REAL-TIME PARTIAL TRACKING IN AN AUGMENTED ADDITIVE SYNTHESIS SYSTEM

Jeremy J. Wells and Damian T. Murphy

Department of Electronics
University of York, Heslington, York, YO10 5DD, UK
jjw100@ohm.york.ac.uk, dtm3@ohm.york.ac.uk

ABSTRACT

This paper describes an approach to real time partial tracking in an analysis\transformation\resynthesis system using a combination of linear and bi-linear time-frequency techniques. Tests of the system have been made using both natural and synthetic sounds. Results are presented and areas for further research and development are discussed.

1. INTRODUCTION

Much of the recent work on analysis and resynthesis for spectral modeling of sound has concentrated on developing a sinusoidal plus residual signal model. Description of the residual signal has ranged from the parameters of a filtered noise source [1],[2] to those of harmonic band wavelets [3]. Augmented Additive Synthesis (AAS) is a proposed system which seeks to describe timbre as a linear combination of perceptually orthogonal sound types. This combination may, or may not, include sinusoids or filtered noise as well as other sound types to offer a meaningful palette of descriptions for as wide a range of timbres as possible, encompassing both real and synthetic targets in real time.

For harmonic sounds it is likely that a linear combination of sinusoids will be one of the sound types that contributes to the AAS model. Additive resynthesis of sinusoids has traditionally been based on phase vocoder analysis of sounds with subsequent peak detection and continuation, known as *partial tracking*, such as that used in the spectral modeling synthesis (SMS) system. Although recent versions of SMS software offer real-time interaction with the analysis data during resynthesis simultaneous analysis/interaction/resynthesis is not possible. One of the reasons for this is that traditional partial tracking algorithms require acquisition of the entire audio signal before analysis can begin. With knowledge of the entire signal events can be analysed from their steady state towards their transient onsets (i.e. backwards) and such an algorithm can take a revisionist approach to existing analysis data by adapting it as the analysis progresses.

2. PARTIAL TRACKING

One of the aims of AAS is to provide a more intuitive system for musicians to use by offering parameters which relate more directly to perceptual effect. Therefore parameters relating to loudness and frequency are considered as having greater perceptual relevance than the complex and abstract mathematical data used in the transforms themselves. It is also important to reinforce intuition and usefulness through realtime interaction (action by the player, feedback by the system) with these parameters. As such these requirements demand a partial tracking algorithm that can operate in realtime (or within the limitations of analysis windowing if implemented). Therefore it is useful to investigate possible computationally efficient methods for such a system that might offer additional information within a single frame about the type of signal under analysis.

The advantages and limitations of the short time Fourier transform (STFT) are well documented [4]. The STFT is a linear analysis technique that attempts to extend the Fourier analysis of stationary signals to non-stationary signals by dividing the input into overlapping windows of shorter signals. Each windowed signal is therefore assumed to be stationary. AAS makes use of the STFT and derives amplitude and frequency for spectral peaks by comparing the magnitude data for the DFT of the windowed signal with that of its derivative [5]. Simply interpreting spectral peaks for signals which contain transients and/or few harmonically related elements may lead to incorrect identification of parts of such signals as sinusoids. Such systems may therefore attempt to use sinusoids to recreate broad band signal components which is inefficient, counter-intuitive and difficult to achieve if the number of sinusoidal oscillators is limited (which may well be the case for real time operation). It is for this reason that systems such as SMS with partial tracking and synthesis of the residual with filtered noise have been developed.

3. THE WIGNER BI-LINEAR TIME-FREQUENCY DISTRIBUTION

When considering individual frames in isolation the STFT is not able to provide enough information with regards to the type of signal within a bin to determine which resynthesis technique should be used to reproduce it. Bi-linear (or quadratic) time-frequency analysis techniques applied to the same windowed frames can give much higher time resolution than the STFT, since they do not assume a stationary signal within a frame. The disadvantage of such techniques is that they introduce interference (or cross) terms due to their inherent non-linearity [4]. A well known example of these techniques is the Wigner distribution (WD).

The WD is defined for all time and for real time analysis of sampled signals a running window is applied to the input resulting in the pseudo discrete WD (PDWD). Exploiting symmetry and recasting in a form which can be evaluated using the FFT the PDWD of a discrete signal *x(n)* is as follows:

$$PDWD(m,k) = 4\mathrm{Re}\left\{ \sum_{n=0}^{(N/2)-1} y(m+n)\, y^*(m-n)\, e^{j4\pi nk/N} \right\} - 2y[m]\,y^*[m] \quad (1)$$

Where *y(m)* is *x(n)* upsampled by a factor of 2 and interpolation filtered. We can see from this equation that the PDWD is the DFT of a form of auto correlation function performed on the windowed signal. The time resolution of the PDWD is determined by the step size of *m* and reaches its upper limit when *m*=1. The frequency resolution is determined by the windowing function used and the window length and is proportional to that of the STFT for the same window function and sequence length.

