AUTOMATIC PARTIAL EXTRACTION FROM THE MODAL DISTRIBUTION

Thomas Lysaght
Department of Computer Science, NUI Maynooth
Co. Kildare, Ireland
Tom.Lysaght@nuim.ie

Joseph Timoney
Department of Computer Science, NUI Maynooth
Co. Kildare, Ireland
Joseph.Timoney@nuim.ie

Victor Lazzarini
Department of Music, NUI Maynooth
Co. Kildare, Ireland
Victor.Lazzarini@nuim.ie

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. The problem of signal synthesis from bilinear time-frequency representations such as the Wigner distribution has been investigated [1,14] using methods which exploit an outer-product interpretation of these distributions. Methods of synthesis from the MD based on a sinusoidal-analysis-synthesis procedure using estimates of instantaneous frequency and amplitude values have relied on a heuristic search 'by eye' for peaks in the time-frequency domain [2,7,8]. An approach to detection of sinusoidal components with the Wigner Distribution has been investigated in [15] based on a comparison of peak magnitudes with the DFT and STFT. In this paper we propose an improved frequency smoothing kernel for use in MD partial tracking and adapt the McCauley-Quatieri sinusoidal analysis procedure to enable a sum of sinusoids synthesis. We demonstrate that the improved kernel enhances automatic partial extraction and that the MD estimates of instantaneous amplitude and frequency are preserved. Suggestions for future extensions to the synthesis procedure are given.

1. INTRODUCTION

The MD was introduced by Pielemeier and Wakefield [2] as a member of the Cohen general class of time-frequency distributions [10] for the analysis of music signals. It is primarily a Wigner distribution, or more specifically, a 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 neighborhood 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. However, one drawback of the Wigner distribution in relation to modal analysis of quasi-harmonic signals is the existence of both inner and outer cross terms [11] amounting 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 outer cross terms for music signals. This time smoothing reduces the bandwidth of the distribution in the time direction and so facilitates 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 instrumental sounds such as the frequency modulation of attack transients or 'rogue' piano partials [7], whereas normal spectrogram smoothing would obfuscate such detailed characteristics. While the relative lack of smoothing in the MD is its strength, when compared with the spectrogram, the existence of cross terms and added artifacts complicate any interpretation of the MD surface as representing a sinusoidal plus residual noise model as in SMS [4]. Therefore, the application of useful partial tracking methods proves difficult. In this paper we construct a novel frequency smoothing kernel which provides better noise suppression in the MD while conserving the accuracy of parameter estimates in the distribution. The paper is organized as follows. Section 2 gives the theoretical background to the MD. Part 3 describes the new frequency kernel and details its application of partial tracking with the MD. Part 4 describes testing for both synthetic and real signals. Part 5 gives test results and conclusions are drawn and suggestions for future work in Part 6.

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:

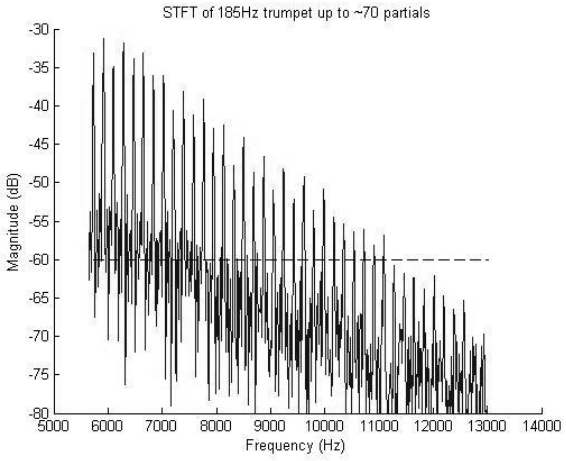


Figure 1: *Spectrogram frame for a trumpet tone 185Hz sampled at 44.1kHz showing partials beyond 13kHx (> $f_s/4$) in which case normal MD sampling would produce aliasing.*

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

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:

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

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 prior band-limiting is incurred. An example is shown in Fig. 1 for an F#3 trumpet tone. In these cases the analytic version of the signal may be generated which avoids such aliasing.

