

HIGH ACCURACY FRAME-BY-FRAME NON-STATIONARY SINUSOIDAL MODELLING

Jeremy J. Wells, Damian T. Murphy

Audio Lab, Intelligent Systems Group, Department of Electronics
University of York, YO10 5DD, UK
{jjw100|dtm3}@ohm.york.ac.uk

ABSTRACT

This paper describes techniques for obtaining high accuracy estimates, including those of non-stationarity, of parameters for sinusoidal modelling using a single frame of analysis data. In this case the data used is generated from the time and frequency reassigned short-time Fourier transform (STFT). Such a system offers the potential for quasi real-time (frame-by-frame) spectral modelling of audio signals.

1. INTRODUCTION

Spectral modelling (SM) for transformation of musical signals is a well established area of digital audio effects [1]. Whereas the STFT represents a signal as grains with stationary magnitude and phase, which overlap in time and frequency, spectral models attempt to infer more intuitive and flexible representations of sound from such data. Whilst the original signal cannot be perfectly reconstructed from such models they are generally more amenable to feature extraction and perceptually meaningful transformations such as pitch shifting and hybridisation (cross-synthesis). The spectral modelling synthesis system (SMS) of Serra represents signals as the combination of sinusoids with slowly varying amplitude and frequency and filtered noise [2]. Other systems have extended the component set to include transients [1].

Whilst SM systems exist that can perform transformations and resynthesis from model data in real-time, the possibility of *generating* model data in real-time has received little attention. Since the complete analysis-modification-resynthesis cycle cannot be currently implemented in real-time audio processors that use SM, such a modelling paradigm is unavailable in the traditional real-time studio effects unit. A real-time/streaming system has recently been described but there is more than a single frame's delay between input and output while a minimum number of 'track points' are acquired [3]. There have been investigations into both single-frame sinusoidal discrimination and non-stationarity but this has been applied to the improvement of offline analysis. The system described in this paper produces non-stationary 'sinusoidal plus residual' model data on a frame-by-frame basis. Once a single frame of data has been acquired, sinusoids and non-sinusoids can be separated and the sinusoids described and synthesized with non-stationary amplitude and frequency (*i.e.* these parameters can change on a sample by sample basis).

Since this paper builds on existing work relating to single-frame non-stationary modelling of sinusoids an overview of this work is given in section 2. The limitations of existing non-stationary analysis and an improved system, which is adapted to function with reassigned STFT data and reduce parameter interdependence, are discussed in section 3. Section 4 describes how estimates of intra-frame amplitude and frequency change obtained using the

methods described in the previous section can improve estimates of the mean amplitude and frequency. Section 5 presents a frame-by-frame spectral modelling system which uses the analysis techniques described in the previous sections.

2. EXISTING METHODS FOR PARAMETER ESTIMATION

The non-stationary sinusoids discussed in this paper are assumed to be of the form:

$$s(t) = A(t) \sin \left(\int_{\tau=0}^{\tau=t} 2\pi f(\tau) d\tau + \phi \right) \quad (1)$$

where, for a single frame, $A(t)$ is an exponential function describing the amplitude trajectory and $f(t)$ is a linear function describing the frequency trajectory. ϕ is the phase of the sinusoid at the start of the frame. Therefore in this model the amplitude is piecewise exponential, the frequency is piecewise linear and the phase is piecewise quadratic. The 'piecewise' nature of these trajectories is inherent in the frame-by frame approach and existing cubic phase modelling techniques require more than a single frame of data to have been acquired [4].

