IOSR Journal of VLSI and Signal Processing (IOSR-JVSP)
Volume 5, Issue 6, Ver. I (Nov -Dec. 2015), PP 63-67
e-ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197
www.iosrjournals.org

Marathi Text-To-Speech Synthesis using Natural Language Processing

Sangramsing Kayte[1], Monica Mundada[1, 2]Dr. CharansingKayte
[1, 1]*Department of Computer Science and Information Technology*
Dr. BabasahebAmbedkarMarathwada University, Aurangabad
[2]*Department of Digital and Cyber ForensicAurangabad, Maharashtra, India*
bsangramsing@gmail.com , monicamundada5@gmail.com,charankayte@gmail.com

Abstract: *Research on Text-to-speech technology has received the interest of professional researchers in many languages which is a consequence of wide range of applications where Text-To-Speech is implemented. However, Maharashtra Marathi pronunciation: locally is a state in the western region of India and is the nation's third largest state and also the world's second-most populous sub-national entity. It has over 110 million inhabitants and its capital, Mumbai, has a population of approximately 18 million. Nagpur serves as second capital as well as winter capital of the state. Maharashtra's business opportunities along with its potential to offer a higher standard of living attract migrants from all over India. This paper is a development of a rule based text- to- speech synthesis system which is a combination formant and concatenation techniques with acceptable naturalness. The simulation results of the system shows good quality in handling word, phrase, and sentence level compared to other available Marathi TTS systems. The accuracy of the overall system is 91%. Further improvements need to be done for stressed syllable position and intonation.*

Keywords: *Rule-based text-to-speech (TTS); speech synthesis; Marathi phonology.*

I. Introduction

Speech is the primary means of communication between people. Speech synthesis is an automatic generation of speech waveforms. The dream of producing a talking machine started at the 18th century while recent progress in speech synthesis has produced synthesizers with high intelligibility but the sound quality and naturalness still remain a major problem. This fact makes speech synthesis an important field for investigation and improvement for the major languages including Marathi [1].

Speech synthesizer or as it known Text-to-Speech system is one of the important technology in the current time due to the expanding field of applications, it is used in multimedia applications to read e-mail, mobile messages, or in any kind of human-machine interaction. It is helpful and common among visually impaired people as a simple reading machine, gives the deafened and vocally handicapped an opportunity to communicate with people who do not understand the sign language. It can be used also in many educational tasks like spelling and pronunciation teaching aid for different languages. Maharashtra Marathi pronunciation: locally is a state in the western region of India and is the nation's third largest state and also the world's second-most populous sub-national entity. It has over 110 million inhabitants and its capital, Mumbai, has a population of approximately 18 million. Nagpur serves as second capital as well as winter capital of the state. Maharashtra's business opportunities along with its potential to offer a higher standard of living attract migrants from all over India [2].

Ancient and medieval Maharashtra included the empires of the Satavahana dynasty, Rashtrakuta dynasty, Western Chalukyas, Mughals and Marathas. Spread over 118,809 sq mi (307,710 km2), it is bordered by the Arabian Sea to the west and the Indian states of Karnataka, Telangana, Goa, Gujarat, Chhattisgarh, Madhya Pradesh and the Union territory of Dadra and Nagar Haveli. The major rivers of the state are Godavari, Krishna, Narmada and Tapi. The state has several tourist destinations including the popular Hindu places of pilgrimage, Pandharpur, Dehu and Alandi. Places with wide appeal include Hazur Sahib Nanded at Nanded, and Saibaba shrine at Shirdi. Maharashtra is the most urbanized state in India, with large cities besides the capital Mumbai such as Pune, Nagpur, Nashik and Aurangabad [3].

So, it's an important issue to build Marathi TTS which is reliable, intelligent and user friendly system to give those people a chance to use the technologies like text messages, emails, and web sites using their native language.

The text-to-speech (TTS) synthesis procedure consists of two main phases. The first one is text analysis, where the input text is transcribed into a phonetic or some other linguistic representation, and the second one is the generation of speech waveforms, where the acoustic output is produced from this phonetic and prosodic information. The main techniques used in speech synthesis design are Articulator synthesis, Formant

synthesis, and Concatenative synthesis. Articulatory synthesis attempts to model the human speech production system directly. Formant synthesis, which models the pole frequencies of speech signal or transfer function of vocal tract based on source-filter-model. Concatenative synthesis, which uses different length pre-recorded samples derived from natural speech [10].

