IJCSN International Journal of Computer Science and Network, Volume 4, Issue 4, August 2015
ISSN (Online) : 2277-5420 www.IJCSN.org
Impact Factor: 0.417

Marathi Text to Speech Synthesis – Using Matlab®

[1] Darshna Badhe, [2] P. M. Ghate

[1] Digital Systems, Rajarshi Shahu College of Engineering, Savitribai Phule Pune University, Pune, Maharashtra-41033, India

[2] Digital Systems, Rajarshi Shahu College of Engineering, Savitribai Phule Pune University, Pune, Maharashtra-41033, India

Abstract - In this paper, we present the unit selection based concatenative text-to-speech synthesis. Here the unit of selection is syllable. The paper presents system for text to speech synthesis for Marathi language. This paper gives complete idea that how to convert Marathi text into speech right from text processing to audio processing. The proposed system requires audio data base and very limited text data base. The TTS system has been developed on Matlab® R2014a. Matlab® is Unicode software therefore UTF-8 encoding has used to read the Marathi text.

Keywords - *Text To Speech, Unicode, transliteration, Syllabification, Structure, UTF-8(universal characterset transformation function 8 bit)*

1. Introduction

The Text to speech synthesis (TTS) system is a system which is used to convert any given text as an input into sound as an output. Number of TTS system has been developed for foreign languages but from the literature studied it was realized that less work has been done in TTS for Marathi language.Marathi is widely spoken in Maharashtra and Goa, the states in western part of India, where it has the status of official and co-official language respectively. It is recognized by government of India as one of the 23 official languages. In India it has fourth largest number of native speakers. Marathi globally has around close to 73 million speakers, when surveyed in 2001. In terms of most spoken languages, globally, Marathi ranks 19th. This makes the TTS for Marathi a desirable requirement for communication to masses. The main feature of TTS is the intelligibility and naturalness, which means that the output sound that is generated at the end of the process should be easily understood and at the same time it should sound natural.

2. Literature Review

Tapas Kumar Patra [1] "Text to Speech Conversion with Phonetic Concatenation"The paper describe about how data base can be reduced using phoneme they also mention matlab command that can be used to implement TTS.

Mrs. Madhavi R. Repe [2] "Prosody Model for Marathi Language TTS Synthesis with Unit Search and Selection Speech Database"The paper describe about the how prosody can be generated for a text to speech system using posla.

Shreekanth.T, [3]"An Unit Selection based Hindi Text To Speech Synthesis System Using Syllable as a Basic Unit "The paper describe about implementation of TTS using Unicode values for Hindi.

H. Segi, R. Takou, N. Seiyama and T. Takagi[4], "An automatic broadcast system for a weather report radio program", Paper presents technique for weather broadcasting.

Shruti Gupta[5], "Hindi Text To Speech System", The paper describe implementation TTS for Hindi based on JAVA frameworks.

3. Text To Speech Synthesis System for Marathi Language

The Marathi script uses Devanagari script. This script contains a set of 12 vowels and 36 consonant, which are known as स्वर and व्यंजन respectively in the language. It also contains dependent vowels, which are known asमात्रा. All the vowels, consonants and dependent vowels have been stored in the database in the form of ASCII values with their English transliteration code. This is because Matlab® being Unicode software, it first converts Marathi text to its equivalent English translation.

3.1 Selection Of Unit For Concatenate Synthesis

Among all the synthesis techniques that have been studied, it has been observed that the output of unit selection synthesis has higher naturalness and intelligibility. The selection of unit is very important. Different units for selection that are used are diphone, phoneme, syllable, words or even sentences. Creation of

database for all the words or sentences / phrases is a very challenging task as it requires huge memory , however the quality of output sound by this method will be much closer to natural and intelligible output as it will have less break points. While using diphone or phoneme as a unit of selection the memory requires less space but requires more audio processing as it contains higher number of break points. The quality of output sound will have comparatively a lower effective naturalness and intelligibility. Therefore selection of syllable as unit of selection results in a situation of tradeoff between memory and quality of output. Based on this in this proposed system we have used syllable based speech synthesis.

3.2 Creation Of Database

The proposed system required two database, which are as follows.