Many methods exist for the estimation of the mean instantaneous frequency of components in the Fourier domain. These include measuring the phase difference between successive frames, interpolation of the magnitude spectrum and time-frequency reassignment. Reassignment is used in the system described here since estimates are obtained from a single analysis frame and it provides better estimates than other single-frame methods [5]. Time-frequency reassignment estimates the deviation of component energy from the centre of an analysis bin (frequency) and analysis frame (time) by taking two additional DFTs per frame. The first DFT uses a time ramped version of the original window function, the second uses a frequency ramped (time domain first order difference) version of the window. The estimate of frequency deviation from the centre of an analysis bin is given by:

$$f_{\text{deviation}} = -B\Im \left(\frac{DFT_{\text{frequency ramped window}}}{DFT_{\text{original window}}} \right) \quad (2)$$

where B is the width of a single analysis bin and \Im denotes the imaginary part of a complex number. The estimate of time deviation (in seconds) from the centre of an analysis frame is given by:

$$t_{\text{deviation}} = -\frac{1}{F_S} \Re \left(\frac{DFT_{\text{time ramped window}}}{DFT_{\text{original window}}} \right) \quad (3)$$

where F_S is the sampling rate and \Re denotes the real part of a complex number [6].

Systems for the single frame estimation of the parameters of non-stationary sinusoids have recently been proposed. These include the use of direct analytical methods for Gaussian windows and Fresnel integral approximations or empirical adaptation of the Gaussian methods for other window types [7, 8, 9]. The technique adopted and adapted here is phase distortion analysis (PDA) [10]. This method uses phase differences either side of a zero-padded spectral peak to provide a measure of intra-frame linear frequency change and exponential amplitude change within a single frame. The relationship between these measures and the actual amplitude change (dB per frame) and frequency change (bins per frame) is dependent upon the window type and is empirically determined. This is formally described by:

$$\Delta A_p = g(\phi_{p+1} - \phi_{p-1}) \quad (4)$$

$$\Delta f_p = h(\phi_{p+1} + \phi_{p-1}) \quad (5)$$

where p is the index of a magnitude spectrum peak, ϕ is phase, ΔA and Δf are the intra-frame amplitude and frequency change and $g(x)$ and $h(x)$ are the functions relating the phase difference to the intra-frame parameter changes. Amplitude and frequency non-stationarity produces changes in the window shape in the Fourier domain. Therefore, if these non-stationarities can be estimated then errors in the estimation of amplitude can be corrected and the quality of the model data improved [11].

3. REASSIGNMENT DISTORSION ANALYSIS

In this section we describe the adaptation of PDA to reassignment data, referred to here as reassignment distortion analysis (RDA). PDA uses phase deviations either side of the magnitude peaks in the DFT spectrum. For reassignment these deviations are embedded in the corresponding frequency and time offset estimates, given by Equation (2) and Equation (3). PDA effectively models the phase either side of a magnitude peak as a first order polynomial:

$$y = mx + c \quad (6)$$

where y represents the phase value, x the bin number, c the value from which the intra-frame frequency change Δf is derived, and m the value from which the intra-frame amplitude change ΔA is derived. PDA uses the difference in phase between the peak bin and those either side of the peak giving two data points. RDA directly uses time reassignment offset data across a peak giving three data points. Since three data points are available they can be modelled using a second-order polynomial which better represents the underlying shape of the phase spectrum. The non-stationary measures are therefore given by:

$$y = px^2 + mx + c \quad (7)$$

For RDA the relationship between this polynomial and the non-stationary measures is reversed: c is the value from which ΔA is derived and m is the value from which Δf is derived, p is not used. Figures 1 and 2 show the relationship between the RDA measure m and various combinations of values of ΔA and Δf for non-stationary Hann-windowed sinusoids using first- and second-order polynomials obtained from an 8192 point FFT of a 1025 sample frame. Where the frequency is decreasing the sign of m changes but its magnitude is the same.

Figures 1 and 2 show that, as for PDA, there is a limited range of Δf values for which m is monotonically increasing. This is

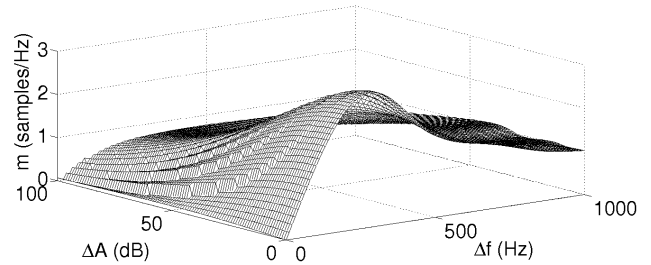


Figure 1: m versus ΔA and Δf , first order polynomial.

