SIGNAL SYNTHESIS FROM THE MODAL DISTRIBUTION USING MINIMUM PHASE

Thomas Lysaght*, Joseph Timoney*, and Victor Lazzarini[+]

*Dept. Of Computer Science, [+]Dept. of Music, NUI, Maynooth, Co. Kildare, Ireland.

tlysaght@cs.nuim.ie; jtimoney@cs.nuim.ie; victor.lazzarini@nuim.ie

Keywords: time-frequency resynthesis, minimum phase.

Abstract

The Modal Distribution (MD) is a time-frequency distribution specifically designed to model the quasi-harmonic, multi-sinusoidal, nature of music signals and belongs to the Cohen general class of time-frequency distributions. Signal synthesis from bilinear time-frequency representations such as the Wigner distribution has been based on methods which exploit an outer-product interpretation of these distributions [1, 2]. Methods of synthesis from the MD based on a sinusoidal-analysis-synthesis procedure using estimates of instantaneous frequency and amplitude only have been investigated in [3, 4, 5]. However, the modal distribution is basically a subsampled version of the smoothed pseudo Wigner distribution and thus does not lend itself easily to direct inversion such as in the outer product methods mentioned above. Furthermore, the modal distribution is real, and the above sinusoidal-analysis-synthesis methods rely on phase estimated as the integral of instantaneous frequency. In this paper, we show that in some cases, this synthesis results in a roughness or phasiness in the synthesized signal and demonstrate that using minimum phase derived from the magnitude spectrum of the distribution produces a timbre closer to the original in the case of certain brass sounds. Suggestions for future work are also given.

1 Introduction

The MD was introduced by Pielemeier and Wakefield [3] as a member of the Cohen general class of time-frequency distribu-tions for the analysis of music signals. It is primarily a Wigner distribution, or more specifically, a subsampled smoothed pseudo-Wigner distribution (SPWD), with a kernel that takes account of the modes present in quasi-harmonic, multi-sinusoidal, music signals. Based on the Wigner distribution, it allows for accurate measurement of instantaneous amplitude and frequency estimates calculated as local averages in the neighbourhood of each partial's bandwidth. Furthermore, it does not suffer from the time-bandwidth trade-off inherent in the spectrogram, one of the key advantages attributed to the Wigner distribution.

One drawback of the Wigner distribution in relation to the analysis of quasi-harmonic signals is the existence of both inner and outer cross terms [6] that amount to beats between partials (outer cross terms) which do not exist in the original signal, and artifacts due to non-linear frequency modulations (inner cross terms). To counteract this drawback, the SPWD and MD utilize both a one-dimensional frequency-smoothing kernel and one-dimensional time-smoothing kernel. The frequency smoothing kernel determines the suppression of artifacts along the frequency axis while the time-smoothing kernel reduces the effect of the outer cross terms. This time smoothing also reduces the bandwidth of the distribution in the time direction and so facilitates the subsampling. This greatly reduces the number of output frames and the number of DFTs that need to be computed. This is a key advantage of the MD over the Wigner distribution.

Based on this innovation, the MD has been utilized as an analysis tool for estimating the detailed amplitude and frequency variations of musical instrument sounds. It has been used to identify phenomenon such as the frequency modulation of attack transients or 'rogue' piano partials [5], whereas under normal spectrogram smoothing such detailed characteristics would be obfuscated. In previous work [7], we have proposed a novel frequency smoothing kernel which provides better noise suppression in the MD while conserving the accuracy of parameter estimates in the distribution. A key issue is now that once the analysis is completed it is often desired to resynthesise the signal displayed in the MD, particularly if post-processing of the MD is enacted to alter the signal's properties. However, this procedure is not straightforward as the MD is not directly invertible so signal approximation approaches must be used. It is the perceptual quality of the output that establishes the differences between these. This goal of this paper is to examine an enhancement ap-plied to one of these approaches, specifically one that involves a minimum phase assumption for the phase trajectories of the signal's harmonics. This paper is organized as follows. Section 2 gives the theoretical background to the

MD. Part 3 describes the proposed method of using a minimum phase approach for resynthesis from the MD. Part 4 gives test results, which then lead to conclusions and suggestions for future work in Section 5.

