Project Report on Synthesizing Natural Sounding Nepali Voice

Roop Shree Ratna Bajracharya

2018-04-15

Contents

1 Abstract

Text-to-Speech (TTS) synthesis has come far from its primitive synthetic mono-
tone voices to more natural and intelligible sounding voices. Festival is one of
such systems that uses a concatenative speech synthesis method to produce nat-
ural sounding voice. This project aims to develop Nepali voice using Festival
system with tools included in Festvox as a part of the Nepali-TTS project cur-
rently being conducted in Information and Language Processing Research Lab
(ILPRL) in Kathmandu University. This project includes studying and improv-
ing different steps and procedures involved in the speech synthesis process and
will aim to produce more natural sounding Nepali voice.

Keywords: Text-To-Speech, Festival Speech Synthesis, Concatenative Speech
Synthesis

2 Introduction

2.1 Text to Speech Synthesis

Text-To-Speech (TTS) system converts the given text to a spoken waveform through text processing and speech generation processes. These processes are connected to linguistic theory, models of speech production, and acoustic-phonetic characterization of language [1]. There are three approaches to TTS system: 1) articulatory-model 2) parameter-based and 3) concatenation of stored speech [1]. This project deals with the third approach to concatenate stored speech segments (units). This approach includes storing and concatenation of speech from huge number of recorded sentences of a language to produce a speech waveform.

References

[1] D. H. Klatt, *Review of texttospeech conversion for English.* The Journal of the Acoustical Society of America, 82(3), pp.737-793, 1987.