Abstract

This paper presents a study on prosody modeling for speech synthesis. Any Text to Speech system comprises of two phases. One is text analysis and second is speech synthesis. The task of text analysis is to find the words and the task of speech synthesis is to generate the speech. To attain this, different models are available such as text as language models, grapheme to phoneme models, full linguistic analysis model and complete prosody generation model. In complete prosody generation model, the quantities like phrasing, stress and the like are determined to generate naturalness bearing synthetic voice. Towards generating such a speech, an explicit prosodic model is required. This makes the speech more understandable. Many researches have been done in this stream, but still better solution is required. In this paper, the strength and weaknesses of different approaches of prosody models are discussed.
Prosody Modeling Techniques for Text-to-Speech Synthesis Systems – A Survey

- X.J. Ma, W. Zhang, W.B. Zhu, Q. Shi and L. Jin, “Probability Based Prosody Model for Unit Selection”, ICASSP 2004, Montreal, Canada
- Wei Zhang, Liang Gu and Yuqing Gao “Recent improvements of probability based prosody model for unit selection in concatenative Text to Speech”, in the proceedings of ICASSP 2009, pp 3777-3780
- Nicolas Obin, Xavier Rodet and Anne Lacheret Dujour,”A multi-level context-dependent prosodic model applied to duration modeling”, in the tenth
- Javier Latorre, Sabine Buchholz, Masami kamine, "Usages of an external duration model for HMM-based speech synthesis", in fifth International conference on Speech Prosody, Chicago, 2010

Index Terms

Computer Science
MultiMedia

Keywords
Prosody model speech synthesis Text to Speech systems