2.2. The Analytic Signal

The analytic signal of $s(t)$ is defined as:

$$s(t)=s(t)+j\hat{s}(t) \qquad (6)$$

where:

$$\hat{s}(t)=s(t)*\frac{1}{\pi t}=\int_{-\infty}^{\infty}\frac{f(\tau)}{\pi(t-\tau)}d\tau \qquad (7)$$

is the Hilbert Transform of $s(t)$, and the * symbol represents convolution. The corresponding analytic spectrum is given by:

$$F_a(f)=2U(f)F(f) \qquad (8)$$

where $U(f)$ is the unit step function given by:

$$U(f)=\begin{cases}0, & f<0,\\ 1, & f>0\end{cases} \qquad (9)$$

This allows the complete spectrum to be used to represent the positive frequencies only, thus doubling the spectral resolution for the distribution. Another advantage of using the analytic signal is the elimination of cross terms between negative and positive frequencies. These can manifest as extra tracks at non partial or partial locations.

2.3. Cross terms

Given a music signal model as follows:

$$s(t)=\sum_{k=1}^{M}A_k e^{j(w_k t+\phi_k)} \qquad (10)$$

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:

$$W_s(t,\omega)=\int_{-\infty}^{\infty}R_s(t,\tau)e^{-j\omega\tau}d\tau$$
$$=\sum_{k=1}^{M}A_k^2\delta(\omega-\omega_k)+ \qquad (11)$$
$$\sum_{k=1}^{M}\sum_{l=1}^{M}A_k A_l\cos([\omega_k-\omega_l]t+\phi_k-\phi_l)$$
$$\times\delta\left(\omega-\frac{(\omega_k+\omega_l)}{2}\right)$$

The *auto* terms of $s(t)$ are given by the first term in Eq. 11. 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. For music signals, this property gives rise to amplitude modulated partials and the possibility of additional artefacts and cross terms at partial frequencies not in the original signal.

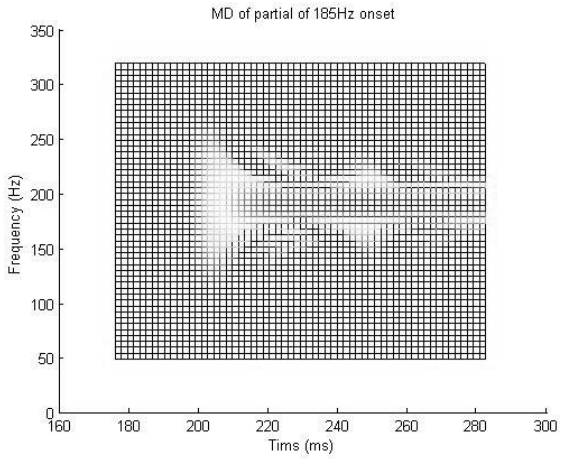


Figure 2: *MD of attack of a fundamental at 185Hz sinusoid. The wideband onset and widening of the auto terms mainlobe at points of amplitude discontinuity around 240msecs is clearly evident.*

2.4. The Modal Distribution (MD)

The MD was designed to minimise these cross terms in Eq. (11) for music signals. 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

$$M(n,k) = \sum_{n=-L+1}^{L-1} R_{s,l}(n,l) g_{LP}(l) e^{\frac{-j2\pi kl}{2L}} \tag{12}$$

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. Following time smoothing the number of MD frames can be decimated by at least half the length of the impulse response of the cross term filter. 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

