

OVERLAP-ADD SYNTHESIS OF NONSTATIONARY SINUSOIDS

Michael Goodwin and Alex Kogon

Center for New Music and Audio Technologies (CNMAT)
University of California, 1750 Arch Street, Berkeley, CA 94709
e-mail: michaelg@cnmat.berkeley.edu, alko@cnmat.berkeley.edu

ABSTRACT: In overlap-add synthesis of sinusoidal partials, amplitude modulation distortion can occur in the overlap region if the partial frequencies in successive frames differ. The frame-to-frame frequency changes can be accommodated by synthesizing linearly time-varying sinusoids (*chirps*) in adjacent frames. In this paper we present a new approach to chirp synthesis in which the distortion is eliminated by using a succession of chirp rates in each frame for each partial.

1. INTRODUCTION

A common mathematical model for a musical signal is a sum of time-varying sinusoids or *partials*:

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

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 parameters of the q -th partial, respectively. For the sake of generality, the argument of the sinusoid is often expressed in the literature as a total phase $\Phi_q[n]$. Analysis methods based on this model derive estimates of the partial parameters, generally on a frame-by-frame basis as in [McAulay & Quatieri]. Since these frame-by-frame estimates provide only a sparse description of the time evolution of the signal, it is necessary for the accompanying synthesis to interpolate the parameter tracks. For instance, in the time-domain speech synthesizer proposed in [McAulay & Quatieri], the interpolation across adjacent frames is linear for the amplitude and cubic for the total phase. In such time-domain approaches with non-overlapping output frames, arbitrary interpolation models can generally be imposed. In synthesizers based on overlapping frames, however, the interpolation model is restricted by the overlap mechanism. For instance, the overlap-add (OLA) process in the inverse FFT (IFFT) synthesizer described in [Rodet & Depalle] imposes a nonlinear frequency interpolation that can result in distortion in the overlap region.

2. FREQUENCY-DOMAIN ADDITIVE SYNTHESIS

Frequency-domain additive synthesis was introduced in [Rodet & Depalle] as an alternative to time-domain approaches. The first stage in the synthesis is the construction of the short-time spectrum of a given frame. For each partial, a spectral *motif* $B(\omega)$ is added into the spectrum centered at frequency ω_q and weighted by $A_q e^{j\phi_q}$. Since time-domain multiplication is equivalent to frequency-domain convolution, the accumulated spectrum corresponds to a time-windowed sum of sinusoids:

$$\sum_{q=1}^Q A_q e^{j\phi_q} B(\omega - \omega_q) = B(\omega) * \sum_{q=1}^Q A_q e^{j\phi_q} \delta(\omega - \omega_q) \xrightarrow{\mathcal{F}^{-1}} b[n] \sum_{q=1}^Q A_q \cos(\omega_q n + \phi_q)$$

where $*$ denotes convolution, \mathcal{F}^{-1} denotes an inverse discrete Fourier transform (implemented as an IFFT), and where the time window $b[n]$ is the inverse transform of the motif $B(\omega)$. The conjugate symmetric components of the spectrum have been omitted for simplicity. The motif $B(\omega)$ is designed to be localized in frequency so that the spectral accumulation of partials is computationally efficient.

After the IFFT, $b[n]$ is divided out, which results in a frame containing a sum of partials with constant amplitudes A_q and frequencies ω_q . The full output is constructed using overlap-add with a triangular window, which provides linear amplitude interpolation. The frequency interpolation is not linear; if the frequency of a partial changes from frame to frame, the OLA involves a sum of sinusoids of different frequency, which may result in distortion. This distortion can be lessened by phase-matching the partials in adjacent frames [Rodet & Depalle, Goodwin & Rodet].

3. CHIRP SYNTHESIS

The OLA distortion can be further alleviated by synthesizing *chirps* in each frame instead of constant frequency partials; a chirp is a sinusoid with a linearly increasing (or decreasing) frequency, which is expressed as $\cos((\omega + \alpha n)n + \phi) = \cos(\omega n + \alpha n^2 + \phi)$. For each partial in each frame a new parameter α that describes the rate of frequency change or *chirp rate* is derived from the partial frequency parameters according to a frequency-matching criterion; this is coupled with phase-matching in the synthesis. The spectral motif for the q -th partial is then a function of its chirp rate α_q ; ideally, it should satisfy $B(\omega, \alpha) = \mathcal{F}\{b[n]e^{j\alpha n^2}\}$ for $\alpha = \alpha_q$. Since computing $B(\omega, \alpha_q)$ for each partial is costly, a piecewise linear estimate of $B(\omega, \alpha)$ is used; the chirped motif $B(\omega, \alpha_p)$ is pre-computed 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 $B(\omega, \alpha_{p+1})$. The distortion improvement provided by this chirp synthesis is illustrated in [Goodwin & Rodet].

Any remaining distortion is a result of mismatched chirp rates in adjacent frames. For a 50% overlap ratio, the distortion can be removed altogether by synthesizing partials with two different chirp rates in each frame. In this technique, the chirp rate of a partial during the first half of a frame matches the final rate of that partial in the previous frame; during the second half, the chirp rate matches the initial rate of the next frame. This results in linear frequency interpolation, which when coupled with the OLA amplitude interpolation implies modeling the signal as a sum of sinusoids with linear amplitude and frequency variations, which are intuitively reasonable components. Note that this approach can be generalized to arbitrary overlap ratios by changing the number of chirps in each frame; the dual-chirp case is adhered to for simplicity.

Dual-chirp synthesis can be implemented as follows. First, the chirp rates for each half of a frame are determined using a frequency-matching constraint at the frame boundaries. Then, spectral representations of the first-half and second-half chirping partials are separately derived using the piecewise linear approximation of $B(\omega, \alpha)$ as in the single-chirp synthesis; in this case, though, $b[n]$ is half the length of a frame. These spectra are processed by two half-size IFFTs, and $b[n]$ is divided out of each of the IFFT outputs. Then, the first-half signal is multiplied by the rising edge of the triangular OLA window and the second-half signal is multiplied by the falling edge. The first and second half are then merged, and the OLA proceeds as normal. A phase-matching constraint is used in the spectral construction to prevent discontinuities at the boundary of the two halves. Since the final chirp rate of a frame matches the initial rate of the next frame, only one half-size IFFT must actually be computed for each frame, meaning that this approach entails a computational savings with respect to the single-chirp synthesis. Finally, it should be noted that the dual-chirp synthesis performs the frequency interpolation in the frequency domain and the amplitude interpolation in the time domain; the fact that this approach eliminates the OLA distortion further verifies the effectiveness of merged time-frequency methods.

4. REFERENCES

- R. J. McAulay and T. F. Quatieri. Speech analysis/synthesis based on a sinusoidal representation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(4), August 1986.
- X. Rodet and Ph. Depalle. A new additive synthesis method using inverse Fourier transform and spectral envelopes. *ICMC*, 1992.
- M. Goodwin and X. Rodet. Efficient Fourier synthesis of nonstationary sinusoids. *ICMC*, 1994.