OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

The Prosody Subsystem and Pitch Pattern for Marathi Text To Speech Synthesis

Sangramsing N. Kayte, Dr.Charansing N. Kayte*, Dr.Bharti Gawali
Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India
**Digital and Cyber Forensic, Aurangabad, Maharashtra, India*

ABSTRACT:- *The demands of interactive approaches to TTS require more freedom to express prosody than current systems allow. Most current TTS systems, including the SCRM Labs TTS system, were designed to operate on text with little information beyond the text. The prosody subsystem was therefore designed conservatively. The application may be intending to convey that a set of words is a single proper noun, that a word is especially important, or that a word needs confirmation. This state information needs to be expressed prosodic ally, so one should think of speech synthesis more in the context of a concept-to-speech system than a text to- speech system. Similarly, there are applications where the simulation of emotions, subtle meanings in speech acts, and stylistic variations are desirable.*

KEYWORDS:- *Di-phones Simulation of emotions, prosody subsystem, Speech synthesis*

I. INTRODUCTION

A speech synthesis system is a computer-based system that produce speech automatically, through a grapheme-to-phoneme transcription of the sentences and prosodic features to utter. The synthetic speech is generated with the available phones and prosodic features from training speech database [1-2]. The speech units is classified into phonemes, diaphones and syllables [3-4]. The output of speech synthesis system differs in the size of the stored speech units and output is generated with execution of different methods. A text-to-speech system is composed of two parts: a front-end and a back-end. The front-end has two major tasks. First, it converts raw text containing symbols like numbers and abbreviations into the equivalent words. This process is often called text normalization, preprocessing, or tokenization. Second task is to assigns phonetic transcriptions to each word, and divides and marks the text into prosodic units like phrases, clauses, and sentences [5]. Although text-to-speech systems have improved over the past few years, some challenges still exist. The back end phase produces the synthesis of the particular speech with the use of output provided from the front end. The symbolic representations from first step are converted into sound speech and the pitch contour, phoneme durations and prosody are incorporated into the synthesized speech [7]. The paper is dedicated to the second part of this issue, i.e. to the computer-aided text-to-speech synthesis, that is recognized as a very urgent problem. Nowadays the solution of this problem can be applied in various fields. First of all, it would be of great importance for people with weak eyesight. In the modern world, it is practically impossible to live without an information exchange. The people with weak eyesight face with big problems while receiving the information through reading. A lot of methods are used to solve this problem. For example, the sound version of some books is created. As a result, people with weak eyesight have an opportunity to receive the information by listening. But there can be a case when the sound version of the necessary book couldn't be found. Therefore, the implementation of the speech technologies for information exchange for users with weak eyesight is of a crucial necessity. Synthesis of speech is the transformation of the text to speech [2]. This transformation is converting the text to the synthetic speech that is as close to real speech as possible in compliance with the pronunciation norms of special language [6]. TTS is intended to read electronic texts in the form of a book, and also to vocalize texts with the use of speech synthesis. When developing our system not only widely known modern methods but also a new approach of processing speech signal was used. In general, synthesis of speech can be necessary in all the cases when the addressee of the information is a person.

II. EXISTING SYSTEM

Text to speech synthesis is converting the text to the synthetic speech that is as close to real speech as possible according to the pronunciation norms of special language. Such systems are called text to speech (TTS) systems.

