Bandwidth Enhanced Sinusoidal Modeling in Lemur

Kelly Fitz and Lippold Haken CERL Sound Group, University of Illinois 103 S. Mathews, Urbana IL 61801, USA lemur@uiuc.edu

Abstract

We present a system for sound modeling and synthesis that preserves the elegance and malleability of a sinusoidal model, while accommodating sounds with noisy (non-sinusoidal) components. We use an enhanced McAulay-Quatieri (MQ) style analysis that extracts bandwidth information in addition to the sinusoidal parameters for each partial. To produce noisy components, we synthesize with sine wave oscillators that have been modified to allow the introduction of variable bandwidth. The enhanced MQ analysis and the bandwidth-ehanced oscillators are implemented in Lemur, a widely distributed MacintoshTM program (free at the demonstration), which will be used to demonstrate the technique.

Introduction

We are developing a homogeneous model for sounds composed of both sinusoidal components and noiselike components. The model is homogeneous in the sense that sinusoidal and noise components share a common representation. Our goal is to design a general tool for sound analysis/modification/synthesis, based on a model that allows manipulation of sounds with both kinds of components. Since sinusoidal and noise components will have the same representation in our model, we will be able to apply the same manipulations and transformations to both.

We base our work on the considerable strength of sinusoidal modeling and synthesis (or additive synthesis). Sinusoidal modeling and synthesis are computationally very expensive, but they are more powerful and more flexible than other methods, because fine control can be exercised over the behavior of each individual frequency component.

The Lemur Platform

We use the Macintosh[™] application, Lemur, as a platform for the development of our sound model. Lemur is an implementation of an extended McAulay-Quatieri (MQ) (McAulay and Quatieri 1986) algorithm for sound analysis and synthesis based on the work of Maher and Beauchamp (Maher 1989). Lemur analysis consists of a series of short-time Fourier spectra from which significant frequency components are selected. Similar components in successive spectra are linked to form time-varying partials, called tracks. The number of significant frequency components, and, thus, the number of tracks may vary over the duration of a sound. Figure 1 shows an example of a cello tone analyzed in Lemur and displayed in the Lemur editor. Synthesis is performed by a bank of oscillators, each oscillator reproducing the frequency and amplitude trajectory of a single track. The Lemur model allows extensive modification of the sound using Lemur's built-in editing functions, or using other customized editors to modify the intermediate analysis file before resynthesis (Fitz, Walker, and Haken 1992, Fitz, Haken, and Holloway 1995).



Figure 1. Example of Lemur analysis data from a cello tone, displayed in Lemur's editing window.

Noise Representation in Sinusoidal Models

Signals with strong noise components pose a problem for any sinusoidal model, since noise (even bandlimited noise) requires an infinite number of sinusoids for accurate representation. Sinusoidal modeling techniques (like the MQ technique) usually represent noise in a signal as many short tracks with widely and rapidly varying frequencies and amplitudes, as shown in Figure 2. In this way, they are capable of producing good quality syntheses of many sounds with noise components that are weak relative to the sinusoidal components. When such signals are stretched in time, however, the tracks representing the noise are stretched and can be heard as rapidly modulated sine waves. Synthesis of noisy signals modeled and stretched in this way has been described as "wormy". In fact, since the noisy character of the sound is carried mostly in the phase contributions from these many short tracks, any time or frequency scale modification (both of which inevitably change the phase portrait of the model) tends to destroy the properties of the noise, and produce wormy syntheses. Moreover, the modeling of noise as a collection of jittery tracks is intuitively unsatisfactory because it provides no means of manipulating compositionally useful parameters of the noise, such as bandwidth and center frequency, and no means of separating the noise into distinct components.



Figure 2. Lemur analysis of a sound with strong noise components, showing the proliferation of short, jittery tracks.