2 Theoretical background

Leon Cohen [10] proposed a general class of time-frequency distributions which are related through linear transformations. The set of all linear transformations of the Wigner distribution has come to be known as the Cohen general class. A two-dimensional kernel determines the linear transformation involved. The Wigner distribution, Eq. 1, in terms of the signal $s(t)$ and the spectrum $S(\omega)$ is given by:

$$W(t,\omega)=\frac{1}{2\pi}\int s^*\left(t-\frac{1}{2}\tau\right)s\left(t+\frac{1}{2}\tau\right)e^{-j\tau\omega}d\tau \qquad (1)$$
$$=\frac{1}{2\pi}\int S^*\left(t-\frac{1}{2}\theta\right)S\left(t+\frac{1}{2}\theta\right)e^{jt\theta}d\theta$$

Here the kernel is 1. The autocorrelation with the lag variable, τ, produces the time-relative-time or instantaneous temporal autocorrelation function:

$$R_s(t,\tau)=s^*\left(t-\frac{1}{2}\tau\right)s\left(t+\frac{1}{2}\tau\right) \qquad (2)$$

An important property of the Wigner distribution is that it is real with $W^*(t,\omega)=W(t,\omega)$.

2.1 The discrete pseudo-Wigner Distribution

The discrete implementation of the pseudo-Wigner distribution with a frequency smoothing kernel $w(k)$, with length $M=2L-1$, $w(k)=0$ for $|k|\geq L$ is then defined as:

$$PWD\left(n,\frac{m\pi}{M}\right)=2\sum_{n=-L+1}^{L-1}g(n,k)p(k)e^{-2jk\frac{mn}{N}}, \qquad (3)$$
$$m=0,\ldots,M$$

where the discrete instantaneous autocorrelation function is:

$$g(n,k)=f(n+k)f^*(n-k) \qquad (4)$$

and the 'pseudo' window is given by:

$$p(k)=w(k)w^*(-k) \qquad (5)$$

Eq. 3 can be interpreted as the discrete Fourier transform of the autocorrelation function $g(n,k)$ with respect to *n* for each value of *k*. Note that the frequency smoothing kernel in Eq. 5 is squared in order to maintain the quadratic nature of the distribution defined in Eq. 4. As autocorrelation samples are only specified at each discrete integer point *k* in Eq. 4, compared with the continuous lag variable $\tau/2$ in Eq 2. The discrete version requires the input signal to be either oversampled by 2, or band-limited to half the Nyquist rate in order to avoid aliasing [5]. This is significant for the analysis of music signals where partials may exist beyond $f_s/4$ unless band-limiting is applied prior to the analysis.

2.2 Cross terms

Given a music signal model as follows:

$$s(t)=\sum_{k=1}^{M}A_k e^{j(w_k t+\phi_k)} \qquad (6)$$

with each partial indexed by k, specified uniquely by partial amplitude A_k, frequency ω_k, and phase ϕ_k, the Wigner distribution can be expanded to:

$$W_s(t,\omega)=\int_{-\infty}^{\infty}R_s(t,\tau)e^{-j\omega\tau}d\tau$$
$$=\sum_{k=1}^{M}A_k^2\delta(\omega-\omega_k)+ \qquad (7)$$
$$\sum_{k=1}^{M}\sum_{l=1}^{M}A_k A_l\cos([\omega_k-\omega_l]t+\phi_k-\phi_l)$$
$$\times\delta\left(\omega-\frac{(\omega_k+\omega_l)}{2}\right)$$

The *auto* terms of $s(t)$ are given by the first term in Eq. 7. The second double summation indicates the cross terms, arising from products between auto terms, which lie between any pair of auto terms. The magnitude of the cross terms is the product $A_k A_l$ of the amplitudes of auto terms k and l and they oscillate at a frequency, $(\omega_k-\omega_l)$ equal to the difference between the frequencies of the two auto terms. For strictly harmonic signals, the cross terms form a partial series an octave below the fundamental, with the consequence that some cross terms fall at the same frequency location as the auto terms. This phenomenon gives rise to amplitude modulated partials and the possibility of additional artefacts and cross terms appearing at the partial frequencies.

