# METHODS FOR SEPARATION OF AMPLITUDE AND FREQUENCY MODULATION IN FOURIER TRANSFORMED SIGNALS

*Jeremy J. Wells*

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

**ABSTRACT**

This paper describes methods for the removal and/or separation of amplitude and frequency modulation of individual components within a Fourier spectrum. The first proposed method has a relatively low cost and works under assumptions about the behaviour of both the local and non-local magnitude and phase of sinusoidal components for these two forms of component non-stationarity. The second method is more expensive and re-synthesizes components either in the Fourier or time domain following a parameter estimation stage. Typical applications are the adjustment of expressive parameters in music signals and conditioning of signals prior to cross-synthesis.

## 1. INTRODUCTION

The Discrete Fourier Transform (DFT) has been widely used in Computer Music and Audio Processing for many years. Applications range from independent time and pitch modification through cross-synthesis and spectral modification to feature extraction for sound modelling [1]. One of the attractive features of Fourier analysis and processing is that individual narrow-band components, such as stationary sinusoids, are clearly and intuitively represented in the transform domain. An assumption of Fourier analysis (implicit in the choice of basis functions) is that individual components are stationary sinusoids. Under ideal conditions (i.e. a rectangular window applied to a stationary sinusoid whose period of oscillation is an integer multiple of the window length) such components appear as a single delta function in the Fourier domain.

Where signals under transformation contain non-stationary components a common approach is to divide them into shorter analysis frames within which those components can be considered to be quasi-stationary. This is known as the short-time Fourier transform [2]. Longer frames increase frequency resolution but at a cost of temporal resolution and the optimum frame length is often determined to be the point at which the assumption of component stationarity breaks down. A problem here is that useful and interesting audio signals tend to be multi-component with different localisation properties in both time and frequency. This leads to a compromise between time and frequency resolution in which the quasi-stationarity assumption is violated for at least some of the components. Information about non-stationarity is not lost in the Fourier domain (since the transform is perfectly invertible) but it is embedded in the relationships between the phase and magnitude of multiple transform bins, rather than being more directly accessible [3]. For Fourier domain processing of such signals which requires separation of, or interaction with, such non-stationarities, these phase and magnitude relationships must be identified and interacted with.

The work described in this paper addresses signals which contain intra-frame non-stationarities. Its aim is to enable the identification of amplitude and/or frequency change of individual signal components and to selectively remove either or both of them. This is either done completely in the Fourier domain or partly in the Fourier and then the time domain. The methods exploit the differences in the phase and magnitude characteristics for stationary, amplitude modulated and frequency modulated sinusoidal components. These differences are described and explored in the next section of this paper. The third section describes the two sets of identification and removal algorithms for both kinds of modulation. Results from the application of the algorithms to different types and combinations of signal components are also presented in this section. A potential application of this process, to polyphonic spectral whitening, is described in Section 4. Finally, conclusions are presented in Section 5.

## 2. FOURIER REPRESENTATIONS OF NON-STATIONARITY

As stated in the previous section, the DFT offers the most compact representation of a stationary sinusoid when its frequency is harmonically related to the analysis frame length. Where this is not the case, discontinuous phase in the time domain will cause spectral leakage into analysis bins in the Fourier domain other than the one in which the sinusoid is centred. The extent of this leakage can be controlled by the use of tapered windows. These reduce or eliminate abrupt phase changes but do so at the expense of the component width: even where the component period and frame length have an integer relationship some energy will exist in bins adjacent to the centre bin. In fact, this energy spreading in the Fourier domain is due to amplitude non-stationarity introduced by the windowing process.

Different types of windows offer different trade-offs between the local (main-lobe) width of the component and the amount of non-local (side-lobe) leakage. Windowing of signals has been the subject of extensive research and discussion (e.g. [4]). For the rest of this paper the example used is the Hann (or raised-cosine) window. However what is presented and discussed can be generally applied to any symmetrical taper, the important distinctions are between local and non-local and phase and magnitude behaviour, whatever the window being used.

The form of amplitude non-stationarity assumed is exponential, either increasing or decreasing. The form of frequency

modulation is linear increase or decrease (chirping) which gives rise to quadratic phase trajectories in the time domain.

### 2.1. Amplitude modulation in the Fourier domain

Exponential intra-frame amplitude change can be interpreted as a change in the window applied to an amplitude-stationary signal. Considering the continuous case, this modified window is described by (adapting equation (3) in [4]) as a function of time *t* by:

$$w(t) = \frac{e^{\alpha t}}{L}\left(\frac{1}{2} + \frac{\cos(2\pi t/L)}{2}\right), |t| \le \frac{L}{2} \tag{1}$$

where *t* is time in seconds, *L* is the window duration and $\alpha$ is the intra-frame amplitude change in Nepers (Np):

$$\alpha = \frac{\Delta A \ln(10)}{20} \tag{2}$$

where $\Delta A$ is intra-frame amplitude change in dB. In the following equations $L = 1$, since this makes the presentation more compact but does not sacrifice generality. With this value of *L* the Fourier transform of this window as a function of frequency is given by:

$$\begin{aligned} W(f) &= \int_{-1/2}^{1/2} e^{\alpha t}\left(\frac{1}{2} + \frac{\cos(2\pi t)}{2}\right) e^{-j2\pi f}\, dt \\ &= \frac{4\pi^2 \sinh\left(\frac{1}{2}(\alpha - j2\pi f)\right)}{(\alpha - j2\pi f)\left(\alpha^2 - j4\alpha\pi f - 4\pi^2\left(f^2 - 1\right)\right)} \end{aligned} \tag{3}$$

