International Journal of
Science and Engineering Investigations vol. 5, issue 51, April 2016
ISSN: 2251-8843

Design of a syllable based Bengali Text-to-Speech System

Md. Kausar Ahmed[1], Abu Salah Mohammad Asif[2], Labiba Jahan[3], Shantanu Mandal[4]
[1,2,3,4] Computer Science and Engineering, Metropolitan University, Sylhet, Bangladesh
([1]kausarahmedpial@gmail.com, [2] salehmdasif@gmail.com, [3]labiba@metrouni.edu.bd, [4]shanto@metrouni.edu.bd)

Abstract- This paper describes the design of a syllable based Text-to-Speech for Bengali language. We divide our system into four different phases-Texts Normalization, Syllable Detection, Syllable Selection and Sound File Collection. In Text Normalization phase, we will split words and mark their type. Different rules of pronunciation to detect syllable will be applied in syllable detection phase. In Syllable Selection phase, we will split each syllable from normalize text according to a remark and search them into database. Finally, in the Sound File Collection phase, we will collect sound file corresponding with syllable and concatenate them to produce speech. We hope that our system will perform better as we need few concatenation points to produce a word than any other Bengali Text-to-Speech synthesis system.

Keywords- *Text Normalization, Syllable Detection, Syllable Selection, Sound File Collection*

I. INTRODUCTION

Text to speech is a modern computer system which converts normal language text into its speech by applying some linguistic rules and algorithm. It has a broad research and application area in the modern Human Computer Interaction systems. [1] It has significant importance in education, entertainment, business, and especially for the people with visual impairment and dyslexia. [2]

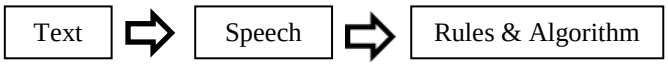

Figure 1. Block diagram of text-to-Speech Synthesis System [3]

The two primary technologies for speech synthesis are formant synthesis [4] and concatenative synthesis [5]. Formant synthesis converts text to its speech based on an acoustic model. On the other hand, concatenative synthesis converts text to its speech based on human speech samples. [6]

In this paper we have described the design of a Bengali text-to-speech where syllable is the single unit to produce output.

II. LITERETURE REVIEW

BRAC developed a Bengali Text-to-Speech synthesis system [8] "Kotha" using festival [5, 7]. Festival is a multilingual speech synthesis system which helps to provide general framework for building speech synthesis.

Shahjalal University of Science & Technology developed a diphone based Bengali Text-to-Speech synthesis system named "Subachan". It uses a minimum diphone set (527) for Bengali Text to Speech Synthesis. [9]

Kotha is based on Festival, a multilingual speech synthesis system where Subachan is based on diphone, a small unit of sound. But there are no syllable based Bengali Text-to-Speech in Bangladesh. Our proposed system is entirely focused on syllable. A syllable based Text-to-Speech can produce sound with less concatenation point than a diphone based Text-to-Speech system. So, we can expect more reliable outcomes from our proposed model of Text-to-Speech than all previous systems.

III. SYSTEM ARCHITECTURE

There are four major phases in our system named Text Normalization, Syllable Detection, Syllable Selection and Sound File Collection. Text Normalization phase includes splitting word, identifying type of each word. Syllable Detection phase provides different rules of pronunciation for detecting each syllable. Syllable Selection phase includes splitting syllable, searching database. Finally, collection and concatenation of syllable are performed in the Sound File Collection phase. Thus we produce speech from text by concatenating syllables. The architecture of our system is shown in Figure 2.

A. Text Normalization

In the first phase of our system we split each word from text. We will use some symbols including comma, high-pen, white space etc. as a delimiter. Then we detect the category of each word because word detection is important for removing ambiguity problem, expanding words etc. Moreover, word detection is helpful to make correct pronounce of each word. Thus we normalize our raw text by splitting words, expanding according to their types, elaborating abbreviated words etc. Some rules of text normalization are listed here.

- Number will be normalized as their spelling in Bengali. Example: ৩২ -> বত্রিশ
- Abbreviated words will be normalized with expansion. Example: মোঃ -> মোহাম্মদ
- 'কার' related word will be normalized by specific rules. Example: মৌমাছি -> মউমাছি
- Words with joint letter will be normalized by specific rules. Example: যুক্ত -> যুকতো
- 'ফলা' related word will be normalized by specific rules. Example: নিত্য -> নিতত্যো
- Sentence analysis is also needed in text normalization phase to solve ambiguity problem. Most of the ambiguity problems are arisen between noun and verb. So, we will detect noun and verb by n-gram model, a renowned algorithm of machine learning. Example: তারা বল (noun) খেলে। Here বল pronounced as বল্ । তুমি কথা বল(verb)। Here বল pronounced as বলো ।

