PhD Work: Text-To-Speech Synthesis System for Marathi Language Using Concatenation Technique

Language is the structural form of sharing thoughts and emotions in humans. The research motivates to stroke up for the Human-computer interaction. The overall intention of my PhD research program is focused to design Concatenation and Hidden Markov Model (HMM) based speech synthesis for the Marathi language. This will facilitate to correspond to the system and extend the technology for assertive devices based on the Marathi language. The advantage and attractive feature of the HMM system are that the voice alteration can be performed without large databases. To understand the detailed study of Synthesis techniques, I have also implemented the system for Unit Selection method. The *Marathi Talking calculator* is published at Play store using the technique of concatenation. This calculator performs the basic arithmetic operations and additionally speaks out the numeral in Marathi as the key is pressed. The result box synthesis the voice and speaks out the result in Marathi with correct place value of digits. The weakness of USS is it requires a large database and at joins, the quality is affected. To overcome these issues, the study reveals the built-up of a system with a phonetic based approach for Marathi Language using Concatenation and HMM.

The Marathi recording corpus of 1000 sentences is constructed and labelling is completed manually using PRAT with the step-by-step procedure. With the help of festival and festvox packages, the Marathi TTS system is generated. The Clustergen module generates the spectrum, pitch and duration parameters of input speech. The Prosody is indulged with the context-dependent models calculated during the synthesis phase of the system. The algorithm re-synthesis speech with Mel-cepstral coefficients using MLSA (Mel Log Spectrum approximation) filter. The system is demonstrated for the speech parameters i.e. naturalness and understanding using subjective test i.e. Mean opinion score (MOS) and Objective test i.e. Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR) values. The MOS test result verified the system to the accuracy rate of 85%. While the Objective test evaluated for the performance of the system as 80%.

Note:- Code available in code folders (folder name PHD)

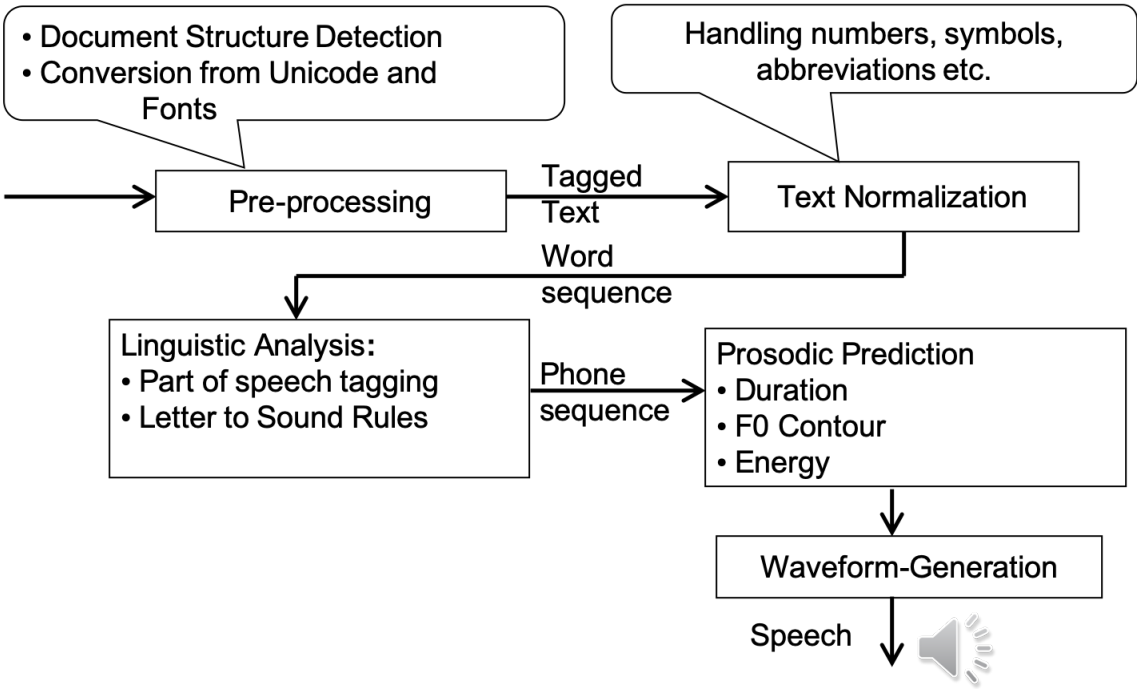

Figure 1: System architecture of Text-to-Speech System

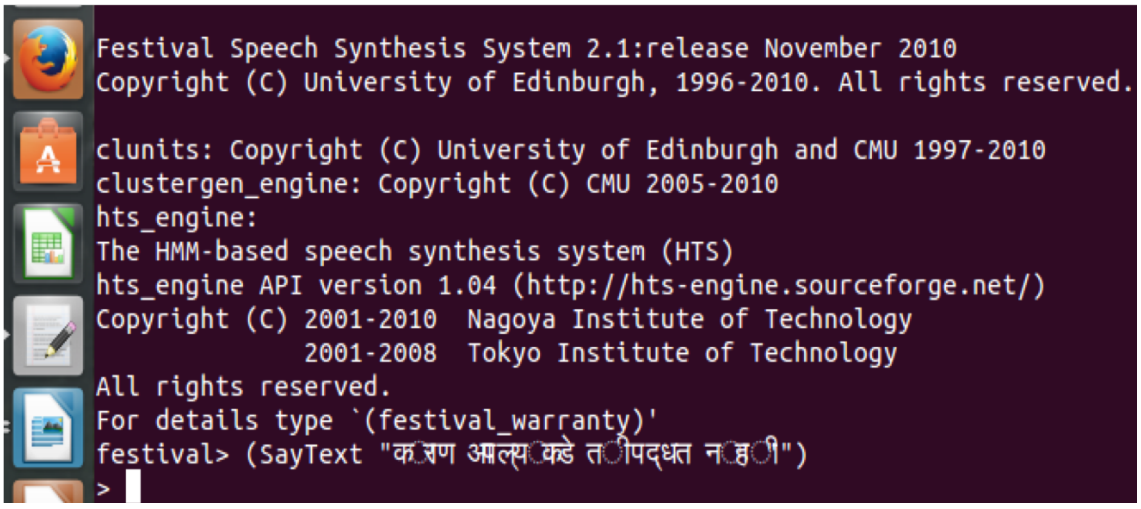

Figure 2: Output Screen of TTS Voice