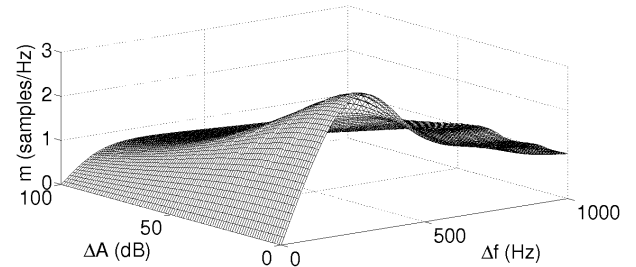


Figure 2: m versus ΔA and Δf , second order polynomial.

related to the length of the input frame and for a 1025 sample frame at 44.1 kHz this range of values is 0 to approximately 260 Hz/frame. Secondly, it can also be seen that a second order polynomial provides a smoother relationship between the polynomial coefficients and the non-stationary measures. Thirdly, these measures are not independent of each other as has been assumed in previous applications of PDA for non-stationary analysis [10, 11]. An intuitive explanation for the latter observation is that amplitude change effectively changes the shape of the analysis window and hence alters this relationship, since the relationship between m and Δf is dependent upon the type of window used.

Figure 3 shows the relationship between the RDA measure c and various combinations of values of ΔA and Δf . As for Δf and m , if the amplitude is decreasing then the sign of c changes but its magnitude is the same. From this figure it can be seen that c is monotonically increasing for all values of ΔA , although the relationship is not linear, as has been assumed in previous work, if the range of amplitude change values is wide enough to account for full onset or offset of a component within a single frame (≈ 96 dB in a 16 bit system).

Large values of Δf do not cause a significant change in the effective window shape and so the influence of this parameter over the relationship between c and ΔA is not as great as the opposite situation depicted in Figure 2.

The data presented in Figures 2 and 3 indicates that if highly non-stationary sinusoids are to be accurately quantified the assumption of independence for the two RDA measures is no longer valid. The relationship between Δf and m is affected by ΔA and the relationship between ΔA and c is affected, to a lesser extent, by Δf . We deal with this by using iterative 2D table look-up. Two modestly sized (100 by 100 element) arrays are filled with the data obtained for Figures 2 and 3. Small arrays and linear interpolation can be used since the functions they represent are smooth. The range of values for Δf is chosen to be that over which m is monotonically increasing and the range of values for ΔA is chosen as

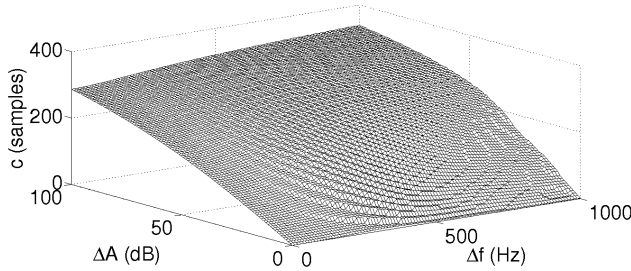


Figure 3: c versus ΔA and Δf , second order polynomial.

the largest range of values that can be represented in a linear 16 bit system. These arrays are then used to look up values of ΔA and Δf using the values of m and c and the current estimates of ΔA and Δf (which are assumed to be zero if no estimate is yet available, *i.e.* we are at the first iteration). The steps of the algorithm are as follows:

1. Obtain values for m and c by fitting a second order polynomial to the time reassignment offset data for the spectral component of interest.
2. Estimate ΔA from the amplitude change array, assuming that Δf is 0 Hz since changes in Δf have a smaller effect on c than those in ΔA have on m .
3. Estimate Δf from the frequency change array, assuming that ΔA is the value estimated in the previous step.
4. Estimate ΔA from the amplitude array, assuming that the value of Δf is that obtained in the previous step.
5. Repeat steps 3 and 4 until the algorithm is terminated (see below).