In theory, the most accurate method is articulatory synthesis which models the human speech production system directly, but it is also the most difficult approach. So, the available TTS systems mostly use either concatenative or formant synthesis technique. Each technique has its own points of strength and weakness and suits a specific language while doesn't for others [4],[5] [7][11].

An interesting approach is to use a hybrid system where the formant and concatenative methods have been applied in parallel to phonemes where they are the most suitable [6]. In general, combining the best parts of the basic methods is a good idea, but in practice, controlling of synthesizer may become difficult [7][10].

II. Marathi Phonetic System

The phonetic system of Modern Standard Marathi has basically 71 phonemes, which consists of Marathi language has 43 consonants and 28 vowels phonemes.

The phoneme inventory of Marathi is similar to that of many other Indo-Aryan languages. An IPA chart of all contrastive sounds in Marathi is provided below.

Table 2: Consonant phoneme inventory[2][3][12]

Consonants							
	Labial	**Dental**	**Alveolar**	**Retroflex**	**(Alveolo) palatal**	**Velar**	**Glottal**
Nasal	**plain**	m	n̪		ɳ	(ŋ)	(ɳ)
	murmured	mʱ	n̪ʱ		ɳʱ		
Stop	**voiceless**	p	t̪	t͡s	ʈ	t͡ɕ~t͡ʃ	k
	aspirated	pʰ	t̪ʰ		ʈʰ	t͡ɕʰ~t͡ʃʰ	kʰ
	voiced	b	d̪	d͡z~z	ɖ	d͡ʑ~d͡ʒ	g
	murmured	bʱ	d̪ʱ	d͡zʱ~zʱ	ɖʱ	d͡ʑʱ~d͡ʒʱ	gʱ
Fricative			s	ʂ	ɕ~ʃ		h~ɦ
Approximant	**plain**	ʋ		l	ɭ	j	
	murmured	ʋʱ		lʱ		(jʱ)	
Flap/Trill	**plain**			ɾ	ɭ̆		
	murmured			ɾʱ			

Older aspirated *tsʰ, dzʱ have lost their onset, with *tsʰ merging with /s/ and *dzʱ being typically realized as an aspirated fricative, [zʱ]. This /ts, dz, zʱ/ series is not distinguished in writing from /tʃ, tʃʰ, dʒ, dʒʱ/.

A. Syllables

The syllabic structures in Marathi are limited in number and easily detectable. Unlike other languages, every syllable in Marathi begins with a consonant followed by a vowel which is called the nucleus of the syllable. Short vowels are denoted by (V) and long vowels are denoted by (VV). It is obvious that the vowel exists in the second place of the syllable. These features facilitate the process of syllabification [2][3].

Marathi syllables can be classified either according to the length of the syllable or according to the end of the syllable. Short syllable occur only in CV form, because it is ending with a vowel so it is open. Medium syllable can be in the form of open CVV, or closed CVC. Long syllable has two closed forms CVVC, and CVCC [8][2][14].

Marathi words contains at least one syllable, most contain two or more syllables. The longest word is combined of five syllables. Table II illustrates Marathi syllables. Some of the Marathi words are spelled together forming new long words with 6 syllables like ( ), or 7 syllables like ( ).

Table Ii. Marathi Syllables Types [2][3][13]

Devanagari	अ	आ	इ	ई	उ	ऊ	ऋ	ए	ऐ	ओ	औ	अं	अः
Transliterated	a	ā	i	ī	u	ū	ṛ	e	ai	o	au	aṃ	aḥ
IPA	/ə/	/a/	/i/	/u/	/ru/	/e/	/əi/	/o/	/əu/	/əm/	/əhə		
Syllable type	CV	CV	CVV	CVV	CV	CVC	CVC	CVC	CV	CV	CVCC	CVCC	CVCC

III. System Description

The complete design of the proposed synthesis system is shown in Fig. 1. The Text-To-Speech system includes preprocessing stage for text normalization. Natural language processing (NLP) decides the stress

pattern and converts the letters to its equivalent phoneme representation. DSP stage produces speech out of the resultant phonemes.