From this it can be shown that the magnitude response of the window function is given by [5]:

$$|W(f)|^2 = \frac{8\pi^4\left(\cosh(\alpha) - \cos(2\pi f)\right)}{\alpha^6 + 4\alpha^4\left(2 + 3f^2\right)\pi^2 + 16\alpha^2\left(1 + 3f^4\right)\pi^4 + 64f^2\left(f^2 - 1\right)^2\pi^6} \tag{4}$$

The phase response (the arctangent of the ratio of the imaginary and real parts of equation (3)) does not reduce to quite such a compact expression. However its first derivative at $f$=0 does, which provides useful information about the phase behaviour around a peak in the Fourier spectrum. This first derivative is given by:

$$\left.\frac{d(\arg(W))}{df}\right|_{f=0} = \pi\left(\frac{2}{\alpha} + \frac{4\alpha}{\alpha^2 + 4\pi^2} - \coth\left(\frac{\alpha}{2}\right)\right) \tag{5}$$

This is in fact an analytical derivation of the amplitude modulation estimator empirically described in [3] and used subsequently in, for example, [6]. To demonstrate this, Figure 1 shows the first-order phase difference plotted against the continuous phase derivative for a sinusoid whose frequency is exactly at the centre of an analysis bin. The slight difference in the plotted values for the non-zero padded Fourier spectrum is due to the fact that the phase derivative is not constant around the peak and so the first-order difference is not exactly equivalent to the actual derivative. The important fact to note here is that, taking the peak as the origin, the local phase is an odd function where there is intra-frame amplitude change and it is 0 where the component has stationary amplitude.

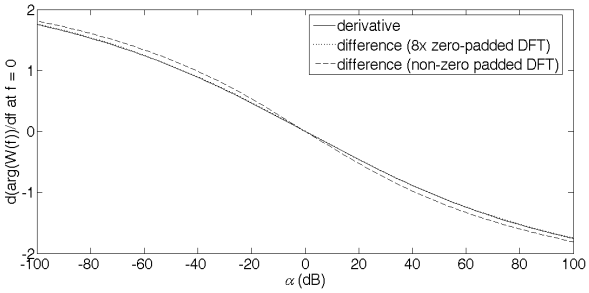



Figure 1: *Derivative and difference (Masri phase distortion estimator for amplitude change) of the phase at f=0.*

Figure 2 shows the magnitude response of the amplitude-modified window, calculated using equation (4), for different values of $\Delta A$. It can be seen that much of the energy spreading is into non-local bins and the main lobe (i.e. the local magnitude) remains quite similar to that for a non amplitude-modified window. Therefore a simple rule-of-thumb for amplitude modulation is that the non-local magnitude increases relative to the local magnitude, and the local phase is an odd function.

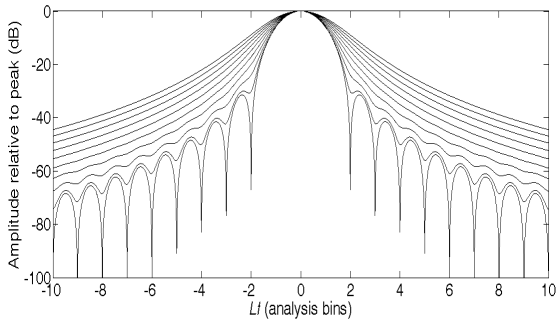



Figure 2: *Normalised magnitude response of the Hann window multiplied by an exponentially changing amplitude function. The amplitude change is in 10 dB increments from 0 to 90 dB.*

### 2.2. Frequency modulation in the Fourier domain

It is not possible to directly derive expressions for components of non-stationary frequency in the Fourier domain, except where the window function is a Gaussian [7]. This means that there are no analytic equivalents to equations (4) and (5) for frequency modulation. However it has been shown empirically [3] and analytically [8, 9] that the phase is concave at a peak in the Fourier spectrum due to a linearly chirping component. In [9] the first derivative of the phase at the peak is shown to be 0 and the magnitude of the second derivative is shown to be inversely proportional to the chirp rate. (However the assumed approximations in that work do not hold for low chirp rates and aliasing occurs, which is discussed further below). Therefore, for frequency modulation, the phase is an even function around the peak.

Figure 3 shows the magnitude response of the Fourier transform of the Hann window applied to sinusoidal components with different chirp rates. The rates are integer multiples of *Lf* from 0 to 5*Lf*. Here it can be seen that the energy spreading is more local than is the case for amplitude modulation. For a high chirp rate the half-height of the magnitude response is approximately proportional to the chirp rate [8]. For a sampling rate of 44.1 kHz and a 1024 point DFT, a chirp rate of 5*Lf* corresponds to 215 Hz per frame (almost 10 kHz/s). For chirp rates greater than about 6 *Lf* the second order difference of the phase begins to alias [3, 8]. This is illustrated in Figure 4 where the magnitude

and phase are shown for sinusoids each with different chirp rates. Each has the same second order phase difference around the peak but quite different magnitude responses and it can be seen that the phase curvature across a greater range of bins is different too. Note that where the chirp rate is negative the magnitude response remains the same but the phase response is inverted.