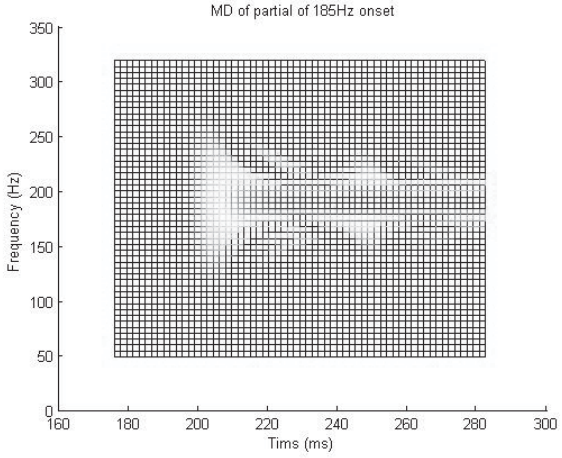

Figure 1: MD of attack of a synthetic monocomponent signal at 185Hz. The wideband onset and widening of the auto terms mainlobe at points of amplitude discontinuity around 240msecs is clearly evident.

2.3 The Modal Distribution (MD)

The MD was designed to minimise these cross terms in Eq. 7. The MD kernel consists of two different filter functions. The time-smoothing window, $h_{LP}(p)$, has the effect of smoothing the cross terms in the time direction, and the frequency-smoothing window, $g_{LP}(l)$, implements cross term

suppression in cases of frequency modulation as well as defining the frequency resolution of the distribution. The discrete form of the MD is defined by

$$M(n,k) = \sum_{n=-L+1}^{L-1} R_{s,l}(n,l) g_{LP}(l) e^{\frac{-j2\pi kl}{2L}} \quad (8)$$

where $R_{s,l}(n,l) = R_s(n-p,l) h_{LP}(p)$ is the time-smoothed temporal autocorrelation function. Both $h_{LP}(p)$ and $g_{LP}(l)$ form a separable kernel, however, they are interdependent for parameter choice [7]. $h_{LP}(p)$, is chosen to be a low pass filter with an upper cut-off just below the minimum frequency spacing between auto terms, $\Delta f_{\min}$, this being set to slightly less than the fundamental frequency for quasi-harmonic signals. This allows for any modulation in the input signal which would narrow the minimum separation between auto terms. The frequency resolution of $\Delta f_{\min}$ is, in turn, defined by the length of $g_{LP}(l)$, chosen so as to avoid overlapping auto term main lobes. The main lobe width of $g_{LP}(l)$ determines the estimation limits for these auto terms. However, the width of the auto terms characteristically exhibits variations due to large amplitude changes or discontinuities. The result is the creation of broadband artefacts in the distribution. Fig 1 shows the attack of a monocomponent synthetic signal illustrating the signature wideband MD onset and the widening of the auto term near the peak of attack around 240msecs where the amplitude is discontinuous. Typically, the wideband onset lasts for the duration of the impulse response of $h_{LP}(p)$ called the 'end-effect' region in [5] where the estimates have been shown to be extremely biased [8]. Inner interference cross terms are also visible between the broadband artefacts along the contour of the auto term main lobe.

2.4 MD Synthesis Parameters

Signal synthesis parameters of amplitude and frequency are calculated as local averages of the MD that are centred around the local instantaneous frequency of the auto terms or partials. The bandwidth for these moments is determined by the main lobe width of $g_{LP}(l)$. These local moments can be written as follows. Given:

$$p(n) = \sum_{l=-L}^{L} M(n,l) \quad (9)$$

where $p(n)$ is the instantaneous power estimate given $g_{LP}(l)$ with the main lobe width being $2L$, the amplitude estimate is then given by:

$$A(n) = 4\sqrt{p(n)} \quad (10)$$

and the instantaneous frequency by:

