CCIT Report #799 December 2011

Quality Preserving Compression of a Concatenative Text-To-Speech Acoustic Database

Tamar Shoham, David Malah, and Slava Shechtman

Abstract

A Concatenative Text-To-Speech (CTTS) synthesizer requires a large acoustic database for high quality speech synthesis. This database consists of many acoustic leaves, each containing a number of short, compressed, speech segments. In this paper we propose two algorithms for re-compression of the acoustic database, by re-compressing the data in each acoustic leaf, without compromising the perceptual quality of the obtained synthesized speech. This is achieved by exploiting the redundancy between speech frames and speech segments in the acoustic leaf. The first approach is based on a vector polynomial Temporal Decomposition. The second is based on 3D Shape-Adaptive DCT, followed by optimized quantization. In addition we propose a segment ordering algorithm in an attempt to improve overall performance. The developed algorithms are generic and may be applied to a variety of compression challenges. When applied to compressed spectral amplitude parameters of a specific IBM small footprint CTTS database, we obtain a re-compression factor of 2 without any perceived degradation in the quality of the synthesized speech.

Index Terms

Concatenative Text-To-Speech (CTTS), Acoustic Leaf Compression, Temporal Decomposition (TD), Discrete-Cosine-Transform (DCT), Shape-Adaptive-DCT (SADCT).

Tamar Shoham and David Malah are with the Department of Electrical Engineering, Technion - Israel. Institute of Technology, Haifa 32000, Israel. Slava Shechtman is with the IBM Haifa Research Labs. [e-mails: tshoham3@gmail.com; malah@ee.technion.ac.il; SLAVA@il.ibm.com]