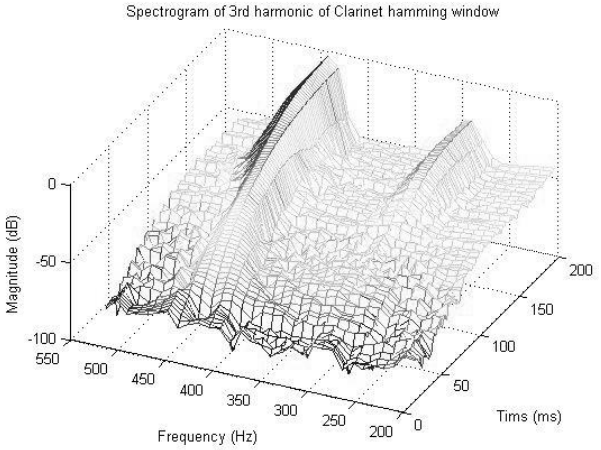


Figure 3: *Characteristic smoothing and sidelobe suppression to -42dB for the spectrogram showing the second and third harmonics of a 146Hz clarinet fundamental.*

for these auto terms. However, the width of auto terms characteristically exhibits variations due to large amplitude changes or discontinuities. The effect is the creation of broadband artefacts in the distribution. Fig 2. shows the attack of a monocomponent synthetic signal illustrating the signature wideband MD onset and 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 [7] 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 atuoterm main lobe.

2.5. MD Synthesis Parameters

Signal synthesis parameters of amplitude and frequency are calculated as local averages in the MD centered 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 are given as follows. Given:

$$p(n) = \sum_{l=-L}^{L} M(n,l) \tag{13}$$

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

$$A(n) = 4\sqrt{p(n)} \tag{14}$$

and the instantaneous frequency by:

$$F(n) = \sum_{l=-L}^{L} \frac{lM(n,l)}{p(n)} \tag{15}$$

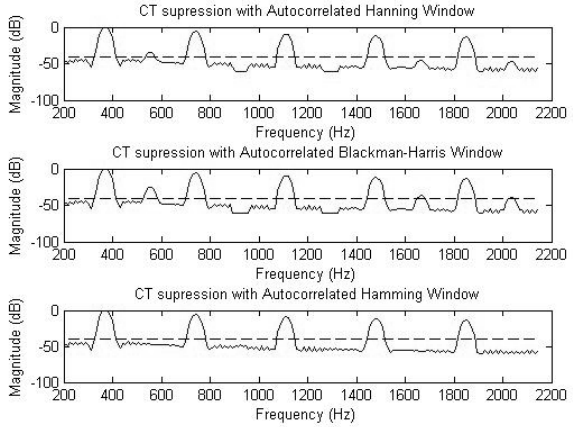


Figure 4: Comparison of cross term smoothing effect for three autocorrelated window, Hamming, Hanning and Blackman-Harris. The lower plot for the Hamming window performs best, lowering the cross terms below 40dB below neighbouring auto terms.

2.6. MD Kernel Choice

In [2], both $h_{LP}(p)$ and $g_{LP}(l)$ are chosen to be an autocorrelated Hamming window for the purpose of obtaining sidelobe attenuation of -42dB comparable with the spectrogram shown in Fig. 3. For the generalized cosine family of windows, the main lobe width is known to be 8 bins or $8\pi/M$ radians per sample [9]. Autocorrelating $g_{LP}(l)$ maintains this main lobe width. Fig. 4 illustrates that using an autocorrelated Hamming window for $h_{LP}(p)$ reduces the effect of the cross terms below -40dB between auto terms compared with either the BlackmanHarris or Kaiser-Bessel windows, both known for their superior harmonic identification properties [13].

3. PARTIAL TRACKING WITH 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 [3] 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. 14 & 15, and calculated around the bandwidth of each candidate peak, this bandwidth being determined by the main lobe width of the frequency smoothing kernel $g_{LP}(l)$. Due to the prominence of noise terms in the MD pertinent pruning of all tracks based on the assumption of a quasi-harmonic series may be required prior to synthesis.