$$F(n) = \sum_{l=-L}^{L} \frac{lM(n,l)}{p(n)} \quad (11)$$

3 Signal synthesis from Modal analysis

Methods of analysis and synthesis from time-frequency distributions are well documented [12, 13, 14, 15] and have been com-pared in [16]. In the case of the MD, partials can be interpreted as time-varying salient ridges in the MD surface from which the synthesis parameters of instantaneous amplitude and frequency are estimated. We use the well established McCauley-Quatieri procedure [9] for peak identification and track formation from the MD surface. Interpolation of amplitude and frequency estimates is replaced by the MD parameter estimates defined in Eqs. 10 & 11, and calculated around the bandwidth of each candidate peak this bandwidth being determined by the main lobe width of the frequency smoothing kernel $g_{LP}(l)$.. However, because the MD is real, no phase information is available from the MD parameters. As already explained, a difficulty with the MD is the existence of noise terms. To facilitate partial extraction from the MD, we employ a novel frequency smoothing window described in [7], the squared autocorrelated frequency smoothing kernel. This significantly reduces the number of possible candidate artifacts/peaks in partial tracking. One approach to phase estimation is to calculate phase as the integral of the instantaneous frequency values over time. Subsequent synthesis using this cumulative phase produces, for certain instrumental sounds, a rough and 'phasey' characteristic timbre while other instrumental sounds appear much less affected. We therefore explore a method to re-cover coherence in the phase as explained in the following section.

3.1 Minimum-Phase from Magnitude

It is well known that phase information can be extracted from the magnitude of the Fourier transform [17]. Creating a minimum-phase desired frequency-response from a given magnitude response can be achieved by reflecting all the zeros of zeros of z_i for which $|z_i| > 1$, back inside the unit circle, i.e., replacing z_i by $1/z_i$ [10]. In practice this is achieved by implementing a simple matlab function given in [10]. Given a magnitude spectrum consisting of partial magnitudes, *s*, the minimum phase spectrum, *sm*, is given by:

```
sm=exp(fft(fold(ifft(log(clipdb(s,-100))))));
```

where `fold` converts non-minimum-phase spectral zeros to minimum-phase spectral zeros. This function works well as long as the desired frequency re-sponse is smooth which is partially guaranteed by clipping mag-nitude response below its maximum (`clipdb`) [10]. The minimum phase is extracted as, `angle(sm)`, and then the phase shift for each partial is calculated as the difference between the original fundamental phase and the corresponding first value of `angle(sm)`. The resulting phases are then used in the sum-of-sinusoids synthesis.

4. Results

We used instrumental samples from the McGill University Master Samples, sampled at 44.1 kHz to evaluate the various synthesis approaches. From informal listening tests, we found that for a number of instrumental sounds (trumpet, trombone

and saxo-phone) exhibit a roughness in timbre when synthesized from the Modal distribution using the McCauley-Quatieri sum of sinusoids approach based on the MD estimates with the instantaneous phase simply calculated as the integral of the instantaneous fre-quency, i.e. a cumulative phase approach. In contrast for the clarinet the results were somewhat better with this cumulative phase approach.

Applying the minimum phase approach was found to produce more perceptually pleasing results for the three brass instruments, overcoming some of the roughness of the previous examples. The results are first illustrated by the waveform plots, shown in figures 2 to 5, and then by the phase spectrograms shown in figures 6 to 9 for each instrument sound respectively. The spectrograms were computed using a 512-point Hanning window with a 50% overlap and then zero-padded to 2048 points before the FFT is taken and its phase found at each frame. Examining the wave-form plots first: In figures 2 to 5 the top panel is the original waveform, the middle panel is the waveform using the cumulative phase method while the lower panel is for the minimum phase approach. It can be seen that the minimum phase approach produces a waveform that is a closer visual match to the original waveform than the cumulative phase method. Only for the clarinet sound does the cumulative phase method seem to give a waveform that more resembles the original. The appearance of the phase spectrograms in figures 6 to 9 support these observations as for the cases of the trumpet, trombone and saxophone that phase spectrogram for the minimum phase approach is more close to the phase spectrogram of the original signal. Distinctive patterns can be seen in those spectrograms that match each other well. Only in the case of the clarinet is the cumulative phase spectrogram more like that of the original. These plots thus tend to support the aural observations.

