IOSR Journal of VLSI and Signal Processing (IOSR-JVSP)
Volume 5, Issue 6, Ver. II (Nov -Dec. 2015), PP 44-49
e-ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197
www.iosrjournals.org

Artificially Generatedof Concatenative Syllable based Text to Speech Synthesis System for Marathi

Sangramsing N. Kayte[1], Monica Mundada[1], Dr. Charansing N. Kayte[2], Dr.Bharti Gawali*

*1,*Department of Computer Science and Information Technology Dr. Babasaheb Ambedkar Marathwada University, Aurangabad*

2Department of Digital and Cyber Forensic, Aurangabad, Maharashtra

Abstract: *This research paper addresses the problem of improving the intelligibility of the synthesized speech in Marathi TTS synthesis system. The human speech is artificially generated by Speech synthesis. The normal language text will be automatically converted into speech using Text-to-speech system. This research paper deals with a corpus-driven Marathi TTS system based on the concatenative synthesis approach. Concatenative speech synthesis involves the concatenation of the basic units to synthesize an intelligent, natural sounding speech. In this paper syllables are the basic unit of speech synthesis database and the modification of syllable pitch by time scale modification. The speech units are annotated with associated prosodic information about each unit, manually or automatically, based on an algorithm. An annotated speech corpus utilizes the clustering technique that provides way to select the suitable unit for concatenation, depends on the minimum total join cost of the speech unit. The entered text file is analyzed first, this syllabication is performed based on the linguistics rules and the syllables are stored separately. Then the syllable corresponding speech file is concatenated and the silence present in the concatenated speech is removed. After that discontinuities are minimized at syllable boundaries without degrading the quality. Smoothing at the concatenatedsyllable boundary is performed and changing the syllable pitches by time scale modification.*

Keywords: *Marathi TTS, Concatenative Speech Synthesis, Text to speech synthesis, Syllable based synthesis.*

I. Introduction

Over the past years, there has been an immense development in Speech technologies. Among the applications of speech technology, the automatic speech production, which is referred to as TTS system is the most natural sounding technology[1]. TTS synthesis is the process of converting ordinary orthographic text into speech signal which is indistinguishable from human speech [1][2]. It can be widely classified into front end and back end as shown in Fig.1.The conversion of natural language text to a structured linguistic representation is associated with front end. From the raw text this front end identifies a sequence of segments called target segments. These target segments have a different features estimated from the text. The back end is referred as the second part of the system which modifies these target segments into a speech waveform.

There are two main methods are used for speech production. These methods are format synthesis and concatenation synthesis is illustrated in [3]. The format synthesizer utilizes a simple model of speech generation and a set of rules to generate speech. While these systems can achieve enhanced intelligibility, their naturalness is typically low, since it is very tedious to perfectly describe the process of speech produced in a set of rules. The TTS has been the main research focus automatic speech production in Indian languages nowadays. Some of TTS systems for Indian languages like Hindi, Telugu, Marathi and Bengali have been developed using the unit selection and festival framework in [4] and [5]. Listeners are able to clearly perceive the message

