A NEW ALGORITHM FOR BANDWIDTH ASSOCIATION IN BANDWIDTH-ENHANCED ADDITIVE SOUND MODELING

Kelly Fitz, Lippold Haken, and Paul Christensen

CERL Sound Group, University of Illinois at Urbana-Champaign
www.cerlsoundgroup.org
loris@cerlsoundgroup.org

ABSTRACT

The Bandwidth-Enhanced Additive Model represents sound as a collection of partials having sinusoidal and noise-like characteristics. Bandwidth Association is the process of constructing the bandwidth envelopes for partials in the bandwidth-enhanced model by associating noise energy not captured by conventional sinusoidal modeling processes. We present analytical methods for collecting and associating noise energy by extracting additional information from the short-time analysis spectra.

1. NOISE IN SINUSOIDAL MODELS

In representations derived from purely sinusoidal analyses, sounds are described by a collection of sinusoidal components called partials. Sinusoidal partials are defined by time-varying amplitude and frequency envelopes formed by linking spectral energy peaks extracted from short-time Fourier spectra [1, 2, 3]. For sounds that are locally nearly periodic, having very highly concentrated short-time spectral energy, a perceptually complete representation can be constructed using sinusoidal methods.

Signals with significant noise energy are problematic for sinusoidal models. The conditions established by Karhunen-Loève are met by associating noise energy not captured by conventional sinusoidal modeling processes. We present analytical methods for collecting and associating noise energy by extracting additional information from the short-time analysis spectra.

2. THE BANDWIDTH-ENHANCED ADDITIVE MODEL

The Bandwidth-Enhanced Additive Model is similar in spirit to traditional sinusoidal sound models in that a sound is modeled as a collection of components, called partials, having time-varying amplitude and frequency envelopes. Bandwidth enhancement expands the notion of a partial to include the representation of both sinusoidal and noise energy by a single component type.

Bandwidth-enhanced partials are defined by a trio of synchronized breakpoint envelopes specifying the time-varying amplitude, center frequency, and noise content (or bandwidth) for each component. The bandwidth envelope allows us to represent a mixture of sinusoidal and noise energy with a single component.

Figure 1 shows a 3D spectrogram plot of a flute tone. Breath noise is a significant component of this sound. This noise is visible between the strong harmonic components in the spectrogram, particularly at frequencies above 3 kHz. The absence of the breath noise is apparent in the spectrogram plot of a sinusoidal reconstruction from quasi-harmonic, non-bandwidth-enhanced analysis data, shown in Figure 2. The breath noise is captured in bandwidth-enhanced analysis, and faithfully reproduced in the bandwidth-enhanced reconstruction plotted in Figure 3, even though the analysis data include only partials near harmonic frequencies.

Bandwidth Association is the process of constructing the bandwidth envelopes for partials in the bandwidth-enhanced model, that is, determining how much noise energy should be represented by each bandwidth-enhanced partial. In Section 3 we discuss the exchange of energy between partials, specifically the redistribution of energy embodied bypartials that are removed from the representation. This technique allows us to prune or “clean up” representations of noisy sounds cluttered by jittery noise partials. In Section 4 we present analytical methods for associating energy not captured by the sinusoidal modeling process. These analytical methods are enhancements to the short-time sinusoidal analysis, and as such, extract additional information from short-time transforms during the analysis process.

The Bandwidth-Enhanced Additive Model could easily be coupled with an FET synthesizer [8, 11, 12] or with any other synthesis engine capable of generating both noise and sinusoids. We introduced of a new kind of oscillator, called...
Bandwidth-Enhanced Additive Model

The use of bandwidth-enhanced oscillators for additive synthesis adds another dimension to the basic sinusoidal model. Whereas sinusoidal partials are described by time-varying amplitude and frequency, bandwidth-enhanced partials are described by time-varying amplitude, center frequency, and bandwidth coefficient. A model based on bandwidth-enhanced partials, the Bandwidth-Enhanced Additive Model, can represent a great variety of sounds at high fidelity without sacrificing the intuitive sense of the sinusoidal model.

The time-variant parameters of bandwidth-enhanced partials can be used to manipulate both sinusoidal and noisy components of sound in an intuitive way, using a familiar set of controls. The encoding of noise associated with a bandwidth-enhanced partial is robust under partial parameter transformations, and is independent of other partials in the representation. Bandwidth-enhanced partials can be modified without destroying the character of the noise or introducing audible artifacts related to the representation of noise. Even changes in time or frequency scale, which degrade sinusoidal reconstructions of noisy sounds, do not adversely or unpredictably affect the character of noise synthesized from bandwidth-enhanced partials.

3. ENERGY REDISTRIBUTION

Figure 5 is a plot of sinusoidal analysis data for a breathy flute sound, partitioned into sinusoidal and noise representations. Par-
The post hoc redistribution algorithm described in Section 3 does not increase the energy in the representation; it only redistributes energy already captured in undesirable sinusoidal partials. Energy redistribution cannot compensate for an energy shortage in part of the frequency spectrum. It is not always feasible to obtain an energy-complete sinusoidal representation, especially for sounds that have substantial nonsinusoidal energy. Since noise partials are indistinguishable from other partials, sinusoidal representations of noisy sounds are difficult to clean up using post hoc methods. A better representation can be constructed by tailoring the sinusoidal analysis parameters to the deterministic part of the sound, and augmenting the analysis procedure with an algorithm for associating spectral energy not captured by the sinusoidal analysis.

