International Journal of
Emerging Research in Management &Technology
ISSN: 2278-9359 (Volume-4, Issue-10)

Research Article | **October 2015**

Study of Marathi Phones for Synthesis of Marathi Speech from Text

Sangramsing Kayte, Monica Mundada
Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University,
Aurangabad, Maharashtra, India

Abstract—

Text-To-Speech is the automatic production of speech, through a grapheme-to-phoneme transcription of the sentences to utter. This paper describes the study of marathi phones which considers the transformation of text into respective speech signal. The phonemes helps to build the database which can be considered further for building of TTS. The collection of phonemes for particular langauge makes the basic architecture for the system.

Keywords-TTS, speech corpus, Marathi phonemes.

I. INTRODUCTION

Text to Speech (TTS) is a system in which sequence of words are taken as input and converts them to speech. In conversion process of speech synthesis method, vowels and consonants are most important in marathi language [1]. Each phonemes are made of combination of consonants and vowels. There are different concatenation methods like unit selection,diaphone or domain specific method for speech synthesis. In all these methods the voices are sampled from real recordedspeech and speech synthesis is handled by computers. Here researchers main focus is to designing a database of phonemes in such a way that it speedup searching and retrieving process in marathi TTS.

II. LITERATURE REVIEW OF SPEECH SYNTHESIS TECHNIQUES

A. Concatenate Synthesis

Concatenate speech synthesis is a method where speech is generated by concatenating speech units one after the other as per the requirement. There are three different types of concatenate speech synthesis they are Domain specific synthesis, Di-phone synthesis and Unit selection synthesis [2][9][10] The focus of the paper is Unit selection Synthesis.

In this method the database is built up with all phones present in the particular language. The design of such database includes well labeled phones with high quality utterances. The synthesized speech output signal generated with the concatenated parts from the database [2],[8],[10]. The synthesized speech naturally sounds equal to the recorded utterances with respect to intonation, emotion and style. Nevertheless unit selection synthesis has shown itself to be capable of producing high quality natural sounding synthetic speech when constructed from large databases of well-ordered and well-labeled speech. The USS(Unit Selection Synthesis) extracts the prosodic and spectral part from input speech signal during the training part. In synthesis part, the analysis of text is done and prosody is incorporated with the use of algorithm and artificial speech is produced [2]. In USS, intially the text is converted into phones of the particular segment. Then the phones are assigned the labels like vowels, semivowels and consonants. With the help of acoustic trees the ID's are generated for the given input. At the final step, the speech is synthesized with the USS algorithm with the incorporation of the needed prosody elements.

B. Hidden Markov Model Based Speech Synthesis

Hidden Markov Model synthesis is also called statistical parametric synthesis of speech. It is similar to the unit selection synthesis, trained with natural speech, however due to the parametric model approach; it allows a better modeling of variation. The work can be divided into two main parts which represent the topology of the synthesizer itself. For the synthesizer parametric models have to be trained first in order to estimate parameters from the models. The HTS system simultaneously models excitation and duration of speech using context dependent hidden markov model 's and thus generate speech waveforms from HMM's themselves [3]. The main advantages of the referred hidden markov model – based synthesis techniques is that the voice alteration can be performed without large databases, being at par with quality[2].

In the training part, spectrum and excitation parameters are extracted from speech database and modeled by context dependent hidden markov model. In the synthesis part, context dependent hidden markov model are concatenated according to the text to be synthesized. Then spectrum and excitation parameters are generated from the hidden markov model by using a speech parameter generation algorithm. Both the Techniques are implemented using Festival framework [4].

