Spectral Envelopes and Additive + Residual Analysis/Synthesis

XAVIER RODET AND DIEMO SCHWARZ

1 Introduction

The subject of this chapter is the estimation, representation, modification, and use of spectral envelopes in the context of sinusoidal-additive-plus-residual analysis/synthesis. A spectral envelope is an amplitude-vs-frequency function, which may be obtained from the envelope of a short-time spectrum (Rodet et al., 1987; Schwarz, 1998). [Precise definitions of such an envelope and short-time spectrum (STS) are given in Section 2.] The additive-plus-residual analysis/synthesis method is based on a representation of signals in terms of a sum of time-varying sinusoids and of a non-sinusoidal residual signal [e.g., see Serra (1989), Laroche et al. (1993), McAulay and Quatieri (1995), and Ding and Qian (1997)]. Many musical sound signals may be described as a combination of a nearly periodic waveform and colored noise. The nearly periodic part of the signal can be viewed as a sum of sinusoidal components, called partials, with time-varying frequency and amplitude. Such sinusoidal components are easily observed on a spectral analysis display (Fig. 5.1) as obtained, for instance, from a discrete Fourier transform.

In consequence, some of the first attempts at sound synthesis were based on the additive synthesis method, i.e., the summation of time-varying sinusoidal components [e.g., Risset and Mathews (1969)]. This signal-modeling approach inherits a rich history of signal processing techniques. For example, harmonic or inharmonic partials are easy to characterize and easy to synthesize. Also, there exist many methods to automatically analyze sounds in terms of partials and noise that can then be used directly for additive synthesis [e.g., Serra and Smith (1990)].

Another interesting aspect of additive synthesis is its ease for mapping partial parameters (frequency and amplitude) into the human perceptual space. Also, these parameters are meaningful and easily understood by musicians. Furthermore, because independent control of every component is available in additive synthesis, it is possible to implement models of perceptually significant features of sound such as inharmonicity and roughness. Thus, additive synthesis is accepted as perhaps the most powerful and flexible sound synthesis method available.

A drawback of the classical sinusoidal oscillator (i.e., simple addition of sine waves) implementation of additive synthesis (Moore, 1990) is its computational
cost that can easily be seen by considering a sound such as a low-pitched piano
tone, which can sometimes require more than a hundred partials to properly repre-
sent it. However, another additive synthesis technique (Rodet and Depalle, 1992)
will be examined in Section 6.3. This method, named FFT$^{-1}$, is based on the
inverse fast Fourier transform and allows an efficiency gain of 10–30 compared
to the classical method. A second drawback of the oscillator method of additive
synthesis is its difficulty for introducing precisely controlled noise components
that are very important for realistic sounds and musical timbres such as speech
or Japanese shakuhachi flute sounds, which cannot be created without noise. The
FFT$^{-1}$ technique and spectral envelopes make noisy components easy to describe
and cheap to compute. Last but not least, controlling hundreds of sinusoids is a
great challenge for the computer musician. Spectral envelopes render this con-
rol more simple, direct, and user-friendly, and are easily implemented with the
FFT$^{-1}$ method.

As mentioned above, speech and musical sounds always have random compo-
nents, often heard as a noise, superposed on the harmonic or inharmonic parts. Since
a second assumption underlying the sinusoidal additive model is that the number
of sinusoidal partials is limited, a purely sinusoidal model $d(t)$ with slowly vary-
ing parameters cannot completely represent a real signal $s(t)$ and therefore must
be complemented by a non-sinusoidal residual part $r(t)$ aimed at representing the
random components:

$$r(t) = s(t) - d(t).$$  \hspace{1cm} (5.1)

Even though the term spectral envelope is commonly used only for the envelope
of the magnitude of the short-time spectrum, we will also consider envelopes that
include the phase of the STS and even the frequencies of nearly harmonic partials
as a function of their harmonic number. Such envelopes are called generalized