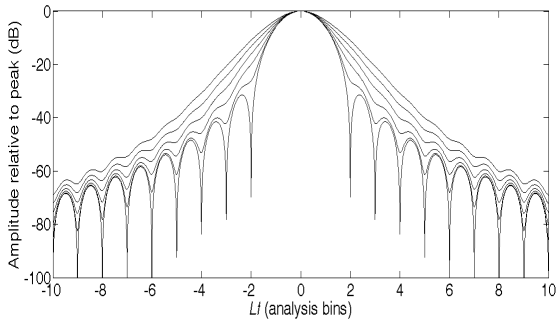



Figure 3: *Normalised magnitude response of Hann windowed linear chirp. The chirp rate is in integer Lf increments from 0 to 5Lf.*

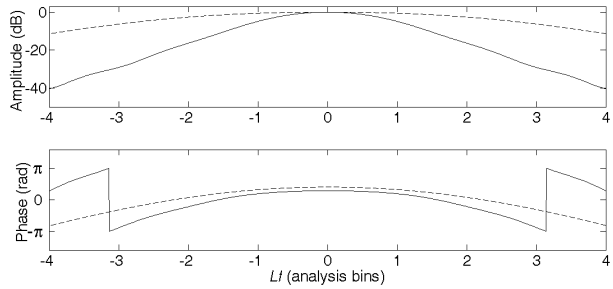



Figure 4: *Magnitude and phase responses of Hann windowed linear chirps. The chirp rates are 2.9 Lf (solid lines) and 11.7Lf (dashed lines).*

### 2.3. Interdependence of amplitude and frequency modulation estimators

The previous two sub-sections have shown how frequency and amplitude modulation affect both the magnitude and frequency response of a windowed component in different ways. This suggests that it is possible to separate the effects of amplitude change from those of frequency change and vice versa. Whilst this is the case for mild non-stationarities, this is not so where the intra-frame changes are more extreme. Large amplitude modulation has a considerable effect on phase (or reassignment) based estimators of frequency change, since it drastically alters the effective window shape. An approach to improving the independence of such estimators using recursive 2D lookup was described in [10].

## 3. ALGORITHMS FOR SEPARATION OF MODULATION TYPES

In the previous section the Fourier domain behaviour of Hann windowed linear chirps and exponential amplitude change was considered. This section outlines two different methods for separating these two kinds of modulation. First a simple, low-cost method is described, then, in the next sub-section, a more sophisticated but costly approach is presented.

### 3.1. Removal of intra-frame amplitude and frequency change by spectral modification

The effectiveness of some audio processing applications, such as cross-synthesis, can be improved by the separation of amplitude from frequency information. The following algorithm is designed to remove intra-frame amplitude change or intra-frame frequency change from components within a Fourier spectrum. The input to the DFT must be zero-phase windowed otherwise stationary sinusoids will have a linear, rather than flat, phase. For the rest of this sub-section it is assumed that the analysis frame is *not* zero-padded prior to the DFT. The algorithm is based on the following assumptions:

1. A component with intra-frame amplitude change is represented by phase which is an odd function and by non-local magnitude that decays much more slowly than that for a stationary component.
2. A component with intra-frame frequency change is represented by phase which is an even function and by a more local change in magnitude.

#### *3.1.1. Initial stages of algorithm*

These assumptions, although crude (particularly where there is a high degree of non-stationarity), do lead to a reasonably effective method for eliminating either frequency or amplitude change, or both. The following steps are common to both amplitude and frequency removal:

1. Identify individual components in the spectrum. A single component is classified as the region between two magnitude minima within which the magnitude is either monotonically increasing or decreasing.
2. Estimate the exact centre of component within the peak bin. Various methods exist for this which are both phase-based (e.g. frequency reassignment [13]) or magnitude based (e.g. parabolic interpolation) [11]. Phase-based methods are generally more accurate but parabolic interpolation is used here for its relatively low computational cost.
3. Fit a second-order polynomial to the local unwrapped phase, treating the position estimated in step 2 as the origin. The definition of local is the width of the main lobe of the window for a stationary signal. For a non zero-padded DFT of a Hann window this taken as being the peak magnitude value and the three highest and nearest neighbours.

#### *3.1.2. Removal of intra-frame amplitude change*

4. Set the slope (first order coefficient) of the phase polynomial to 0, this will ensure that the local phase is an even function.
5. Set the non-local magnitudes (i.e. those bins that are between the two minima, but outside of the four centre bins) to those of a stationary sinusoid. This is done using equation (4) with $\alpha = 0$ and the scaling the result by the ratio of the actual to the synthesized peak magnitude value.

#### *3.1.3. Removal of intra-frame frequency change*

6. Set the curvature (second order coefficient) of the phase polynomial to 0, this will ensure that the local phase is an odd function.
7. Set the local magnitudes to those of a stationary sinusoid, using equation (2) and scaling to the magnitude of the actual peak.