The termination point may be determined by the processing power available (particularly in a real-time context), the required accuracy of the estimates or the number of iterations before the final estimates of Δf and ΔA are no longer improved by repeated steps but begin to oscillate either side of their correct values. Increasing the number of iterations beyond 3 and taking the resulting single estimate does not improve the accuracy of the method. However, a small improvement over the accuracy obtained with 3 steps can be achieved by taking the mean of the estimates after 3 and 4 steps.

Figure 4 shows the percentage error in the estimation of Δf where $\Delta A = 90$ dB for existing applications of PDA and that for the iterative 2D interpolated RDA method described here. Whereas the estimation error for the former method rises from 58% to almost 80% as Δf increases from 0 to 260 Hz, the error for the new method is close to zero between 0 and 150 Hz, only rising above 10% for 220 Hz or greater with a maximum error of 20% at 260 Hz.

Figure 5 shows the percentage error in the estimation of ΔA where $\Delta f = 250$ Hz for the two methods. A logarithmic scale has been used for the error since there is such a large difference between the two methods. Even for such a large change in frequency within a single bin the error for the iterative method is never greater than 1%.

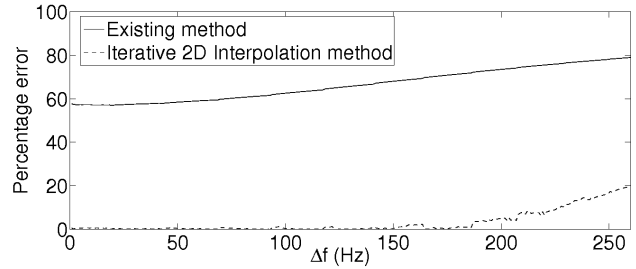


Figure 4: Δf estimation error for $\Delta A = 90$ dB.

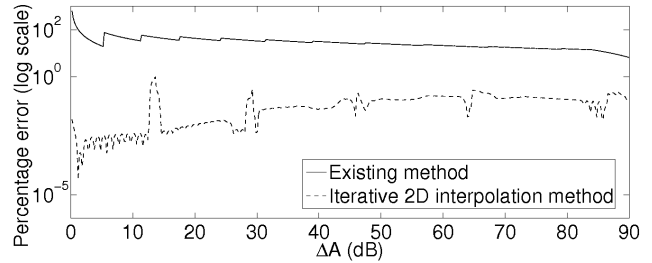


Figure 5: ΔA estimation error for $\Delta f = 250$ Hz.

4. FREQUENCY AND AMPLITUDE ESTIMATION

Like most DFT based frequency estimators, reassignment gives an estimate of the mean instantaneous frequency of a component across an entire analysis frame. This mean is amplitude weighted which allows the use of a windowing function to bias the estimate towards the instantaneous frequencies which occur nearest the centre of the frame. Where the amplitude of the sinusoid is constant throughout the frame no other biasing will occur. However where there are amplitude and instantaneous frequency non-stationarities this will affect the mean frequency estimate. We refer to this estimate as \bar{f}_{amp} , the amplitude-weighted mean instantaneous frequency. To fully separate the amplitude and frequency functions knowledge of the non amplitude-weighted mean instantaneous frequency, \bar{f} , is required. Previous non-stationary sinusoidal models simply use \bar{f}_{amp} but in the presence of large amplitude and frequency changes, such as at the onset and offset of sounds, large errors in the frequency estimation will result. Here we propose a method to correct this bias using the estimates of \bar{f}_{amp} , obtained directly from frequency reassignment, and ΔA and Δf obtained using the RDA method described in the previous section.