4. PARTIAL TRACKING USING THE STFT AND THE WIGNER DISTRIBUTION

For the STFT, provided correct oversampling of the signal is carried out in the time domain, the magnitude of the assumed components (basis functions) can be determined for each bin. Then the frequency and amplitude of the sinusoidal function can be derived taking into account any energy smearing caused by the windowing function [5]. Whilst techniques such as comparing the peaks in the magnitude spectrum to the transform of the analysis window can be used to distinguish partials from other frequency components for certain signals, existing methods for partial detection also use a peak continuation algorithm to determine which frequency components are ‘well behaved’ partials[2]. We hope to use Wigner analysis of the same windowed data as the STFT to offer greater insight into the nature of signals on a more localised basis.

Partial tracking systems which perform their analysis ‘off line’ can detect a number of peaks for each frame and then, considering all frames together, determine which peaks within each frame are partials via some form of peak continuation. For a real time application where the delay between input (analysis) and output (synthesis) must be as low as possible such an approach is not possible. If possible a partial tracking algorithm which is designed to operate in real time should be able to determine which peaks are due to deterministic partials for each frame as it analysed. If this cannot be done reliably within one frame then such a decision should be made within as few frames as possible.

When a single partial signal is perfectly correlated (i.e. it contains no other components) its PDWD magnitude will be the square of the equivalent (same window function, step size etc.) DFT magnitude. Where the signal is not well correlated with itself in parts of the spectrum (i.e. it contains some noise as well) this relationship is not maintained. Thus we can compare the magnitude data from the PDWD with magnitude squared data from the STFT to determine whether an STFT component is likely to be a sinusoid or not. Figure 1 shows the magnitude correlation for PDWD and STFT analysis of a slowly amplitude modulated sinusoid. Figure 2 shows the correlation for a noise source filtered around 1 kHz to occupy a single DFT bin for a sample rate of 44.1 kHz and a frame size of 1024 samples. In Figure 1 it can be clearly seen that, for a sinusoidal component the PDWD magnitude is the square of the STFT magnitude. In figure 2 the relationship is much less clearly defined. This indiciates that we can test the relationship between the STFT and the PDWD magnitude for analysis bins containing spectral peaks to determine what signal component type might be present.

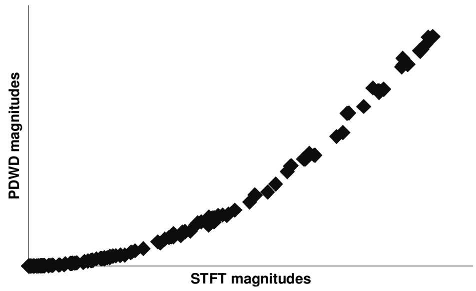


Figure 1: *STFT against PDWD magnitudes for amplitude modulated sinusoid at 1 KHz.*

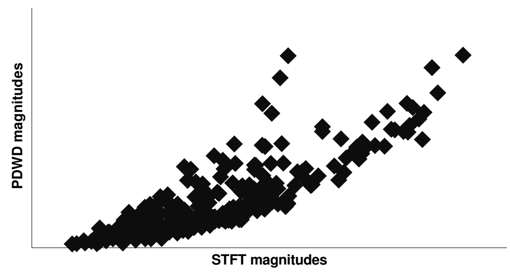


Figure 2: *STFT against PDWD magnitudes for narrowband filtered noise centered 1 KHz.*

There are few real world or synthetic signals which contain a single partial and any partial tracking algorithm should be able to cope with multi-partial (harmonic and non-harmonic) signals. Despite possessing a superior time resolution to the STFT the usefulness of the WD is severely limited by the cross term interference which is introduced in the kernel for all but single partial signals. For N components the number of cross terms generated is given by:

$$N_{Cross} = N_{Term}(N_{Term} - 1)/2 \qquad (2)$$

These terms are large in amplitude compared to the auto terms that have generated them and oscillate rapidly. For all but the simplest signals differentiation between contributions to bin magnitudes from auto and cross terms is very difficult. The WD for complex signals can also be negative within some bins. Certainly a direct comparison between PDWD and STFT magnitudes for a peak can no longer be relied upon as an indicator of the type of signal within that bin.