*3.1.4. Removal of both intra-frame amplitude and frequency change*

8. Set the slope and the curvature of the local phase to 0.
9. Synthesize all magnitude values (local and non-local) using equation (4) and by scaling so that actual and synthesized peak magnitude values match.

*3.1.5. Examples of modulation separation for single and multi-component signals*

To further illustrate the procedure, Figure 5 shows the modifications made to the phase and magnitude of a single sinusoid at a frequency of 1 kHz which undergoes a 48 dB increase in amplitude during a single analysis frame. Figure 6 compares the time-domain output with the windowed input (after the zero-phase windowing has been undone). The amplitude increase has been removed and the shape of the Hann window has been largely restored. An artefact of the process is that there has been a small leftwards circular shift in the overall window shape, but not in the phase of the underlying component. For an amplitude decrease of the same amount there is an equal sized shift but in the opposite direction.

Of course, this kind of correction can be done quite easily for single components by applying the inverse exponential function in the time – where this algorithm is of interest is in the independent correction of multiple components. Figure 7 shows the output for components at 1 and 2 kHz with -48 and +48 dB changes in amplitude respectively. As can be seen the output is similar in shape to the Hann window with the two frequency components preserved (as can be seen by the different oscillation rates at the start and end of the window) but, again, with a small circular shift.

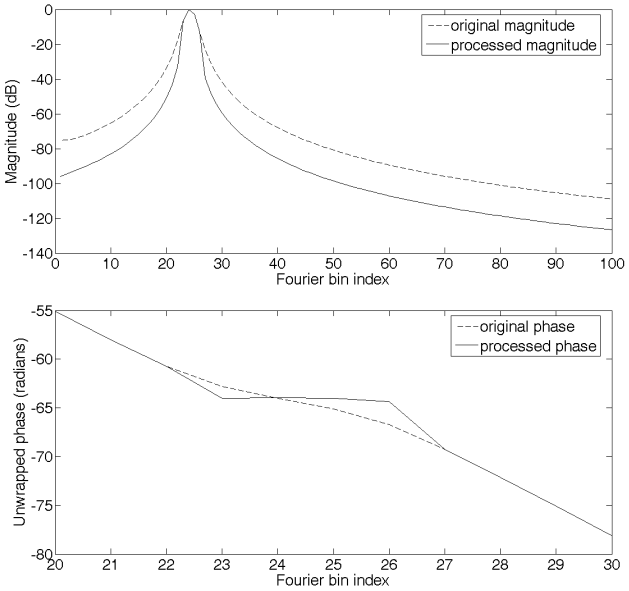



Figure 5: *Original and processed magnitude and phase responses for a single component at 1 kHz with 48 dB exponential amplitude change. The sample rate is 44.1 kHz and the input frame length is 1024.*

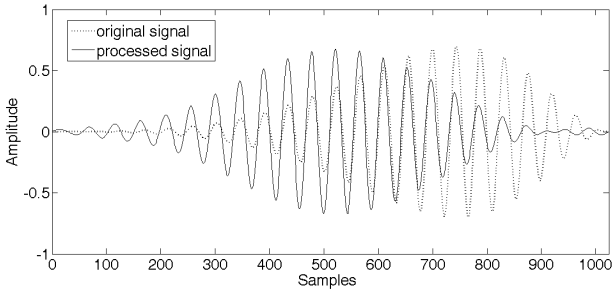



Figure 6: *Original and processed time domain signals.*

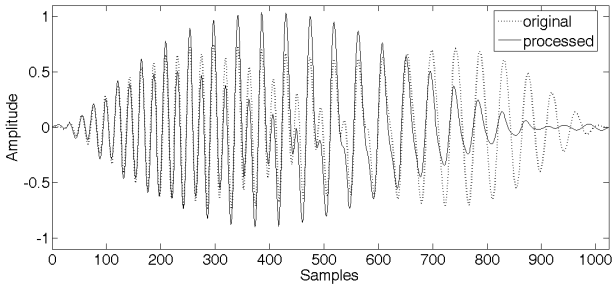
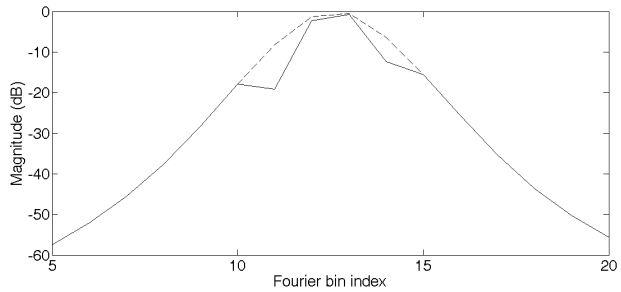



Figure 7: *Original and processed time domain signals for an input signal comprising two components at 1 kHz and 2 kHz with 48 dB falling and rising amplitude.*

Next the removal of frequency non-stationarity is considered. Figure 8 shows the original and processed magnitude and phase spectra of a component whose frequency changes linearly from 400 Hz to 600 Hz during a single frame. Figure 9 compares the input and output via the Hilbert transform. The top panel shows the instantaneous frequencies, derived from the first-order difference of the phase of the analytic signals. The bottom panel shows the amplitudes of the analytic signals. The linear frequency increase for the input can be clearly seen (the errors at the start and end of the frame are due to the significant tapering at extremes of the Hann window). During the centre of the frame the frequency trajectory is much flatter in the output however it is not perfectly constant and nearer the frame edges there is significant variation in the instantaneous frequency. The amplitude plot shows that the shape of the Hann window is largely, but not perfectly, preserved in the output. As for the examples of amplitude change removal, a circular shift is evident in both the amplitude and instantaneous frequency of the output.