Taking the continuous case of a non-stationary sinusoid, as in Equation (1), with the parameters f , ΔA and Δf , where ΔA is given in dB and is assumed to be exponential and Δf is assumed to be linear, the sinusoid has the following amplitude and frequency functions:

$$f(t) = \bar{f} + \frac{\Delta f t}{2} \quad (8)$$

$$A(t) = 10^{at} \bar{A} \quad (9)$$

where

$$a = \frac{\Delta A}{40} \quad (10)$$

and t is in the range -1 to 1 (chosen to simplify the following integration). For a Hann window the amplitude weighted mean

instantaneous frequency is given by:

$$\bar{f}_{amp} = \frac{\frac{1}{2} \int_{-1}^1 10^{at} \bar{A} \left(\bar{f} + \frac{\Delta \bar{f} t}{2} \right) \left(\frac{1}{2} + \frac{1}{2} \cos(\pi t) \right) dt}{\frac{1}{2} \int_{-1}^1 10^{at} \bar{A} \left(\frac{1}{2} + \frac{1}{2} \cos(\pi t) \right) dt} \quad (11)$$

The \bar{A} in the numerator and the denominator of (11) cancel out so solving the integrals and rearranging to find \bar{f} gives:

$$\bar{f} = \bar{f}_{amp} - \frac{\frac{\Delta f}{\kappa^2} [a(\kappa-1) + \frac{1}{a}(\kappa+1)]}{K} - \frac{\left[\frac{\Delta f}{(\pi^2 + \kappa^2)^2} \right] \left[\frac{1}{a} (\pi^2 - \pi^2 \kappa - \kappa^2 - \kappa^3) - a (\pi^2 + \pi^2 \kappa - \kappa^2 + \kappa^3) \right]}{K} \quad (12)$$

where

$$\kappa = \ln(a) \quad (13)$$

and

$$K = \left(a - \frac{1}{a} \right) \left(\frac{1}{\kappa} - \frac{\kappa}{\pi^2 + \kappa^2} \right) \quad (14)$$

Using this equation to improve the estimate of \bar{f} gives a significant improvement in the model accuracy for highly non-stationary components. This is illustrated by comparing figures (6) and (7) which show the error in estimating \bar{f} for different values of ΔA and Δf with and without this bias correction. For both figures estimates of ΔA and Δf obtained using the methods described in section 3, rather than the actual values used to synthesize the sinusoids, were used in the bias correction. Again, the data was derived from an 8 times zero-padded FFT of a 1025 sample frame.

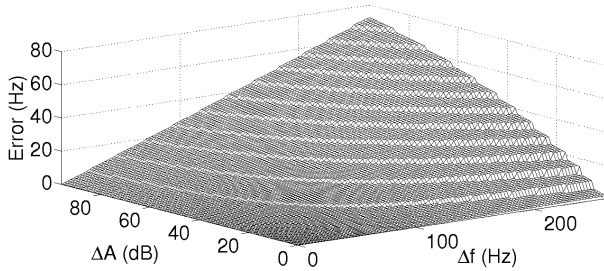


Figure 6: Amplitude biased frequency estimation error.

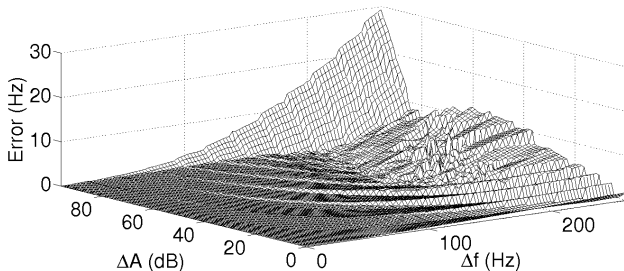


Figure 7: Frequency estimation error after correction.

Mean amplitude estimation is also affected by non-stationarity: frequency change within a frame causes greater spreading of signal energy across bins around a peak, which lowers the magnitude of the peak bin, and amplitude change causes the signal to

be more localised in time thus widening the main lobe. In addition to this the peak magnitude varies with the difference between \bar{f}_{amp} and the actual centre frequency of the bin in which the peak occurs. For stationary sinusoids, knowledge of the magnitude of the window function in the frequency domain allows amplitude estimation errors caused by the deviation from the centre of the analysis bin to be corrected [12]. Since no analytical solution for a Hann windowed sinusoid with non-stationary frequency is known it has been proposed to calculate the magnitude spectrum of the window for each component via FFT. From this an ‘amplitude factor’ is derived which is multiplied by the initial estimate of \bar{A} (the magnitude of the peak bin) [10]. As previously discussed, such an approach is likely to be prohibitively expensive in a real-time context.