3.1. An Improved Modal Kernel

One possible approach to reducing the prominence of such noise terms is to improve the smoothing capabilities of $g_{LP}(l)$. Fig. 5 shows a comparison of the suppression characteristics for $g_{LP}(l)$. The autocorrelated Hamming window (HammingX) has much better sidelobe suppression than the ordinary Hamming window while maintaining the same main lobe width measured between mainlobe minima and detailed in Table I. On the other hand, the squared Hamming window (HammingQ) with a slightly wider main lobe has a monotonically decreasing series of sidelobes. By autocorrelating this squared window we obtain a kernel which significantly lowers the sidelobe attenuation, greater than that of the autocorrelated Hamming window and retains the monotonically decreasing series of sidelobes and mainlobe width. This new window (HammingQX) significantly reduces ridges between the mainlobes of auto terms in the distribution as shown in Fig. 6, thus reducing the number of possible candidate noisy peaks for partial tracking. From Table I, we note that the highest sidelobe level of this new window is lower than that of the autocorrelated window although at the expense of a slightly increased main lobe width by 4 bins. For brevity, in the remainder of the paper, we will refer to the autocorrelated frequency smoothing window as winX, and the autocorrelated squared window as winXS and the corresponding MDs as MDX and MDSX.

Table 1: *Window suppression characterises*

Window Type	MLW (bins)	Highest Sidelobe (dB)
Hamming	8	-42
HammingQ	12	-49.2
HammingX	8	-85.4
HammingQX	12	-93.3

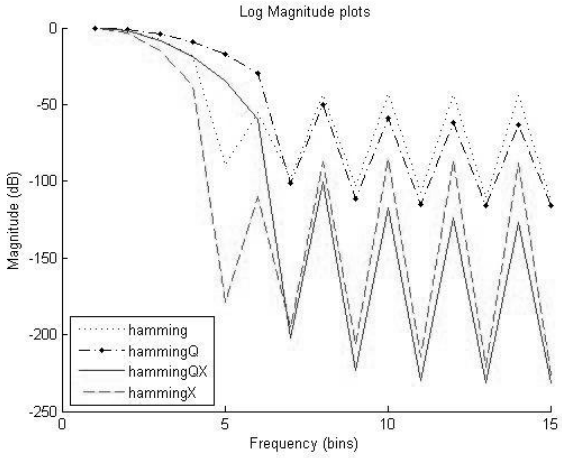
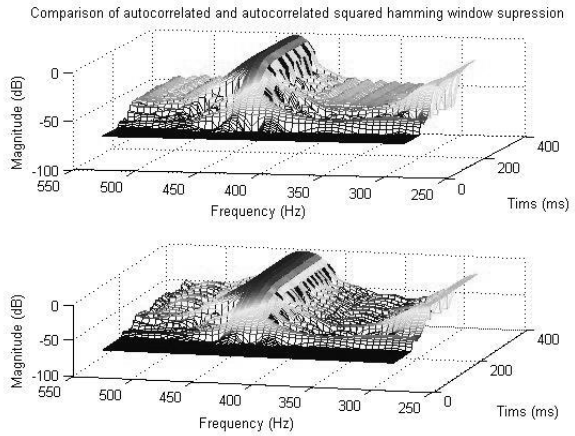


Figure 5: *Main lobe width and sidelobe suppression characteristics for normal, squared (HammingQ), autocorrelated (HammingX), and autocorrelated squared (HammingQX) Hamming windows*

4. TESTING

We test the accuracy of the MD estimates by comparing results for a synthetic test signal and for various instrumental tones using samples from the McGill University Master Samples, sampled at 44.1 kHz. Synthesis of tones is carried out using a sum of sinusoids approach based on the MD estimates for instantaneous amplitude and frequency. Instantaneous phase is calculated as the integral of the instantaneous frequency. The synthetic test signal has a fundamental (f_c) of 146Hz with 42 exact harmonics of



