Text to Speech System for Malayalam

Riya Rajan¹,Arunima S V², Rahul C³

Department of Computer Science and Engineering, Mes College of Engineering, Kuttipuram, L B S College of Engineering, Kasaragod
Email: riyarajan88@gmail.com , sv.arunima@gmail.com , rahulshreys@gmail.com

Abstract: Text-to-speech (TTS) systems which mainly meant for speech synthesis are, used for one of the South Indian languages called Malayalam. The paper makes a brief study on, Malayalam linguistics, and also gives a comparison between two prominent methodologies for speech synthesis, viz Concatenative based synthesis and HMM based synthesis. As a result, the paper mentions some of the problems facing by Concatenative based TTS systems and thereby, the research goes on with HMM synthesis. The paper also done a proposal for TTS system for Malayalam which is statistical based using HMM’s (Hidden Markov Model).

Keywords: TTS, HMM Synthesis, Phoneme, Prosody, Concatenation Synthesis

1 Introduction

Natural language processing (NLP) is a field of computer science, artificial intelligence (also called machine learning), and linguistics concerned with the interactions between computers and human (natural) languages.

Why NLP?

Internet domain consists of huge amount of various data. So, we need applications for processing this large amount of texts. Thus we requires of NLP expertise usually called computational linguistics. Speech Synthesis which is a prominent area under NLP that is having so much importance in researches, has introduced Text to Speech Systems (TTS) for almost all foreign and Indian languages. Among the applications of speech technology, the automatic speech production, which is referred to as text-to-speech (TTS) system is the most natural sounding technology. The text-to-speech (TTS) system will convert ordinary orthographic text into acoustic signal which is indistinguishable from human speech [1].
2 Underlying Technologies

2.1 TTS System [2]

A Text-To-Speech (TTS) synthesizer is a computer-based system that should be able to read any text aloud. We can define the Text-To-Speech as the automatic production of speech, through a grapheme-to-phoneme transcription of the sentences to utter. As for human reading, a TTS system comprises a Natural Language Processing module (NLP), which is capable of producing a phonetic transcription of the text read, together with the desired intonation and rhythm, and a Digital Signal Processing module (DSP), which transforms the symbolic information it receives into speech.

2.1.1 Significance of TTS System

Nowadays researches in NLP have widened to include TTS processing since; it allows people with visual impairments or reading disabilities to listen to written works on a home computer. It can be implemented in software or hardware. It helps to improve human interaction with computer.

Physically challenged people find computers, difficult to use. So, TTS synthesis is used in software applications that enable access to text for, visually-impaired and individuals with specific learning difficulties. TTS systems in NLP are mainly meant for Speech Synthesis.

2.2 Speech Synthesis [2]

The artificial production of human speech is known as speech synthesis and the computer system used for this purpose is known as Speech synthesizer. There are two levels of synthesis, High level synthesis and Low level synthesis. Low-level synthesis is the production of sound to simulate human speech whereas, High-level synthesis deals with the conversion of written text or symbols into an abstract representation of the desired acoustic signal, suitable for driving a low level synthesis system.

2.3 TTS Malayalam [3]

TTS systems for many foreign languages and also some of the Indian languages have been introduced. But, TTS system for the South Indian based language Malayalam has not been introduced so far effectively. So, here discusses the basic terminologies that needed to consider, making computers speak to Malayalam users around the world.

2.3.1 Malayalam Language

Malayalam is the mother tongue of the state, Kerala. It is one among the 22 official languages of India. Malayalam is a syllabic alphabet in which all consonants have an inherent vowel. Each word is pronounced according to the phonetic units comprising the word. There are 36 consonants and 15 vowel characters in Malayalam.
2.3.2 Malayalam Script
Appearance of the characters in the Malayalam script is affected by the following factors:
- Ordering of the concerned character with respect to other characters
- The font employed
- The application or system environment
These variables can cause the appearance of the Malayalam characters to differ from their nominal glyphs.

2.3.3 Malayalam Phonology
Phonology is the study of how sounds are organized and used in natural languages. Phonology analyzes the sound patterns of a particular language by,
- Determining which phonetic sounds are significant
- Explaining how these sounds are interpreted by the native speaker
With the help of Malayalam phonology a phone model for Malayalam words are developed. Malayalam is described as a philologist’s paradise. The distinctive segments of Malayalam are consonants and vowels.

2.3.4 Vowels
Vowel sounds can be distinguished in terms of their quality which means, the precise combination of sound frequencies which make it up, by the position of tongue, lips and other articulators. Also by quantity which deals with, how long the sound lasts.