Many spectral processing methods re-window the signal after re-synthesis by inverse DFT, in order to avoid discontinuities at frame boundaries. This re-windowing should be applied for this method if artefacts due to these circular shifts become audible.



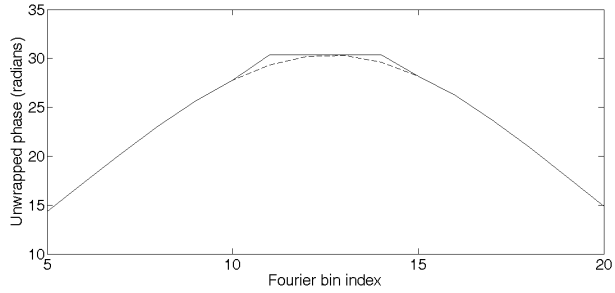



Figure 8: *Original and processed magnitude (above) and phase (previous page) responses for a single sinusoid with linearly increasing frequency from 400 to 600 Hz.*

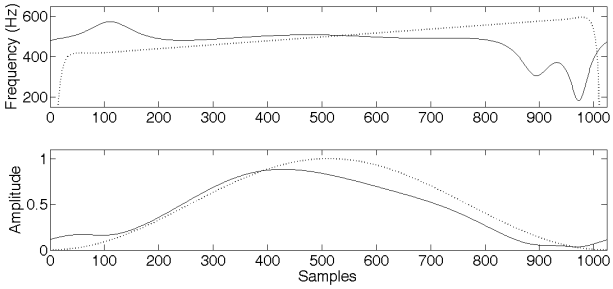



Figure 9: *Instantaneous frequency (top) and amplitude (bottom) of Hilbert transformed input (dotted) and output (solid line) signals.*

### 3.2. Adjustment of intra-frame amplitude and frequency change by analysis and re-synthesis

The method described in the previous sub-section is crude but reasonably effective given its computational cost. Higher quality methods for achieving the same goals are described in this section. These algorithms work by analysing and then re-synthesizing each component, either wholly in the Fourier domain or in the Fourier (for analysis) and then in the time (for synthesis) domain. They assume that each component within a single frame can be wholly described as sinusoid with the parameters $A$, $f$, $\varphi$, $\Delta A$ and $\Delta f$:

$$s(t) = A10^{\left(\frac{\Delta A}{20L}\right)t} \sin\left(\phi + 2\pi\left(ft + \frac{\Delta f t^2}{L}\right)\right), t \le \left|\frac{L}{2}\right| \quad (6)$$

Since the model is much more sophisticated than that described in 3.1, adjustment of $\Delta A$ and $\Delta f$ rather than just elimination of one, the other or both is possible.

The estimation of parameters uses methods described in [10] and [12]. These methods provide highly accurate estimates of the parameters $A$, $f$, $\varphi$, $\Delta A$ and $\Delta f$. Additionally here, a similar approach (interpolated 2D table look-up) is taken to the estimation of $\varphi$. This is in order to reduce biasing, by amplitude and frequency non-stationarities, of the value derived directly from Fourier analysis. The table used for this phase correction is shown in Figure 10.

These methods require a zero-padded Fourier spectrum for accurate estimation and in this section the analysis frame is 1025 samples, zero-padded to 8192. The initial analysis steps of the algorithm described in this sub-section are:

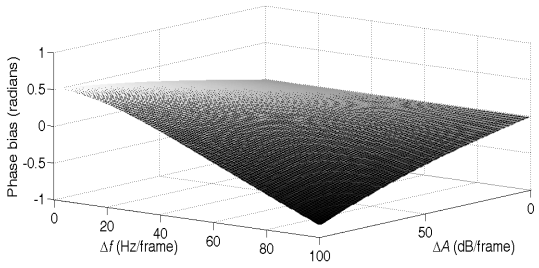



Figure 10: *Bias in phase estimation due to non-stationarity. The frame length is 1025 samples zero-padded to 8192. The sample rate is 44.1 kHz.*

1. Identify individual components in the spectrum. As previously, a single component is classified as the region between two magnitude minima within which magnitude is either increasing or decreasing. Since the spectrum is now zero-padded care must be taken to ensure that local minima due to side lobes are not interpreted as global minima.
2. Frequency reassignment is used to estimate the exact component centre within the peak bin.
3. The parameters of the component are estimated, as described in [10]. As for $f$ and $A$, bias in the estimation of $\varphi$ is corrected, once estimates for $\Delta A$ and $\Delta f$ have been obtained, by the use of the interpolated 100 x 100 2D lookup table shown in Figure 10.

Although in previous work this analysis method has been used in a 'sinusoids + noise' system, here all components are classified as sinusoidal, since the goal here is Fourier-based processing rather than generation of a spectral model (i.e. the resynthesis is overlap-add).