III. FEATURES OF MARATHI CHARACTERS

Marathi language uses Devanagari, a characterbased script. A character represents one vowel and zero or more consonants. Consonant clusters are represented by combination of ligatures; so, there are hundreds of characters [5]. Many characters share glyphs. To take advantage of such shared glyphs,some true type font designers have used codes corresponding to glyphs rather than phonemes. This makes the task of rendering Marathi text in Devanagari script easier. However, this causes problems for us, since we are interested in statistical analysis of phonemes. The grapheme-to-phoneme mapping of text in true type fonts is not straightforward, especially because of non-linear nature of Devanagari script. For example, the word "ki" is a sequence two phonemes: /k/ and /i/. However, the corresponding character in Devanagari script has the glyph corresponding to phoneme.There are 50 Marathi phonemes [3]. About 2/3rd of the sentence sets contain at least one token ofeach and every Marathi phoneme. Even the least phonetically rich sentence set contains at least 45 phonemes out of 50. It would be interesting to discuss about the phonemes that did not occur in some sentence set. Out of 10 rare phonemes, 6 are asphirated stop consonants; 2 are retroflex sounds; the rest are velar and palatal nasals. Such a tilt of the distribution of rare phonemes towards asphirated stops is to be expected since asphirated phonemes are known to occur rarely in natural text. There are 10 asphirated stop consonants in Marathi. So, it is difficult to have all of the 10 rare phonemes in addition to other phonemes in just 10 sentences. The most rare phoneme is the velar nasal (/ng/); it does not occur in 64 out of 1000 sentence sets.

IV. MARATHI CONSONANTS/VOWELS WITH ASCII VALUE

TABLE I. THE TABLE BELOW SHOWS THE ASCII VALUE FOR THE VOWELS AND CONSONANTS

Vowels

Character Name	**Character**	**ASC-II value**
DEVANAGARI LETTER A	अ	0905
DEVANAGARI LETTER AA	आ	0906
DEVANAGARI LETTER I	इ	0907
DEVANAGARI LETTER II	ई	0908
DEVANAGARI LETTER U	उ	0909
DEVANAGARI LETTER UU	ऊ	090A
DEVANAGARI LETTER	ऋ	090B
VOCALIC R	ऌ	090C
DEVANAGARI LETTER	ऍ	090D
VOCALIC L	ऎ	090E
DEVANAGARI LETTER	ए	090F
CANDRA E	ऐ	0910
DEVANAGARI LETTER	ऑ	0911
SHORT E	ऒ	0912
DEVANAGARI LETTER E	ओ	0913
DEVANAGARI LETTER AI	औ	0914
DEVANAGARI LETTER	ॠ	0960
CANDRA O	ॡ	0961
DEVANAGARI LETTER		
SHORT O		
DEVANAGARI LETTER O		
DEVANAGARI LETTER AU		
DEVANAGARI LETTER		
VOCALIC RR		
DEVANAGARI LETTER		
VOCALIC LL		

Consonants

TABLE II. TABLE MARATHI CONSONANTS WITH THEIR ASCII VALUE

Character Name	**Character**	**ASC-II value**
DEVANAGARI LETTER	क	0915
KA	ख	0916
DEVANAGARI LETTER	ग	0917
KHA	घ	0918
DEVANAGARI LETTER	ङ	0919
GA	च	091A
DEVANAGARI LETTER	छ	091B
GHA	ज	091C
DEVANAGARI LETTER	झ	091D
NGA	ञ	091E
DEVANAGARI LETTER	ट	091F
CA	ठ	0920
DEVANAGARI LETTER	ड	0921
CHA	ढ	0922
DEVANAGARI LETTER	ण	0923