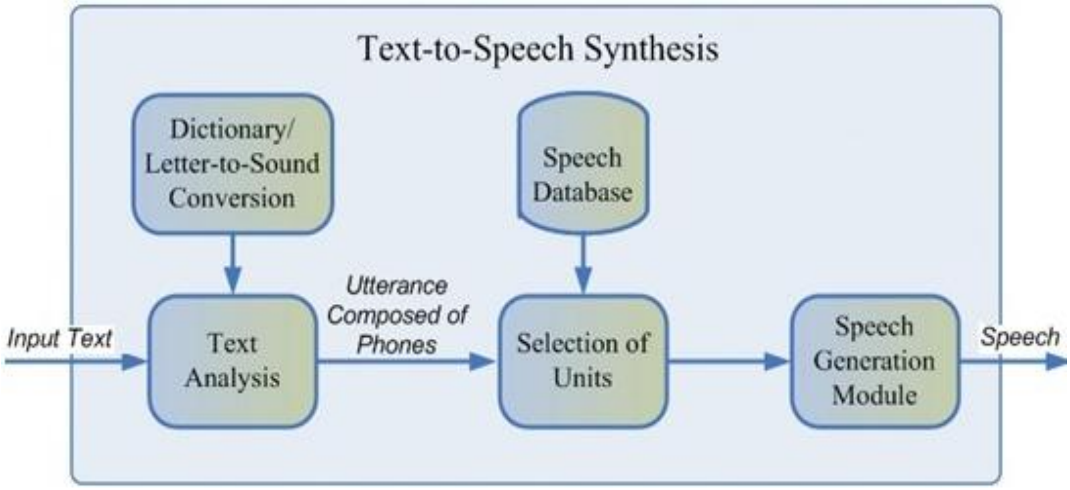
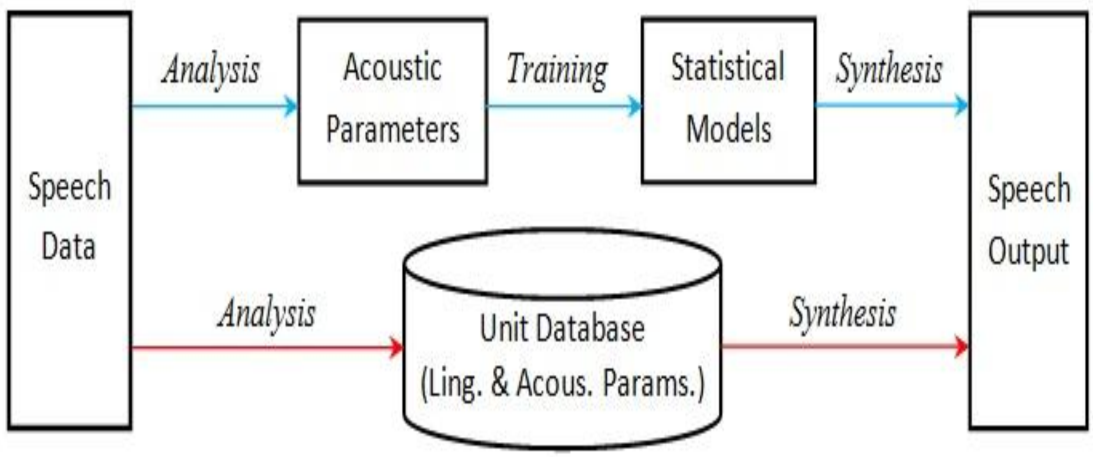


Fig.1 Block diagram for existing system

III. PROPOSED SYSTEM

A. Block of linguistic processing



B. Text input

The sounded text can be entered in any form. The size or font type is of no importance. The main requirement is that the text must be in Marathi language.

C. Initial text processing

For forming of transcriptional record, the input text should be shown as sequence of accentuated spelling words separated by space and allowed punctuation marks. Such text can conditionally be named as "normalized". Text normalization is a very important issue in TTS systems. The general structure of normalizer is explained in Figure 4. This module has several stages as it is shown in the figure.

Stage 1: Spell-checking of the text the spell-checkers are used in some cases (modules of correction of spelling and punctuation errors). The module helps to correct spelling errors in the text thereby to avoid voicing of these errors.

Stage 2: A pre-processing module a pre-processing module organizes the input sentences into manageable lists of words. First, text normalization isolates words in the text. For the most part this is as trivial as looking for a sequence of alphabetic characters, allowing for an occasional apostrophe and hyphen.

It identifies numbers, abbreviations, acronyms, and transforms them into full text when needed.