3.2.1 Audio Database

An Audio database, that is created contains prerecorded sound of all the "Barakhadi" (Barakhadi is the phonetic chart that enables one to recognize "akshar" (letters) and **मात्रा**(vowel sounds) used in Marathi language. All sounds can be recorded into single audio file later it can cut down into multiple.

3.2.2 Text Database

The text database contains words,the ASCII values of **स्वर**,**व्यंजन**,**मात्रा** are stored in text database also English transliteration of Marathi is also stored in text database for text processing.

3.2.3 Process Flow of the Proposed System

As shown in the block diagram the input to the system is a Marathi text. The first step now is to break given Marathi text into words. These words are then mapped with their respective English transliteration.The system now will check for the related audio file in the database.If audio file is present in the database it will concatenate files and play them.

If word is not present in database the system will break words into syllable and check weather related audio file is present in the database.Matched file will then be identified, which is will then concatenate and played.

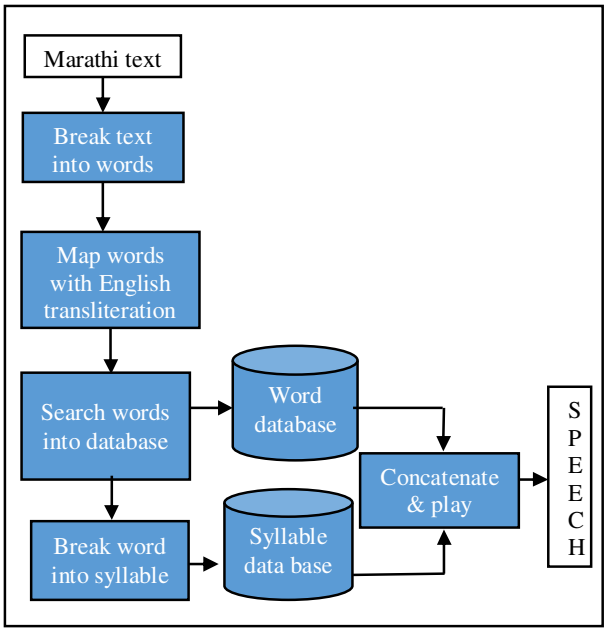

Fig 1: Block Diagram for proposed system

3.3 Different CV Structure For Breaking Text Are As Follows:

CV structure	CV Break
CVCVCVC	CV+CV+CVC
CVCV	CV+CV
CVCCVCVCCV	CVC+CV+CVC+CV
CVCVCCVCCVCCV	CV+CVC+CVC+CVC+CV
VCCVCVC	VC+CV+CVC
VCV	V+CV
CVCVCVCV	CV+CV+CV+CV
CVCVCVCVCVC	CV+CV+CV+CV+CVC

3.4 Algorithm for the Proposed System

The proposed system requires an excel data base, two text files ,one is for giving Marathi text as an input while another is for exporting audio file names. These audio files are directly stored in the same folder of Matlab®. For audio database complete sentences have been were recorded. These were recorded in authors' own voice. These sentences were then cut down into syllable using ".wav" cutter. This enabled, avoiding the recording of an individual syllable separately. This process helped us to preserve naturalness of the output sound helping in generating a better quality of output.

3.4.1 Text Processing

Step-1 Creating a text file, insert some Marathi text.
e.g. "**मीभारतीयआहे**"

Step-2 Create database using excel sheet with column containing hexadecimal values of all vowel consonant

and dependent vowel, with their equivalent transliteration in English of Marathi text.

e.g. "**मीभारतीयआहे**"

Step-3 Compare hex values that have been calculated with hex values stored in database. After this the matching values were mapped with their English transliteration codes. This will get us the required cell values. These cell values are then joined which then converts Marathi text into English as shown below:
"Meebharateeyaaahe".

Step-4 This text is then broken into syllable by applying different Consonant Vowel structure (CV structure) and rules, which gives us the will give:
e.g. 'mee', 'bha', 'ra', 'tee', 'ya', 'aa', 'he'.

Step-5 Add '.wav' string to these syllable will give strings e.g. 'mee.wav', 'bha.wav', 'ra.wav', 'tee.wav', 'ya.wav', 'aa.wav', 'he.wav'.

Step-6Export above strings to another text file.
Here the text processing is completed. Further will be audio processing.