5. Conclusions

In this paper we have presented a minimum phase solution for the recovery of phase in Modal Distribution synthesis. We have described how certain sounds exhibit roughness when synthesized with a McCauley-Quatieri sum-of-sinusoids synthesis technique using a cumulative phase method. Use of phase calculated from a minimum phase approach instead appears to recover the original timbre of brass sounds more faithfully. In contrast, for the clarinet the distinction is less obvious. These preliminary results suggest that the assumption of minimum phase can be applied successfully to certain instrumental signals when resynthesizing from the Modal Distribution.

Of immediate interest for future work is an extension of this method to include a mixture of both minimum and maximum phase elements as described in [11] that may give better and more consistent results across a broader range of musical instruments. Furthermore, a more comprehensive suite of perceptual tests would help to set the boundaries on where the minimum phase works best as opposed to any alternative technique.

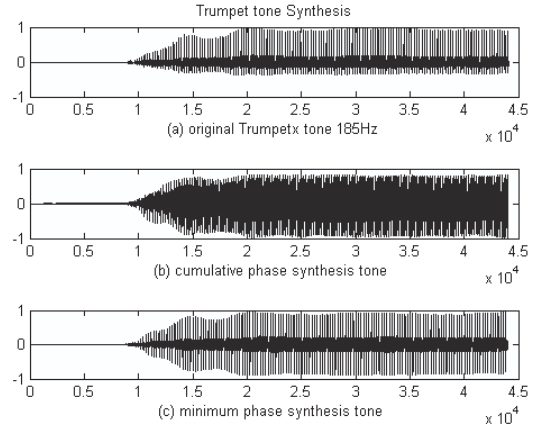

Figure 2: Plots for trumpet tone 185Hz for (a) original (b) syntheized using the cumulative phase (the integral of the instantaneous frequency) and (c) synthesized using minimum phase.

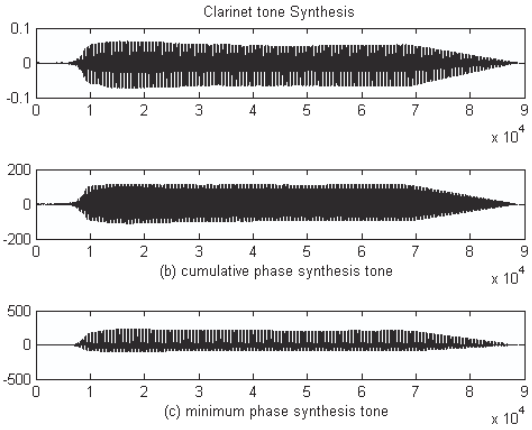

Figure 3: Plots for clarinet tone 146Hz for (a) original (b) syntheized using cumulative phase and (c) synthesized using minimum phase.

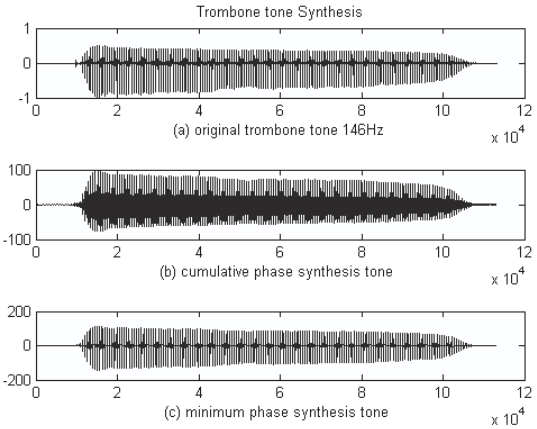

Figure 4: Plots for Trombone tone 146Hz for (a) original (b) syntheized using cumulative phase and (c) synthesized using minimum phase.

