The International Journal Of Engineering And Science (IJES)
|| Volume || 5 || Issue || 4 || Pages || PP -11-14|| 2016 ||
ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

Normalization of Non Standard Words for Marathi Speech Synthesis

[1]Imtiyaz Khan, [2]Akshay Khilare, [3]Ganesh Bobde,[4]Prof.S.R.Gulhane
[1,2,3,4]Department of E&TC, D Y Patil College of Engineering

ABSTRACT

This paper considers the task of text normalization in concatinative Text To Speech (TTS) synthesis for Marathi language. This deals on how Non-standard Marathi words - acronyms, abbreviations, proper names derived from other languages, phone numbers ,decimal numbers, fractions, ordinary numbers, sequence of numbers, money, dates, measures, titles, times and symbols - are pre-processed before passing it to the TTS system as an input. The paper also discusses about the methodology used to normalize the non-Marathi text present in the input text to get an equivalent Marathi as output. We are doing our project in MATLAB.

Keywords *- Concatenation, Non Standard Words (NSW), phoneme, Text-To-Speech (TTS) Synthesis.*

Date of Submission: 10 March 2016 Date of Accepted: 14 April 2016

I. Introduction

A TTS synthesizer can generates the speech from a given text. Although TTS is not able to replicate the quality of recorded human speech or sound, it has improved greatly in recent years. There exist many different synthesis technologies which is suitable for different applications. A non-general system could have a limited vocabulary and limitations in depend on the length of spoken utterances. A text to speech synthesizer is now an important part of information technology because it has integrated language and speech for human computer interaction. Creation of synthetic voice from text is usually referred with the general term text-to-speech though it requires a wide range and variety of procedures. Processing the given text to readable form consists in expanding abbreviations, converting names, numbers, acronyms, dates etc. into their speech form. In real text, many non-standard representations of Marathi words appear, for e.g., numbers, abbracronyms, currency, dates. All these non-standard words representations must typically be normalized to standard words before it go for synthesis. For example, Moreover, certain numbers have to be pronounced as individual digits or as a whole. a phone number such as 88234567809 will be pronounced *eight eight two three four five six seven eight zero nine*, but it will be pronounced as *eight thousand eight hundred twenty three crores forty five lakhs sixty seven thousand eight hundred and nine.*

1.1 Text-to-Speech (TTS) Synthesizer:

It is a computer based system that which is able to read any text aloud whether it was directly introduced in the computer by an operator or scanned . The process of synthesizing speech is to be divided into two broad**:** Analysis and then Synthesis and this is the two mostly known methods of speech synthesis .We are discussing these in this section. However, analysis is same for either technique and is therefore discussed independent.

II. Speech Synthesis

2.1 What Is Speech Synthesis

For getting computers to read out loud by the given text and it is about three things: reading process, speaking process, and the issues related using computers and this field of study is known as speech synthesis, which is “synthetic” (computer) generation of speech, and text-to-speech or TTS which is used to convert written text into speech or generates speech from a given text. For TTS or speech synthesis we can also say that it is automatic production of speech, by ‘grapheme to phoneme’ transcription.G2P generates phonemic transcription of a word given its spelling. There are two commonly used approaches in G2P for conversion: letter to sound rules (LTS), lookup dictionaries.

2.2 Phonetics

In most of the languages the words is not pounced as it should be. So to pounce it correctly some type of symbolic presentation is needed. Every languages have different phonetics alphabets and phoneme and their combination. The phonetics alphabets are divided into two categories: vowels and consonants. Vowels are voiced sounds which is produced by vocal cords in vibration. Consonants can be voiced or unvoiced. Vowels have high amplitude and are more stable than consonants. We have a many numbers of small units which can be distinguish from one another and from this unit different sequences can be created to from large numbers of words and this units is called as phonemes. The numbers of phonemes is varies from one language to other language but in all language the sets of in units ranging from size 15 to 50 is similar.

2.3 Synthesizer Technology

The synthesizer technology depends on *naturalness* and *intelligibility*. In which naturalness is how much closely the output sound like human speech and intelligibility is ease of sound understood. There are many type of synthesizer technology: concatenative synthesis, domain specific synthesis, diaphone synthesis.

2.3.1 Concatenative Synthesis:

It is based on concatenation of segments of recorded speech. This is also called as *cut and paste* synthesis in which the short segments of the speech are selected and cut from a pre-recorded database and are paste or joined to produce the desired utterances of speech. The real speech is very high quality, which is serious limitations in practice, due to the memory capacity required by a system. Connecting the pre-recorded utterances is the easiest way to produce naturalness and intelligible sound. It is limited to one voice and one speaker and it required more memory capacity. The most important thing that concatenative synthesis do is finding the correct length unit. More the unit length high the naturalness but memory is increased, and for small unit length memory needed is less but naturalness and concatenation is high.