Stage 3: Number Expansion Text normalization then searches for numbers, times, dates, and other symbolic representations .These are analyzed and converted to words. Someone needs to code up the rules for the conversion of these symbols into words, since they differ depending upon the language and context [8].

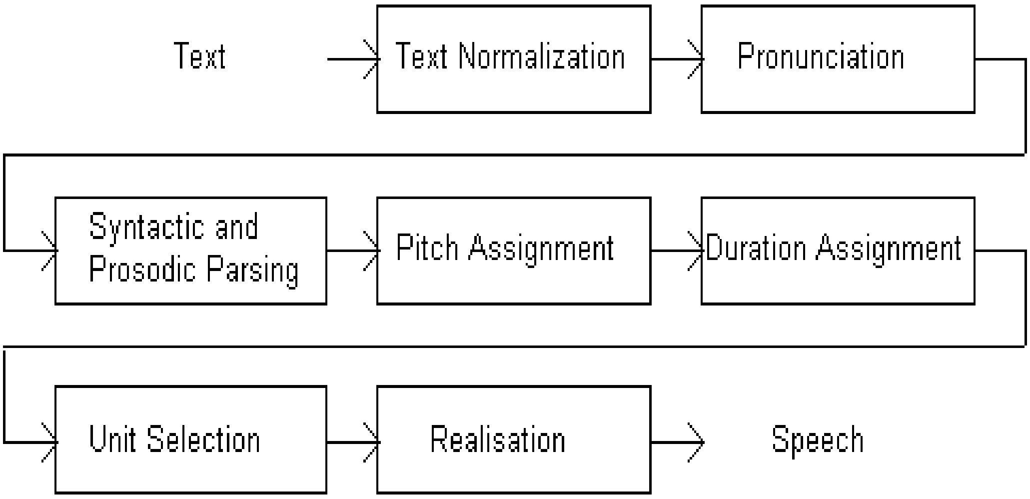


Fig.3 Text Normalization System

Stage 4: Punctuation analyze

Whatever remains is punctuation. The normalizer will have rules dictating if the punctuation causes a word to be spoken or if it is silent. In normal writing, sentence boundaries are often signaled by terminal punctuation from the set: full stop, exclamation mark, question mark or comma {.!?,} followed by white spaces. In reading a long sentence, speakers will normally break up the sentence into several phrases, each of which can be said to stand alone as an intonation unit. If punctuation is used liberally so that there are relatively few words between the commas, semicolons or periods, then a reasonable guess at an appropriate phrasing would be simply to break the sentence at the punctuation marks though this is not always appropriate [14-26]. Hence, determining the sentence break and naming the type of sentence has to be done so as to apply the prosodic rules. In natural speech, speakers normally and naturally give pauses between sentences

D. Unit Selection:

The unit selection is the basic steps for synthesis. They can be

- ➢ Low concatenation distortion
- ➢ Low prosodic distortion
- ➢ General nature, if it's not restricted then text-to-speech is used

On the basis of the issue discussed before [9], we are in need of considering the syllabus that has to be used as basic units. There are many possibilities which include phonemes, di-phones or tri-phones. The possibilities of the unit are V, CV, VC, VCV, VCCV and VCCCV, where V stands for a vowel and C stands for a consonant [13]. During concatenation the duration rules are applied which leads the database to contain only long vowels. The total number of units is around 1000 [9-12].

E. Grapheme to phoneme conversion

The Grapheme to phoneme conversion in which the form of the word is given as input and thus the letter is converted into sound by using sound rules (That is the Phonetics)

F. Prosody Modelling:

The speech database is developed for prosody modeling.

The work that can be carried in the prosody modeling include following parameters

- Pitch
- Duration
- Intonation of the speech

The quality of the speech is improved by varying parameters the quality refers to the naturalness of the speech and pause that is given by the each word is developed [13].