3.4.2 Audio Processing

Step-1 Read the files with .wav extension that has been exported to text file one by one. The audio files are stored in same Matlab® folder of application programme.

Step-2 Remove silence part from the signal formed after concatenation.

Step-3 Signal is ready to play.

Step-4 Further signal can be processed for prosody generation.

3.4.3 Prosody Generation

Prosody generation means adding emotions to the signal. There are number of parameters like linear predictive coefficient (LPC), pitch period, energy, formants. By observing either of parameter in frequency domain and varying their values prosody can be generated. These will further change the output and the recognition of the speech for the audience.

Prosody depends upon four factors are as follows:

1. Speaker Characteristics
   - Age
   - Gender
2. Feelings
   - Anger
   - Happiness
   - Sadness
4 Fundamental Frequency
   - Stress
   - Duration
5 Meaning Of Sentence
   - Neutral
   - Imperative
   - Question

4. Result

The feedback of the output from the proposed system was taken from a sample of 30. Table below shows the Mean Opinion Score (MOS) conducted among people with different age and gender.

Table 1: MOS Test Results

Listeners	Age	Gender	Natural-ness	Intelligi-bility
15	>20	M	8 of 15	14
15	>50	F	9of 15	12

The sample was equally divided in the gender to avoid any gender bias in the results. Further this was conducted across the spread of the age to capture the output based on the listening capacity, which varies with the aging factor of person.

5. Conclusion

It can be concluded from the results of the MOS test for checking the intelligibility, shows that around 86% accuracy was reported. However in terms of naturalness the level of accuracy was around 56%, which was lower, but acceptable. Hence prosody generation is the desired future scope from the study to come closer to higher level of naturalness, which cannot be calculated with this system. The main advantage of the proposed system is that it requires very less text data base. Further the memory required for audio database is trade-off amongdiphone, phoneme, and words as a unit of selection. Directions for future work: The prosody generation is an important module in TTS for increasing the naturalness of the output of the speech.

References

[1] Mr.S.D.Shirbahadurkar, "*Marathi Language Speech Synthesizer Using Concatenative Synthesis Strategy (Spoken in Maharashtra, India)*", Second IEEE International Conference on Machine Vision 2009.

[2] Mrs. Madhavi R. Repe, "Natural Prosody Generation in TTS for Marathi Speech Signal", IEEE

International Conference on Signal Acquisition and Processing 2010.

[3] Mrs. Madhavi R. Repe," Prosody Model for Marathi Language TTS Synthesis with Unit Search and Selection Speech Database", IEEE International Conference on Recent Trends in information, Telecommunication and Computing 2010

[4] H. Segi, R. Takou, N. Seiyama and T. Takagi, "An automatic broadcast system for a weather report radio program", *IEEE Trans. on broadcasting*, vol. 59, no 3, September 2013.

[5] Shruti Gupta, "Comparative study of text to speech system for Indian language",International Journal Of Advances In Computing And Information Technology

[6] Shruti Gupta, "Hindi Text To Speech System", Computer Science And Engineering Department Thapar University Patiala June 2012

[7] Tapas Kumar Patra, "Text to Speech Conversion with Phonetic Concatenation", International Journal of Electronics Communication and Computer Technology (IJECCT) Volume 2 Issue 5 (September 2012)

[8] MrsMinaksheepatil, " Syllable" Concatenation for Text to Speech Synthesis for Devnagari Script", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 9, September 2012

[9] Shreekanth.T, " An Unit Selection based Hindi Text To Speech Synthesis System Using Syllable as a Basic Unit", IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 4, Issue 4, Ver. II (Jul-Aug. 2014)

[10] Mr. S. B. Chaudhari, " A Review on Multilingual Text to Speech Synthesis by Syllabifying the Words of Devanagari and Roman", International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169,Volume: 2 Issue: 11

[11] Snehali. K. Nandurkar, ZakirM.Shaikh, "Speech Generation Of Transliterated Hindi Text",International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 3, Issue 10, October 2014 ISSN 2319 - 4847

[12] Hiroyuki Segi, "An Automatic Broadcast System for a Weather Report Radio Program", IEEE Transactions on Broadcasting, Vol. 59, No. 3, September 2013