III. Flowchart

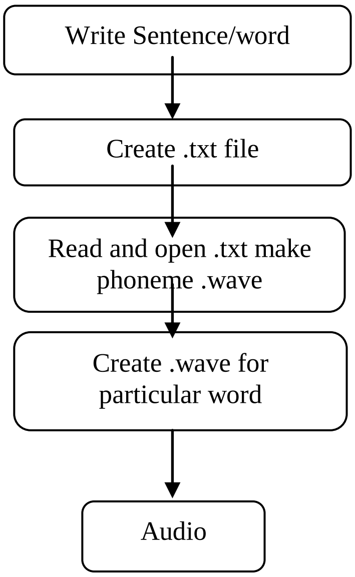
Write Sentence/word
Create .txt file
Read and open .txt make phoneme .wave
Create .wave for particular word
Audio

Fig 3.1: Flowchart of text to speech

IV. Algorithms

4.1. Character –To- Voice:

Let start with the simple conversion of text to speech. We required the database in which character-to-voice conversion is recorded alphabets (a-z) and digits (0-9) all are in wave files (.wav). First we have to convert text to speech to create text file (.txt).After conversion then the created file is open and read in MATLAB.

Algorithm

1: First create a database of all various wave files
2: Create a text file (.txt)
3: Open the .txt file in MATLAB.
4: Read .txt file opened.
5: play the wave (.wav) of character read.

Some delay are produced as default while recording a sound and this delay should be removed for continuous utterance of speech.

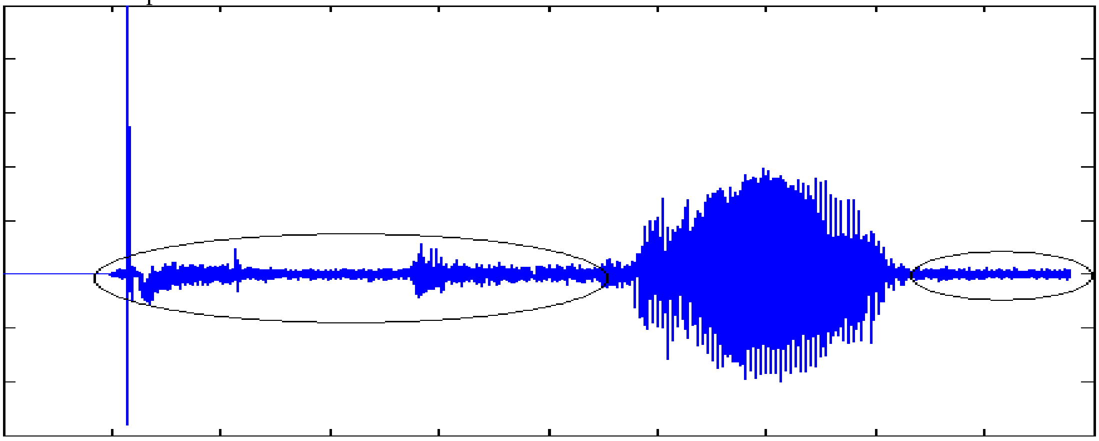

Fig 4.1 Sound with delay

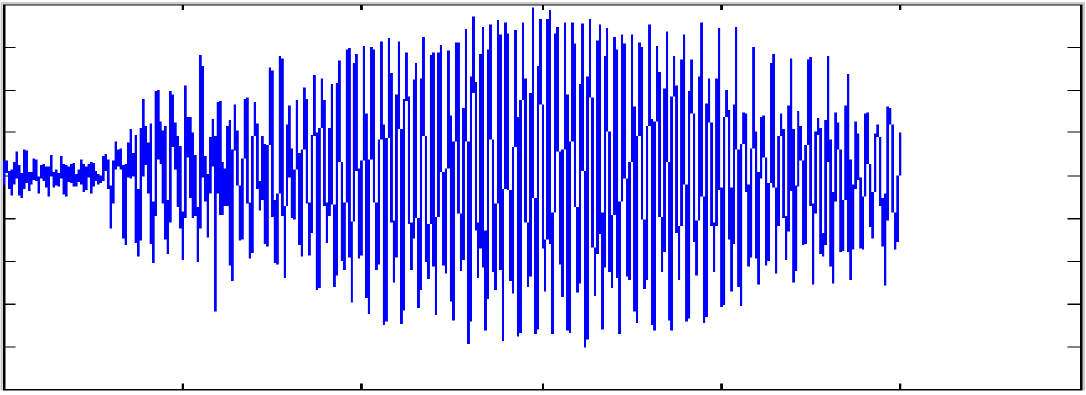

Fig 3.2 Sound after removal of delay

See the figure above the Fig3.1 is the sound with some delay which is produces when sound is recorded. Fig 3.2 show the fig when the delay is removed from file.

