SINUSOIDALITY ANALYSIS AND NOISE SYNTHESIS IN PHASE VOCODER BASED TIME-STRETCHING

Dr. Ted Apel

New Zealand School of Music
Victoria University of Wellington, New Zealand
ted.apel@nzsm.ac.nz

ABSTRACT

A novel extension to the phase vocoder method of representing sound is presented in which the "sinusoidality" of spectral energy is estimated during analysis and employed to add noise to a time-stretched phase vocoder synthesis. Three methods of sinusoidality analysis are presented as well as a sinusoid and noise synthesis method which extends the phase vocoder method. This method allows for the noise characteristics of the original sound to be better maintained during time-stretching.

1. INTRODUCTION

The phase vocoder technique of sound analysis and synthesis is well known for its ability to "time stretch" a sound (Moorer 1978). However, when a sound is time-stretched with this technique, the noise aspects of the sound tend to become pitched. Under extreme time lengthening, all noisy aspects of the original sound are transformed into stable sinusoidal components. This behavior is consistent with the phase vocoder's modeling of short-time Fourier transform (STFT) energy as exclusively sinusoidal energy. The purpose of this research is to extend the phase vocoder (PV) representation of monophonic sounds to allow for the original noise characteristics to be maintained during PV lengthening.

In order to derive the noise characteristics of sound in a PV representation, methods of analyzing the "sinusoidality" of STFT channels is employed. STFT channels with high sinusoidality are composed of predominantly sinusoidal energy and should exhibit sinusoidal characteristics during synthesis. STFT channels with low sinusoidality have predominantly noise energy and should be synthesized with noise characteristics. In this paper, three techniques for measuring the sinusoidality of STFT channels are presented, and a new method of combining pitched and noisy components during PV synthesis using the sinusoidality is presented.

2. SINUSOIDALITY ANALYSIS

The phase vocoder is built upon the Fourier transform. The Fourier transform analyzes a signal for sinusoidal components, that is, the output consists of coefficients to a sinusoidal basis function. The sinusoidal nature of Fourier analysis is ill-suited to the analysis of noisy sounds or sounds with significant noise components because these components are analyzed as many rapidly varying sinusoidal components. While this analysis can be used to reconstruct the original sound, temporal manipulation of this Fourier energy using phase vocoder techniques typically results in the noise characteristics of the sound taking on a pitched or sinusoidal character.

In order to alleviate this problem of noisy components of a sound being misinterpreted in phase vocoder manipulations, we will attempt to analyze the phase vocoder representation for its noise characteristics, and to create these noise levels during a phase vocoder synthesis. This analysis will consist of calculating the sinusoidality of each Fourier analysis channel and, using this measure, determining noise levels for each channel during synthesis.

The sinusoidality of an STFT channel represents the degree to which the energy of each spectral bin consists of sinusoidal based energy. A low sinusoidality measure indicates that the energy in that band is based on random or noise signals. Here we will follow the notation and conventions set by Peeters and Rodet (Peeters and Rodet 1998) where $\Gamma(n,k)$ is the sinusoidality of spectral bin k at time frame n, and $\Gamma(n,k)$ varies between high sinusoidality of 1 and 0 for low sinusoidality or high-noise content.

In the next section, three methods of estimating sinusoidality are presented which are used for estimating noise levels. Two of these methods (Charpentier and Phase Acceleration) are adapted from established algorithms, while the third is a novel method based on the harmonic structure of musical sounds.

2.1. Charpentier Sinusoidality

The phase vocoder method calculates an approximation of the instantaneous frequencies of spectral components from the difference in phase between consecutive spectral frames. These frequency values can be thought of as refinements to the nominal center frequency values of each bin. Charpentier devised a pitch detection algorithm that groups spectral bins based on their similar instantaneous frequencies (Charpentier 1986). Charpentier notes that a sinusoid will exhibit energy in at least three adjacent spectral bins, and that the instantaneous frequency of these bins will be correlated around the true frequency of that sinusoidal component. Figure 1 shows the amplitude spectrum

