Abstract

Speech is the most natural way for humans to communicate with each other. Synthesis of artificial human speech provides efficient human-computer communication. Nowadays, the speech synthesis systems are widely used in various applications such as screen readers for visually challenged people, speech interface for mobile devices, navigation, and personal guidance gadgets. As humans are very sensitive in perceiving even the slightest distortions in the speech signal, speech synthesizers with suboptimal quality make them unfit for usage in commercial applications. The main goal of this thesis is to improve the quality of HMM-based speech synthesizer by efficiently modeling the source or excitation signal. The excitation signal used in synthesis should preserve all natural variations so that the synthesized speech is close to natural quality. The work presented in this thesis confines its scope to (i) accurate estimation of pitch (F_0) and (ii) precise modeling of excitation signal. For modeling the excitation signal, both parametric and hybrid approaches are explored. In this work, creaky voice has been synthesized at appropriate places by proposing appropriate methods and models.

Major contributions of this thesis are:

1. A robust voicing detection and F_0 estimation method is proposed for enhancing the quality of synthesized speech from HMM-based speech synthesis.
2. A parametric source modeling method is proposed based on principal component analysis of pitch-synchronous residual frames of excitation signal.
3. A parametric source modeling method is proposed by modeling excitation signal in terms of deterministic and noise components.
4. A hybrid source modeling method is proposed by utilizing optimal residual frames extracted from the excitation signal of a phone.
5. A hybrid source modeling method is proposed based on time-domain deterministic plus noise model.
6. A method based on epoch parameters is proposed for accurate detection of creaky regions.
7. A source model is proposed by incorporating creaky excitation at appropriate places.