TRANSIENT PRESERVATION UNDER TRANSFORMATION IN AN ADDITIVE SOUND MODEL

Kelly Fitz, Lippold Haken, and Paul Christensen

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

ABSTRACT

We introduce the use of the method of reassignment in sound modeling to produce a sharper, more robust additive representation. The *Reassigned Bandwidth-Enhanced Additive Model* follows ridges in a time-frequency analysis to construct partials having both sinusoidal and noise characteristics. This model yields greater resolution in time and frequency than is possible using conventional additive techniques, and preserves the temporal envelope of transient signals, even in modified reconstruction, without introducing new component types or complicated phase interpolation algorithms.

1. INTRODUCTION

The method of reassignment has been used to sharpen spectrograms to make them more readable [1], to measure sinusoidality, and to ensure optimal window alignment in analysis of musical signals [2]. We use reassignment to improve our bandwidth-enhanced additive modeling representation. Our representation is similar in spirit to traditional sinusoidal models [3, 4, 5] in that a waveform is modeled as a collection of components, called *partials*, having time-varying amplitude and frequency envelopes. Our partials are not strictly sinusoidal, however. We employ a technique of *Bandwidth Enhancement* to combine sinusoidal energy and noise energy into a single partial having time-varying amplitude, frequency, and bandwidth parameters [6, 7, 8]. Our model uses the method of reassignment to improve the time and frequency estimates used to define our partial parameter envelopes, thereby improving the time-frequency resolution of our representation, and improving its phase accuracy.

2. TIME-FREQUENCY REASSIGNMENT

The short-time Fourier transform is often used as the basis for a time-frequency representation of time-varying signals, and is defined as a function of time index n and frequency index k as

$$X_{h;k,n} = \sum_{l=-\infty}^{\infty} h_{n-l} x_l e^{\frac{-j2\pi lk}{N}} \tag{1}$$

where the subscript h specifies that the short-time waveform is windowed by the function h_n.

Typically only the short-time magnitude spectrum is considered in the time-frequency representation. The short-time phase spectrum is sometimes used to improve the frequency estimates in the time-frequency representation of quasi-harmonic sounds [9], but it is often omitted entirely, or used only in reconstruction [3].

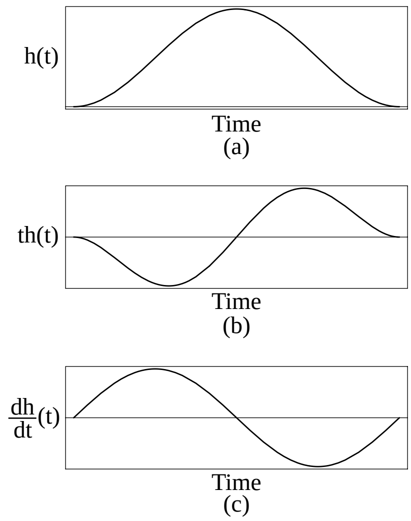

Figure 1: Window functions used by the three short-time transforms used to compute reassigned times and frequencies. Waveform (a) is the original continuous-time window function, $h(t)$, the waveform (b) is time-weighted window function, $th(t)$, and waveform (c) is the frequency-weighted window function, computed in the time domain by $\frac{dh}{dt}(t)$.

It has been shown that partial derivatives of the short-time phase spectrum can be used to substantially improve the short-time time and frequency resolution [10], and an efficient short-time Fourier transform-based implementation has been developed that does not require complex division [1].

The method of reassignment computes time and frequency values for each spectral component from the spectral data. Instead of locating each component at the center of the analysis window in time and frequency, as in traditional short-time spectral analysis, components are reassigned to the corresponding center of gravity in time and frequency, computed according to the principle of stationary phase [10].

Time and frequency coordinates can be computed using two additional short-time Fourier transforms, one employing a window function equal to h_n, the short-time window function in Equation 1, weighted by a time ramp and one employing a window function having Fourier transform equal to the transform of h_n weighted by a frequency ramp. The windows employed by the three transforms are shown in Figure 1. Note that weighting by a frequency ramp in the frequency domain is equivalent to a time derivative in the time domain, so the frequency-weighted window function can be computed as the time derivative of the original