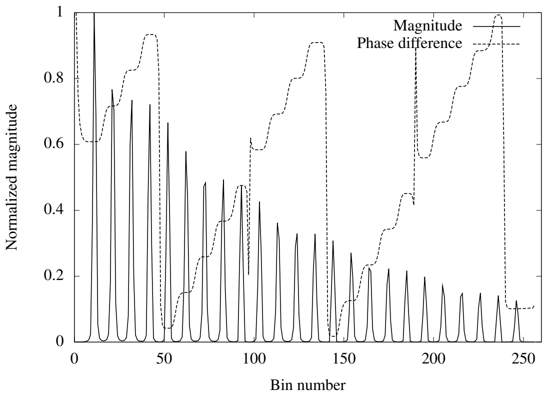

Figure 1. Amplitude spectrum and phase difference spectrum of a harmonic sound. Visible are the similar phase difference values around the sinusoidal energy peaks.

of a synthetically generated harmonic sound and the corresponding phase difference spectrum both scaled between 0 and 1. It can be clearly seen that the phase difference values stabilize around the areas with sinusoidal energy. Dressler (Dressler 2006) formalizes Charpentier's phase difference method as a decimal offset $\kappa(k)$ to the integer bin number k. Here $\kappa(k)$ is calculated from the offset between the measured phase $\theta_n(k)$ and the expected phase calculated from the prior phase of that bin, here called $\theta_{(n-1)}(k)$.

$$\kappa(k) = \left(\frac{N}{2\pi(SR)}\right)(\theta_n(k) - \theta_{(n-1)}(k)), \qquad (1)$$

Where N is the hop size and SR is the sampling-rate. Charpentier uses each value of $\kappa(k)$ to detect harmonics of a signal by declaring a bin k to contain a harmonic when $\kappa(k-1)$ and $\kappa(k+1)$ are "sufficiently close" to $\kappa(k)$. We interpret this idea here to mean that for each spectral bin k,

$$SC(k) = |\kappa(k) - \kappa(k-1)| + |\kappa(k) - \kappa(k+1)|. \qquad (2)$$

Here we can see that three points that cluster around a particular instantaneous frequency value will produce a low coefficient for that bin, and conversely, large variations in instantaneous frequency between spectral bins will produce a larger coefficient.

We can put this measure into our sinusoidality framework by normalizing the results and subtracting from one.

$$\Gamma_c(k) = 1 - \frac{SC(k)}{\underset{k}{\operatorname{argmax}}(SC(k))}. \qquad (3)$$

As discussed below, the Charpentier sinusoidality method devised here exhibits the lowest error on one of our test signals with a single sine tone and broadband noise.

2.2. Phase Acceleration Sinusoidality

Our second phase spectrum based sinusoidality measure also employs the phase difference method of instantaneous frequency computation. Settel and Lippe devised a "band-limited frequency dependent noise gate" as a method of separating stable STFT channels from non-stable channels (Settel and Lippe 1994). Their method defines a threshold for changes in the phase vocoder's instantaneous frequency values between spectral frames. Because sinusoidal components should have a relatively constant instantaneous frequency, these slowly changing components are typically sinusoidal and pitched. A mathematical presentation of the Settel and Lippe method is presented by Arfib, Keiler, and Zölzer (Arfib, Keiler, and Zölzer 2002). Here we formulate this idea in terms of a variable sinusoidality instead of a threshold for each bin, and in terms of "phase acceleration" because the difference between instantaneous frequency values can be thought of as the acceleration of the phases of that frame.

The instantaneous frequency is calculated for each channel from the phase difference between frames n and $(n-1)$ by,

$$\Delta\theta_n(k) = \theta_n(k) - \theta_{(n-1)}(k) \qquad (4)$$

where $\theta(k)$ are the principle values of the phase and are bounded by π and $-\pi$. The phase acceleration, $\Delta\Delta\theta_n(k)$, is the difference between the instantaneous frequency and the prior instantaneous frequency,

$$\Delta\Delta\theta_n(k) = \Delta\theta_n(k) - \Delta\theta_{(n-1)}(k). \qquad (5)$$