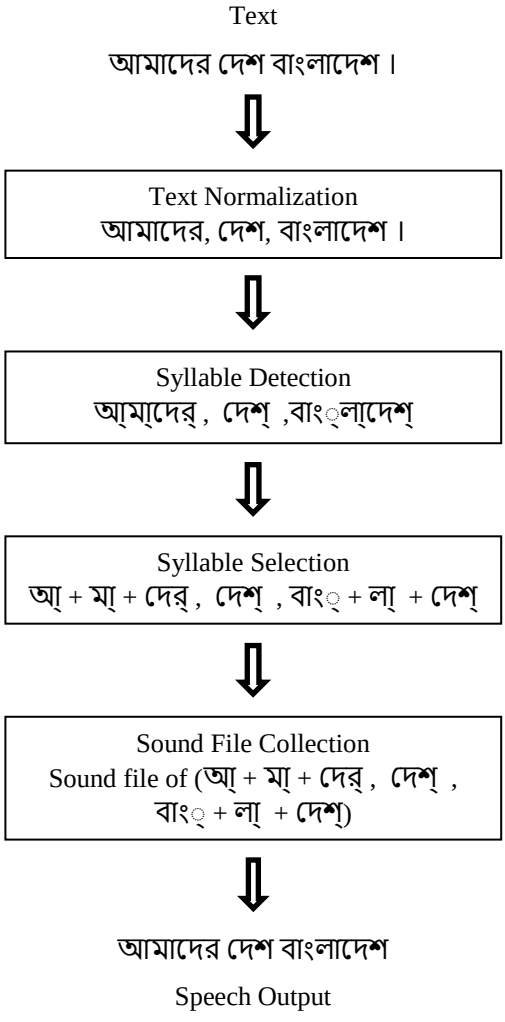

Figure 2. Syllable based Bengali Text to Speech Synthesis System

B. Syllable Detection

In syllable detection phase we detect all syllable of each word and use a remark (্) at the end letter of each syllable. Generally we try to construct a word with only mono or di-syllable. We use some specific rules for mono and di-syllable detection. When we find a syllable according to our rules we mark them with an end marker (্). This will be helpful for our next phase syllable selection. We categorize all rules based on number of letter in a word. There are many rules for detecting a syllable in words. These are some examples for each category.

Rules for a word consist of one letter:

- All one lettered words are considered as mono-syllable.

Rules for a word consist of two letters:

- If two letters have no 'কার' then the word will be a di- syllable.
  Example: এক -> এক্, বল -> বল্ etc.
- If first letter has a 'কার' but second letter has no 'কার' then the word will be a di- syllable.
  Example: কাজ -> কাজ্ etc.
- If first letter has no 'কার' but second letter has a 'কার' then the word consists of two mono-syllable.
  Example: কলা -> ক্ + লা্, নদী -> ন্ + দি্ etc.
- If two letters have 'কার' then the word consists of two mono-syllable.
  Example: আসা -> আ্ + সা্, ছেলে -> ছে্ + লে্ etc.

Rules for the word consists of three, four, five letter will follow the rules of one and two.

C. Syllable Selection

In this phase we split each syllable from a word when we get a remark (্) and search it in our database which is consists of about 5,700 sound files for mono and di-syllable. Following table shows the list of approximate number of syllables we need to develop our system. Here, we assume number of vowels =6 (অ, আ, ই, উ, এ, ও),vowel diacritics=5(া, ি, ু, ে, ো) and number of consonant = 27 (ক, খ, গ, ঘ, চ, ছ, জ, ঝ, ট, ঠ, ড, ঢ, ত, থ, দ, ধ, ন, প, ফ, ব, ভ, ম, র, ল, শ, হ, য়).

V defines mono syllable consists of vowel.

C defines mono syllable consists of consonant.

CD defines mono syllable consists of consonant with vowel diacritic

VC defines di-syllable consists of vowel and consonant.

CV defines di-syllable consists of consonant and vowel.

VV defines di-syllable consists of vowel and vowel.

CC defines di-syllable consists of consonant and consonant.

CDV defines di-syllable consists of consonant with vowel diacritic and vowel

CDC defines di-syllable consists of consonant with vowel diacritic and consonant.

TABLE I. SYLLABLE LIST

Syllable type	Number of Syllable
V	6
C	27
CD	27*5=135
VC	6*27=162
CV	27*6=162
VV	6*6=36
CC	27*27=729
CDV	27*5*6=810
CDC	27*5*27=3645
Total	5712

- V type Syllable: অ, আ, ই etc.
- C type Syllable: ক, খ, গ etc.
- CD type Syllable: কা, খা, গা etc.
- VC type Syllable: অক, আক, ইক etc.
- CV type Syllable: কই, কউ, খই etc.
- VV type Syllable: অই, আই, উই etc.
- CC type Syllable: কক, খক, গক etc.
- CDV type Syllable: কাই, খাই, গাই etc.
- CDC type Syllable: কাক, খাক, গাক etc.

