EFFICIENT FOURIER SYNTHESIS OF NONSTATIONARY SINUSOIDS

Michael Goodwin[†] and Xavier Rodet[*]

[†]Center for New Music and Audio Technologies (CNMAT)
University of California, 1750 Arch Street, Berkeley, CA 94709
Tel: (510) 643 9990, Fax: (510) 642 7918
e-mail: michaelg@cnmat.berkeley.edu

[*]IRCAM, 31 rue Saint Merri, 75004, Paris, France
Tel: (33 1) 44 78 48 45, Fax: (33 1) 42 77 29 47
e-mail: rod@ircam.fr

ABSTRACT

This paper discusses an additional signal processing procedure for use in a recently proposed frequency-domain additive synthesizer [1, 2]. The overlap-add approach used in the frequency-domain synthesis can result in amplitude modulation in the overlap region between successive time frames if the partial frequencies in those frames are significantly different. A new method is proposed in which sinusoids of linearly time-varying frequency (*chirps*) are used to accommodate the frame-to-frame frequency changes.

1. INTRODUCTION

In analysis-based synthesis, the synthesis is driven by a set of analysis parameters that describe the time evolution of the input signal. For additive music synthesis, these parameters typically include the amplitude, frequency, and phase of each sinusoidal component or *partial* of the input. A time-domain synthesizer uses these parameter *tracks* as control inputs to a bank of oscillators whose respective outputs are accumulated to generate the final output; this method is a realization of the standard musical signal model:

$$d[n] = \sum_{q=1}^{Q} A_q[n] \cos\left(\omega_q[n]n + \phi_q[n]\right)$$

where q is the partial index, n is the discrete time index, and $A_q[n]$, $\omega_q[n]$, and $\phi_q[n]$ are the amplitude, frequency, and phase of the q-th partial, respectively. These partial parameters are assumed to vary at a rate substantially slower than the sampling rate. In this model description, a stochastic signal component has been omitted since it is not relevant to this discussion.

2. FREQUENCY-DOMAIN ADDITIVE SYNTHESIS

In a recently proposed system, the additive synthesis is done in the frequency domain instead of the time domain so as to enhance the modification capabilities and improve the computational efficiency[1, 2]. The synthesis is based on partial parameters derived by a short-time Fourier transform (STFT) analysis followed by a spectral peak-picking algorithm. Since the STFT is a frame-by-frame analysis, the partial amplitude, frequency, and phase parameters are determined at the frame rate; the parameters are thus constant within each frame.

The synthesis relies on the equivalence of time domain multiplication and frequency domain convolution. The spectrum of the time-windowed signal $b[n]d[n]$ is a convolution of $D(\omega)$ with the window transform $B(\omega)$. Since $d[n]$ consists of cosines at frequencies ω_q, $D(\omega)$ consists of delta functions at $\pm\omega_q$ weighted by amplitude $\frac{1}{2}A_q$ and phase factor $e^{\pm j\phi_q}$. Convolution of $B(\omega)$ with $D(\omega)$ thus modulates the baseband window transform $B(\omega)$ to the frequencies $\pm\omega_q$. The spectrum of $b[n]d[n]$ is then the sum of these modulated window transforms. This relationship is the foundation of the frequency-domain synthesis. To synthesize a frame of the signal $d[n]$, a weighted version of $B(\omega)$ centered around ω_q is accumulated into the spectrum for each partial; the computation is reduced by exploiting the conjugate symmetry of the spectrum and by choosing $b[n]$ such that $B(\omega)$ can be well-represented by only the bins in its main spectral lobe. Then, the spectrum undergoes an inverse fast Fourier transform (IFFT) to yield the signal $b[n]d[n]$, from which the frame of $d[n]$ can be extracted by dividing out the window $b[n]$. The full output signal is constructed using overlap-add (OLA) with a triangular window to provide lin-

ear amplitude interpolation between frames.

In the synthesis, the amplitude and frequency parameters are constant in each frame. Frame rate changes in the amplitude are smoothed by the interpolation provided by the triangular OLA window. Frequency variations from frame to frame, however, pose a considerable difficulty since the addition of different frequency sinusoids in the overlap region can result in amplitude distortion due to phase cancellation. This distortion can be minimized by adjusting the phase of each partial in frame $i + 1$ such that the partials in frame i and $i + 1$ are in phase in the middle of the overlap. For a frame size of N and a 50% overlap:

$$\cos\left(\omega_{q,i}\frac{3N}{4} + \phi_{q,i}\right) = \cos\left(\omega_{q,i+1}\frac{N}{4} + \phi_{q,i+1}\right)$$

$$\phi_{q,i+1} = \phi_{q,i} + \frac{N}{4}(3\omega_{q,i} - \omega_{q,i+1})$$

Even with the proper phase alignment, distortion still occurs as indicated in figure 1A.