The above two equations are combined to show the phase acceleration in terms of phase:

$$\Delta\Delta\theta_n(k) = \left(\theta_n(k) - \theta_{(n-1)}(k)\right) - \left(\theta_{(n-1)}(k) - \theta_{(n-2)}(k)\right), \qquad (6)$$

so that,

$$\Delta\Delta\theta_n(k) = \theta_n(k) - 2\theta_{(n-1)}(k) + \theta_{(n-2)}(k). \qquad (7)$$

We can normalize this phase acceleration to our nominal sinusoidality range of zero to one to create our final sinusoidality coefficients:

$$\Gamma_{pa}(n,k) = \left(1 - \left(\frac{\Delta\Delta\theta_n(k)}{2\pi}\right)\right)^p, \qquad (8)$$

where $\Gamma_{pa}(n,k)$ is the resultant sinusoidality spectrum for frame n. This spectrum shows values near 1 for stable channels and values tending toward 0 for unstable channels. This sinusoidality measure tends to produce sinusoidality coefficients near one for both stable and non-stable components. The variable p in the above equation is used to counteract this tendency. By raising each coefficient to a small integer power of itself, typically $p = 4$, this overall tendency is significantly abated.

Duxbury et al. proposed a frequency dependent threshold to improve the Settel and Lippe method (Duxbury et al. 2001). They note that for any given value of Γ_{pa}, low frequency channels tend to be selected as stable and high frequency channels selected as non-stable. They alleviate this tendency by choosing a different threshold value for each octave sub-band created with six constant-Q filters.

Here we propose a new method of reducing frequency dependency in terms of our sinusoidality spectrum $\Gamma_{pa}(k)$. In order to reduce the sinusoidality measure $\Gamma_{pa}(k)$ for high k, we will change the slope of the sinusoidality spectrum by multiplying each $\Gamma_{pa}(k)$ by a scaled version of the channel number k. Our new stability spectrum is:

$$\Gamma_{pa}(k) = \left(1 - \left(\frac{\Delta\Delta\theta_n(k)}{2\pi}\right)\right)^p (M)\left(\frac{k}{k_{max}}\right), \qquad (9)$$

where M is a slope constant with a heuristically determined value of approximately -1.024. This correction to the stability spectrum avoids the constant-Q filter calculation and multiple threshold calculations of the Duxbury et al. method.

2.3. Harmonic Sum Spectrum Sinusoidality

The measure of the overall "harmonicity" of a STFT frame is a frequently calculated feature, that typically requires the explicit calculation of a single fundamental frequency (f_0) by peak picking or other parametric techniques (Sarlo 2004). As our project here is based on the traditional PV parameters of amplitude, phase and/or instantaneous frequency, those methods that require higher order analysis such as spectral peak picking or fundamental frequency analysis are precluded. With these restrictions, we have created a new non-parametric sinusoidality analysis technique for STFT frames that relies on the harmonic relationships present in the sound, but without requiring explicit computation of these frequencies.

Musical instrument sounds typically have a predominantly harmonic structure. The spectral components of a musical tone that are in a harmonic relationship are the predominantly sinusoidally based energy of a spectral frame. Conversely, we can consider spectral energy that is not in a harmonic relationship to be noise based. This is the basis of the sinusoidality coefficient measure presented here.

The harmonic product spectrum (HPS) technique is a method of finding the fundamental frequency energy of a STFT by combining the energy of the harmonics with each other at the location of the fundamental frequency (Schroeder 1968). The harmonic product spectrum is the product of the power spectrum multiplied by successively compressed versions of the same spectrum. These spectra are down-sampled along the frequency axis by consecutive integer amounts. The resultant product reinforces the amplitude of the fundamental frequency. A related spectrum, the harmonic sum spectrum (HSS), with similar output, sums these down-sampled spectra instead of multiplying them (Alonso, David, and Richard 2003). The harmonic sum spectrum is given by:

$$HSS(k) = \sum_{r=1}^{K} |X(k_r)|^2 \quad \text{where} \quad k_r = (r)(k). \qquad (10)$$

Here K is the number of down-sampled spectra typically around six.