2.3.5 Consonants
In consonant system of sounds, the International Phonetic Association (IPA) descriptions distinguish three aspects or parameters of sounds. These parameters are its voicing, its place of articulation and its manner of articulation.

2.4 General Comparison Over Prominent Speech Synthesis Techniques
To overcome the challenges, one must have chosen an appropriate methodology for generating highly natural speech output. Nowadays, researchers using two prominent methodologies for efficient speech synthesis, they are:

2.4.1 Concatenative Synthesis
Concatenative synthesis is based on the concatenation of segments of recorded speech. Generally, concatenative synthesis produces the most natural-sounding synthesized speech. However, differences between natural variations in speech and the nature of the automated techniques for segmenting the waveforms sometimes result in audible glitches in the output [3].

2.4.2 HMM Based Synthesis
HMM-based synthesis is a synthesis method based on hidden Markov models, also as called Statistical Parametric Synthesis. In this system, the frequency spectrum (vocal tract), fundamental frequency (vocal source), and duration (prosody) of speech are
modelling simultaneously by HMMs. Speech waveforms are generated from HMMs themselves, based on the maximum likelihood criterion [6].

A general comparison on these two prominent output methodologies has been taken and is illustrated in the following table:

<table>
<thead>
<tr>
<th>Advantage</th>
<th>Drawback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good naturalness</td>
<td>Less intelligibility</td>
</tr>
<tr>
<td>Simpler and faster synthesis</td>
<td>Requires more memory</td>
</tr>
<tr>
<td>Unit-based synthesis requires small amount of</td>
<td>Presence of audible glitches</td>
</tr>
<tr>
<td>DSP to recorded speech</td>
<td></td>
</tr>
<tr>
<td>HMM-based synthesis</td>
<td>Requires large database of speech samples</td>
</tr>
</tbody>
</table>

Aim of this paper is to develop a text-to-speech system for Malayalam. When we try to propose a TTS system for Malayalam using either any one of the methodology, we want to consider certain facts since it affects the naturalness of speech sound. Some of these problems with concatenative synthesis are listed below [4]:

- Discontinuities in concatenation points in the synthesized speech
- Memory requirements are usually very high, especially when long concatenation units are used, such as syllables or words
- Data collecting and labeling of speech samples is usually time-consuming
- Strong dependency of output speech on the chosen database

The discontinuity at the concatenation points arises because; the speech units that are to be concatenated are extracted from different locations in recorded sound. Thus the pitch cannot be exactly same throughout the recording and, also different word sounds have different pitch levels. The discontinuity at concatenation points need to be removed which can be done by, equalizing the pitch of the two speech units that are to be concatenated.

### 3 Proposed System

We must consider an effective technology of speech synthesis, which generates highly natural, comprehensive speech output. HMM-based synthesis which is a statistical-based method and, there have been several attempts proposed to utilize HMM for constructing TTS systems. In one of the latest approaches, speech parameter sequences are generated from HMM directly based on maximum likelihood criterion. By considering the relationship between static and dynamic parameters, smooth spectral sequences are generated according to the statistics of static and dynamic parameters modeled by HMMs. As a result, natural sounding speech can be synthesized. This paper suggests a HTK (Hidden Markov Model Tool Kit) to build and manipulate the speech models.
HMM’s. HTK consists of a set of library modules and tools available in C source form. It has the components like, data preparation tools, training tools and recognition tools [5].

Since speech parameter sequences are generated directly from HMMs, it is possible to covert voice characteristics of synthetic speech to a given target speaker. In the current HMM-based TTS system, dynamic features play an important role in generation of speech parameter sequences. Without dynamic features, generated spectral sequences have discontinuities at the state transitions, which result in clicks in synthetic speech [6]. HMM based synthesis, which can be applied to a TTS system for Malayalam, is very effective since it is a phonetically as well as prosodically rich language.

4 Conclusion and Future Work

Here, it has been concentrating on TTS systems for one of the South Indian based language, Malayalam. From the literature survey done, we can access a comparison between mainly two prominent speech synthesis technologies, Concatenative synthesis and HMM based synthesis. Thus, some of the existing problems using Concatenative synthesis have been detailed. Since it has been shown that, the HMM-based speech synthesis system have an ability to synthesize speech with arbitrarily given text and speakers voice characteristics, this can be considered to be applicable to generate a better speech output with desirable features like, good naturalness, quality, and intelligibility. As a solution for the referred problems, we can adopt the HMM speech synthesis technique and therefore proposes an effective TTS system for Malayalam. The proposed system discussed in this paper is under the process of implementation.

References