

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 analysis specify that a very large number of sinusoids are needed in each short-time analysis frame to adequately represent even narrowband noise [4]. Sinusoidal analysis algorithms generally retain too few noise partials to adequately characterize noisy sounds, and the resulting reconstructions are often described as “tinny,” “wormy,” “flanged,” or “reverberant.”

For sounds with noise energy that is weak relative to the sinusoidal energy, a large number of short, jittery partials may be adequate to synthesize the noise in the analyzed waveform, but the representation is sensitive to changes in time and frequency scale. Furthermore, this representation provides no means of distinguishing noisy components from deterministic components, and no means of independently manipulating the noise. The fragility and inaccessibility of the noise representation make purely sinusoidal models unsatisfactory for manipulating many sounds of interest.

Hybrid sinusoidal models, incorporating additional, non-sinusoidal components to represent noise, often yield high fidelity reconstructions, even from modified data [2, 5, 6, 7], but the synthesized noise sometimes fails to fuse with the sinusoids into a single sound, particularly under transformations of time or frequency scale [8]. Moreover, for non-harmonic or extra-musical waveforms, in which it is difficult to distinguish noise energy from sinusoidal energy and to prevent the former from corrupting the sinusoidal representation, even hybrid models may yield fragile or immutable representations.

Our Bandwidth-Enhanced Additive Model is a robust, high-fidelity, homogeneous (employing a single component type) representation of sounds having significant nonsinusoidal energy. The homogeneity of the representation is essential for implementing certain kinds of model-domain behaviors and real-time timbral manipulation. For example, real-time morphing of quasi-harmonic sounds is greatly facilitated by a representation consisting of one partial per harmonic, and can be achieved at no cost to fidelity using bandwidth-enhanced partials [9, 10].

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 by partials 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 FFT^{-1} 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

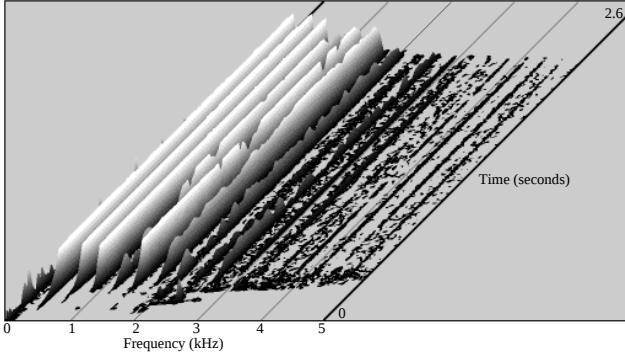


Figure 1: 3D spectrogram plot for a breathy flute tone (D above middle C). Audible low-frequency noise and rumble from the recording is visible. Strong low-frequency components are clipped and appear to have unnaturally flat amplitude envelopes due to the high gain used to make low-amplitude high-frequency partials visible. Spectrograms were made using SoundMaker by Alberto Ricci.

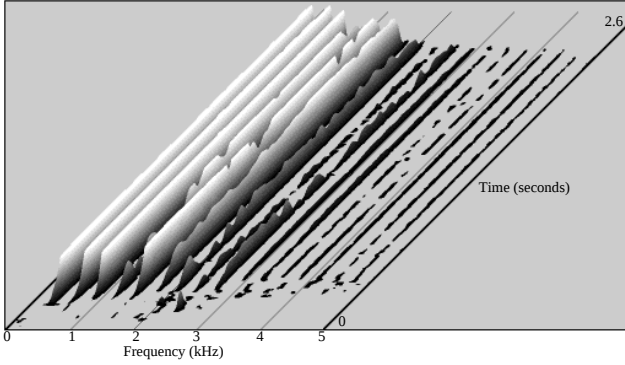


Figure 2: 3D spectrogram plot for a reconstruction of the breathy flute tone plotted in Figure 1 from purely sinusoidal, quasi-harmonic analysis data. The breath noise and the low-frequency rumble are absent in the reconstruction, since only sinusoids near harmonic frequencies are present in the analysis data.

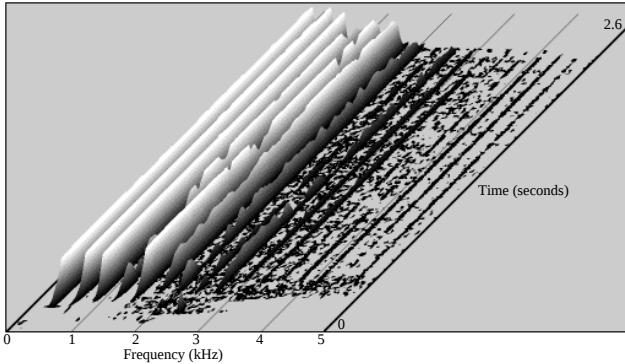


Figure 3: 3D spectrogram plot for a reconstruction of the breathy flute tone plotted in Figure 1 from bandwidth-enhanced analysis data. The breath noise visibly absent in Figure 2 is faithfully reproduced in reconstructions from bandwidth-enhanced analysis data, even though the analysis data include only partials near harmonic frequencies.