D. Sound File Selection

Finally we collect all corresponding sound file for each syllable and concatenate them to produce an output speech. We need to implement advanced search algorithm to reduce complexity as our database is too large (approximate 7000 sound files). We can categorize our sound files according to 'কার'. Some categories are listed here.

- Mono syllable with no 'কার':
  অ, আ, ই, উ, ক, খ, গ, ঘ etc.
- Mono syllable with 'কার':
  কা, কি, কু, কে, কো, ক্যা
  খা, খি, খু, খে, খো, খ্যা etc.
- Di-syllable where two letters have no 'কার':
  কক খক গক ঘক চক ছক জক ঝক
  কখ খখ গখ ঘখ চখ ছখ জখ ঝখ etc.
- Di-syllable where first letter has 'কার':
  কাই খাই গাই ঘাই চাই ছাই যাই ঝাই
  কাউ খাউ গাউ ঘাউ চাউ ছাউ জাউ ঝাউ
  কাক কিক কুক কেক কোক ক্যাক
  কাখ কিখ কুখ কেখ কোখ ক্যাখ etc.

In our database a huge number of syllables are CDC type where first letter has 'কার'. So, we can categorize them according to different 'কার' for reducing searching complexity. Some categories are listed here.

- CDC type di-syllable where first letter has 'আ-কার'
  কাক খাক গাক ঘাক চাক ছাক জাক ঝাক
  কাখ কাগ কাঘ কাচ কাছ কাজ কাঝ etc.
- CDC type di-syllable where first letter has 'ই-কার':
  কিক কিখ কিগ কিঘ কিচ কিছ কিজ কিঝ
  খিক খিগ খিঘ খিচ খিছ খিজ খিঝ etc.
- CDC type di-syllable where first letter has 'উ-কার':
  কুক কুখ কুগ কুঘ কুচ কুছ কুজ কুঝ
  খুক খুখ খুগ খুঘ কুচ খুছ খুজ খুঝ etc.
- CDC type di-syllable where first letter has 'এ-কার':
  কেক কেখ কেগ কেঘ ক্রচ কেছ কেজ কেঝ
  খেক খেখ খেগ খেঘ খেচ খেছ খেজ খেঝ etc.
- CDC type di-syllable where first letter has 'ও-কার':
  কোক কোখ কোগ কোঘ কোচ কোছ কোজ কোঝ
  খোক খোখ খোগ খোঘ খোচ খোছ খোজ খোঝ

IV. CONCLUSION

The architectural design of our system is completely new approach for a Bengali text-to-speech. Satisfactory outcomes are expected from our system if we can record corpus in suitable environment. Though Bengali grammar is not that much controllable under some general rules but we can develop a system which performs better in most of the situations. We can also attach a small dictionary to handle some undesirable situations. Our next goal is to implement our system architecture in real world and develop it day by day.

REFERENCES

[1] Human computer interaction http://en.wikipedia.org/wiki/Human%E2%80%93computer_interaction.

[2] Abu Naser, Devojyoti Aich, Md. Ruhul Amin,"Implementation of Subachan: Bengali Text To Speech Synthesis Software "

[3] Muhammad Masud Rashid, Md. Akter Hussain, M. Shahidur Rahman, "Text Normalization and Diphone Preparation for Bangla Speech Synthesis"

[4] Center for Research on Bangla Language Processing - http://crblp.bracu.ac.bd/

[5] Festival Speech Synthesis, Speech Tools & documentation – http://www.festival.org/

[6] Accents, Symbols & Foreign Scripts http://symbolcodes.tlt.psu.edu/bylanguage/bengalichart.html

[7] A. G. Ramakrishnan, G. L. Jayavardhana Rama,R. Muralishankar and R Prathibha , A COMPLETE TEXT-TO-SPEECH SYNTHESIS SYSTEM IN TAMIL, Proceedings of IEEE Workshop on Speech Synthesis, 191-194, 2002

[8] Firoj Alam, Promila Kanti Nath and Mumit Khan, "Text To Speech for Bengali Language using Festival"

http://www.bracuniversity.ac.bd/research/crblp/

[9] Abu Naser, Devojyoti Aich and Md. Ruhul Amin, "Architectural Design of Bengali Text to Speech Synthesis Software " with Sentence Analysis using Advanced Linguistic Processing Modules: Stemming, Phrase Analysis and Expansion Rules.