JA	त	0924
DEVANAGARI LETTER	थ	0925
JHA	द	0926
DEVANAGARI LETTER	ध	0927
NYA	न	0928
DEVANAGARI LETTER	ऩ	0929
TTA	प	092A
DEVANAGARI LETTER	फ	092B
TTHA	ब	092C
DEVANAGARI LETTER	भ	092D
DDA	म	092E
DEVANAGARI LETTER	य	092F
DDHA	र	0930
DEVANAGARI LETTER	ऱ	0931
NNA	ल	0932
DEVANAGARI LETTER	ळ	0933
TA	ऴ	0934
DEVANAGARI LETTER	व	0935
THA	श	0936
DEVANAGARI LETTER	ष	0937
DA	स	0938
DEVANAGARI LETTER	ह	0939
DHA	क़	0958
DEVANAGARI LETTER	ख़	0959
NA	ग़	095A
DEVANAGARI LETTER	ज़	095B
NNNA	ड़	095C
DEVANAGARI LETTER	ढ़	095D
PA	फ़	095E
DEVANAGARI LETTER	य़	095F
PHA		
DEVANAGARI LETTER		
BA		
DEVANAGARI LETTER		
BHA		
DEVANAGARI LETTER		
MA		
DEVANAGARI LETTER		
YA		
DEVANAGARI LETTER		
RA		
DEVANAGARI LETTER		
RRA		
DEVANAGARI LETTER		
LA		
DEVANAGARI LETTER		
LLA		
DEVANAGARI LETTER		
LLLA		
DEVANAGARI LETTER		
VA		
DEVANAGARI LETTER		
SHA		
DEVANAGARI LETTER		
SSA		
DEVANAGARI LETTER		
SA		
DEVANAGARI LETTER		
HA		
DEVANAGARI LETTER		
QA		
DEVANAGARI LETTER		
KHHA		

DEVANAGARI LETTER GHHA DEVANAGARI LETTER ZA DEVANAGARI LETTER DDDHA DEVANAGARI LETTER RHA DEVANAGARI LETTER FA DEVANAGARI LETTER YYA		

IV. RECORDING OF UTTERANCES OF MARATHI PHONEMES

Speech for each phoneme should be recorded from the native speakers of the language [7]. The recorded sound files are then named and stored by using the phoneme name itself. For example the sound file of /ka/ (क) is named ka.wav. All the sound files recorded are named and stored in the similar way using PRAAT tool. Each uttered sound is recorded at sampling rate of 44100Hz, 16 bit Stereo quality format.

V. CONCLUSION

One of the factors regarding success of a good TTS system depends upon how speech database is organized. Here authors have construct speech database in such a way that speech units are associated with phoneme and its ASCII code. Author found two benefits by applying this methodology:
1] phoneme matching and speech unit searching becomes very easy.
2] As Indian languages are phonetic in nature, one can easily apply same methodology for other Indian languages. This research will be useful as a base to recognise Marathi text pattern in speech conversion process.

REFERENCES

[1] Taylor P., Text to Speech Synthesis. University of Cambridge, 2007, DRAFT available: http://mi.eng.cam.ac.uk/pat40/ttsbook_draft_2.pdf, Retrieved on Feb 01, 2015.
[2] Sangramsing Kayte, Kavita waghmare ,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.
[3] T.Yoshimura, K.Tokuda, T. Masuko, T. Kobayashi and T. Kitamura,“Simultaneous Modeling of Spectrum, Pitch and Duration in HMM-Based Speech Synthesis”In Proc. of ICASSP 2000, vol 3, pp.1315-1318, June 2000.
[4] A. Black, P. Taylor, and R. Caley, “The Festival Speech Synthesis System System documentation Edition 1.4, for Festival Version 1.4.3 27th December 2002.
[5] Dan T K, Datta A K & Mukherjee, B, “Speech synthesis using signal concatenation”, J ASI, vol. XVIII (3&4), 1995, pp 141-145.
[6] Kishore S. P., Kumar R & Sanghal R, “A data driven synthesis approach for Indian language using syllable as basic unit”, Proc ICON 2002, Mumbai, 2002.
[7] Agrawal S. S. 2010, “Recent Developments in Speech Corpora in Indian Languages: Country Report of India”, O-COCOSDA, Nepal.
[8] Monica Mundada, Bharti Gawali, Sangramsing Kayte "Recognition and classification of speech and its related fluency disorders" Monica Mundada et al, / (IJCSIT) nternational Journal of Computer Science and Information Technologies, Vol. 5 (5) , 2014, 6764-6767
[9] 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
[10] Sangramsing Kayte, Kavita Waghmare, 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

