

SYLLABLE BASED INDIAN LANGUAGE TEXT TO SPEECH SYSTEM

Sanghamitra Mohanty
Department of Computer Science and Application,
Utkal University, Bhubaneswar – 751004
sangham1@rediffmail.com

ABSTRACT

This paper describes the Text-To-Speech System for four of the Indian Languages namely, Hindi, Odiya, Bengali and Telegu. In this system we have taken care of the articulation as well as co-articulation system. Earlier we have developed the Text-To-Speech (TTS) system for Odiya and Hindi. For this we have selected the phonetic units of the languages from many spoken vocabulary and then have concatenated the appropriate phones to get the TTS systems. This has inherited the problem of concatenation junctions and time delay in concatenating larger units in case of generating polysyllables. To overcome these problems we have made the syllable based concatenation of the speech units for the Odiya TTS system. Observing the success we have tried this for the other three languages like Hindi, Bengali and Telegu as people of these languages are available with us in our place of research.

Keywords: *Indian Language TTS System, Di-phone concatenation, Syllable based concatenation, Consonants and Vowels.*

I. Introduction

Text-To-Speech system has been developed by many scientists so far due to its potential application in solving the human machine interfacing issues.¹⁻⁵ Some of the scientists are also trying for the syllable based TTS systems using Festival system⁶⁻⁹. In phone based concatenation system of the TTS, the database for the Odiya and Hindi languages are created taking the phone units from the speech corpora collected from the native speakers of these languages. The pure C (consonant) and V (vowel) are kept apart and a database was formed. While considering the vowel it is observed that the duration of the vowel matters for the formation of a word by the method of concatenation. The duration of the vowel utterance with respect to its position i.e. at the beginning, middle and the last differ quite a lot. So the database of vowels was tagged accordingly and the respective part was selected as per the text required to be uttered. In this TTS system a word is formed by the concatenation of V, VC, CV, CCV, CCVC and CCCV. CV is a simple syllable while CCV and CCCV are complex syllables or conjuncts. Since most of the Indian languages are derived from Sanskrit the formation of conjunct is unavoidable. While dealing with pure phone concatenation²⁻³ we have observed that the time complexity of these CCV, CCVC and CCCV utterance after concatenation is more and also the final utterance is convolved with noise besides the issues of phase differences during concatenation.

The above complexity of the TTS system for Indian languages have provoked us in developing a syllable based TTS system for Indian Languages as all Indian languages are linear in utterance while non linear in writing. Vowels with different position value like beginning, middle and end are also considered here. Many scientists⁶⁻⁹ have tried for such system and our efforts are put before for the development of such systems not using Festival system for four of the Indian languages namely Hindi, Odiya, Bengali and Telegu. We have selected these languages as Hindi is our National Language, Odiya is the official language of our state Odisha, which is in the eastern part of India, Benagli, which is similar to Odiya in many form being the official language of West Bengal one of the neighbouring states of Odisha and Telegu is one of the Dravidian Languages and is the official language of the state of Andhra Pradesh which is also one of the neighbouring states of Odisha.

Section 2 contains the description on the Phonetic behaviour of the four languages, while Section 3 has the design part for the TTS system. Section 4 has description of the implementation of the system while Section 5 has the Conclusion.

II. Vowels and Consonants of Indian Languages

Indian languages are phonetic in nature. The grapheme to phoneme mapping is linear. All the languages have their own set of Vowels and Consonants. They have a good mapping from one language to the other with respect to the features like pitch or fundamental frequency, duration of utterance of a standard CV unit and moreover the absence of some of the consonants in the typical Dravidian language Tamil. The Vowels and Consonants of these four languages are given below.

HINDI ORIYA BENGALI TELUGU

VOWEL

अ	ଅ	অ	అ
आ	ଆ	আ	ఆ
इ	ଇ	ই	ఇ
ई	ଈ	ঈ	ఈ
उ	ଉ	উ	ఉ
ऊ	ଊ	ঊ	ఊ
ऋ	ଋ	ঋ	ఋ
ॠ			
ऐ			ఎ
ॐ	ଏ	এ	ఏ
े	ଐ	ঐ	ఐ
ओ			ఒ
ओ	ଓ	ও	ఓ
औ	ଔ	ঔ	ఔ

CONSONANTS

क	କ	ক	క
ख	ଖ	খ	ఖ
ग	ଗ	গ	గ
घ	ଘ	ঘ	ఘ
ङ	ଙ	ঙ	ఙ
च	ଚ	চ	చ
छ	ଛ	ছ	చ

स	श	ष	ः
ह	ड	इ	ः

It is observed that the fundamental vowels and consonants for all the four languages are same besides some speciality with respect to the consonants for each of the languages. These special characters (C or V) are taken into consideration while designing the TTS for that language.