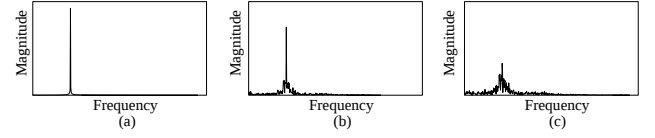


Figure 4: Spectra for partials having different amounts of spectral line widening due to bandwidth enhancement. (a) corresponds to a partial with no line widening, (b) corresponds to a partial with a moderate amount of line widening, and (c) corresponds to a partial with a large amount of line widening. (Figure reproduced from [14].)

the *Bandwidth-Enhanced Sinusoidal Oscillator*, or simply the *Bandwidth-Enhanced Oscillator*, suitable for additive synthesis of a broad class of sounds, including noisy sounds, from bandwidth-enhanced analysis data [13, 14, 15]. Bandwidth-enhanced oscillators affect spectral line widening [16] by using stochastic modulation to spread spectral energy away from the partial’s center frequency. The bandwidth-enhanced oscillator is described by

$$y_n = \tilde{A} \left(\sqrt{1 - \kappa} + \sqrt{2\kappa} [\zeta_n * h_n] \right) \cdot e^{j\omega_c n} \quad (1)$$

where $\tilde{A}$ represents the local average partial energy, κ represents the fraction of total partial energy that is attributable to noise, ω_c is the center frequency, and ζ_n is a noise sequence that excites a filter with low-pass impulse response h_n . The bandwidth coefficient κ assumes values between 0 for a pure sinusoid and 1 for pure noise.

The time-varying value of κ for a partial controls the balance of sinusoidal and noise energy, while the time-varying value of $\tilde{A}$ controls the total partial energy. As the energy in the bandlimited stochastic modulator is increased, the partial bandwidth, or spectral linewidth, increases relative to its peak spectral amplitude (though the bandwidth of the stochastic modulator is fixed). The line widening effect is shown in Figure 4, where the leftmost spectrum (a) corresponds to a partial with no line widening, the middle spectrum (b) corresponds to a partial with a moderate amount of line widening, and the rightmost spectrum (c) corresponds to a partial with a large amount of line widening.

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-

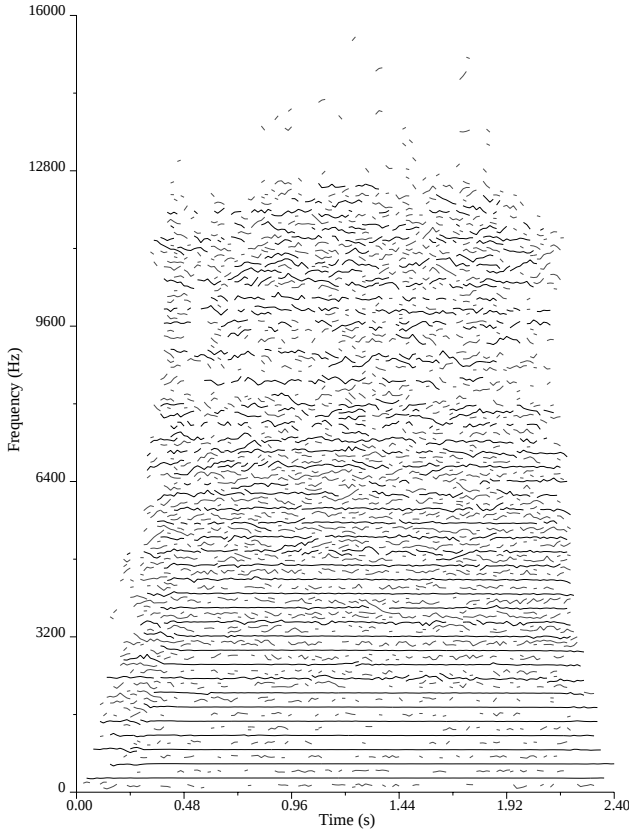


Figure 5: Graph of non-bandwidth-enhanced sinusoidal partials for the breathy flute sound. Partial amplitude is not indicated on this plot. Partial selected to represent sinusoidal energy in the flute sound, mostly near harmonic frequencies, are drawn in black, and make up about 40% of the data. The 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.

tials near harmonic frequencies were selected (somewhat arbitrarily in the case of the high-frequency partials) to represent the sinusoidal energy in the flute sound, and are shown in black. The remaining partials, about 60% of the data, are characterized as noise partials, and are shown in light grey. To improve the representation of the noise energy, the energy in the noise partials is converted from a sinusoidal representation to a noise representation by moving it all to the partial bandwidth envelopes. The encoding of noise energy in partial bandwidth envelopes is more robust than a sinusoidal representation, and yields higher-fidelity reconstructions than the purely sinusoidal representation.

Since small changes in the center frequency of narrow, overlapping noise bands are inaudible, the noise partials can be removed from the representation altogether, and their energy encoded in the bandwidth envelopes of the remaining nearby partials. *Energy redistribution* is the process of redistributing energy already captured in the representation. The energy in the noise partials (light grey) in Figure 5 can be redistributed among the quasi-harmonic partials (black) without degrading the reconstruction.