A. **Preprocessing Stage**

Preprocessing phase is required to prepare the raw text to be suitable to enter the processing stage. It detects the words of each sentence, spaces between them, punctuation marks, or any other non-Marathi symbols

B. **Natural Language Processing**

This phase relies on the developed data base to map each word to its exact phonemic representation equivalent. The data base is formed of three main parts:

Exceptions Dictionary: It contains list of words whose pronunciations are given explicitly, rather than determined by the Pronunciation Rules. If the Pronunciation rules are applied to a word and indicate a standard prefix or suffix, then the word is again looked up in Exception Dictionary List after the prefix or suffix has been removed. Rather than a full pronunciation, just the stress flag may be given, to change where it would be otherwise placed by the Pronunciation Rules. A pronunciation may also be specified for a group of words, when these appear together. Up to four words may be given,

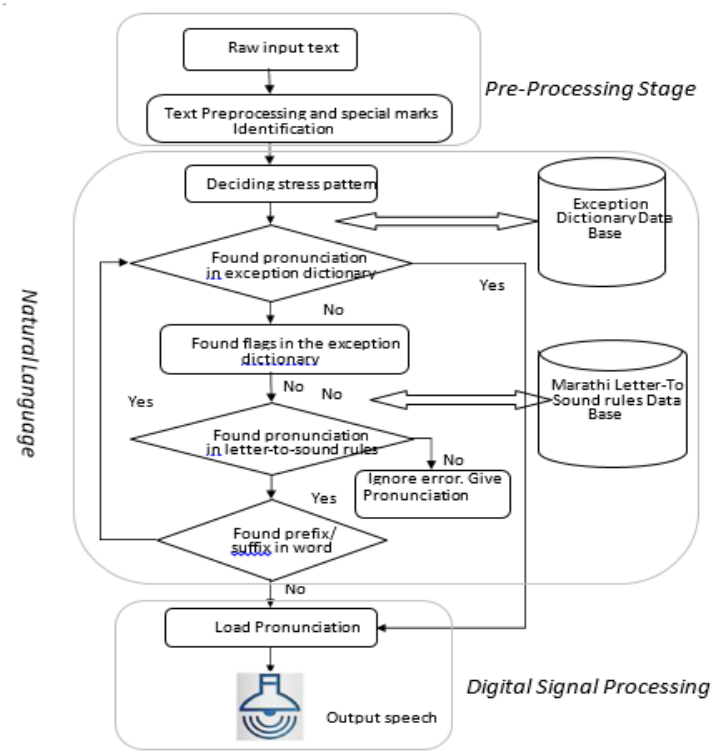

Figure 1. Marathi Synthesis System Structure

Enclosed in brackets. This may be used for change the pronunciation or stress pattern when these words occur together, run them together, pronounced as a single word, or to give them a flag when they occur together. Common set of words mostly need no vowelization [8][2].

2) Marathi Pronunciation Rules (letter-to-sound): This part of the data base specifies the phonemes which are used to pronounce each letter, or sequence of letters. Some rules only apply when the letter or letters are preceded by, or followed by, other specified letters [9], [10]. To find the pronunciation of a word, the rules are searched and any which match the letters in the word will be given a score depending on how many letters are matched. The pronunciation from the best matching rule is chosen. The pointer into the source word is then advanced past those letters which have been matched and the process is repeated until all the letters of the word have been processed. It is important to mention that the Exception Dictionary is searched first while it contains the words with special pronunciation.

3) Marathi phonemes: Defines all the phonemes which are used by Marathi language, together with their properties and the data for their production as sounds. Vowels are generated based on formant synthesis, while special Marathi consonants like: ($ # " !) are pre-recorded wav files based on concatenation synthesis.