III. Design of the TTS Systems

a. Speech Corpora Creation:

Native speakers of all the four languages were identified. Text consisting of most of the syllables (Monosyllable, Bisyllable, Trisyllable and some of the Polysyllables) are selected. Texts were read by the speakers and recording were done under the following environment. All the speakers were male members. The recording is done in the laboratory environment using noise cancellation microphone. The sampling rate is 16 bit in single channel of 16000 Hz.

b. Database Creation:

Different Syllables from the text are chunked and kept in the database. Individual polysyllables are also stored in the database. All these are .wav files and named after the V, CV, CCV, CCVC nomenclature appropriately. These files are searched using intelligent search techniques.

c. Speech database

In the following wave file hindibha.wav which has text */bhArat hAmara desh hai/* the */ha/* part of the wave form may be taken into consideration. Here it is well marked that the

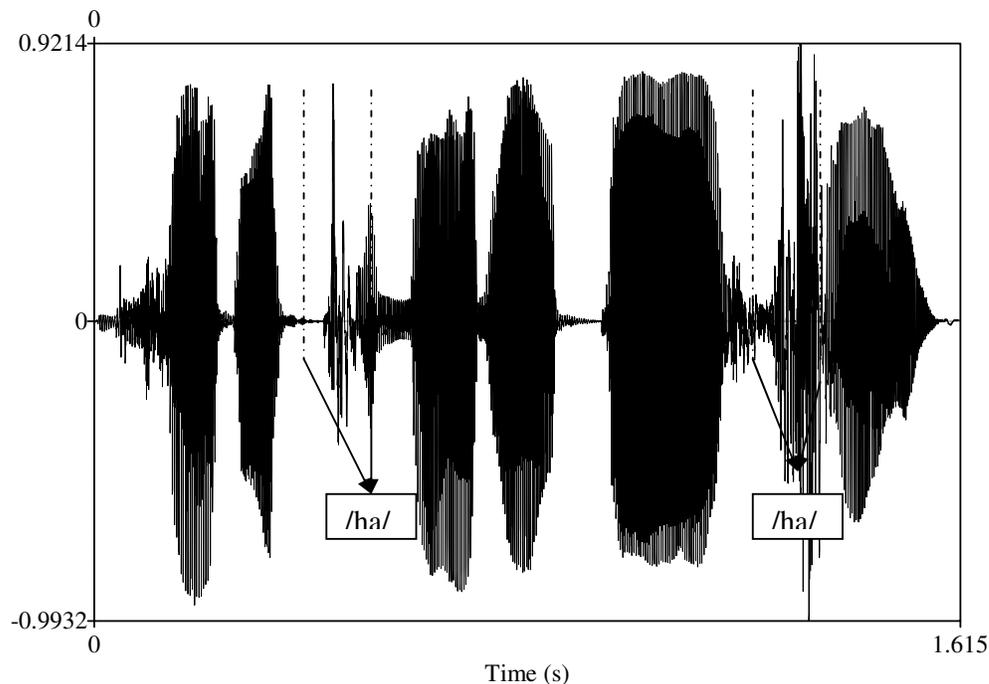


Figure 1. This figure shows the wave form of the text utterance */bhArat hAmara desh hai/*.

/ha/ part of wave form of the sentence having two /ha/ are different with respect to the onset point, which is quite critical for any phone as it completely depends of the prior and posterior phone or syllable.

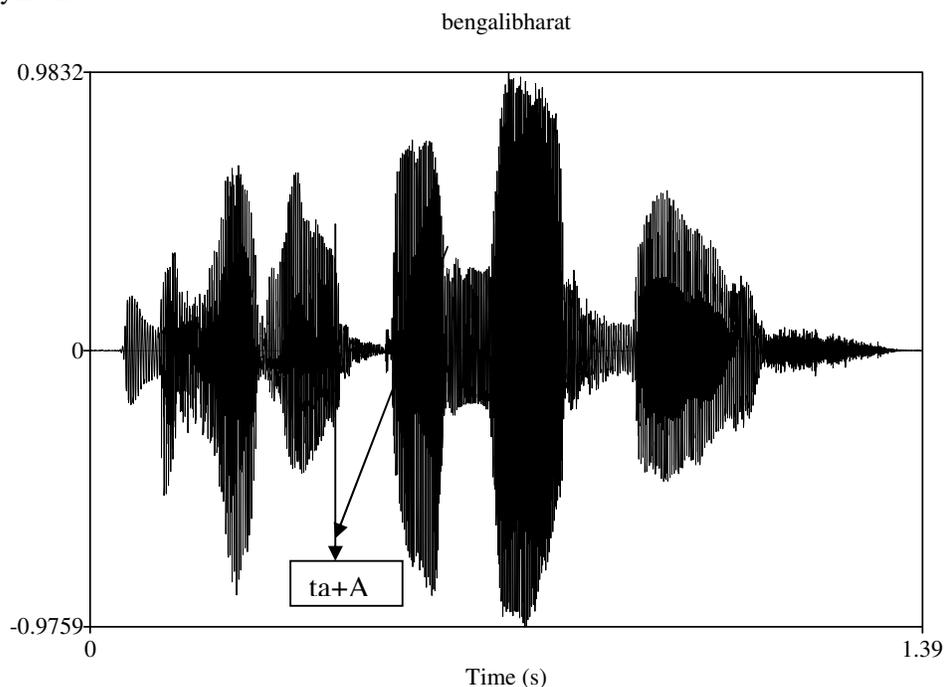


Figure 2. This figure shows bengalibha.wav file

In the above bengalibha.wav file which is for the text /bhArat Amarx deshX/ the appearance of short vowel at the end of *bhArat* and *Amarx* is the place where syllable based concatenation is unavoidable to have smooth utterance as they appear very close and the onset point is confusing to identify for phone level concatenation.