Removal of intra-frame frequency change can be achieved wholly in the Fourier domain, since an analytic representation of $W(f)$ exists where there is only amplitude non-stationarity (equation (3)). However, where there is frequency change then no such solution exists. A large-limits derivation of the local spectrum is given in [8] but is only valid for very large chirp rates, a Taylor series expansion which is even remotely tractable is only valid for low chirp rates and very close to the centre of the main lobe. Thus, synthesis in the Fourier domain of components with frequency change 'from scratch' is not possible. One approach to eliminating $\Delta A$ where there is frequency non-stationarity might be to examine the difference between the Fourier spectrum of the component with $\Delta A$ and $\Delta f$, with the spectrum synthesized just with $\Delta f = 0$. In practice, this is not viable since it would require deconvolution of the two spectra in the Fourier domain which, without perfect parameter estimates for the component (across the whole spectrum – which would only be possible for a single component) would very likely result in instability. The solution is to replace Fourier synthesis followed by inverse DFT with direct synthesis of equation (6) for each component in the time domain. This solution offers considerable flexibility, including independent adjustment as well as simple elimination, but in terms of computational cost is certainly at the other extreme to the methods presented in the previous sub-section. In summary, the removal of frequency change, whilst the values of $\Delta A$ are retained (or adjusted, if required) is achieved in the Fourier domain by:

4. For each component, resynthesize the Fourier spectrum using equation (3), shifting so that the component is centred at $f$ and

normalising the energy so that it is the same as for the component prior to $\Delta f$ removal. It is important to note that the value of $f$ used should not simply be the reassigned frequency (which occurs at the reassigned time) but the value at the centre of the frame (referred to as the non amplitude-weighted mean instantaneous frequency in [10]). Also the phase correction should assume that $\Delta f = 0$ and not the value measured in the analysis (i.e. only $\Delta A$ should be used to correct the phase).

5. Once all components have been resynthesized transform back to the time domain via the inverse DFT.

$\Delta f$ removal by this method is shown in Figure 11. The parameters are the same as those in Figure 9 except $\Delta A$ = 48 dB, rather than 0 dB. Clearly this method is more effective than the one used for Figure 9, since the shape of the amplitude modulated window is perfectly retained and the frequency trajectory is more uniformly flat (except where artefacts of the Hilbert transform are observed due to tapering by the window function). To illustrate the method working independently on two combined components (with parameters $f$ = 500 Hz, $\Delta A$ = 48 dB, $\Delta f$ = 200 Hz/frame and $f$ = 1 kHz, $\Delta A$ = -48 dB, $\Delta f$ = -200 Hz/frame) Figure 12 shows the input and output. Also shown is the sum of the two components synthesized with $\Delta f$ = 0 Hz/frame, but all other parameters the same. It can be seen that the output from the algorithm is indistinguishable from this signal synthesized in the time-domain using *a priori* knowledge of the parameters.

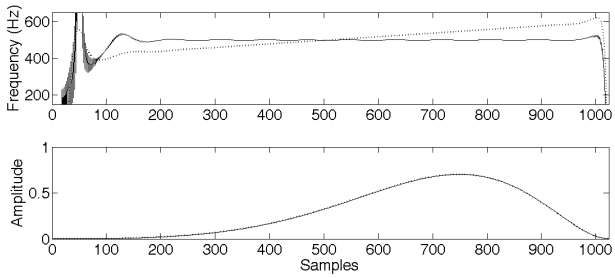



Figure 11: *Instantaneous frequency (top) and amplitude (bottom) of Hilbert transformed input (dotted) and output (solid line) signals.*

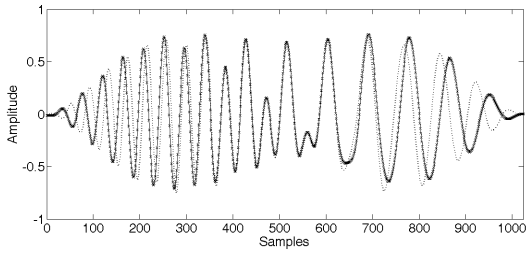



Figure 12: *Original (dotted) and processed (solid line) time domain signals. For comparison the ideal output is also shown (crosses).*

The removal of amplitude change is achieved in the Fourier and time domains by:

6. For each component resynthesize in the time domain using equation (6) with $\Delta A$ = 0, but with all other parameters as estimated in steps 2 and 3.

7. Sum all components and apply Hann window.

Whilst more costly than the Fourier domain method, this time domain synthesis approach can also be used for elimination of $\Delta f$ whilst retaining $\Delta A$. In fact, it offers total flexibility over the independent modification (within the limitations of the parameter estimation) of both of these forms of non-stationarity. This offers the possibility of, for example, increasing vibrato in signals whilst reducing tremolo in others.

Figure 13 shows the Hilbert transformed input and output for a signal with the same parameters as for Figure 11. The amplitude change has been successfully removed, restoring the shape of the Hann window whilst the frequency trajectory has been retained (although close inspection reveals a slight overestimation of $\Delta f$, due to the interdependency of the estimators of this parameter and $\Delta A$). A final example given demonstrates the capacity of this algorithm to handle multiple component signals successfully. Figure 14 shows the inputs (top panels) and outputs (bottom panels) from the algorithm for two frames of Gaussian white noise with $\Delta A$ of 48 dB (left panels) and -96 dB (right panels) respectively. In the 48 dB case 110 components are separately identified and re-synthesized with $\Delta A$ = 0 dB, in the -96 dB case there are 98 components. It can be seen that the recombination of synthesized components has a Hann-like amplitude profile in both cases.