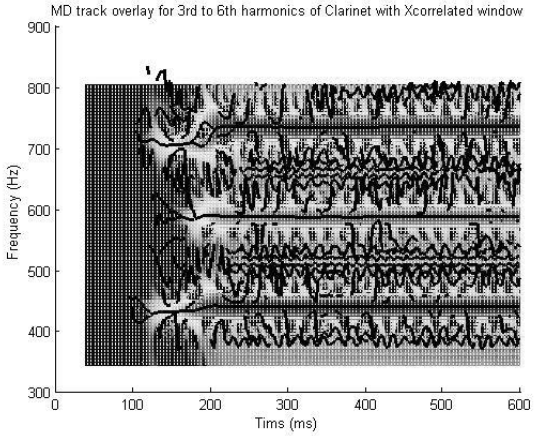
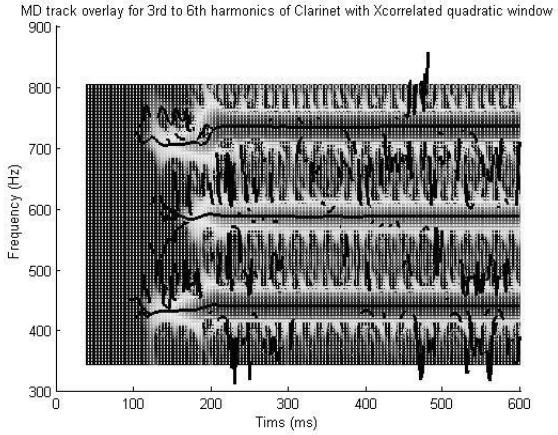
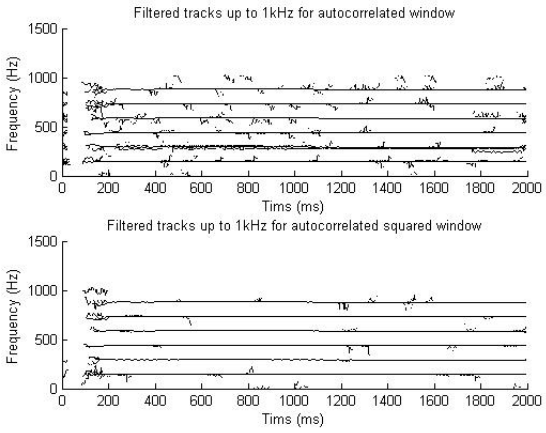


Figure 6 & 7: *Suppression characterises of Hamming and Hamming QX windows for MDX (upper plot) and MDXS for 3rd harmonic of clarinet tone with fundamental at 146Hz. There are a significantly greater number of sidelobe ridges visible between the auto terms for MDX. Fig. 7 shows tracks overplayed on MDX - a large number of noise peaks are present and prominent sidelobes evolve as parallel tracks to the auto terms.*





Figures 8&9: *Fig. 8 shows tracks overlaid on MDX – there is a clear identification of the auto terms as tracks and the absence of parallel sidelobe tracks. Fig. 9 illustrates filtered partial tracks for MDX (upper plot) and MDXS. A prominent parallel sidelobe track close to the centre frequency of the second partial is evident in MDX and absent in MDSX.*

decreasing amplitude and increasingly delayed attacks. In calculating the MD we lower $\Delta f_{\min}$ to 120Hz and calculate $h_{LP}(p)$ as $1.75\times f_s/f_c$, as this was found to guarantee greater than -40dB suppression. This results in a length for $h_{LP}(p)$, L_h, of about 30msecs. We set the MD decimation step size to give MD estimates at about every 2 msecs. We then search for tracks at a depth of 70dB in the MD surface in order to recover the low amplitude attacks of the harmonics. As the number and position of the partials are known we filter the tracks returned to extract the 42 harmonic partials. We then compute the minimum, maximum and average frequency and amplitude estimates for each partial. The estimates are calculated from beyond the end-effects region of each partial to exclude spurious results. We next test a clarinet tone, also 146Hz with the same MD settings as for the synthetic tone. Again we set the search depth at 70dB in the MD surface in order to recover the low amplitude attacks of the harmonics and any very low amplitude partials. With winX the number of tracks returned at a depth of 70dB was over 3000 and with winXS about 1900 or slightly more than half. Figs 7.and 8 show an overlay of the tracks on the MD for harmonics 3-6. We can see an increased number of tracks on each side of the main lobes for MDX compared with MDXS. In a further pruning stage, Fig. 9 shows tracks remaining after filtering for harmonics partials up to 1000Hz. Clearly visible is a strong extra partial close to the center frequency of the second harmonic for MDX. This is audible in the subsequent synthesis as a modulation artifact.

