

Abstract

A TEXT-TO-SPEECH SYSTEM FOR INDIAN LANGUAGES

Arnab Ghoshal
and
Ayaskant Shrivastava
Bachelor of Technology (Hons.), 2002

Dissertation directed by
Prof. Anupam Basu
Department of Computer Science & Engineering
Indian Institute of Technology, Kharagpur

The problem for developing a *TTS* (*text-to-speech*) is a very active field of research. As the *Human-Computer Interfaces* (*HCI*) come of age, the need for a more ergonomic and natural interface than the current one (*keyboard, mouse, etc.*) is being constantly felt. Talking of natural interfaces, what comes to mind, is sound (*speech*) and sight (*vision*). These form the basis of many intelligent systems research like robotics. Moreover, speech can also serve as an excellent interface for sightless people, or people with motor neuron disorders.

In this dissertation we attempt at developing a TTS system for *Indian Languages*. A lot of commercial systems are available for many foreign languages (mostly English), but there is yet to be a competitive system available for Indian languages. Although the task of building very high quality, unlimited vocabulary text-to-speech (TTS) system is still a difficult one, with many open research questions, we believe the building of reasonable quality voices for many tasks can serve our needs. Here we have worked with *Hindi*, the most commonly spoken Indian language. We hope to easily extend the system to other languages, since there are a lot of underlying similarities between Indian languages. Indian languages being highly phonetic, result in simple letter-to-sound rules. We used

the standard concatenative synthesis. The main problem faced by us was to make the synthesized speech sound natural. We investigated the reasons for the mechanical sounding speech and developed different synthesis models to overcome some of those problems. Moreover, we implemented some standard and also novel *intonation* and *duration* modification algorithms, which can be incorporated into the TTS at a later stage. Our main achievement was reasonably legible speech with an unlimited vocabulary.

The following thesis presents a brief overview of the main text-to-speech synthesis problem and its subproblems, and the initial work done in building a TTS for Hindi.