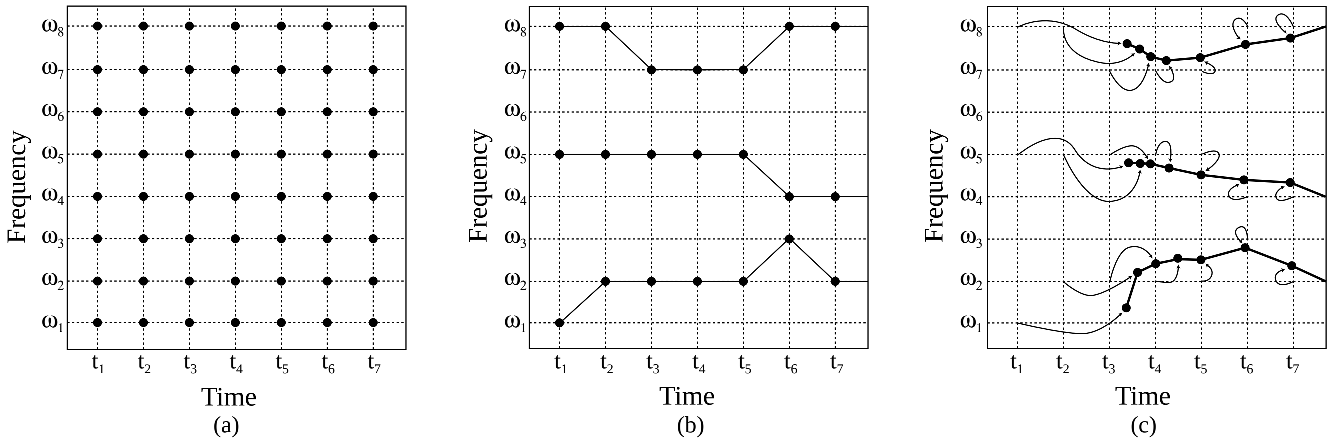

Figure 2: Comparison of time-frequency data included in common representations. Only the time-frequency orientation of the data points is shown. The short-time Fourier transform (a) retains data at every time t_n and frequency ω_k. The McAulay-Quatieri method (b) retains data at selected time and frequency samples [3]. The reassigned bandwidth-enhanced analysis data (c) is distributed continuously in time and frequency, and retained only at time-frequency ridges. Arrows indicate the mapping of short-time spectral samples onto time-frequency ridges due to the method of reassignment.

window function.

The corrected time, $\hat{t}_{k,n}$ for the short-time transform component at time n and frequency k is [1]

$$\hat{t}_{k,n} = t_n - \Re\left\{ \frac{X_{th;k,n} X^*_{h;k,n}}{\left|X_{h;k,n}\right|^2} \right\} \tag{2}$$

where $X_{th;k,n}$ denotes the short-time transform computed using the time-weighted window function th shown in Figure 1b, and t_n is the nominal time associated with the nth short-time transform, normally nH for a short-time hop size of H samples.

The corrected frequency, $\hat{\omega}_{k,n}$ is [1]

$$\hat{\omega}_{k,n} = \omega_k + \Im\left\{ \frac{X_{dh;k,n} X^*_{h;k,n}}{\left|X_{h;k,n}\right|^2} \right\} \tag{3}$$

where $X_{dh;k,n}$ denotes the short-time transform computed using the frequency-weighted window function $\frac{dh}{dt}$ shown in Figure 1c and ω_k is the nominal frequency for the kth short-time coefficient, $\frac{2\pi k}{N}$ for a transform of length N.

Note that, since the short-time Fourier transform is invertible, and the original waveform can be exactly reconstructed from an adequately-sampled short-time Fourier representation, all the information needed to precisely locate a spectral component within an analysis window is present in the short-time coefficients, $X_{h;k,n}$. Temporal information is encoded in the short-time phase spectrum, which is very difficult to interpret. The method of reassignment is a technique for extracting information from the phase spectrum.

3. REASSIGNED BANDWIDTH-ENHANCED ANALYSIS

Reassignment transforms our analysis from a frame-based analysis into a "true" time-frequency analysis. Whereas the discrete short-time Fourier transform defined by Equation 1 orients data according to the analysis frame rate and the length of the transform, the time and frequency orientation of reassigned spectral data is solely a function of the data.

The method of analysis we use in our research models a sampled audio waveform as a collection of partials having sinusoidal and noise-like characteristics. The partials are defined by a trio of synchronized breakpoint envelopes specifying the time-varying amplitude, center frequency, and noise content (bandwidth) for each component. The bandwidth-enhanced oscillator is described by

$$y_n = \tilde{A}\left(\sqrt{1-\kappa} + \sqrt{2\kappa}\left[\zeta_n * h_n\right]\right) \cdot e^{j\omega_c n} \tag{4}$$

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 partial that is entirely narrowband noise. The breakpoints for the partial parameter envelopes are obtained by following ridges on the time-frequency surface.

We use the method of reassignment to improve the time and frequency estimates for our partial parameter envelope breakpoints. Thus, our algorithm shares with traditional sinusoidal methods the notion of temporally connected partial parameter estimates, but by contrast, our estimates are non-uniformly distributed in both time and frequency, as shown in Figure 2.