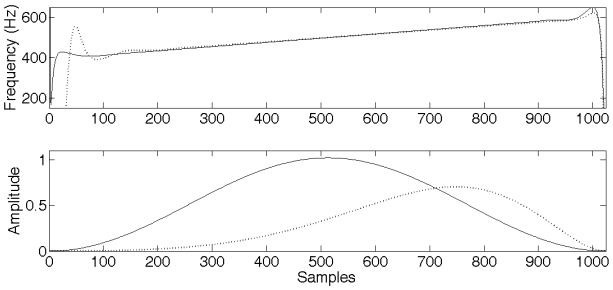



Figure 13: *Instantaneous frequency (top) and amplitude (bottom) of Hilbert transformed input (dotted) and output (solid line) signals.*

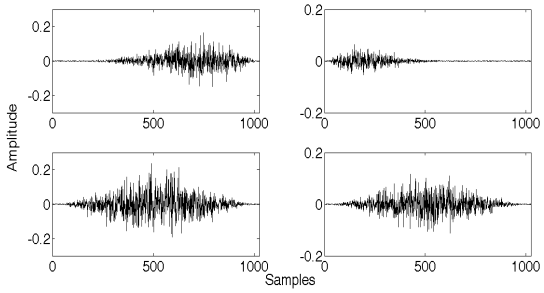



Figure 14: *Input noise with amplitude ramps (top panels, left 48 dB, right -96 dB), output with amplitude ramps removed (bottom panels).*

## 4. APPLICATION TO FREQUENCY SHAPING AND POLYPHONIC SPECTRAL WHITENING

In the previous section two methods for removing either amplitude or frequency change from a single Fourier analysis frame. In this section a related application area for this work is described, inspired by Christopher Penrose' *Shapee* algorithm [14].

### 4.1. Frequency shaping

Many cross-synthesis applications employ the short-time Fourier transform (STFT) with the frame length comparable to the period of the lowest frequency audible by humans (20 Hz, 50 ms). A

reasonable, albeit simplistic assumption, is that the magnitude of the transform data represents the spectral envelope of the signal and the phase represents the exact location of individual frequency components. The most straightforward STFT-based cross synthesis method combines the magnitude from one input signal (the resonator, or the 'formant reference') with the phase from another signal (the excitation, or 'frequency reference') [1]. The output is intended to resemble a perceptual hybrid of the two input sounds. However, good separation between excitation and resonance is not always achieved with such a basic approach. The process of frequency shaping was developed to improve the transfer of frequency information between sounds [15]. It recognises that the frequency content of a signal in the Fourier domain is described by both the magnitude and phase around a peak due to a component. The process divides the Fourier spectrum into 'shaping regions' of equal width from DC to Nyquist. It is the width of these regions which determines how frequency information is transferred between signals. The recommended default width is that of the main lobe of the window function used. For each frame, the hybrid spectrum $X_{\text{hybrid}}$ is calculated according to [14]:

$$X_{\text{hybrid}}(k) = R(k)\left|X_{\text{freq}}(k)\right| e^{j\angle\left(X_{\text{freq}}(k)\right)}, \quad k = 0,1,2..N-1 \tag{7}$$

where $X_{\text{freq}}$ is the Fourier transformed frequency reference, $k$ is the analysis bin, $N$ is the frame length and $R$ is given by:

$$R(k) = \frac{\sum_{n=0}^{w}\left|X_{\text{formant}}\left(\left\lfloor \frac{k}{w} \right\rfloor + w\right)\right|}{\sum_{n=0}^{w}\left|X_{\text{freq}}\left(\left\lfloor \frac{k}{w} \right\rfloor + w\right)\right|}, k = 0,1,2..N-1 \tag{8}$$

Practical implementations, as the processor *Shapee*, are available in various forms, by Penrose and Eric Lyon (Max/MSP object [16]) and by this author (Steinberg VST plug-in and Matlab [17]). The UNIX command line version (part of *PVNation*) is no longer available.

### 4.2. Polyphonic spectral whitening

Implicit in both frequency shaping and the more straightforward combination of magnitude and phase data, is a spectral whitening process. Where magnitude is combined with phase then all of the magnitudes of the phase reference are effectively set to 1, creating a white spectrum. For frequency shaping, the whitening stage of the process is equivalent to equation (8) with the numerator set to 1. Since the process does not require pitch detection (as is the case for some cross-synthesizers based, for example, on linear predictive coding (LPC)) and works on a wide range of harmonic and enharmonic signals, it can be considered a polyphonic whitening process [14]. Considering cross-synthesis as a two stage process: whitening of the frequency reference followed by the application of the spectral envelope of the formant reference, offers more flexibility. For example, a frequency reference that has been whitened by the process described in this section could then be filtered by the infinite impulse response filter derived via LPC.