5. OSCILLATOR ITERATION WITH PDWD AND STFT MAGNITUDES

5.1. Algorithm

Since the WD satisfies the time marginal (i.e. integration along its frequency axis yields the instantaneous power of the signal) we can still make useful comparisons between the PDWD and DFT magnitudes obtained. Although cross term interference makes a simple comparison between PDWD and DFT magnitudes unreliable other statistical analyses of the data both processes produce are more useful. A method for estimating the number of sinusoidal oscillators required for resynthesis of the deterministic part of the signal is presented below. This method uses the standard deviation of PDWD – DFT magnitude ratios as a ‘smoothed’ measure of the signal correlation within bins containing signal maxima.

Analysis of how the ratio between the STFT magnitude squared and the PDWD varies for peaks in a signal has yielded a method for determining how many sinusoidal oscillators are required for resynthesising the deterministic part of a signal at a particular frame. Maxima in the DFT spectrum are identified and the ratio of the magnitude values produced by both analysis methods for these frequency bins is derived. The standard deviation of these ratios is then calculated. This process is repeated reducing the number of maxima searched for by 1 each time. When the optimum numer of deterministic maxima has been reached the standard deviation falls dramatically.

5.2. Results

Figure 3 shows the a test signal waveform consisting of a harmonic tone followed by white noise. Figure 4 shows how the standard deviation varies for this signal when 6 and 3 maxima are considered. Observing Figure 4 for the faded portions of the signal we see that the standard deviation is unaffected by signal level as expected. When the number of maxima considered falls below the optimum of oscillators the standard deviations obtained are similar to those for the optimum number.

Figure 3: *Waveform of test signal. The first portion consists of three harmonically related partials. The second portion is broad band noise.*

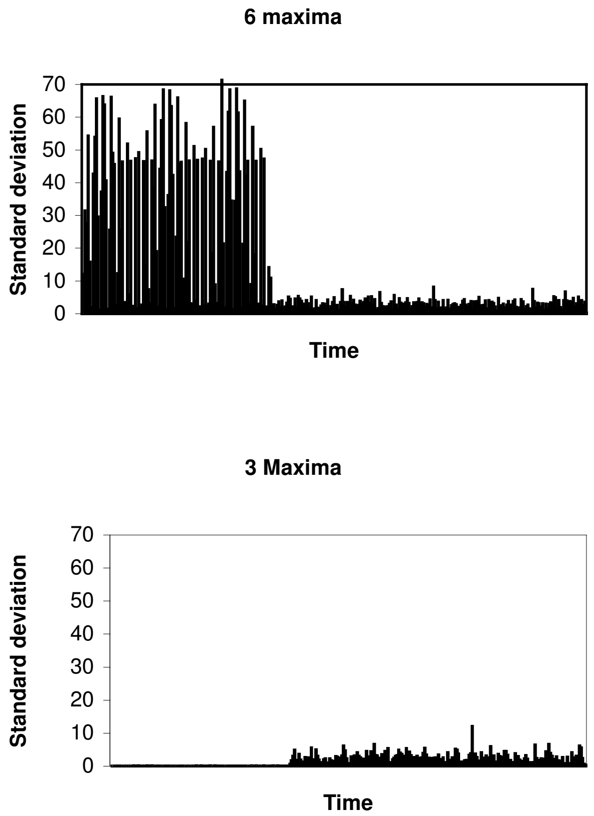


Figure 4*: Standard deviation of ratios of PDWD magnitude and STFT magnitude squared for 6 and 3 maxima.*

For sounds with harmonically related partials it is very likely that cross terms will coincide with actual partials. A similar signal to that in Figure 3 was generated with non-harmonically related sinusoids and noise to examine the effect that such a signal would have the standard deviations present in its deterministic portion. Here the fall in standard deviation is more dramatic when the number of maxima reaches the optimum number of oscillators suggesting that this method might also indicate how close the deterministic part of a signal is to harmonicity.

We have started to investigate the usefulness of this method for real world signals. Figure 5 shows the waveform of a recording of a male treble chorister singing the word “vox”. The deterministic and stochastic part of the utterance are shown. Trial and error, by fixing the number of oscillators and auditioning the result, suggests that the deterministic part of the sound is optimally reproduced with 2 oscillators and this is borne out by the data retrieved with our iterative method. Figure 6 shows the variation in standard deviation during this sound when evaluating for 2 and 6 maxima. Again, the stochastic part of the signal produces similar results for both plots. For all signal types it can be seen that the standard deviation fluctuates to a certain extent for both determinstic and stochastic portions. Currently we only sample the PDWD at the same rate as the STFT (1024 samples with an overlap of 4). It is hoped that by sampling the PDWD at a greater rate will allow smoothing of the standard deviations obtained.

