Abstract

The primary objective of this paper is to provide an overview of existing Concatenative Text-To-Speech synthesis techniques. Concatenative speech synthesis can be broadly categorized into three categories, Diphone Based, Corpus based and Hybrid. Diphone based speech synthesis relies on different signal processing techniques such as PSOLA, FD-PSOLA etc. These signal processing techniques introduce unwanted artifacts in the synthesized speech. The most popularly used method is the Unit selection synthesis which is a corpus based synthesis method. This method produces the most natural sounding synthetic speech.

References


30. Size of Speech Corpora ( As on july 2014) . [Online]
http://www.ldcil.org/resourcesSpeechCorpor.aspx


32. [Online],CMU ARCTIC speech synthesis databases, http://festvox.org/cmu arctic/

33. Online],CMU FAF speech synthesis databases, http://festvox.org/cmu_faf/

34. [ Online],CMU SIN speech synthesis databases, http://festvox.org/cmu_sin/


38. Heiga Zen, Norbert Braunschweiler, Sabine Buchholz, Mark J. F. Gales, Kate Knill,


**Index Terms**

Computer Science

Signal Processing

**Keywords**

TTS; PSOLA; TD-PSOLA; FD-PSOLA; ESNOLA; MOS; SUS; DRT; HMM.