A straightforward approach to bandwidth association attempts to match the short-time spectral energy and its approximate distribution in frequency. The total short-time spectral energy can be computed from Parseval’s theorem. The discrete statement of Parseval’s theorem is

\[
\sum_{n=0}^{N-1} |x_n|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X_k|^2
\]

where \(x_n\) is a sampled waveform and \(X_k\) is its discrete Fourier transform.

Short-time spectral energy not represented in the extracted sinusoidal components is treated as noise energy and distributed among the sinusoidal components as bandwidth. To approximate the frequency distribution of the noise energy, the short-time magnitude spectrum is partitioned into frequency regions, and the spectral energy matched in each region in order to approximate the short-time energy distribution. To the extent that spectral smearing due to the analysis window is minimized, the energy in a region of the spectrum, bounded by \(\omega_{\text{lower}}\) and \(\omega_{\text{upper}}\), is approximately

\[
E \approx \frac{1}{N} \sum_{k \in R} |X_k|^2
\]

where

\[
R = \{k \mid \omega_{\text{lower}} < \omega_k < \omega_{\text{upper}} \}.
\]

For strongly periodic sounds, the energy will be concentrated in the main lobes of the window spectra centered at the sinusoidal frequencies. For noisy sounds having less concentrated spectra, energy will be distributed over larger regions of the spectrum, and Equation (3) can be used to approximate the energy in an arbitrary frequency region.

The energy computed for a region is normalized for the effects of the analysis window, and compared with the energy represented by the sinusoidal components extracted from that spectral region (the energy of a sinusoid is proportional to the square of its amplitude). The difference is the excess energy, not attributable to sinusoids, that should be distributed as bandwidth.

If narrow frequency regions are used, such that few partials occupy each region, then slight changes in the spectrum of the analyzed waveform yield bandwidth envelopes that are so erratic that they introduce modulation artifacts, unexpectedly increasing the audible bandwidth of the reconstructed partials, making the noise difficult to control and degrading the representation.

High-fidelity reconstructions are obtained, however, by using wide and overlapping regions, occupied by many partials. To avoid boundary problems without sacrificing frequency localization, the bandwidth association regions overlap in frequency, and each region has a corresponding weighting function, much like a window function in spectral analysis. Components falling in two or more regions make a weighted contribution to the various overlapping association regions, and make the greatest contribution to regions centered near the component’s frequency. Wide, overlapping regions are not sensitive to the changes in spectral shape and

![Figure 5: Graph of non-bandwidth-enhanced sinusoidal partials for the breathy flute sound. Partial amplitude is not indicated on this plot. Partial amplitude is not indicated on this plot.](Image 406x443 to 414x444)

Figure 5: Graph of non-bandwidth-enhanced sinusoidal partials for the breathy flute sound. Partial amplitude is not indicated on this plot. Partial amplitude is not indicated on this plot. The remaining partials, drawn in light grey, are assigned to represent the noisy part of the sound, and their energy is converted from sinusoidal to noise energy, or bandwidth.
partial density that produce erratic bandwidth envelopes in algorithms employing narrow association regions.

The component weighting for a set of wide, overlapping bandwidth association regions is shown in Figure 6a. The regions are distributed uniformly in (linear) frequency, and their width is equal to twice the distance between their centers, so that the weighting functions are triangular (when plotted on a linear frequency scale) and every spectral component divides its energy between two regions according to its proximity to the regions’ center frequencies.

Employing loudness as a metric in place of energy, we designed a bandwidth association algorithm that matches the perceived distribution of spectral energy by matching loudness (perceived signal intensity) in overlapping frequency regions distributed uniformly in bark frequency.

For narrow-bandwidth tone complexes, loudness, \( L \), is a function of the total signal intensity:

\[
L = C(\omega) \sqrt{I_1 + I_2 + \ldots + I_N}
\]

where \( I_n \) is the intensity due to the \( n \)th tone and \( C(\omega) \) is a constant that depends on the center frequency of the aggregate [17].

If we define bandwidth association regions to be sufficiently narrow, then we can compute the combined loudness of all the components in a region from their aggregate intensity using Equation (5). The component weighting for one such set of bandwidth association regions is shown in Figure 6b. The regions are distributed uniformly on a bark scale, and their width is equal to twice the distance, in bark frequency, between their centers.

We obtain high-fidelity representations of a variety of noisy and non-noisy instrument tones with both energy-matching and loudness-matching bandwidth association algorithms. We retain the bow scraping of cello tones and the characteristic breathiness of flute and clarinet tones in the bandwidth-enhanced representation, and to reproduce them at high fidelity and without artifacts, even under modifications such as pitch shift, time dilation, and morphing. The noise energy was perceived to be appropriately distributed, and the bandwidth envelopes were well behaved and easy to manipulate.

5. CONCLUSION

The bandwidth-enhanced additive model, though not strictly sinusoidal, retains many of the desirable characteristics of the basic sinusoidal model, specifically its homogeneity and manipulability. Using bandwidth association, we obtain manipulable, high-fidelity, artifact-free representations of a variety of noisy and non-noisy sounds.

References