Conceptually, our proposed method is the reverse of the HSS. To the normalized power spectra we add each integer multiple up-sampled copy of the spectra. These up-sampled spectra are calculated by interpolating new spectra at integer multiples of the original spectra. That is:

$$\Gamma_h(k) = \sum_{r=1}^{K} |X(k_r)|^r \quad \text{where} \quad k_r = k/r. \qquad (11)$$

Spectral peaks in each of these up-sampled spectra are proportionately wider than those in the original power spectra. For example, the spectral peaks in the $r = 2$ spectra are twice as wide as the original. In order to slim these peaks back to an approximation of their original width, the up-sampled spectra are raised to increasing powers and normalized to their highest peak. In order to heuristically calculate this power term, a Hann window is interpolated to twice its length and raised to integer powers. These are normalized to the original height of the Hanning window. It was found that the window raised to the 4th power was slightly wider than the original window and was slightly narrower when raised to the 5th power. So, the $r = 2$ spectra would be taken to the 4th power in order to reduce the width of its peaks to approximate the original width. However, this 4th power was found to produce banks that are too narrow to use in practice. The first up-sampled spectrum is set to the 2nd power in our algorithm, and each subsequent up-sampled length is raised to the next power as can be seen in equation 11. This produces approximately equal width peaks for six to twelve up-sampled spectra K. Finally, the slope correction devised for the phase acceleration sinusoidality method is used to reduce the sinusoidality of higher harmonics. Figure 2 shows a normalized power spectrum for a violin tone and the corresponding harmonic sum sinusoidality for this spectra. The harmonics higher than those of the power spectra are visible in the sinusoidality spectrum.

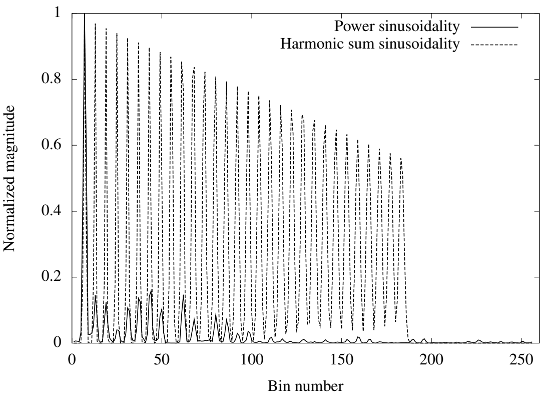

Figure 2. Power spectrum sinusoidality and harmonic sum sinusoidality for a a single frame of a violin tone.

2.4. Other Sinusoidality Analysis Methods

In addition to these three methods, many other sinusoidality analysis methods have been developed and compared in a prior study (Apel 2008). In addition, several other methods have been proposed and not adapted to the PV time-stretching framework discussed here (Zivanovic et al. 2007; Dubnov 1999).

The three sinusoidality measures presented here have been shown to be effective in separating sinusoidal from noise energy for differing synthetic test signals (Apel 2008). The Charpentier method performs well with a single sine tone in noise. The phase acceleration measures performs well for inharmonic tones in noise. The harmonic sum measure performs better than the others for a harmonic tone. They are each suited to analyzing different types of sounds. Even when the sounds are restricted to monophonic tones, no one sinusoidality measure presents itself as ideal. In the following section, methods of adding noise to a PV representation will be presented that rely on the sinusoidality coefficients for each PV frame in order to determine the amount of noise to be added to each channel. We employ the three different measures depending on the type of sound being analyzed.

3. SINUSOID AND NOISE PHASE VOCODER SYNTHESIS

A method of synthesizing time-stretched sounds from phase vocoder analyses and sinusoidality coefficients for each spectral frame is presented in this section. In this method, the percentage of energy of each spectral bin that is noise-based is multiplied by a spectral domain noise signal. Before describing our sinusoid and noise synthesis algorithm, two existing alternative methods will be considered.

A method of segregating spectral energy separates the spectral channels into two groups, one the "sinusoidal" channels, and the other the "noisal" channels. For each frame, the individual bins are determined as belonging to one of two groups. This is the scheme proposed by Lippe and Settle for segregating bins (Settel and Lippe 1994; Magnus 2001).