Marathi words that are not found in the exception dictionary with their stress flag will follow a special pattern for syllable lexical stress. It is designed to apply stress according to a set of rules. Syllable has three lexical stress levels: primary (1), secondary (2), and weak or no stress (3), as reported in [7] and [11] the rules that

determine the stress are: • When a word is made up of a string of the CV type syllables, the first syllable receives the primary stress and the remaining syllables receive no stresses, e.g. (&'% : '(1))(3)(3) : CV(1)CV(3)CV(3) "he wrote"

• When a word contains only one long syllable, the long syllable receives the primary stress and the rest of the syllables go unmarked, receiving no stresses. The final long syllable never receives a primary stress.(&)'% : '(1)&)(3) CVV(1)CVC(3) "writer"

• For polysyllabic word, a stress is placed on the first long syllable counting from the penultimate. The nearest long syllable to the beginning of the word receives the secondary stress. (-) *,+ % :
, (3).(2)/(3)(1)0 (3) (3)1(3)
CVC(3)CV(2)CV(3)CVV(1)CV(3)CVC(3) "their savings"

C. Digital Signal Processing

In this part of the synthesis system the resultant phonemic representation of the input text with the special stress and pause flags will be transformed into the proper utterance and can be saves as a wave file. Table III shows some Marathi words and their resultant phonemic representation and stress flags taken from our system [8][2].

IV. Testing And Results

The simulation results shows 91% accuracy for the phoneme representation Letter-To-sound, all the diacritics of the standard Marathi script including were successfully mapped into their equivalent pronunciation rules. The words with special pronunciation were listed in the exception data base, since that data base is checked before the rules data base, the system ensures error free phoneme representation. In order to assess the quality of our output speech, a subjective test was performed. A set of 25 phonetically balanced sentences was used as the test material. The test sets were played to ten volunteer listeners who were asked to rate the system intelligibility, naturalness and overall voice quality on a scale of 1 to 5. The volunteer listeners were asked to give the rating based on how good they thought the sentences were without any further definition of the word "good". The listeners were encouraged to use the full five point scale. The average scores obtained are 4.2, 3.7, and 3.8 for intelligibility, naturalness and overall voice quality respectively [10][11].

TABLE 3. Phonemic representation of Marathi WORDS [9]

English	Marathi	IPA	Proposed System
Hello	2		/m'a2RHa2ba2n/
Goodbye	3 4 56	/maassala:mah/	/ma2?a2s1sa2l'a2:ma2/
House	7	/bai:t/	/ba2jt/
Thank you	8 9	/	/S'UkRa2n11la2k/
Morning	:2;	/saba:</	/s2a2ba2:H/
Noon	=	/ðuhr/	/D2Uhr/

V. Conclusion

A rule-based synthesis Marathi TTS system was developed. Phonemes were the essential elements of the synthesizer, our Marathi TTS system is vocabulary independent with intelligible output speech, so it can handle all types of input text. It has the flexibility of changing the speaker from male to female and other sound variants like whispering. The standard Marathi text is mostly vowelized; hence the need of vowelization, our system omits the need to some of the vowelization symbols like sukoon and has the ability to enrich the exception dictionary by listing the exact pronunciation of the common words, here, no need to vowelize them. Comparing with other available Marathi TTS systems, our Hybrid TTS has small size, high accuracy, and vocabulary independence features which make it in general more reliable than other TTS systems. The system is free for distribution and for development.

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] Sangramsing Kayte, Monica Mundada, Dr. CharansingKayte "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

[3] 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

[4] S. Lemmetty, Review of Speech Synthesis Technology, Master Thesis. Helsinki University of Technology, 1999.

[5] M. G. Fatima Chouireb, "Towards a high quality Arabic speech synthesis system based on eural networks and residual excited vocal tract model," Springer, Oct. 2008.

[6] G. Fries, "Hybrid time- and frequency-domain speech synthesis with extended glottal source eneration," in IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 1, 1994, pp. 581-584.
[7] 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
[8] Sangramsing Kayte, Monica Mundada, Dr. CharansingKayte "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 www.iosrjournals.org
[9] 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
[10] Sangramsing Kayte, Monica Mundada, Dr. CharansingKayte "A Review of Unit Selection Speech Synthesis International Journal of Advanced Research in Computer Science and Software Engineering -Volume 5, Issue 10, October-2015
[11] 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
[12] 13)SangramsingKayte, 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
[13] 14)SangramsingKayte, 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
[14] 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