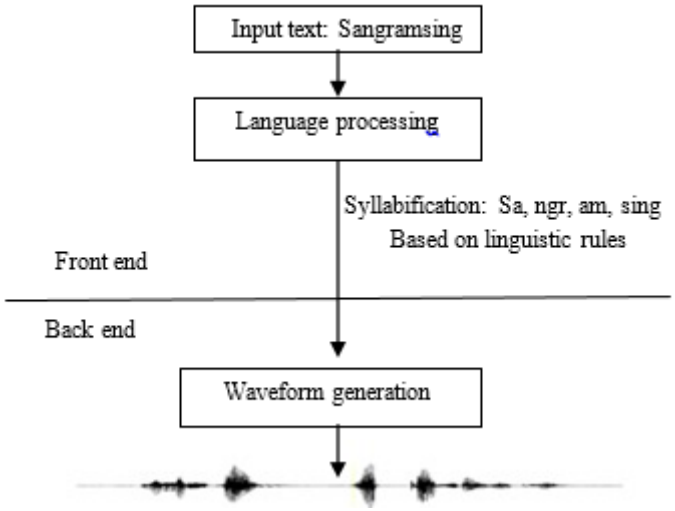

Fig. 1. Parts of speech synthesis system

With little attention, and act on synthesized speech of a command correctly and without perceptible delay in noisy environments. Although many TTS approaches, the intelligibility, naturalness, comprehensibility, and recall ability of synthesized speech is not good enough to be widely accepted by users. There is still considerable rule for further improvement of performance of the text-to speech production system.This paper proposed corpus driven TTS system [6]. In this paper the concatenative-based approach is used to produce desired speech through pre-recorded speech waveforms. Over past decades, this proposal was very complicate to implement because of limitation of computer memory. With the advancements in computer hardware and memory, a large quantity of speech corpus can be stored and utilized to produce high quality speech signal for a given text. Thus, the synthesized speech preserves the naturalness and intelligibility. Here the given input text is analyzed first. Based on the linguistics rules, syllabication is performed. The syllables are called basic speech units. The repository of these units is created with its prosodic information.The pitch value for syllable is changed by performing time scale modification. During the synthesis, these units are selected and concatenated with lowest join cost. After performing concatenation the waveform is smoothened at concatenation joints using LPC[16-24].

II. Concatenative Speech

Synthesis

Concatenative speech synthesis utilizes phones, di-phones, syllables, words and sentences as basic units. Based on selecting these units from the database speech are synthesized, called as a speech corpus. Many researches have been made, selecting each separate unit as the basic unit. When phones are selected as basic units, for Indian languages the size of the database will be less than 50 units. Database may be small, but phones gives very poor co-articulation data's across neighboring units, thus falling to model the dynamics of speech sounds. Di-phones and tri-phones as basic units, it will minimize the discontinuities at the concatenation points and captures the co-articulation effects. But a single example of each di-phone is not enough to generate precious quality speech. So this paper presented a syllable as a basic unit. Indian languages are syllable centered, where pronunciations' are based on syllables.In Marathi, vowels are added to consonants. Most of them are easy to pronounce, ऋ is slightly challenging. Marathi vowels retain much of their original Sanskrit pronunciation making some of them different from their Hindi counterparts. A notable example is औ (au), pronounced as owl in Marathi but as Oxford in Hindi. ऑ (Ao) is a special vowel used for loan English words, andis pronounced as in doctor[16-24].

Aspiration means with a puff of air, and is the difference between the sound of the letter p in English pin (aspirated) and spit (unaspirated). Retroflex consonants, on the other hand, are not really found in English. They should be pronounced with the tongue tip curled back. Practice with a native speaker, or just pronounce as usual — you'll usually still get the message across. There are 36consonants and 15 vowels in Marathi languages. There are defined set of syllabification rules formed by researchers, to generate computationally reasonable syllables. Some of the rules used to perform grapheme to syllable conversion [7] are:

- Nucleus can be Vowel (V) or Consonant (C)
- If onset is C then nucleus is V to yield a syllable of type CV
- Coda can be empty of C

- If character after CV pattern are of type CV then the syllables are split as CV and CV
- If the CV pattern if followed by CCV then syllables are split as CVC and CV
- If CV pattern is followed by CCCV then the syllables are split as CVCC and CV
- If the VC pattern is followed the V then the syllables are split as V and CV
- If the VC pattern is followed by CVC then the syllables are split as VC and CVC

The new rules have been implemented in grapheme to syllable conversion

- If character after CV pattern are of type CV then the syllables are split as CVCV
- Similarly If character after CV pattern are of type CVCV then the syllables are split as CVCVCV
- If the CV pattern if followed by CVC then syllables are split as CVCVC
- If the CV pattern if followed by CCV then syllables are split as CVCCV

This recommended combinations to achieve the best acceptable for synthesis is:

- Monosyllables at the beginning of a word and bisyllables at the end.
- Bisyllables at the beginning of a word and monosyllables at the end.
- Monosyllables at the beginning and trisyllables at the end of a word.
- Trisyllables at the beginning and monosyllables at the end of a word.