A threshold is set for each bin over which the bin is labeled as sinusoidal. Conversely, if the sinusoidal measure is below the threshold, the channel is labeled as noise based energy. Thus,

$$X_s(k) = (\Gamma_n(k) \geq T), \tag{12}$$

where $\Gamma_n(k)$ is a frame of sinusoidality measures between 0 and 1 for each channel, T is a stability threshold between 0 and 1, and $X_s(k)$ is a phase vocoder spectral frame with only the stable channels above the threshold not set to zero. Noise is then added to the spectra by a method shown in section 3.1. This method is less suited than the method proposed below because of the sharp cutoffs between adjacent bins created from the thresholding of energy.

A second method of adding noise to a PV spectrum involves noise modulation of each spectral bin to a degree determined by the sinusoidality of each bin. This method provides a single or unified spectral representation, as opposed to the dual representations in which the sinusoidal spectra and noise spectra are separated for processing and recombined during synthesis. While the technique is well suited for creating a unified representation of noise in sinusoidal modeling synthesis, it is ill sited for PV based systems in which sinusoidal energy is necessarily spread across several bins of the spectrum [1]. For example, a single sinusoid in a PV spectra will occupy at least three consecutive bins. Increasing the noise "bandwidth" of each of these bins would put energy in adjoining channels. A system of partitioning the STFT spectral bins, as proposed by Dolson and Laroach (Laroche and Dolson 1999) could perhaps be of use here, but we preclude it here based on its increased parametric nature.

A system based on this second method has been developed by Liao, Roebel, and Su in order to time-stretch gaussian noise as a test case in the creation of a phase vocoder based sound texture time-stretching method (Liao, Roebel, and Su 2012). Employing a temporal correlation function to maintain the statistical properties of the STFT phase spectrum, their method successfully time-stretches purely noise signals suggesting that it may become suitable for time-stretching complex sounds with noise and pitched components.

3.1. Dual Model Synthesis

Our method of segregating sinusoidal energy from noise based energy in a spectral representation divides the energy of each spectral bin into two parts corresponding to the sinusoidal energy and the noise-based energy. These separated spectra are processed separately and recombined during synthesis. Figure 3 shows our complete analysis synthesis system for time-stretching noise and pitched sounds. This process is discussed below.

The process starts with the STFT analysis data as a series of amplitude and phase spectra frames along with corresponding sinusoidality coefficients for each spectral bin. For each frame n, a new amplitude spectrum is created by using a sinusoidality coefficient spectra $\Gamma_n(k)$ to scale the amount of sinusoidal energy present in each bin. For each bin k of each spectral frame n,

$$|S_n(k)| = (\Gamma_n(k))\,(|X_n(k)|), \tag{13}$$

where $|S_n(k)|$ is a new magnitude spectrum, here called "sinusoidal magnitude spectrum", in which each bin's amplitude is scaled by the corresponding sinusoidality of that bin.

Next, the corresponding "noise magnitude spectrum," $|N_n(k)|$, is calculated by subtracting the sinusoidal magnitude spectrum from the unaltered magnitude spectrum,

$$|N_n(k)| = |X_n(k)| - |S_n(k)|. \tag{14}$$

As can be seen here, the original magnitude spectrum can be recreated by adding the noise magnitude spectrum and the sinusoidal magnitude spectrum.

[1] This distinction is analogous to the distinction between how noise and sinusoids are modeled in a dual manner in SMS modeling (Serra and Smith 1990) and how they are represented in a unified representation in the Fitz "bandwidth enhanced" method (Fitz and Haken 1995). In Fitz's system the individual sinusoids of a sinusoidal model of sound are modulated with noise to increase their bandwidth. This process increases the noise level of each sinusoidal component.