4. BANDWIDTH ASSOCIATION

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

$$\sum_{n=0}^{N-1} |x_n|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X_k|^2 \quad (2)$$

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 ω_{upper} and ω_{lower} , is approximately

$$E \simeq \frac{1}{N} \sum_{k \in R} |X_k|^2 \quad (3)$$

where

$$R = \{k \mid \omega_{lower} < \omega_k < \omega_{upper}\}. \quad (4)$$

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

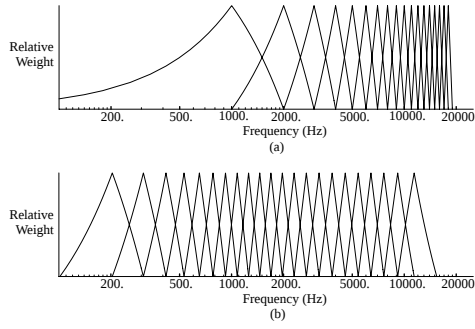


Figure 6: Tapered weighting functions due to overlapping bandwidth association regions. In (a), the regions are 2 kHz wide and their center frequencies are separated by 1 kHz. In (b), the regions are two barks wide, and their centers are separated by one bark.

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:

$$L = C(\omega_c) \sqrt[3]{I_1 + I_2 + \dots + I_N} \quad (5)$$

where I_n is the intensity due to the n th tone and $C(\omega_c)$ 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

- [1] Robert J. McAulay and Thomas F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-34, no. 4, pp. 744 – 754, Aug. 1986.
- [2] Xavier Serra and Julius O. Smith, "Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition," *Computer Music Journal*, vol. 14, no. 4, pp. 12 – 24, 1990.
- [3] Kelly Fitz and Lippold Haken, "Sinusoidal modeling and manipulation using Lemur," *Computer Music Journal*, vol. 20, no. 4, pp. 44 – 59, 1996.
- [4] Harry L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, chapter 3, John Wiley and Sons, Inc., New York, 1968.
- [5] Tony S. Verma and Teresa H. Y. Meng, "An analysis/synthesis tool for transient signals," in *Proc. 16th International Congress on Acoustics/135th Meeting of the Acoustical Society of America*, June 1998, vol. 1, pp. 77 – 78.
- [6] T. S. Verma, S. N. Levine, and T. H. Y. Meng, "Transient modeling synthesis: A flexible analysis/synthesis tool for transient signals," in *Proc. ICMC*, 1997, pp. 164 – 167.
- [7] Geoffroy Peeters and Xavier Rodet, "SINOLA: A new analysis/synthesis method using spectrum peak shape distortion, phase and reassigned spectrum," in *Proc. ICMC*, 1999, pp. 153 – 156.
- [8] Adrian Freed, "Spectral line broadening with transform domain additive synthesis," in *Proc. ICMC*, 1999, pp. 78 – 81.
- [9] Lippold Haken, Kelly Fitz, and Paul Christensen, "Beyond traditional sampling synthesis: Real-time timbre morphing using additive synthesis," in *Sound Of Music: Analysis, Synthesis, And Perception*, James W. Beauchamp, Ed. Springer-Verlag, to appear.
- [10] Lippold Haken, Kelly Fitz, Paul Christensen, et al., "The Continuum Fingerboard and real-time bandwidth-enhanced additive synthesis," Available on the world wide web at <http://www.cerlsoundgroup.org/Continuum/Cont.html>.
- [11] Adrian Freed, Xavier Rodet, and Phillipe Depalle, "Performance, synthesis and control of additive synthesis on a desktop computer using FFT^{-1} ," in *Proc. ICMC*, 1993.
- [12] Xavier Rodet and Phillipe Depalle, "Spectral envelopes and inverse fft synthesis," in *93rd Convention of the Audio Engineering Society*, San Francisco, Oct. 1992.
- [13] Kelly Fitz and Lippold Haken, "Lemur: A bandwidth-enhanced sinusoidal modeling system," in *Proc. of the 16th International Congress on Acoustics/135th Meeting of the Acoustical Society of America*, June 1998, vol. 1, pp. 77 – 78.
- [14] Kelly Fitz and Lippold Haken, "Bandwidth enhanced sinusoidal modeling in Lemur," in *Proc. ICMC*, 1995, pp. 154 – 157.
- [15] Kelly Fitz, Lippold Haken, and Bryan Holloway, "Lemur - a tool for timbre manipulation," in *Proc. ICMC*, 1995, pp. 158 – 161.
- [16] Jean-Claude Risset and David Wessel, "Exploration of timbre by analysis and synthesis," in *The Psychology of Music*, Diana Deutsch, Ed., pp. 26 – 58. Academic Press, Inc., New York, 1982.
- [17] Juan G. Roederer, *Introduction to the Physics and Psychophysics of Music*, Springer-Verlag, New York, 2nd edition, 1975.