Stochastic Modeling

Xavier Serra proposed a method for separating the noise components from a sinusoidal model (Serra, 1989). Serra's algorithm performs an MQ analysis and resynthesis of the signal, and then computes the spectral difference between the original signal and the resynthesized signal. He inverts the difference spectrum to produce a difference signal he calls the "residual". The residual may be stored and used in future resyntheses, or its short-time spectra may be stored, and synthesis performed using inverse spectral analysis (stochastic modeling). While this method yields very high fidelity resyntheses (unmodified resynthesis with the residual, by definition, yields resyntheses indistinguishable from the original sound), it does not address all of the concerns of our research.

A fundamental problem we encountered with this approach is that the MQ analysis is being applied to a noisy signal that it cannot possibly model well, and its failure is being collected as the residual. The sinusoidal analysis is, therefore, cluttered with jittery tracks that are trying to model noise, and any stretched synthesis from this analysis will produce the wormy artifacts described above. Serra's algorithm addresses problem by attempting to remove the wormy tracks from the analysis data before performing spectral subtraction and residual construction. In our experience, however, it is not always possible to remove the entire noise contribution from a MQ-style analysis. The worminess corrupts the tracks representing the sinusoidal components.

Stochastic models do not to fully address the problem of time and frequency scale modification of the noise components in the modeled sound. Clearly, the raw residual signal cannot easily be modified, because it is simply an audio signal of the same length as the original signal. Neither can the short-time spectra be stretched without introducing audible artifacts caused by phase discontinuities at frame boundaries. A separate process is needed for modifying the stochastic part of the model. The problem of performing modified synthesis from short-time spectra has a long history (Allen 1977, Crochiere 1980, Portnoff 1980). Until the introduction of the MQ algorithm, all the solutions to this problem were pitch tracking algorithms that worked well only for monophonic, harmonic sounds.

The greatest shortcoming we find in this and other stochastic methods of accommodating noise, including those which represent noise energy in fixed frequency bands, is that they do not provide a unified representation of sinusoidal and noise components, and therefore don't allow all components of the sound to be edited and manipulated together.

Bandwidth Enhancement

In our enhanced sinusoidal model, we associate a bandwidth with each track in our sinusoidal model. Like frequency and amplitude, the bandwidth will vary over the life of the track. Bandwidth will be measured from the spectrum for each peak found during analysis. For synthesis of these bandwidth enhanced partials, we will use a sinusoidal oscillator modified by the addition of frequency modulation by band-limited noise.

Sinusoidal frequency modulation, implemented using the complex phase modulation equation,

$$x(t) = e^{j(\omega_c t + I\sin(\omega_m t))}$$

generates spectral sidebands at frequencies that are the sum and difference of the carrier frequency, ω_c , and integer multiples of the modulating frequency, ω_m (i.e. frequencies $\omega_c \pm n\omega_m$ n = 0, 1, 2, 3...) (Dodge and Jerse 1985). For very small indices of modulation, only the first sidebands (at frequencies $\omega_c \pm \omega_m$) are of significant amplitude (Stremler 1982). We use narrowband (small modulation index) frequency modulation with a filtered noise modulator to implement a bandwidth enhanced oscillator for additive synthesis. We compute the contribution of the noise modulation by expressing it in the Taylor series expansion,

$$x(t) = e^{j(\omega_0 t + Is(t))} = e^{j\omega_0 t} e^{jIs(t)} = e^{j\omega_0 t} \left(1 + jIs(t) - \frac{1}{2}I^2 s^2(t)...\right)$$

where s(t) is the noise modulator, and I is the index of modulation. For small indices of modulation, we can neglect the higher order terms in the Taylor expansion. For I = 0 (no modulation, corresponding to zero bandwidth), x(t) reduces to a simple sinusoid. The spectral contribution from noise modulation can be computed by taking the Fourier transform of the time domain equation, giving

$$X(\omega) = \delta(\omega - \omega_0) * \left(\delta(\omega) + jIS(\omega) - \frac{1}{2}I^2S(\omega) * S(\omega)...\right)$$

where $\delta(\omega - \omega_0)$ is the delta function produced by a pure sinusoid of frequency ω_0 , and $S(\omega)$ is the spectrum of the noise modulator. For very small modulation indices, only the first $S(\omega)$ term will be significant, so the effect of using modulating with noise is to reshape the spectrum of the oscillator, by adding approximately a scaled copy of the noise modulator's spectrum centered at the carrier frequency, ω_0 .