Analysis windows normally overlap in both time and frequency, so time-frequency reassignment often yields time corrections greater than the length of the short-time hop size and frequency corrections greater than the width of a frequency bin. Large time corrections are common at strong transients. Since we retain data only at time-frequency ridges, we generally observe large frequency corrections only in the presence of strong noise components.

4. SHARPENING TRANSIENTS

Time-frequency representations based on traditional short-time Fourier analysis techniques fail to distinguish transient components from sustaining components. A strong transient waveform,

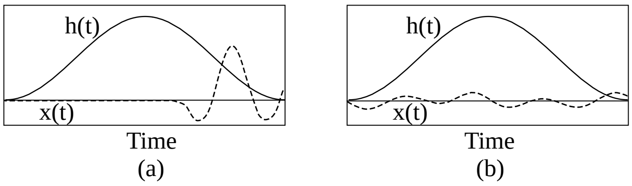

Figure 3: Two windowed short-time waveforms (dashed lines) that are not distinguished in traditional short-time analysis methods. Both waveforms are represented by low-amplitude spectral components, but the strong transient on the left (a) yields off-center components, having large time corrections (positive in this case because the transient is near the right tail of the window), while the sustained quasi-periodic waveform on the right (b) yields time corrections near zero.

as shown in Figure 3a, is represented by a collection of low amplitude spectral components in early short-time analysis frames, that is, frames corresponding to analysis windows centered earlier than the time of the transient. A low-amplitude periodic waveform, as shown in Figure 3b, is also represented by a collection of low amplitude spectral components. As mentioned in Section 2, the information needed to distinguish these two critically different waveforms is encoded in the short-time phase spectrum, and is extracted by the method of reassignment.

Reassignment greatly improves time resolution by relocating spectral peaks closer to the time of the transient events, so that transients are not smeared out by the length of the analysis window. Components extracted from early or late short-time analysis windows are reassigned to times near the time of the transient, yielding clusters of time-frequency data points, as shown in the reassigned analysis depicted in Figure 2.

Short-time components having large time reassignments, that is, components having centers of gravity far from the temporal center of the analysis window, are referred to as *off-center* components. Since they represent events (usually transient events) that are far from the center of the analysis window, and are therefore poorly represented in the windowed short-time waveform, these off-center components introduce unreliable spectral parameter estimates that corrupt our representation, making the model data difficult to interpret and manipulate.

Large time corrections make off-center components easy to identify and remove from our model. By removing the unreliable data embodied by off-center components, we make our model cleaner and more robust. Moreover, thanks to the redundancy inherent in short-time analysis with overlapping analysis windows, we do not sacrifice information by removing the unreliable data points. The information represented poorly in off-center components is more reliably represented in well-centered components, extracted from analysis windows centered nearer the time of the transient event. Figure 4 shows reassigned bandwidth-enhanced model data from the onset of a bowed cello tone before and after the removal of off-center components.

5. PHASE MAINTENANCE

Preserving phase is important for reproducing some classes of sounds, particularly transients and short-duration complex audio events having significant information in the temporal envelope [11]. The McAulay-Quatieri (MQ) sinusoidal algorithm [3] is phase correct. That is, it preserves phase at all times in unmodified reconstruction. In order to match short-time spectral frequency

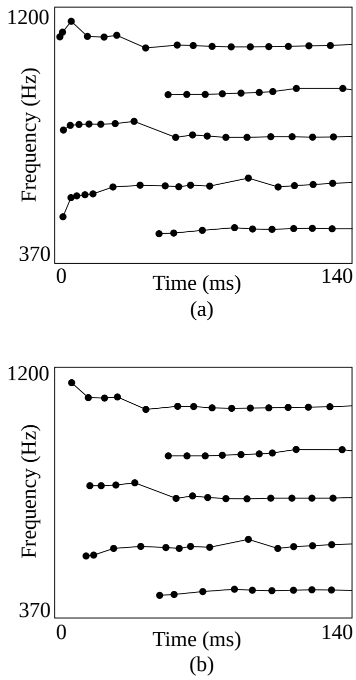

Figure 4: Time-frequency coordinates of data from reassigned bandwidth-enhanced analysis before (a) and after (b) removal of off-center components clumped together at partial births. The source waveform is a bowed cello tone.

and phase estimates at frame boundaries, the MQ method uses cubic interpolation of the partial phase.

Cubic phase has many undesirable properties for implementation. It is difficult to maintain and difficult to manipulate compared to linear frequency envelopes. However, in unmodified reconstruction, cubic interpolation prevents the propagation of phase errors introduced by unreliable parameter estimates, maintaining phase accuracy in transients, where the temporal envelope is important.

It is not desirable to preserve phase at all times in modified reconstruction. Because frequency is the time derivative of phase, any change in the time or frequency scale of a partial must correspond to a change in the sampled phases.

