 'Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones', Text-to-speech algorithms based on FFT synthesis, pp.667-70

[20] Hamon, Christian and Mouline, E and Charpentier, Francis (1989 'A diphone synthesis system based on time-domain prosodic modifications of speech', Acoustics, Speech, and Signal Processing, 1989. ICASSP-89., 1989 International Conference on, IEEE, pp.238-241

[21] Taylor, Paul and Black, Alan W and Caley, Richard (1998 'The architecture of the Festival speech synthesis system', International Speech Communication Association

[22] Black, Alan W and Lenzo, Kevin A (2001 'Flite: a small fast run-time synthesis engine', 4th ISCA Tutorial and Research Workshop (ITRW) on Speech Synthesis

[23] Accessibility features on your iPhone, iPad, and iPod touch (Including VoiceOver, Zoom and Invert Colors), https://support.apple.com/en-us/HT204390, (Accessed 11 September 2017)

[24] Accessibility features built into Windows and Microsoft Office, https://www.microsoft.com/en-us/accessibility/, (Accessed 13 September 2017)

[25] Festival Speech Synthesis System, http://festvox.org/festival/, (Accessed 5 August 2017)

[26] Working Group on Speech Understanding and Aging (1988) 'Speech understanding and aging', The Journal of the Acoustical Society of America,vol.83,No.3,pp.859–895

[27] MBROLA project voice database for speech synthesis, http://tcts.fpms.ac.be/synthesis/mbrola.html, (Accessed 12 September 2018)

[28] Data files, https://github.com/Mamunahmed33/Bangla-Text-to-Speech/tree/master/Data%20files

## Authors' Profiles

**Khandaker Mamun Ahmed** is a lecturer in the Department of Computer Science and Engineering at BRAC University. He has completed his B.Sc in Software Engineering from Institute of Information Technology, University of Dhaka. His core research areas of interests are machine learning, natural language processing and software engineering.

**Prianka Mandal** completed her B.Sc and M.Sc in Software Engineering from Institute of Information Technology, University of Dhaka. She is now working as a lecturer in Daffodil International University in Software Engineering department. Her core areas of interest are software engineering, machine learning and natural language processing.

**Dr. B. M. Mainul Hossain** is Associate Professor at the Institute of Information Technology (IIT), University of Dhaka, Bangladesh. He received his Ph.D. degree in computer science from University of Illinois at Chicago, USA. Before that, he earned his Bachelor of Science and Master degrees from the Department of Computer Science & Engineering, University of Dhaka, Bangladesh. He worked as a Software Engineer in Microsoft Corporation (Redmond, USA). His core areas of interest are software engineering, security, data mining and machine learning.

**How to cite this paper:** Khandaker Mamun Ahmed, Prianka Mandal, B M Mainul Hossain, "Text to Speech Synthesis for Bangla Language", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.11, No.2, pp. 1-9, 2019. DOI: 10.5815/ijieeb.2019.02.01