3. CHIRP SYNTHESIS

The distortion discussed above can be alleviated by synthesizing *chirps* in each frame instead of constant frequency partials; a chirp is a sinusoid with a linearly increasing (or decreasing) frequency:

$$c[n] = \cos((\omega_0 + \alpha n)n + \phi) = \cos(\omega_0 n + \alpha n^2 + \phi)$$

In order to synthesize chirps, for each partial in each frame an analysis parameter $\alpha_{q,i}$ describing the rate of frequency change must be derived. An intuitively reasonable value for a 50% frame overlap is given by

$$\alpha_{q,i} = \frac{\omega_{q,i+2} - \omega_{q,i}}{N}$$

which corresponds to the frequency slope across the length N frame. Using these α values and dropping the q subscript without loss of generality, the chirping sinusoids in adjacent frames can be aligned as in the previous section, yielding the phase expression:

$$\phi_{i+1} = \phi_i + \frac{N}{4}(3\omega_i - \omega_{i+1}) + \left(\frac{N}{4}\right)^2(9\alpha_i - \alpha_{i+1})$$

This phase is applied to the spectral representation of the frame $i + 1$ chirp before the IFFT.

In the constant frequency case of section 2, the additive synthesis is done by accumulating modulated versions of the window transform. In the chirp case, however, the time-windowing does not correspond to a simple modulation in frequency. The spectrum of a windowed chirping partial changes as a function of the chirp rate. For synthesis, then, the
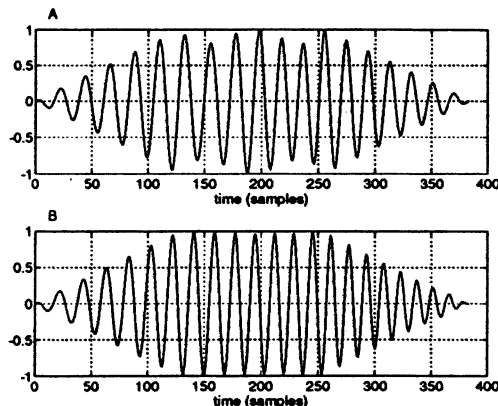

Figure 1: A. Distortion in OLA of adjacent frames for $f_i = 2000Hz$ and $f_{i+1} = 2300Hz$; B. Removal of distortion (with $f_{i+2} = 2350Hz$ and $f_{i+3} = 2750Hz$). The sampling rate is $44.1kHz$, the window length is 256, and the overlap factor is 50%.

accumulated components must be versions of $B(\omega)$ modified according to the chirp rate; for computational feasibility, a first order dependence is desirable. This suggests the application of a Taylor series approximation of $B(\omega, \alpha)$ around $\alpha = 0$. Unfortunately, the behavior of $B(\omega, \alpha)$ as a function of α is highly nonlinear even for small α, meaning that a linear Taylor series is inadequate; such a linear approximation yields significant time-domain artifacts. Thus, a piecewise linear approach has been adopted. Here, the chirped window transform $B(\omega, \alpha_p)$ is precomputed and stored for a set $\{\alpha_p\}$ spaced throughout the desired range of chirp rates. Then, for $\alpha_p \leq \alpha \leq \alpha_{p+1}$, the function $B(\omega, \alpha)$ is approximated as a linear interpolation between $B(\omega, \alpha_p)$ and $B(\omega, \alpha_{p+1})$. This approach removes the distortion effectively, as indicated in figure 1B. The improvement is a result of both the phase-matching of ϕ_{i+1} and the frequency-matching imposed by the choice of α. A similar improvement can be achieved by simply rounding α to the nearest of the $\{\alpha_p\}$, thereby removing the interpolation and reducing the computational cost.

4. REFERENCES

[1] X. Rodet and Ph. Depalle. A new additive synthesis method using inverse fourier transform and spectral envelopes. *International Computer Music Conference*, 1992.

[2] A. Freed, X. Rodet, and Ph. Depalle. Synthesis and control of hundreds of sinusoidal partials on a desktop computer without custom hardware. *International Conference on Signal Processing Applications and Technology*, 1993.