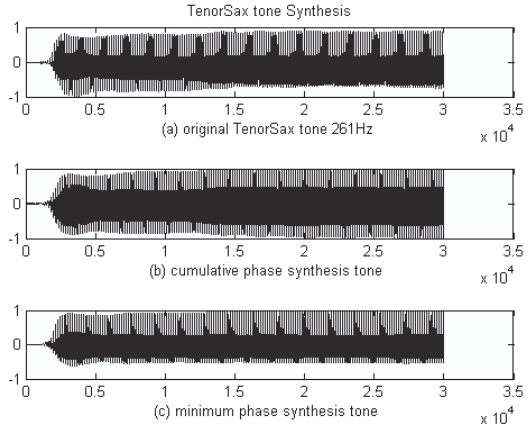

Figure 5: Plots for TenorSax tone 261Hz for (a) original (b) syntheized using cumulative phase and (c) synthesized using minimum phase.

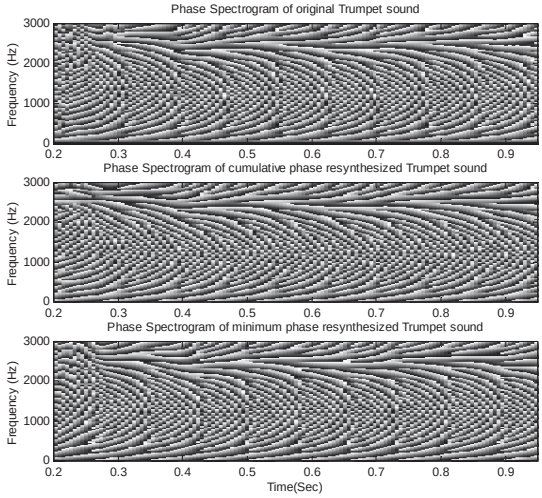

Figure 6: Phase Spectrogram for Trumpet tone for (a) original (b) syntheized using cumulative phase and (c) synthesized using minimum phase.

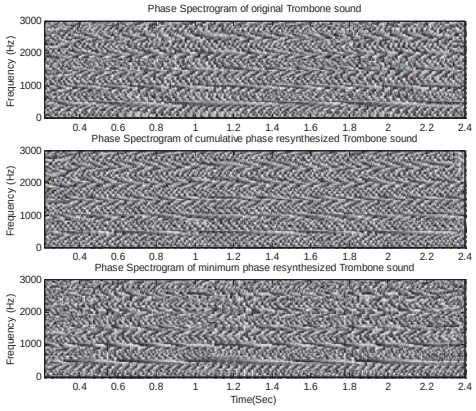

Figure 7: Phase Plots for Trombone tone for (a) original (b) syntheized using cumulative phase and (c) synthesized using minimum phase.

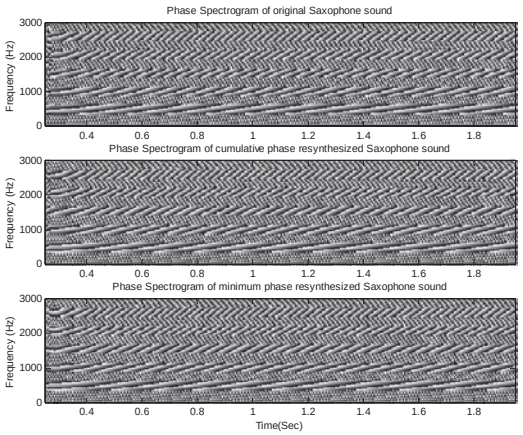

Figure 8: Phase Plots for TenorSax tone for (a) original (b) syntheized using cumulative phase and (c) synthesized using minimum phase.

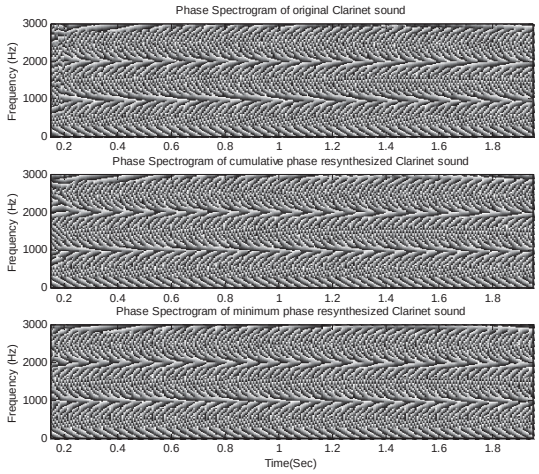

Figure 9: Phase Plots for Clarinet tone for (a) original (b) syntheized using cumulative phase and (c) synthesized using minimum phase

Acknowledgements

We wish to acknowledge the support of the SFI International Strategic Collaboration Award-China .

