Chapter III

Spectral-Based Analysis and Synthesis of Audio Signals

Paulo A.A. Esquef, Nokia Institute of Technology, Brazil

Luiz W.P. Biscainho, Federal University of Rio de Janeiro, Brazil

Abstract

This chapter reviews audio signal processing techniques related to sound generation via additive synthesis. Particular focus will be put on sinusoidal modeling. Each processing stage involved in obtaining a sinusoidal representation for audio signals is described. Then, synthesis techniques that allow reconstructing an audio signal based on a given parametric representation are presented. Finally, some audio applications where sinusoidal modeling is employed are briefly discussed.
There is a family of real-world sounds to which the idea of “pitch” can be associated. According to psychoacoustics (Meddis & Hewitt, 1991; Hartmann, 1996), pitch can be loosely seen as the perceived frequency, so that one can speak of low- or high-pitched signals. The voiced part of speech, singing voice, and most of musical instruments have discernible pitches, to which musical notes are directly linked. The “tonal” aspect of these audio signals reflects into peaky spectra; that means slowly decaying periodic components, which can be modeled as non-decaying oscillations during short periods of time.

Additive synthesis is a method specially suited to accurately generate pitched sounds. Its strength comes from the inherent generality of the idea behind the method, that is, that sounds can be generated by a sum of sinusoids with varying frequency, amplitude, and phase. A diagram of such scheme is depicted in Figure 1 (Cook, 2002). It can be noted that, in addition to the controllable oscillators, there is a noise source that can be shaped in frequency as well. The latter accounts for representing the stochastic components present in the signal to be synthesized (Serra & Smith, 1990).

Naturally, depending on the complexity of the sound to be synthesized, determining manually the amplitude and frequency control functions for each oscillator over time can be an
Visual Speech Processing and Recognition
www.igi-global.com/chapter/visual-speech-processing-recognition/31071?camid=4v1a