When I = 0, the convolution reduces to the delta function, $\delta(\omega - \omega_0)$, the spectrum of a sine wave at frequency ω_0 .

If the spectrum of the modulator rolls off smoothly, as in the case of low pass filtered noise (we have used a four tap averaging lowpass filter), then the frequency modulation will produce an approximately bell-shaped spectrum centered at the carrier frequency, as shown in Figure 3. The exact shape will be determined by the spectrum of the modulator, and the bandwidth will be proportional to the index of modulation.



Figure 3. Spectral snapshot of the signal produced by a bandwidth enhanced oscillator. The spectrum on the left corresponds to a partial with zero bandwidth (a sinusoid). The spectrum on the right corresponds to a partial with non-zero bandwidth.

Another Dimension

The use of bandwidth-enhanced oscillators for synthesis adds another dimension to the Lemur sinusoidal model, without sacrificing the intuitive sense of the model. Where we previously used with tracks with time varying amplitude and frequency, we now use tracks with time varying amplitude, frequency, and bandwidth. Bandwidth-enhanced oscillators allow a composer to manipulate noise-like components of sound in an intuitive way, using a familiar set of controls. The control parameters for the bandwidth enhanced sinusoidal model, amplitude, (center) frequency, and bandwidth, can be used to manipulate and transform both sinusoidal and noise-like components of sound.

Lemur Pro for the Macintosh or PowerMac is available by anonymous ftp at medusa.cerl.uiuc.edu, in pub/lemur. The authors may be contacted at lemur@uiuc.edu.

References

Jont B. Allen, "Short Term Sectral Analysis, Synthesis, and Modification by Discrete Fourier Transform," *IEEE Trans. Acous, Speech, Signal Processing*, vol ASSP-25, pp. 235-238, 1977.

R.E. Crochiere, "A Weighted Overlap-Add Method of Short-Time Fourier Analysis/Synthesis," *IEEE Trans. Acous, Speech, Signal Processing*, vol ASSP-28, pp. 99-102, 1980.

Charles Dodge and Thomas Jerse, Computer Music, New York, NY: Schirmer Books, 1985.

Kelly Fitz, William Walker, and Lippold Haken, "Extending the McAulay-Quatieri Analysis for Synthesis With a Limited Number Of Oscillators," *Proc. Intl. Computer Music Conf.*, 1992, pp. 381-382.

Kelly Fitz, Lippold Haken, and Bryan Holloway, "Lemur - A Tool for Timbre Manipulation," to appear in *Proc. Intl. Computer Music Conf.*, 1995.

Adrian Freed, Xavier Rodet, and Phillipe Depalle, "Synthesis and Control of Hundreds of Sinusoidal Partials on a Desktop Computer without Custom Hardware," *Proc. ICSPAT*, 1993.

Robert C. Maher, "An Approach for the Separation of Voices in Composite Musical Signals," Ph.D. dissertation, Dept. of Computer Science, University Of Illinois at Urbana-Champaign, 1989.

Robert J. McAulay and Thomas Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation," *IEEE Trans. Acous, Speech, Signal Processing*, vol ASSP-34, pp. 744-754, 1986.

Michael R. Portnoff, "Time-Frequency Representation of Digital Signals and Systems Based on Short-Time Fourier Analysis," *IEEE Trans. Acous, Speech, Signal Processing*, vol ASSP-8, pp. 55-69, 1980.

Xavier Serra, "A System for Sound Analysis/Transformation/Synthesis Based on a Deterministic Plus Stochastic Decomposition," Ph.D. dissertation, Dept. of Music, Stanford University, Stanford CA, 1989.

Ferrel G. Stremler, *Introduction to Communication Systems*, Reading, MA: Addison-Wesley Publishing Co., 1982.

MacintoshTM is a registered trademark of the Apple Computer Corporation.