5. RESULTS

Fig. 10 graphs the difference in minimum, maximum, and average frequency and amplitude estimates for both winX and winXS for the synthetic test signal. There is less than a 1 cent difference in the frequency estimates for both windows and a fraction of a decibel for the steady-state amplitude estimates. For the clarinet, the results are graphed in Fig. 11. The difference in frequency and amplitude estimates for 25 of the strongest clarinet partials are shown. Again the difference in the frequency estimates is less than one cent while the difference in amplitude estimates is less than 1dB. The outliers at harmonics 18 and 25 can be explained

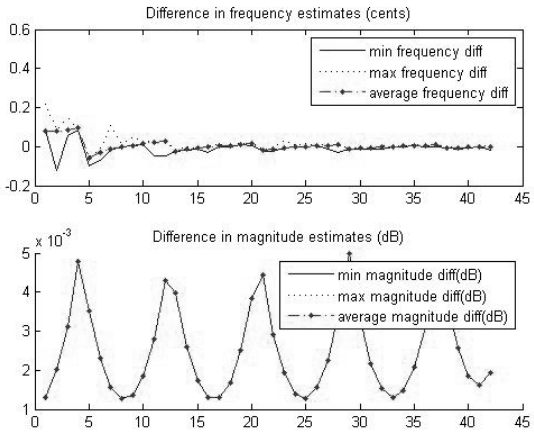


Figure 10: *Difference in Frequency (upper plot) and Amplitude estimate for MDX and MDXS for a synthetic signal of 42 exact harmonics. Frequency differences are within 1 cent tolerance while amplitude differences are a fraction of 1dB.*

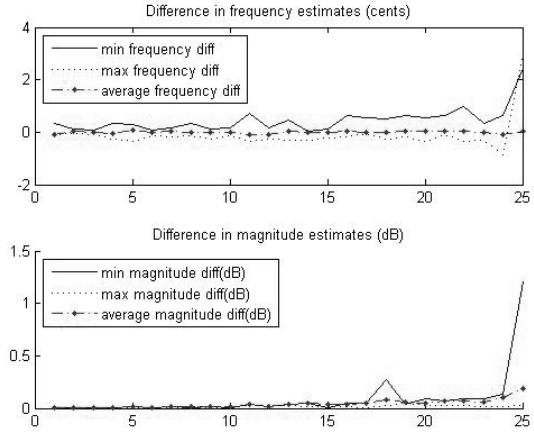


Figure 11: *Difference in Frequency (upper plot) and Amplitude estimate for MDX and MDXS for a clarinet signal of fundamental 146Hz and strongest 25 partials. Frequency differences are within 1 cent tolerance while amplitude differences are a fraction of 1dB, apart from high frequency low amplitude partials.*

by extra power in the estimates for MDX due to excursions of the tracks due to sidelobe artefacts. Synthesis of the resulting tracks was achieved by a sum of sinusoids method with phases computed as the integral of instantaneous frequency. Cubic phase interpolation was implemented between frame boundaries.