III. Proposed Marathi Tts Synthesis System

3.1 Text analysis

To implement the proposed TTS system, the MATLAB 2014 has been used. In text analysis, first stage is text normalization then performs removing of punctuations such as double quotes, full stop, and comma. A pure sentence is synthesized at the end of text analysis. Then all the abbreviations present in the input text are expanded and also unwanted punctuation like (:,; ‘ $ `) etc. are removed to avoid confusion and not to give any disturbance in the naturalness of the speech. The next step in the text normalization is normalizing nonstandard words like abbreviations and numbers. The next stage in the text analysis is sentence splitting. In this stage, the given paragraph will be spitted as sentences. From these sentences, words are separated out. The last stage is Romanization which is the representation of written words with a roman alphabet. In this system Romanized form of Marathi word/syllables are generated[16-24].

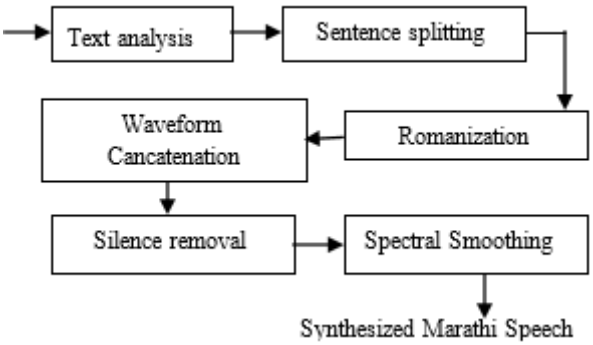

Fig. 2 Block diagram of proposed Marathi text to speech synthesis system

3.2. Speech corpus

Building a speech corpus for Indian languages is a difficult task than that of English speech corpus. Prosodic information such as pitch, duration and intonation prediction has to be done in the corpus development stage itself, some more information has to be specified with the basic speech units after storing them in the corpus. The problem such as mispronunciation, untranscribed speech units, phrase boundary detection, pronunciation variants are to be identified and addressed. For corpus creation we selected one person for recoding these basic units, who has uniform characteristics of speaking, pitch rate and energy profile and developed speech corpus in [8]. The digitized speech signal with sampling rate of 16 KHz and 16-bit resolution (Pulse Code Modulation uncompressed data format) proposed in [8]. The speech wave files are saved according to the requirement. The speech wave files corresponding to the Marathi words are named according to their corresponding Romanized names. The words collected comprises dictionary words, commonly used words, Marathi newspapers and story books, also different domain such as sports, news, literature and education for building unrestricted TTS initiated in [9] [10][16-24].

3.3 Waveform concatenation

In the final stage of the concatenation process, the required syllables are retrieved from the corpus based on the text analysis and arranged to produce the speech. Then all the arranged speech units are concatenated using a concatenation algorithm. The main problem in concatenation process is that there will be glitches in the joint. These are removed in the waveform smoothening stage. The concatenation process combines all the speech files which are given as an output of the unit selection process and then making in to a single speech file.

3.4 Spectral smoothing

The time scale modification is carried out for each syllable to produce individual smoothness for syllable in Marathi TTS. The time scale modification is used to change the pitch value for Marathi syllable. Praat software is used to calculate the Duration value for each syllable [11]. Smoothing at concatenation joints are performed using Linear Predictive Coding (LPC). The LPC is used for representing the spectral envelope of a digital signal of speech using the information of a linear predictive model. It is one of the most powerful methods for encoding enhanced quality speech at a low bit rate and gives extremely accurate estimates of speech features in [11],[12].Now we are getting the improved quality speech for the given input text. It can be played and stopped any where needed. The main aim of the proposed scheme is to achieve good naturalness in output speech. Fig.3 shows the smoothened output waveform[16-24].

IV. Quality Test

For developing a Marathi TTS, we have considered 1000 sentences for recording the speech corpus. These are selected from various domains from newspaper, Wikipedia, news broadcast and story books. After formulating the data, speech corpus will be recorded in a studio environment with a suitable trained speaker. Recorded 1000 sentences of speech corpus is classified into two parts: (i) part-1 contains 1000 sentences for building the training corpus. (ii) part-2 contains 500 sentences for evaluating the established Marathi TTS system. One speakers' voices are selected for constructing prototype TTS systems. From this proposed work the following performance are inferred that the speech of five speakers with respect to

- The quality of the synthesized speech
- Variations in natural prosody and
- The perceptual distortion with respect to prosodic and spectral modifications.