IV. CONCLUSION

On the above mentioned grounds, the voicing of words of any text in Marathi language is carried out with the help of a limited database set. In this study the framework of a TTS system for Marathi language is built. Although the system uses simple techniques it provides promising results for Marathi language, since the selected approach, namely the concatenative method, is very well suited for Marathi language. The system can be improved by improving the quality of the speech files recorded. In particular, the work on intonation is not finished because segmentation was made manually and there is noticeable noise in voicing. It is planned to apply independent segmentation and to improve the quality of synthesis in the future. The punctuations are removed in the preprocessing step just to eliminate some inconsistencies and obtain the core system. In the future versions of the TTS, the text can be synthesized in accordance with the punctuations for considering the emotions and intonations as partially achieved in some of the researches. The synthesis of a sentence ending with a question mark can have an interrogative intonation and synthesis of a sentence ending with an exclamation mark can be an amazing intonation. In addition to these, other punctuations can be helpful for approximating the synthesized speech to its human speech form such as pausing at the end of the sentences ending with full stop and also pausing after the punctuation comma. Major issues considered in developing TTS are text corpus collection, recording and labeling the speech corpus, deriving letter to sound rules and prosody modeling for Marathi language can be solved and attempts to produce a naturalness in speech.

REFERENCES

[1]. Sangramsing Kayte, Monica Mundada,Dr. Charansing Kayte" Speech Synthesis System for Marathi Accent using FESTVOX" International Journal of Computer Applications (0975 – 8887) Volume 130 – No.6, November2015

[2]. Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte "Implementation of Marathi Language Speech Databases for Large Dictionary" IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 6, Ver. I (Nov -Dec. 2015), PP 40-45e-ISSN: 2319 –4200, p-ISSN No. : 2319 –4197

[3]. Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte "Di-phone-Based Concatenative Speech Synthesis System for Hindi" International Journal of Advanced Research in Computer Science and Software Engineering -Volume 5, Issue 10, October-2015

[4]. Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte "Di-phone-Based Concatenative Speech Synthesis Systems for Marathi Language" OSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 5, Ver. I (Sep –Oct. 2015), PP 76-81e-ISSN: 2319 –4200, p-ISSN No. : 2319 –4197

[5]. Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte "A Review of Unit Selection Speech Synthesis International Journal of Advanced Research in Computer Science and Software Engineering -Volume 5, Issue 10, October-2015

[6]. Sangramsing Kayte "Transformation of feelings using pitch parameter for Marathi speech" Sangramsing Kayte Int. Journal of Engineering Research and Applications ISSN: 2248-9622, Vol. 5, Issue 11, (Part -4) November 2015, pp.120-124

[7]. Sangramsing Kayte "Duration for Classification and Regression Tree for Marathi Text-to-Speech Synthesis System" Sangramsing Kayte Int. Journal of Engineering Research and Applications ISSN: 2248-9622, Vol. 5, Issue 11, (Part-4)November2015

[8]. Sangramsing Kayte, Dr. Bharti Gawali "Marathi Speech Synthesis: A review" International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 3 Issue: 6 3708 – 3711

[9]. Sangramsing Kayte, Monica Mundada "Study of Marathi Phones for Synthesis of Marathi Speech from Text" International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-4, Issue-10) October 2015

[10]. Sangramsing Kayte, Monica Mundada,Dr. Charansing Kayte " Performance Evaluation of Speech Synthesis Techniques for Marathi Language " International Journal of Computer Applications (0975 – 8887) Volume 130 – No.3, November 2015

[11]. Sangramsing Kayte, Monica Mundada, Santosh Gaikwad, Bharti Gawali "PERFORMANCE EVALUATION OF SPEECH SYNTHESIS TECHNIQUES FOR ENGLISH LANGUAGE " International Congress on Information and Communication Technology 9-10 October, 2015

[12]. Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte " Performance Calculation of Speech Synthesis Methods for Hindi language IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 6, Ver. I (Nov -Dec. 2015), PP 13-19e-ISSN: 2319 –4200, p-ISSN No. : 2319 –4197