Taking into consideration these behaviours of the wave form of all the text files different syllables namely monosyllable, bisyllable, trisyllable and polysyllables are selected, cut and maintained in the databases.

IV. Implementation of the Syllable based TTS systems

a. Concatenation Algorithm

Few of the developers have used Festival based TTS systems^{8,9} but we have used Visual C++ to develop our system. The codes written for Phone based concatenation system developed by us earlier are modified with respect to the syllable bases system. Appropriate GUIs are also developed for the individual languages. Help files are developed to type text for utterance. These have the phonetic transcription help as Indian Languages are Phonetic in nature and Indian uses this concept in writing mostly. This phonetic transcription helps us to differentiate between different type of utterance of the appropriate character like /sa, sha and Sha/.

At the beginning when a text line is fed to the system, the Word Identifier identifies the words then the words are divided to appropriate syllables. These syllables are then passed on to the utterance algorithm to utter an appropriate word and then the sentence. With the same principle we have developed the Speech data base for all the four languages and the TTS systems work well.

V. Result and Discussion

It is observed that phone level concatenation is leading to complexity roughness of sound utterance while syllable level concatenation has smoothness in the utterance of the syllable. The syllable level Odiya TTS is compared with the earlier phone level concatenation system developed by us at Utkal University named as Priyambada available in the website www.ildc.in.

The syllables (mono, bi and Polysyllables are stored in a database of Wav files named after the language concerned. After the sentence is broken to words and words to Syllables the search algorithm based on the alphabetic appearance of the words in the dictionary are search.

VI. Conclusion

The TTS systems for the four languages Hindi, Odiya, Bengali and Telegu developed here are much better than our earlier TTS system based on Phones. We have made comparison with our earlier Odiya TTS named as *Priyambada* available in the web¹⁰ and Hindi TTS available with us. These systems are having the advantage of avoiding the complexity of unknown word utterance. Inj word based concatenation system though the smoothness is there for utterance but the problem arises when a new word appears in the input. Efforts are on to incorporate Prosody to the systems. However we are also trying to incorporate Statistical tool based unit selection system for our next phase of the TTS systems.

References:

- [1]. T. Dutoit, "An introduction to text-to-speech synthesis", Kulwer Academic Publishers, 1997.
- [2]. Mohanty S., Bhattacharya S. and Bose S. "A Novel Approach for Designing a Speech Synthesis System for Indian Languages", Proceedings of SIMPLE'04, IIT Kharagpur, India, 2004.
- [3]. Mohanty S., Bhattacharya S. and Bose S., "An Approach to Proper Speech Segmentation for Quality Improvement in Concatenative Text-To-Speech System for Indian Languages", Int. Jour. of Computer Processing of Oriental Languages 18(1). 2005.
- [4]. S.P. Kishore and A.W. Black, "Unit size in unit selection speech synthesis", Proceedings of EUROSPEECH, 2003.
- [5]. M. Nageshwara Rao, S. Thomas, T. Nagarajan and Hema A. Murthy "Text-to-speech synthesis using syllable like units", Proceedings of National Conference on Communication (NCC) 2005, pp. 227-280, IIT Kharagpur, India, 2005.
- [6]. Samuel Thomas, M. Nageshwara Rao, Hema A. Murthy and C.S. Ramalingam, "Natural Sounding TTS based on Syllable-like Units", Proceedings of 14th European Signal Processing Conference, Florence, Italy, Sep. 2006.
- [7]. Venugopalakrishna.Y.R., Vinodh.M.V., Hema A. Murthy and C.S. Ramalingam, "Methods for Improving the Quality of Syllable based Speech Synthesis", Proceedings of Spoken Language Technology (SLT) 2008, workshop, Goa, India. 2008.
- [8]. Sreekanth.M and A.G.Ramakrishnan, "Festival based maiden TTS system for Tamil language", Proceedings of 3rd Language and Technology Conference, pp. 187-191, Poznan, Poland, Oct 5-7, 2007.
- [9]. Alan W. Black, John Kominek, "Optimizing segment label boundaries for statistical speech synthesis", IEEE International Conference on Acoustics, Speech and Signal Processing, 2009.
- [10]. www.ildc.in for Odiya Language.

AUTHORS PROFILE

Prof. (Dr.)Sanghamitra Mohanty is the Professor and Head of the Department of Computer Science and Application, Utkal University, Bhubaneswar, Odisha, India. Her research interests are Speech Processing, Image processing and Natural Language Processing. She has published 129 papers in International Journals and Conferences. She has guided many Ph. D. students and has 10 IPRs to her credit.