6. CONCLUSIONS

In this paper we have presented the squared autocorrelated Hamming window as a new MD frequency smoothing kernel which reduces the prominence of noise artifacts. It has better suppression capabilities than the autocorrelated Hamming window due to a monotonically decreasing series of sidelobes and increased sidelobe rolloff. Although this new kernel has a slightly increased mainlobe width, we have demonstrated that the MD estimates for amplitude and frequency are preserved. Although the MD surface is difficult to adapt for automatic partial tracking, the application of the new kernel results in far fewer artifacts. Subsequent informal listening tests for synthesis show that the synthesized tones are very close to the originals for the new kernel whereas audible artifacts are present for the autocorrelated window. Future work will focus on testing the kernel for a greater range of instrumental tones and also on streamlining the automatic sinusoidal analysis/ synthesis to include other and more recent partial tracking methods such as use of hidden Markov models [11,16] and DESAM [12].

7. REFERENCES

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

[2] Pielemeier, W. J., and Wakefield, G. "A high-resolution time-frequency representation for musical instrument signals", *Journal of Acoustical Society of America",* 99(4), Pt. 1, April 1996.

[3] 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.

[4] X. Serra, Musical Signal Processing, chapter Musical Sound Modeling with Sinusoids plus Noise, pp. 91–122, G. D. Poli, A. Picialli, S. T. Pope, and C. Roads Eds. Swets & Zeitlinger, Lisse, Switzerland, 1996.

[5] Mecklenbrauker, W. F. G., and Classen, T. A. C. M, "The Wigner Distribution – A Tool for time-frequency signal analysis; part II: discrete time signals", *Philips J. Res.*, Vol. 35, pp. 276-300, 1980.

[6] Maureen Mellody and Gregory H. Wakefield, "The time-frequency characteristics of violin vibrato: Modal distribution analysis and synthesis", *J. Acoust. Soc. Am.* Volume 107, Issue 1, pp. 598-611 (2000).

[7] Rowena Cristina L Guevara, "Modal distribution analysis and sum of sinusoids synthesis of piano tones", Dissertation (Ph.D.)--University of Michigan, 1997.

[8] W.J., Pielemeier and G.H. Wakefield, ''MultiComponent Power and Frequency Estimation for a Discrete TFD," *Proc. of the IEEE*-SP Intl. Pp. 620-623, 1994.

[9] J.G. Proakis and D.G. Manolakis, "*Digital Signal Processing principles, algorithms, and applications,* 2nd ed., MacMillan Publishing Company, New York.

[10] Cohen, L. *Time Frequency Analysis*. Prentice-Hall, New Jersey, 1995.

[11] Corey Kereliuk, Philippe Depalle, Improved hidden Markov model partial tracking through time-frequency analysis'. *In Proceedings of the Digital Audio Effects (DAFx-08),* September 1-4, 2008, Espoo, Finland. 111-116

[12] Mathieu Lagrange, Sylvain Marchand, Martin Raspaud, and Jean-Bernard Rault. Enhanced Partial Tracking Using Linear Prediction. *In Proceedings of the Digital Audio Effects (DAFx'03)* Conference, London, United Kingdom, September 2003. Queen Mary, University of London.

[13] F. J. Harris, "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform." *Proceedings of the IEEE.* Vol. 66 (January 1978). pp. 51-84.

[14] G.F.Boudeaux-Bartels and T.W.Parks, "Time-varying filtering and signal estimation using Wigner distribution synthe-

sis techniques," *IEEE Trans. ASSP,* vol.ASSP-34, no.3, pp. 442-451, June 1986.

[15] Jeremy J Wells, Damien T. Murphy, 'Real-Time Partial Tracking in an Augmented Additive Synthesis System'. *In Proceedings of the Digital Audio Effects (DAFx-02)*, September 26-28, 2002, Hamburg, Germany. 93-96

[16] P. Depalle, G. Garcia, and X. Rodet, "Tracking of partials for additive sound synthesis using hidden Markov models", *Proceedings of the IEEE conference on Acoustics, Speech, and Signal Processing (ICASSP),* vol. 1, pp. 242-245, 1993.