

Exploring the Speech Prosody Manipulation for Dual-Language TTS System

M.Nagamani¹, Syed Abdul Hameed², Dheeraj Kumar³, Y.Vijay Cyril⁴,
M.Saidi Reddy⁵

¹(School of Computer and Information Sciences, University of Hyderabad,Hyderabad-India)

^{2,3,4}(B.tech -IV Year, Department of Computer Science,KG Reddy College of Engg and Tech, Hyderabad-India)

⁵(Department of Computer Science, KG Reddy College of Engg and Tech, Hyderabad-India)

Corresponding Author: M.Nagamani¹

Abstract: Speech is the most comfortable mode of communication in digital media. Miniature in portable devices and IOT technology demand in comfortable mode of communication. The speech technology includes speech-to-text and Text-to-speech modality which helps visually impaired and illiterate persons to get access to digital media with the help of speech modality. They can quarry by spoken term and access information in their mother tongue as a form of text that in turn convert in speech by TTS. The present work explore the research works done in Indian languages in context of speech synthesis and their application with the prosody manipulation to get naturalness in their language context. Also done a preliminary work using speech synthesis to implement language learning application of Telangana language. Hence, the name Dual-Language. The state language of Telangana Telugu and Urdu, teaching in primary school in scope of this work. The limited domain application the implemented Urdu and Telugu speech synthesis and observed spoken term naturalness by human observation. There are many open source tools available for tts but the most commonly used is festival as it is easy for users to modify the code and manipulate it.

Keywords - Dual-Lingual TTS, Speech synthesis,urdu and telugu voices,festival.

Date of Submission: 13-08-2018

Date of acceptance: 30-08-2018

I. Introduction

The word ‘Synthesis’ is defined by the Dictionary as ‘the procedure of combining parts or elements so as to form a whole’. Speech synthesis generally refers to the artificial way of generating human speech for any devices. The devices that are used for synthesis are called as ‘Speech Synthesizer’,it may be either hardware based and software based .A Text-To-Speech synthesizer (TTS) is a computer-based program in which the system processes through the text and reads it aloud. 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. These two phases are usually called as high- and low-level synthesis.

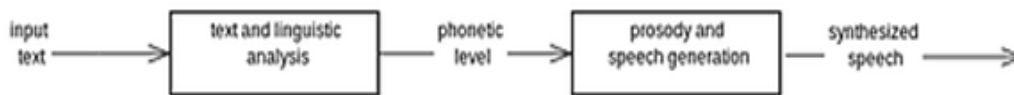


Fig.-1 Simple text-to-speech synthesis procedure.

The Text-To-Speech synthesis (TTS) is implemented entirely in software and only standard audio capability is required . At present ,it contains several components each of which handles a different tasks. For example,the text analysis capabilities of the system detect the ends of sentences ,perform some rudimentary syntactic analysis,expend digit sequence into words, and disambiguate and expand abbreviation into normally spelled words which can then be analyzed by the dictionary-based pronunciation module.[1]In many applications like electronic mail messages,and generating spoken prompts in voice response system, there is a lots of demands of technology ,which produces good and acceptable speech .The performance and quality of the Speech Synthesizer can be measured based on it’s naturalness and ability to be understood by its listener.

There are many speech synthesis techniques but most widely used are unit selection and Hidden Markov Model (HMM).

Unit Selection

It is the most widely used speech synthesis technique in which the text is divided into parts i.e., individual syllables, words, phrases, phones and diphones. During synthesis, the synthesizers utilize the information related to units, and pick the most appropriate unit based on the target cost and the concatenation cost. On the basis of target cost, best match units in the database are identified, whereas the joining cost chooses the units that can be concatenated smoothly. The best optimal selected units are concatenated and speech is synthesized.

At the run-time the

Hidden Markov Model

The HMM-based speech synthesis framework performs simultaneous modeling of pitch and spectrum taking into account the dynamics of both quantities as well. Spectral representation utilizes Mel-based cepstral coefficients while prosody is represented as logF0. Multi Space probability Distribution (MSD) modeling is performed to alleviate the problem of non continuous pitch values in unvoiced regions. Moreover, context clustering is performed using decision trees so as to fully exploit the contextual information in lexical and syntactic level

Festival

Festival is a tool which offers a general framework for building speech synthesis systems (TTS) as well as including examples of various modules. As a whole it offers full text to speech through a number of APIs: from shell level, through a Scheme command interpreter, as a C++ library, and an Emacs interface. The main usage of Festival is to convert the text file or any text input into voice (Speech). When you pass a text file to Festival it converts the contents of a text file into voice. For example, if I want to read a letter (mail) which is residing in a text file (say letter.txt). It can let festival read it out loud for me as follows: `festival >(tts "filename.txt",nil)` or `festival>(SayText "hello world")`. The advantages of this tool is, it is available for free under open source license, and in festival the voice quality and pronunciation are good and understandable.