3.2 Word Pronunciation

The Character to voice is not a big task because there are only 26 characters in English and each and every character has a unique pronunciation. The sound is produced by reading every character but is every difficult to make out the word from the characters read and practically it is impossible to record all the words of a dictionary. So there is another method, in this we read the syllables of word not character.

Algorithm

```
1: create a database of phonemes.
2: create a text file.
3: open the .txt file.
4: read the .txt file.
5: concatenate the .wav files and play them.
```

V. Problem

The speech synthesis problem area is very wide and there are many problems in text pre-processing, like numerals, time, abbreviations, date, and acronyms. The problem which mainly occur first is conversion of text into speech which is grapheme-to-phoneme conversion. The difficulty is conversion of highly language. Conversion of English language is very complicated because there are many different sets of rules and it is needed to produce the correct pronunciation of words by it.

Text pre-processing problems is occur in many language. Digits and numerals is expanded into its full words. For example in English, numeral would be expanded like 659 as *six hundred and fifty-nine* and 1650 as *one-thousand six-hundred and fifty* (measure). Fractions and dates are also problematic. 8/16 is expanded as *eight-seventeen* (fraction) or *august seventeen* (date).

VI. Result

In our project first we make the database and then we have to record the alphabets and numerical digits and convert it into .wave file. Then we open the .txt file and read it in MATLAB. But after testing we find the problem in it that is delay in the sound .The delay in the sound is produced in when we recorded and so we remove it. And then play it .We have two method of conversion character to sound, word voice. In character to sound the one by one character is read and play which is difficult and practically impossible so another method is used word voice in which the not character wise but syllable of word is read and play. So in our project we convert the Non-standard words of Marathi into standard words. We use MATLAB in our project it is easy to operate we can easily understand the MATLAB we only have to use the command.

VII. Conclusion

The paper, presents the complexities' of Marathi language and the method to normalize the NSW of Marathi. speech synthesis have many advantages and many application now a days it used in many field .The presented work is suitable only for some specialized cases of the Marathi language but in future for large amount of complex cases can also be considered. The proposed system does not handle the context specific text.

7.1 Advantages

1. It is used for those peoples who are visually handicapped. It help to listen the written works.
2. It is also used by the people who use sign language to communicate with the other peoples.
3. It is also advantages to the children to learn words pronunciation.

7.2 Drawback

1. It is difficult to achieve high accuracy.
2. It lack in naturalness and sound quality.

7.3 Application

1. It is used for blind people.
2. It is used in railway, airport, and bus station for announcements purpose by domain specific.
3. It is used in telecommunication and multimedia like email, mobile computer.

References

[1] N.Swetha ,K..Anuradha, **"***Text-To-Speech Conversion*", *International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.6,* Pages : 269-278 (2013).

[2] Muhammad Masud Rashid , Md. Akter Hussain, M. Shahidur Rahman "*Text Normalization and Diphone Preparation for Bangla Speech Synthesis", JOURNAL OF MULTIMEDIA, VOL. 5, NO. 6*, DECEMBER 2010.

[3] Rohini B. Shinde, V. P. Pawar, *"A Review on Acoustic Phonetic Approach for Marathi Speech Recognition" , International Journal of Computer Applications (0975 – 8887) Volume 59– No.2*, December 2012.

[4] Dr.K.V.N.Sunitha, P.Sunitha Devi, "*Text Normalization for Telugu Text-to-Speech Synthesis"*, ISSN 2277-3061.

[5] Sproat R., Black A.W., Chen S., Kumar S., Ostendorf M, and Richards C., *Normalization of non-standard words*, *Computer Speech and Language, pp. 287–333*, 2001.

[6] K. Panchapagesan, Partha Pratim Talukdar, N. Sridhar Krishna, Kalika Bali, A. G. Ramakrishnan_. "*Hindi Text Normalization".*

[7] Surendra P. Ramteke, Gunjal Oza, Nilima P. Patil. "*Development of TTS for Marathi Speech Signal Based on Prosody and Concatenation Approach" Vol. 1* Issue 10, December- 2012.