Two new approaches are presented in this paper which do not require additional FFTs to be calculated: estimation of the amplitude correction factor by 2D array look-up (as described in the previous section for estimating ΔA and Δf) and modelling of the relationship between the amplitude correction factor, ΔA and Δf with two polynomials. For the first approach a 100 by 100 element array and linear interpolation is used (as described in section 3). The values for the array are calculated by inverting the normalised magnitude values obtained for sinusoids with \bar{f} which coincides with the centre of an analysis bin for the same range of values of ΔA and Δf . The second approach models the inverted normalised magnitude values with quartic polynomials. The non-stationary amplitude correction factor, α , is then estimated by:

$$\alpha = g(\Delta A) \cdot h(\Delta A) \quad (15)$$

where $g(x)$ and $h(x)$ are the quartic functions. Figures 8 and 9 show the data obtained and the best least squares fit provided by the quartic functions. For both figures an 8 times zero-padded FFT of a 1025 sample frame was used. For Figure 8, $\Delta A = 0$ dB, and for Figure 9, $\Delta f = 0$ Hz.

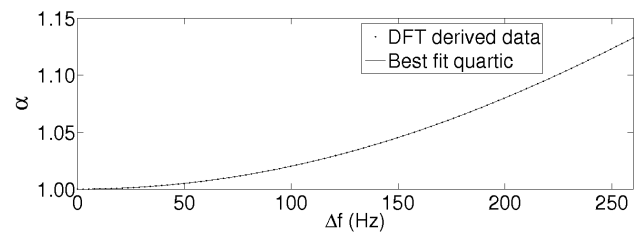


Figure 8: Amplitude correction factor versus Δf .

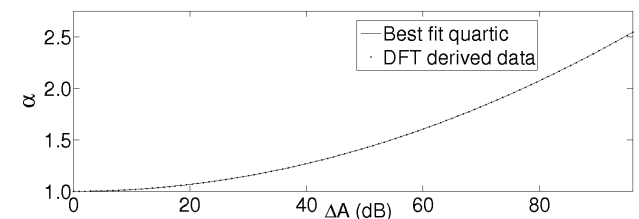


Figure 9: Amplitude correction factor versus ΔA .

Figures 10 and 11 show the percentage error in amplitude estimation for non-stationary sinusoids, whose \bar{f} is at the centre of an

analysis bin, for the 2D lookup and polynomial fit methods respectively. Again, for both figures estimates of ΔA and Δf obtained using the RDA method rather than the actual values used to synthesize the sinusoids, were used in the bias correction. The error without any correction is greater than 75% for $\Delta A = 96$ dB and $\Delta f = 260$ Hz and both methods offer a significant improvement over this. The array lookup performs best out of the two methods, indicating that the effects of amplitude and frequency stationarity are not entirely independent of each other. The first method is the one used in our system but the second may be useful in a system where memory is scarce or where memory lookup is a relatively expensive operation.

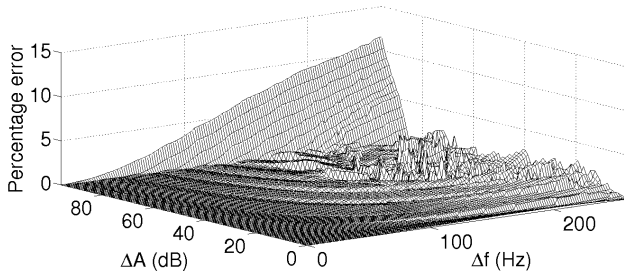


Figure 10: \bar{A} estimation error using 2D array lookup.

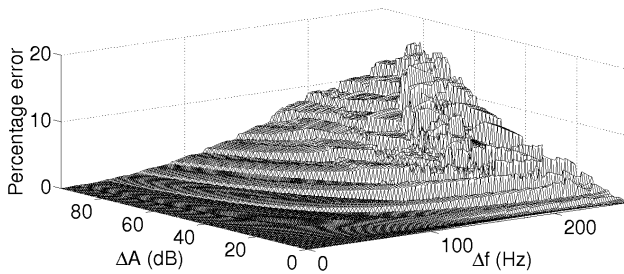


Figure 11: \bar{A} estimation error using polynomial product.