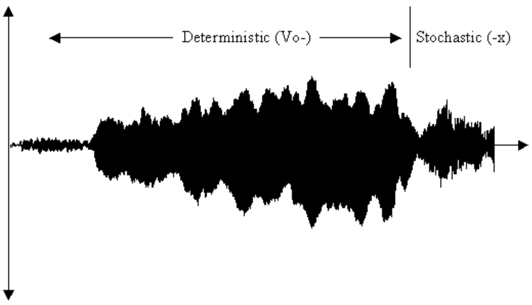
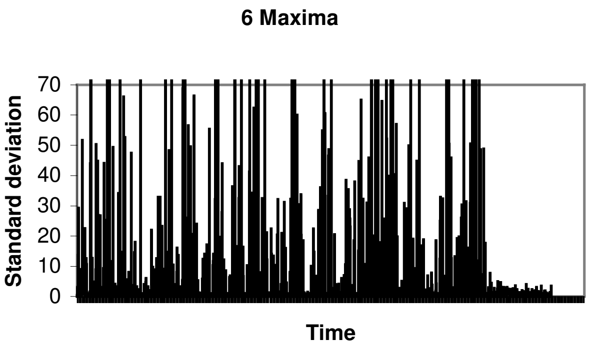


Figure 5: *Waveform of recording of chorister singing the word "vox"*



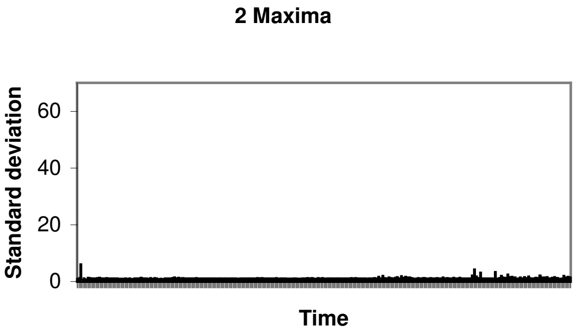


Figure 6: *Standard deviation of ratios of PDWD magnitude and STFT magnitude squared for 6 and 2 maxima.*

6. CONCLUSIONS

The potential usefulness of combining the STFT and PDWD for the detection of sinusoidal signal components and estimation of their parameters has been presented. In particular we have demonstrated how, for various signal types, the optimum number of oscillators required for sinusoidal resynthesis can be determined using a simple process. Our work in implementing this as a real time process makes use of the highly optimised routines for the fast Fourier transform, auto correlation and various statistical analysis techniques which are offered by the Intel signal processing library[6]. We are currently investigating how the PDWD magnitudes might be smoothed by filtering or using some other cross term suppression technique to achieve greater consistency of standard deviation across frames so that this technique can be applied more reliably on a more localised (ideally frame by frame) basis. It is hoped that further investigation will yield more information on how this combination of analysis methods works. A partial tracking and oscillator system which uses the PDWD and STFT of current and acquired data to operate in real time is in development. This will form part of the AAS analysis and resynthesis engine.

7. REFERENCES

[1] Serra, X. and Smith, J., "Spectral modeling synthesis: a sound analysis/synthesis system based on a deterministic plus stochastic decomposition", in *Computer Music Journal 14(4)*, Cambridge, Massachusetts, pp.12-24, 1990

[2] Serra, X. "Musical sound modeling", in *Musical Signal Processing* (ed. De Poli, G., Piccialli, A., Pope, S. T., Roads, C.), Swets and Zeitlinger, 1997

[3] Evangelista, G. and Polotti, P., "Fractal additive synthesis via harmonic-band wavelets", in *Computer Music Journal 25(3)*, pp.22-37, 2001

[4] Pielemeier, W.J., Wakefield, G.H. and Simoni, M.H., "Time-frequency analysis of musical signals" in *Proceedings of the IEEE 84(9)*, pp.1216-1230, 1996

[5] Desainte-Catherine, M. and Marchand, S., "High-precision Fourier analysis of sounds using signal derivatives" in *Journal of the Audio Engineering Society 48(7/8)*, pp.654-667, 2000

[6] Intel Corporation, "Intel signal processing library reference manual", 2000 (web document no longer available).