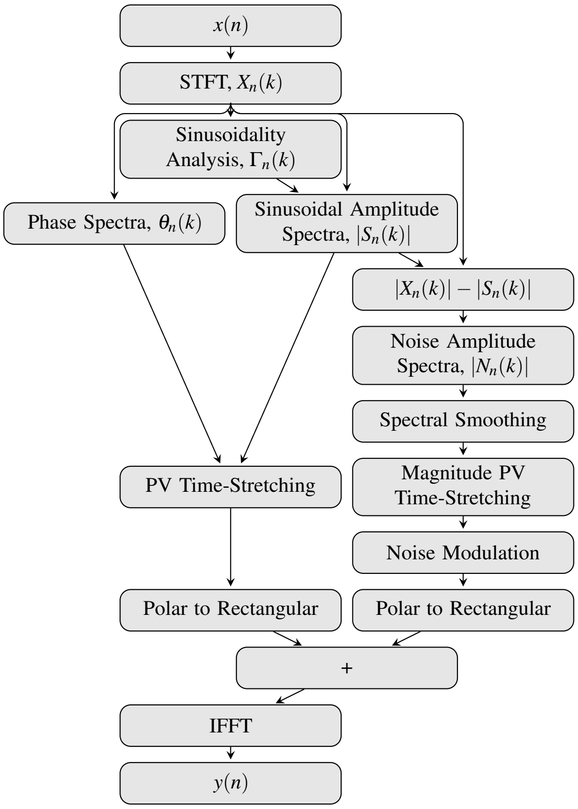
$x(n)$
STFT, $X_n(k)$
Sinusoidality Analysis, $\Gamma_n(k)$
Phase Spectra, $\theta_n(k)$
Sinusoidal Amplitude Spectra, $|S_n(k)|$
$|X_n(k)| - |S_n(k)|$
Noise Amplitude Spectra, $|N_n(k)|$
Spectral Smoothing
Magnitude PV Time-Stretching
PV Time-Stretching
Noise Modulation
Polar to Rectangular
Polar to Rectangular
+
IFFT
$y(n)$

Figure 3. Sinusoidality analysis and noise synthesis system.

The noise magnitude spectrum is subject to a large variance in bin amplitude values between spectral frames. While this behavior accurately reflects the character of noise energy, it produces unwanted sonic artifacts when these random fluctuations are subject to PV time-stretching. For this reason the noise magnitude spectra are filtered in order to smooth each channel in time and frequency. The time frame smoothing is achieved by recursively averaging past frames of $|N_n(k)|$. Martin suggests a spectral noise smoothing filter for use on ambient stationary noise signals (Martin 1994),

$$|NS_n(k)| = \alpha|NS_{(n-1)}(k)| + (1-\alpha)|N_n(k)|, \tag{15}$$

where $|NS_n(k)|$ is the smoothed noise magnitude spectrum, and the smoothing constant α is typically set, according to Martin, between 0.9 and 0.95. As the dynamic character of musical sounds is typically present in the noisy aspects of sound, we use a smoothing constant of 0.4, which is much smaller than the Martin suggestion.

In addition to this temporal smoothing, we smooth the noise magnitude spectrum across bins. This is simply achieved by using a running average of β bins, where $8 < \beta > 12$,

$$|NS_n(k)| = \frac{\sum_{l=-\beta/2}^{\beta/2} |N_n(k+l)|}{\beta}. \tag{16}$$

Both of these techniques are used to smooth our noise magnitude spectra.

Our smoothed noise magnitude spectra $|NS_n(k)|$ are time-stretched separately from the sinusoidal magnitude spectra above. Since they are only amplitude spectra without corresponding phase or instantaneous frequency spectra, the amplitude spectra are simply interpolated between spectral frames. These new noise magnitude spectra are then each multiplied by a different STFT analysis of white noise,

$$N_n(k) = |NS_n(k)|P_n(k), \tag{17}$$

where $P_n(k)$ is a new STFT of white noise for each n of the time stretched sound. Each new noise spectral frame is added to its corresponding sinusoidal spectral frame,

$$Y_n(k) = S_n(k) + N_n(k), \tag{18}$$

where $S_n(k)$ is the complex sinusoidal spectral frame produced by a traditional phase vocoder time stretch of the sinusoidal magnitude spectrum $|S_n(k)|$ with the original phase and derived instantaneous frequency values. $Y_n(k)$ is the resultant spectral frame that is inverted to the time domain by the IFFT and appropriate windowing. The final new sound $y(n)$ is shown at the bottom of figure 3.