To install Festival in our computer or laptops, open the Terminal (Ctrl+Alt+T), then type the command as follows:

```
$ sudo apt-get install festival
```

After installation of festival tool, type `$festival` in Terminal to check whether it is installed or not. If it was installed then it will produce the window as shown below:

```
dheeraj@dheeraj-Aspire-E5-576:~$ cd Desktop/project_HCU/speech/
dheeraj@dheeraj-Aspire-E5-576:~/Desktop/project_HCU/speech$ festival
Festival Speech Synthesis System 2.4:release December 2014
Copyright (C) University of Edinburgh, 1996-2010. All rights reserved.

clunits: Copyright (C) University of Edinburgh and CMU 1997-2010
clustergen_engine: Copyright (C) Carnegie Mellon University 2005-2014
hts_engine:
The HMM-Based Speech Synthesis Engine "hts_engine API"
hts_engine API version 1.07 (http://hts-engine.sourceforge.net/)
Copyright (C) The HMM-Based Speech Synthesis Engine "hts_engine API"
Version 1.07 (http://hts-engine.sourceforge.net/)
Copyright (C) 2001-2012 Nagoya Institute of Technology
                2001-2008 Tokyo Institute of Technology
All rights reserved.

All rights reserved.
For details type '(festival_warranty)'
festival>
```

Implementation of Bilingual System (Telugu and Urdu)

The telugu language package can be directly installed in the festival.

Telugu package:

```
$ sudo apt-get install festival-te-nsk
```

After installing the package change the language to telugu: `$festival>(voice_telugu_NSK_diphone) { //voice_telugu= teluguvoice diphone= phonemes }`

For the speed (duration) of utterance:

```
$festival>(Parameter.set 'Duration_Stretch' <numeric value>)-----{ //lower the value higher the speed rate }
```

To make system repeat same utterances (same sentences)

1)store the utterance by : \$festival>(set! utt<no> (SayText "Hello World"))
{This command will store utterance in utt and also produce the voice }

2)To repeat the same utterance : \$festival>(utt.play utt<no>)

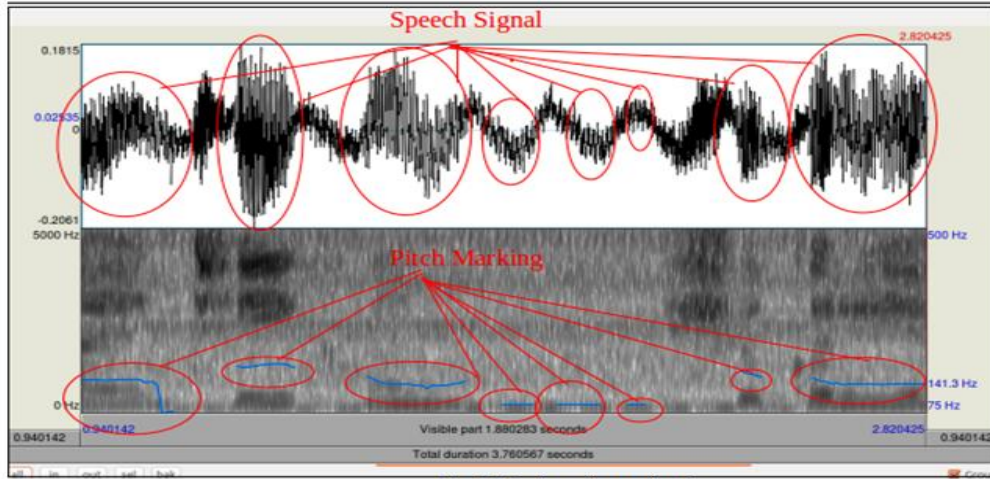


Fig: Urdu Speech synthesis

kushu AmxIx خوش آمدید (Kush amdeed)
mai wlk hu ShukrIya میں ٹھیک ہوں، شکر بہ (Mai theek hu shulriya)

References

- [1]. HMM-Based Speech Synthesis for the Greek Language
- [2]. BHAASHIKA: TELUGU TTS SYSTEM
- [3]. The HMM-based Speech Synthesis System (HTS) Version 2.0

IOSR Journal of Computer Engineering (IOSR-JCE) is UGC approved Journal with Sl. No. 5019, Journal no. 49102.

* M.Nagamani | Exploring the Speech Prosody Manipulation for Dual-Language TTS System." IOSR Journal of Computer Engineering (IOSR-JCE) 20.4 (2018): 20-22.