### 4.3. Application of modulation separation

The aim of frequency shaping is to improve the separation of frequency and spectral magnitude information between two signals that are being cross-synthesized. As for many Fourier-based processes it will be most successful when the signals are stationary during each analysis frame. Where the signals are non-stationary then these amplitude and/or frequency changes are embedded in both the magnitude and phase data of the signals. The algorithms outlined in the previous two sections of this paper are designed to remove one or other of these non-stationarities. By removing the intra-frame amplitude change from the frequency reference and the intra-frame frequency change from the formant reference as pre-processing stage in a cross-synthesis process the separation between amplitude and frequency information will be improved.

The complete elimination of amplitude/frequency change in the formant/frequency references will not always succeed in the separation of frequency information. For example ensembles of acoustic instruments playing in unison will not be perfectly in tune with each other. This combination of very closely spaced partials will produce components that have slow amplitude and frequency change (i.e. that beat). In this case the amplitude modulation *is* a representation of the frequency content of the signal and should not be removed from the frequency reference. This can be avoided by removing intra-frame amplitude change which is above a certain threshold. The final algorithm presented in the previous section offers the possibility of applying this thresholding.

Another consideration is the fact that these processes do not distinguish between 'noisy' and more stable, sinusoidal components. However, it is not clear how a cross-synthesis method should classify noise. Does noise contain information about amplitude or frequency or both? Where a separation between these component types is required methods, such as those surveyed in [18], could employed. Examples of frequency shaping and polyphonic whitening, with and without modulation removal (or suppression) using the methods described in this paper are available online [19].

## 5. CONCLUSIONS

This paper has presented two set of algorithms for the removal of intra-frame amplitude and/or frequency change from Fourier spectra. This is done entirely in the Fourier domain, except for the final algorithm, which uses parameters derived in the Fourier domain for time domain resynthesis. Although costly, this final algorithm offers adjustment, rather than simple removal of non-stationarity, and is highly effective for a wide range of values of $\Delta A$ and $\Delta f$. Matlab code that implements these processes is available online [19].

## 6. REFERENCES

[1] U. Zölzer, Ed., *DAFX – Digital Audio Effects*, J. Wiley & Sons, 2002.

[2] J. Allen, "Short Term Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform", *IEEE Trans. on Acoustics, Speech, and Sig. Proc.*, vol. 25, no. 3, pp. 235-238, June 1977.

[3] P. Masri and A.Bateman, "Identification of Nonstationary Audio Signals Using the FFT, with Application to Analysis-based Synthesis of Sound", in *Proc. of the IEE Colloquium on Audio Engineering*, London, UK, May 1995.

[4] A. Nuttall, "Some Windows with Very Good Sidelobe Behavior", *IEEE Trans. Acoustics, Speech, and Sig. Proc.*, vol. 29, no. 1, February 1981, pp. 84-91.
[5] Derivation provided at http://www.jezwells.org/research/dafx10/DAFx10_mathematica.pdf
[6] M. Lagrange et al., "Sinusoidal Parameter Extraction and Component Selection in a Non-Stationary Model", *Proc. of the Int. Conf. on Digital Audio Effects (DAFx-02)*, pp.59-64, 2002.
[7] M. Abe and J. Smith, "AM/FM Rate Estimation for Time-Varying Sinusoidal Modeling", *Proc. of the 2005 IEEE Conference on Acoustics, Speech and Sig.Proc.*
[8] A. Master, "Nonstationary Sinusoidal Model Frequency Parameter Estimation via Fresnel Integral Analysis", Technical Report, Stanford University, 2002.
[9] Y. Liu and A. Master, "Phase of a Continuous Time Linear-Frequency Chirp Signal: Analysis and Application", Research Note, CCRMA, Stanford University, October 2002.
[10] J. Wells and D. Murphy, "High Accuracy Frame-by-Frame Non-Stationary Sinusoidal Modelling", in *Proc. Digital Audio Effects (DAFx'06)*, Montreal, Canada, Sep. 2006, pp. 253-258.
[11] F. Keiler and S. Marchand, "Survey on Extraction of Sinusoids in Stationary Sounds", in *Proc. Digital Audio Effects (DAFx'02)*, Hamburg, Germany, Sep. 2002, pp. 51-58.
[12] J. Wells, "Real-Time Spectral Modeling of Audio for Creative Sound Transformation", PhD Thesis, Department of Electronics, University of York, 2006. Available: http://www.jezwells.org/Jez_Wells_PhD.pdf
[13] F. Auger and P. Flandrin, "Improving the Readability of Time-Frequency and Time-Scale Representations by the Reassignment Method", *IEEE Trans. on Sig. Proc.*, vol. 43, pp.1068-1089, May 1995.
[14] C. Penrose, "Frequency Shaping of Audio Signals", in *Proc. Intl. Computer Music Conf.*, Havana, Cuba, Sept. 18-22, 2001, pp. 334-337.
[15] C. Penrose, private email correspondence with author, 2002.
[16] http://www.sarc.qub.ac.uk/~elyon/LyonSoftware/MaxMSP/FFTease/
[17] http://www.jezwells.org/Computer_music_tools.html
[18] J. Wells and D. Murphy, "A Comparative Evaluation of Techniques for Single-Frame Discrimination of Nonstationary Sinusoids", *IEEE Trans. on Audio, Speech and Language Proc.*, vol. 18, pp. 498-508, March 2010.
[19] http://www.jezwells.org/research/dafx10/