Several differing sounds both musical and environmental were time-stretched with the new sinusoidality analysis and noise synthesis method. In each case, the original unaltered sound is followed by a traditional phase vocoder time-stretching with 8 times the normal length. Then, the new time stretched sound with the noise characteristics preserved is listed. Sample sounds can be found on the website:

`http://vud.org/fppv/`

4. CONCLUSION

In each case, it can be heard that the generation of noise as part of the phase vocoder time-stretching creates a sound that more closely resembles the noise characteristics of the unaltered sound as compared to the traditional phase vocoder time-stretching method. On occasion, the two parts of the time-stretched sound, sinusoidal and noise based, do not fuse perceptually as they do in the original sound. This is no doubt attributable to the dual nature of our synthesis system. Future work could consist of devising methods of combining the two modes of synthesis into a unified representation. Improvements to the sinusoidality algorithms could consist of developing a single optimized measure for most sound types or automating the selection of sinusoidality measures based on analyzing the source sound.

References

Alonso, M., B. David, and G. Richard. 2003. “A Study of Tempo Tracking Algorithms from Polyphonic Music Signals.” *4th COST 276 Workshop* .

Apel, T. 2008. “Feature Preservation and Negated Music in a Phase Vocoder Sound Representation.” Ph.D. thesis, University of California, San Diego.

Arfib, D., F. Keiler, and U. Zölzer, (editors) . 2002. *DAFX - Digital Audio Effects*. John Wiley and Sons, LTD.

Charpentier, F. 1986. "Pitch detection using the short-term phase spectrum." *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP* 11:113– 116.

Dressler, K. 2006. "Sinusoidal extraction using an efficient implementation of a multi-resolution FFT." *Proceedings of 9th International Conference on Digital Audio Effects (DAFx-06)* :247–252.

Dubnov, S. 1999. "HOS Method for Phase Characterization in Sinusoidal Models with Applications for Speech and Audio." *IEEE Signal Processing Workshop on Higher-Order Statistics* :1–5.

Duxbury, C., M. Davies, and M. Sandler. 2001. "Separation of Transient Information in Musical Audio using Multiresolution Analysis Techniques." *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01)* .

Fitz, K., and L. Haken. 1995. "Bandwidth Enhanced Sinusoidal Modeling in Lemur." *Proceedings International Computer Music Conference* .

Laroche, J., and M. Dolson. 1999. "New phase-vocoder techniques for real-time pitch shifting, chorusing, harmonizing, and other exotic audio modifications." *Journal of the Audio Engineering Society* 47(11):928–936.

Liao, W.-H., A. Roebel, and A. W. Su. 2012. "On Stretching Gaussian Noises with the Phase Vocoder." *Proceedings of the 15th Conference on Digital Audio Effects (DAFx-12)* .

Magnus, C. 2001. "Real-Time Separation of Periodic and Non-periodic Signal Components." Unpublished paper.

Martin, R. 1994. "Spectral Subtraction Based on Minimum Statistics." *Proceedings of EUSIPCO-94 Seventh European Signal Processing Conference* .

Moorer, J. A. 1978. "Use of Phase Vocoder in Computer Music Applications." *Journal of the Audio Engineering Society* 26(1-2):42–45.

Peeters, G., and X. Rodet. 1998. "Signal Characterization in terms of Sinusoidal and Non-Sinusoidal Components." *Proceedings DAFX98* .

Sarlo, J. 2004. "Real-time Pitched/Unpitched Separation of Monophonic Timbre Components." *Proceedings International Computer Music Conference* .

Schroeder, M. 1968. "Period Histogram and Product Spectrum: New Methods for Fundamental-Frequency Measurement." *The Journal of the Acoustical Society of America* 43(4):829–834.

Serra, X., and J. Smith. 1990. "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic Plus Stochastic Decomposition." *Computer Music Journal* 14:12–24.

Settel, J., and C. Lippe. 1994. "Real-time Musical Applications using the FFT-based Resynthesis." *Proceedings International Computer Music Conference* .

Zivanovic, M., A. Roebel, and X. Rodet. 2007. "Adaptive Threshold Determination for Spectral Peak Classification." *Proceedings of Ninth International Conference on Digital Audio Effects* .