In general, preserving phase using the cubic phase method in the presence of modifications (or estimation errors) introduces wild frequency excursions [12]. Phase can be preserved at one time, however, and that time is typically chosen to be the onset, or birth of each partial, although any single time could be chosen. The partial phase at all other times is modified to reflect the new time-frequency characteristic of the modified partial.

Off-center components with unreliable parameter estimates introduce phase errors in modified reconstruction. Since the phase is maintained at only one time, typically the onset, even the cubic interpolation scheme cannot prevent phase errors from propagating in modified syntheses.

By removing the off-center components at the onset of a partial, we not only remove the primary source of phase errors, we also improve the shape of the temporal envelope in modified reconstruction of transients by preserving a more reliable phase estimate at a time closer to the time of transient event. We can therefore maintain phase accuracy at critical parts of the audio waveform even under transformation, and even using linear frequency envelopes, which are much simpler to compute, interpret, edit, and maintain than cubic phase curves. Removing off-center components with large time corrections leaves us with no reason to use cubic phase interpolation.

6. BREAKING PARTIALS AT TRANSIENT EVENTS

A few methods have been proposed for representing transient waveforms in additive sound models. Verma and Meng [13] introduce new component types specifically for modeling transients, but this method sacrifices the homogeneity of the model. A homogeneous model, that is, a model having a single component type, such as the breakpoint parameter envelopes in our Reassigned Bandwidth-Enhanced Model, is critical for many kinds of manipulations [14, 15]. Quatieri [11] proposes a method for preserving the temporal envelope short-duration complex acoustic signals, but it is inapplicable to longer duration waveforms. Neither of these methods addresses the phase problems associated with waveforms having multiple transient events. Peeters and Rodet [2] have developed a hybrid analysis/synthesis system that eschews high-level transient models and retains unabridged *OLA* (overlap-add) frame data at transient positions. The method of removing off-center components described in Section 5 suggests another approach.

Transients corresponding to the birth of all associated partials are preserved in our model by removing off-center components at the ends of partials. If transients always correspond to the birth of associated partials, then that method will preserve the temporal envelope of multiple transient events. In fact, however, partials often span transients. In such cases, it is not possible to preserve the phase at the locations of multiple transients, since, under modification the phase can only be preserved at one time in the life of a partial.

Strong transients are identified by the large time corrections they introduce. By breaking partials at components having large time corrections, we can cause all associated partials to be born at the time of the transient, and thereby enhance our ability to maintain phase accuracy. By breaking partials at the locations of transients, we can preserve the temporal envelope of multiple transient events, even under transformation.

7. CONCLUSIONS

We have found that the method of reassignment dramatically improves our bandwidth-enhanced additive sound model. Temporal smearing is greatly reduced because the time-frequency orientation of the model data is waveform-dependent, rather than analysis-dependent as in traditional short-time analysis methods. Moreover, time-frequency reassignment allows us to identify unreliable data points (having bad parameter estimates) and remove them from the representation. This not only sharpens the representation and makes it more robust, but it also allows us to maintain phase accuracy at transients, while avoiding the problems associated with cubic phase interpolation.

References

[1] François Auger and Patrick Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1068 – 1089, May 1995.

[2] 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.

[3] 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.

[4] 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.

[5] Kelly Fitz and Lippold Haken, "Sinusoidal modeling and manipulation using Lemur," *Computer Music Journal*, vol. 20, no. 4, pp. 44 – 59, 1996.

[6] Kelly Fitz, Lippold Haken, and Paul Christensen, "A new algorithm for bandwidth association in bandwidth-enhanced additive sound modeling," in *Proc. ICMC*, 2000, to appear.

[7] 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.

[8] Kelly Fitz and Lippold Haken, "Bandwidth enhanced sinusoidal modeling in Lemur," in *Proc. ICMC*, 1995, pp. 154 – 157.

[9] Mark Dolson, "The phase vocoder: A tutorial," *Computer Music Journal*, vol. 10, no. 4, pp. 14 – 27, 1986.

[10] Kunihiko Kodera, Roger Gendrin, and Claude de Villedary, "Analysis of time-varying signals with small BT values," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-26, no. 1, pp. 64 – 76, Feb. 1978.

[11] Thomas F. Quatieri, R. B. Dunn, and T. E. Hanna, "Time-scale modification of complex acoustic signals," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 1993, pp. I–213 – I–216.

[12] Yinong Ding and Xiaoshu Qian, "Processing of musical tones using a combined quadratic polynomial-phase sinusoidal and residual (QUASAR) signal model," *Journal of the Audio Engineering Society*, vol. 45, no. 7/8, pp. 571 – 584, July/August 1997.

[13] 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.

[14] 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.

[15] 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.