[13]. Sangramsing N. Kayte, Monica Mundada, Dr. Charansing N. Kayte, Dr.Bharti Gawali "Approach of Syllable Based Unit Selection Text-To-Speech Synthesis System for Marathi Using Three Level Fall Back Technique OSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 6, Ver. II (Nov -Dec. 2015), PP 31-35e-ISSN: 2319 –4200, p-ISSN No. : 2319 –4197

[14]. Sangramsing N.kayte "Marathi Isolated-Word Automatic Speech Recognition System based on Vector Quantization (VQ) approach" 101th Indian Science Congress Jammu University 03th Feb to 07 Feb 2014.

[15]. Monica Mundada, Sangramsing Kayte "Classification of speech and its related fluency disorders Using KNN" ISSN2231-0096 Volume-4 Number-3 Sept 2014

[16]. Monica Mundada, Sangramsing Kayte, Dr. Bharti Gawali "Classification of Fluent and Dysfluent Speech Using KNN Classifier" International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 9, September 2014

[17]. Monica Mundada, Bharti Gawali, Sangramsing Kayte "Recognition and classification of speech and its related fluency disorders" International Journal of Computer Science and Information Technologies (IJCSIT)

[18]. Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte "A Corpus-Based Concatenative Speech Synthesis System for Marathi" IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 6, Ver. I (Nov - Dec. 2015), PP 20-26e-ISSN: 2319 –4200, p-ISSN No. : 2319 –4197

[19]. Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte "A Marathi Hidden-Markov Model Based Speech Synthesis System" IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 6, Ver. I (Nov - Dec. 2015), PP 34-39e-ISSN: 2319 –4200, p-ISSN No. : 2319 –4197

[20]. Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte "Marathi Text-To-Speech Synthesis using Natural Language Processing "IOSR Journal of VLSI and Signal Processing (IOSR-JVSP)Volume 5, Issue 6, Ver. I (Nov -Dec. 2015), PP 63-67e-ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197

[21]. Sangramsing Kayte, Monica Mundada,Dr. Charansing Kayte “Screen Readers for Linux and Windows – Concatenation Methods and Unit Selection based Marathi Text to Speech System” International Journal of Computer Applications (0975 – 8887) Volume 130 – No.14, November 2015

[22]. Sangramsing Kayte, Monica Mundada, Jayesh Gujrathi, “ Hidden Markov Model based Speech Synthesis: A Review” International Journal of Computer Applications (0975 – 8887) Volume 130 – No.3, November 2015

[23]. Sangramsing N. Kayte ,Monica Mundada,Dr. Charansing N. Kayte, Dr.Bharti Gawali "Approach To Build A Marathi Text-To-Speech System Using Concatenative Synthesis Method With The Syllable” Sangramsing Kayte et al.Int. Journal of Engineering Research and Applications ISSN: 2248-9622, Vol. 5, Issue 11, (Part-4) November 2015, pp.93-97

[24]. Sangramsing N. Kayte, Dr. Charansing N. Kayte,Dr.Bharti Gawali* "Grapheme-To-Phoneme Tools for the Marathi Speech Synthesis" Sangramsing Kayte et al.Int. Journal of Engineering Research and Applications ISSN: 2248-9622, Vol. 5, Issue 11, (Part -4) November 2015, pp.86-92

[25]. Sangramsing N. Kayte, Monica Mundada, Dr. Charansing N. Kayte, Dr.BhartiGawali "Automatic Generation of Compound Word Lexicon for Marathi Speech Synthesis" IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 6, Ver. II (Nov -Dec. 2015), PP 25-30e-ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197

[26]. Sangramsing N. Kayte, Monica Mundada, Dr. Charansing N. Kayte, Dr.Bharti Gawali " Artificially Generatedof Concatenative Syllable based Text to Speech Synthesis System for Marathi" OSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 6, Ver. II (Nov -Dec. 2015), PP 44-49e-ISSN: 2319 –4200, p-ISSN No. : 2319 –4197