References

[1] K. B. Yu, and S. Cheng, “Signal synthesis from pseudo Wigner distribution and applications”, *IEEE Trans. Acoust., Speech, Signal Processing*, **Vol. ASSP-35**, pp.1289-1302, Sept. 1987.

[2] G.F. Boudeaux-Bartels and T.W. Parks, “Time-varying filtering and signal estimation using Wigner distribution synthesis techniques,” *IEEE Trans. Trans. Acoust., Speech, Signal Processing*, **Vol. ASSP-34**, pp. 442-451, June 1986.

[3] W. J. Pielemeier, and G. Wakefield, "A high-resolution time-frequency representation for musical instrument signals”, *Journal of Acoustical Society of America,* **99(4)**, pp. 2382-2396, April 1996.

[4] M. Mellody and G. H. Wakefield, "The time-frequency characteristics of violin vibrato: Modal distribution analysis and

synthesis", *Journal of Acoustical Society of America*. **107(1)**, pp. 598-611, 2000.

[5] R. Guevara, "Modal distribution analysis and sum of sinusoids synthesis of piano tones", PhD. Dissertation, University of Michigan, 1997.

[6] C. Kereliuk and P. Depalle, “Improved hidden Markov model partial tracking through time-frequency analysis”. *Proceedings of the Digital Audio Effects (DAFx-08)*, September 1-4, 2008, Espoo, Finland.

[7] T. Lysaght, J. Timoney, and V. Lazzarini, “Automatic Partial Extraction from the Modal Distribution”. *Proc. of the 15th Int. Conference on Digital Audio Effects (DAFx-12)*, September 17-21, 2012, York, UK.

[8] W.J., Pielemeier and G.H. Wakefield, ‘‘MultiComponent Power and Frequency Estimation for a Discrete TFD," *Proc. of the IEEE Intl. Symp. On Time-frequency and Time-scale analysis*, pp. 620-623, 1994.

[9] Robert J. McAulay and Thomas F. Quatieri., “Speech Analysis/Synthesis Based on a Sinusoidal Representation”, *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. 34, no. 4, pp.744-754, 1986.

[10] J. Smith, “Creating minimum Phase filters and signals”, https://ccrma.stanford.edu/~jos/fp/.

[11] N. D’Alessandro A. Moinet T. Dubuisson, and T. Dutoit “Causal/anticausal decomposition for mixed-phase description of brass and bowed string sounds”, *Proc. of the International Computer Music Conference (ICMC’07)*, Copenhagen, Denmark, 2007.

[12] J. Beauchamp, “Unix Workstation software for analysis, graphics, modification, and synthesis of musical sounds”. *AES Convention 94*, March 1993.

[13] K. Fitz, and L. Haken, “Bandwidth enhanced sinusoidal modeling in Lemur”. *Proc. of the International Computer Music Conference (ICMC’95)*, Banff, Canada, 1995.

[14] X. Rodet P. Depalle, and G. Poirot, “Speech analysis and synthesis methods based on spectral envelopes and voiced/unvoiced functions”. *Proc. of the European Conf. on Speech Tech*, Edinburgh, UK, 1987.

[15] X. Serra and J. Smith, “Spectral Modeling Synthesis: a sound analysis/synthesis system based on a deterministic plus stochastic decomposition”, *Computer Music Journal*, **14**(4), pp. 12-24, 1990.

[16] M. Wright, J. Beauchamp, K. Fitz, X. Rodet, A. Rôbel, X. Serra, and G. Wakefield, “Analysis/Synthesis Comparison”. *Organised Sound*, **5**(3), pp. 173-189, 2000.

[17] D.W. Griffin and J.S. Lim, “Signal reconstruction from short-time Fourier transform magnitude”, *IEEE Trans. Acoust., Speech, and Signal Processing.*, **Vol. ASSP-32**, no. 2, pp.: 236-243, 1984.