Once the corrected amplitude estimate has been obtained, the deviation of \bar{f}_{amp} from the centre of the analysis bin can be used to produce a further correction as is performed for stationary sinusoids with knowledge of the window shape [12]. However for $8\times$ zero-padding, as used here, the amplitude estimation difference between a frequency offset of zero and the maximum of half an analysis bin is negligible (< 0.03 dB) and so this step can be omitted.

5. FRAME-BY-FRAME SPECTRAL MODELLING

The prime motivation for the research presented in this paper is the development of a real-time spectral modelling system. Our definition of real-time in this context is:

1. Quasi instantaneous: as close to instantaneous as is allowed by the frame size of the algorithm.
2. Frame-by-frame: this is implied in point 1. Only the current and/or previous frames may be used, waiting for future frames is not permitted.

3. Real-time execution: the execution time of the algorithm must be shorter than the time taken to acquire/replay the data it analyses/produces.

We have developed a frame-by-frame spectral modelling system which uses sinusoids to model the deterministic part of monophonic signals and a bank of parametric equalisers applied to a noise source to model the residual. Complex wavelet analysis is used to determine the centre frequency, bandwidth and gain of the equalisers. Since both signal types are synthesized in the time domain the model can be interacted with for sound transformation on a sample-by-sample basis. The separation of sinusoidal and residual parts of the signal is performed by measuring the goodness of fit of time reassignment offset data around magnitude peaks in the spectrum to the second order polynomial used to produce the RDA data. The benefit of the high accuracy description of sinusoids presented here is that they can be tracked more accurately within a single frame, and across frames, in a real-time system. More accurate modelling of amplitude and frequency trajectories within each frame minimises discontinuities in amplitude and frequency between frames. This is shown for frequency in Figure 12 which shows the partial tracks generated by the system for a synthetic harmonic signal with fast and deep vibrato. This signal is chosen to demonstrate that this method produces accurate frequency tracks even in the presence of highly non-stationary components. In this example the analysis frame length is 513 samples with a window overlap of 2, therefore the synthesis frame length is 256. The system is able to produce such partial tracking with no knowledge of previous or subsequent analysis frames.

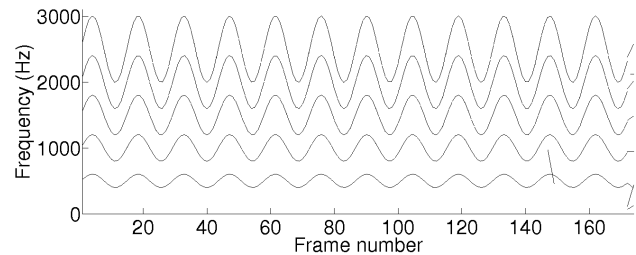


Figure 12: Partial tracks of harmonic sound with vibrato.

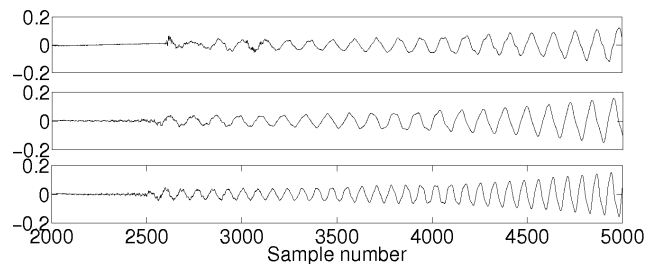


Figure 13: Flute onset: original (top), resynthesized (middle) and pitch shifted up by a perfect fifth (bottom).