Evaluation of quality of the synthesized speech is carried out by subjective measures. An intelligibility and naturalness are estimated from the listening tests. Tests are conducted with 10 research scholars in the age group of 23–28 years. The subjects have sufficient speech knowledge for proper assessment of the speech signals, as all of them have taken a full semester course on speech technology. Speech utterances corresponding to the test sets are synthesized using the developed Marathi TTS system. Each of the subjects was given a pilot test about perception of speech signals by playing the original speech samples of the test files.

Once they are comfortable with judging, they are allowed to take the tests. The tests are conducted in the laboratory environment by playing the speech signals through headphones. In the test, the subjects were asked to judge the distortion and quality of the speech. Subjects are asked to assess the quality and distortion on a 5-point scale for each of the sentences. The 5-point scale for representing the quality of speech and the distortion level is given in Table 1. For evaluating the quality of synthesized speech generated from the developed TTS system, there are three sets of test utterances are considered. Each set consists of 20 sentences.

Set-1: All the words are available from the training data, but the entire word sequences are not present in the training data. Set-2: 50% of the words available in training corpus. Set-3: none of the words are available in the training corpus. Table 2 shows the MOS scores for the three test sets. From the table 2 it is observed that, MOS for set-1 is more compared set-2 and set-3 as all the words of sentences in set-1 is present in database. So it provides the better performance compared to other two sets[13][14][15] [16-24].

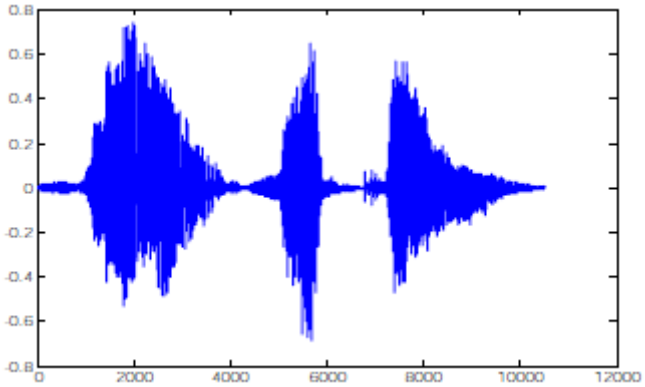

Fig.3. Resulting concatenated waveform after performing silence removal

Table 1. Instructions to evaluators

Score	Subjective perception
1	Poor speech, with distortion and very low intelligibility
2	Poor speech with distortion and intelligible
3	Good speech with less distortion and intelligibility
4	Very good speech quality with less naturalness
5	As good as natural speech

Table 2. Mean opinion scores for three sets

Test Set	MOS
I	5
II	5
III	4

V. Conclusion

In this proposed word, a speech synthesis system has been designed and implemented for Marathi language. A database has been created from various domain words and syllables. Syllable pitch modification is performed based on time scale modification. The speech files present in the corpus are recorded and stored in PCM format in order to retain the naturalness of the synthesized speech. The given text is analyzed and syllabication is performed based on the rules specified. The desired speech is produced by concatenative speech synthesis approach such that spectral discontinuities are minimized at unit boundaries. It is inferred that the produced synthesized speech is preserving naturalness and good quality based the subjective quality test results. The final output speech file is stored in the specified location in the system for further analysis.

References

[1]. 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.
[2]. 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)
[3]. 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
[4]. 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
[5]. 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
[6]. 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
[7]. 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
[8]. 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
[9]. 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

[10]. 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
[11]. Monica Mundada, Sangramsing Kayte "Classification of speech and its related fluency disorders Using KNN" ISSN2231-0096 Volume-4 Number-3 Sept 2014
[12]. 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
[13]. 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
[14]. 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
[15]. 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
[16]. 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
[17]. 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
[18]. 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
[19]. Sangramsing Kayte, Monica Mundada, JayeshGujrathi, " Hidden Markov Model based Speech Synthesis: A Review" International Journal of Computer Applications (0975 – 8887) Volume 130 – No.3, November 2015
[20]. 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
[21]. 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
[22]. 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
[23]. 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
[24]. 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