Figure 13 shows time domain waveforms of the onset of a flute note, its unmodified resynthesized version and the resynthesized shifted in pitch by a perfect fifth (ratio of 3:2). In this example the analysis frame length is 1025 with an overlap of 2, giving a

synthesis frame size of 512. Even with a relatively long analysis frame the temporal envelope of the signal is largely retained, although the onset is not quite as sharp as in the original signal. With the signal modelled in such a way pitch shifting is a trivial operation: the values for \bar{f} are simply multiplied by the pitch ratio prior to resynthesis. For the residual the centre frequencies of the parametric equalisers are also scaled by the same amount. Of course, the spectral envelope can be preserved by frequency-domain interpolation of amplitudes, if required. The system has been implemented in Matlab as a combination of 'm' and MEX files. The sinusoidal analysis, modelling, discrimination and synthesis is executed in faster-than-real-time for all input sound types and the combined sinusoidal and residual modelling takes less than twice real-time when run on a modest general purpose PC. This suggests that a real-time spectral modelling system based on these methods, written entirely in a low level language and/or running on specialised hardware, can be realised.

6. CONCLUSION

A frame-by-frame sinusoidal analysis system which offers high accuracy estimates of the intra-frame change of amplitude and frequency using time reassignment data has been presented. These estimates can, in turn, be used to reduce errors in the estimation of the means of the amplitude and frequency functions. The high accuracy sinusoidal model that these techniques yield can be implemented with much smaller discontinuities in amplitude and frequency trajectories across frames than would otherwise be possible in such a frame-by-frame system. A detailed description and assessment of the sinusoidal discrimination and residual modelling methods used will be the subject of future papers. Further work will investigate how the bin phase affects estimation of the parameters considered in this paper and whether higher-order polynomial modelling of the time reassignment data could improve parameter estimates.

7. REFERENCES

- [1] X. Serra, "Spectral modeling synthesis: Past and present, keynote in *Proc. Int. Conf. on Digital Audio Effects (DAFx-03)*, London, UK. [online] <http://www.iaa.upf.es/~xserra/presentacions/Spectral-Modeling-Synthesis-Past-and-Present.pdf>, 2003," sep 2003.
- [2] —, "A system for sound analysis/transformation/synthesis based on a deterministic plus stochastic decomposition," Ph.D. dissertation, Stanford University, USA, 1989.
- [3] V. Lazzarini, J. Timoney, and T. Lysaght, "Alternative analysis-synthesis approaches for timescale, frequency and other transformations of musical signals," in *Proc. Int. Conf. on Digital Audio Effects (DAFx-05)*, Madrid, Spain, 2005, pp. 18–23.
- [4] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 34, no. 4, pp. 744–754, 1986.
- [5] F. Keiler and S. Marchand, "Survey on extraction of sinusoids in stationary sounds," in *Proc. Int. Conf. on Digital Audio Effects (DAFx-02)*, Hamburg, Germany, 2002, pp. 51–58.
- [6] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Trans. Sig. Proc.*, vol. 43, no. 5, pp. 1068–1089, 1995.
- [7] G. Peeters and X. Rodet, "SINOLA: A new analysis/synthesis method using spectrum peak shape distortion, phase and reassigned spectrum," in *Proc. Int. Comp. Music Conf. (ICMC'99)*, Beijing, China, 1999, pp. 153–156.
- [8] A. S. Master, "Nonstationary sinusoidal model frequency parameter estimation via fresnel integral analysis," Master's thesis, Stanford University, USA, 2002.
- [9] M. Abe and J. Smith, "AM/FM rate estimation for time-varying sinusoidal modeling," in *Proc. IEEE Int. Conf. Acoust., Speech, and Sig. Proc. (ICASSP'05)*, Philadelphia, USA, 2005.
- [10] P. Masri, "Computer modelling of sound for transformation and synthesis of musical signals," Ph.D. dissertation, University of Bristol, UK, 1996.
- [11] M. Lagrange, S. Marchand, and J.-B. Rault, "Sinusoidal parameter extraction and component selection in a non-stationary model," in *Proc. Int. Conf. on Digital Audio Effects (DAFx-02)*, Hamburg, Germany, 2002, pp. 59–64.
- [12] M. Desainte-Catherine and S. Marchand, "High precision Fourier analysis of sounds using signal derivatives," *J. Audio Eng. Soc.*, vol. 48, pp. 654–667, 2000.