REAL TIME SPECTRAL EXPANSION FOR CREATIVE AND REMEDIAL SOUND TRANSFORMATION

Jeremy J. Wells and Damian T. Murphy

Media Engineering Group, Department of Electronics
University of York, Heslington, York. YO10 3DD
jjw100@ohm.york.ac.uk, dtm3@ohm.york.ac.uk

ABSTRACT

In this paper we describe the implementation, use and applications of *WaveThresh,* a real time Fourier\wavelet spectral expander. Expansion and reverse-expansion of spectral components is offered. In order that analysis methods can be better adapted to the signal we offer a combined wavelet\Fourier mode. This mode separates sinusoids from the rest of the signal (residual) and applies Fourier analysis to the sinusoids and wavelet analysis to the residual.

1. INTRODUCTION

Short time spectral gating\attenuation (expansion) with wavelets or Fourier analysis is often the basis for single ended noise reduction and has been implemented as a sound transformation process (e.g. [1] and [2]). Section 3 of this paper discusses some relative strengths and weaknesses of these analysis methods

In signals where both deterministic and stochastic sound types are present a method for separating the two for separate analysis by these two methods is desirable. By offering Fourier or wavelet analysis or an adaptive combination of both for each successive frame of a windowed and overlapped signal a useful real-time sound transformation method (creative and\or remedial) using spectral component reduction or elimination can be realised. The following sections describe how *WaveThresh* has been implemented.

2. EXPANSION

Expansion is a term commonly used in dynamic processing of time domain signals. It describes a process whereby a signal is attenuated if it falls below a certain level and the amount of attenuation applied increases as the signal level decreases according to an expansion ratio (soft thresholding). If this ratio is set to infinity the signal is switched off if it falls below the threshold (hard thresholding). We apply the same principle to the levels of individual spectral components within an analysis frame and offer reverse thresholding whereby components are attenuated if they are higher in level than the set threshold. We refer in this paper to spectral coefficients. For wavelet analysis these are directly derived from the wavelet transform (WT), for Fourier analysis these are the magnitudes of the complex values derived from the discrete Fourier transform (DFT). Fourier coefficients are only positive, wavelet coefficients may be positive or negative. Figure 1 shows the relationship between the

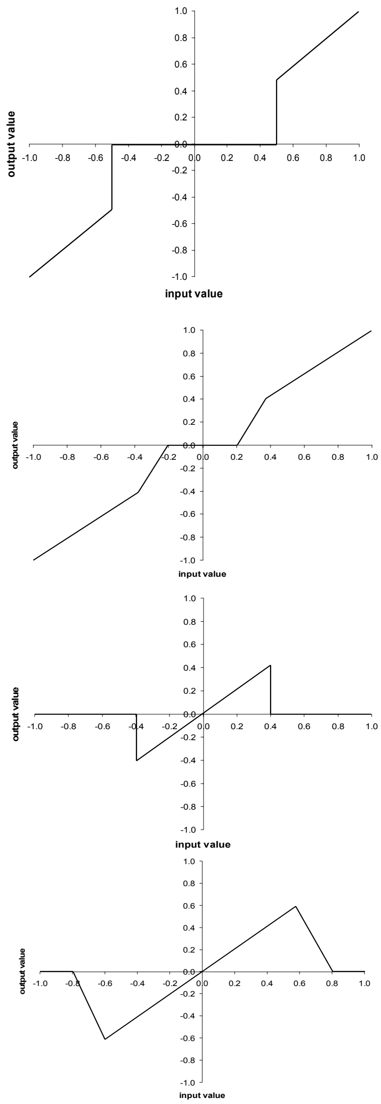

Figure 1:*Input versus output characteristics for spectral analysis coefficients. From top to bottom: hard thresholding (threshold = 0.5), soft thresholding (threshold = 0.4), reverse hard thresholding (threshold = 0.4) and reverse soft thresholding (threshold = 0.6). Soft thresholding ratios are set to 2.*

levels of these coefficients before and after expansion for the four different types of thresholding. All of these types are offered in *WaveThresh* with variable threshold and ratio and optional input level sensing to scale thresholds according to the average signal magnitude within an analysis frame.

3. ANALYSIS

3.1. Fourier and wavelet analysis

Fourier and wavelet transforms are commonly used for spectral analysis of digital audio signals. Short-time Fourier analysis requires that the signal be divided into frames whereas this is not necessarily required for time-scale (wavelet) analysis. However for real-time (or quasi real-time since such analysis cannot be done on a sample by sample basis) operation dividing the signal into frames is required for both. Windowing and overlapping of analysis frames prevents discontinuities at frame boundaries causing excessive spectral smearing in Fourier analysis. Windowing of synthesis frames smoothes out any discontinuities between modified frames.

The short time Fourier transform (STFT) assumes that the signal is stationary for the duration of each frame. For fast changing components in a signal the analysis frame must be short for this assumption to be valid but this will give poor resolution for the description of components that are highly localised in frequency. For acceptable frequency resolution for many signals the time resolution is such that important transient detail is lost in the thresholding process.

The wavelet transform does not assume stationarity but also does not offer very intuitive results for sinusoidal functions. A single sinusoidal function is represented by many wavelet coefficients in a single frame and this produces distortion of the waveform if coefficients with different values are expanded. Figure 2 shows how the thresholding of a sine wave using a Daubechies wavelet of order 10 distorts the signal component rather than attenuating or eliminating it. Clearly in this case we are introducing sinusoidal components rather than eliminating or attenuating them.

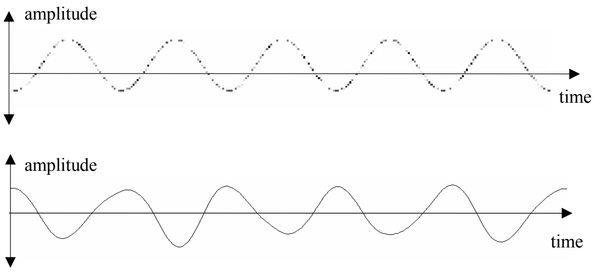

Figure 2*: Sine wave before (top) and after (bottom) wavelet coefficient expansion.*

3.2. Combined analysis

The relative strengths and weaknesses of the two analysis methods discussed previously suggest that separating signal types in each frame, assigning them to the most suitable analysis method and recombining them at the output will produce more desirable results. To this end we extract sinusoids from the input

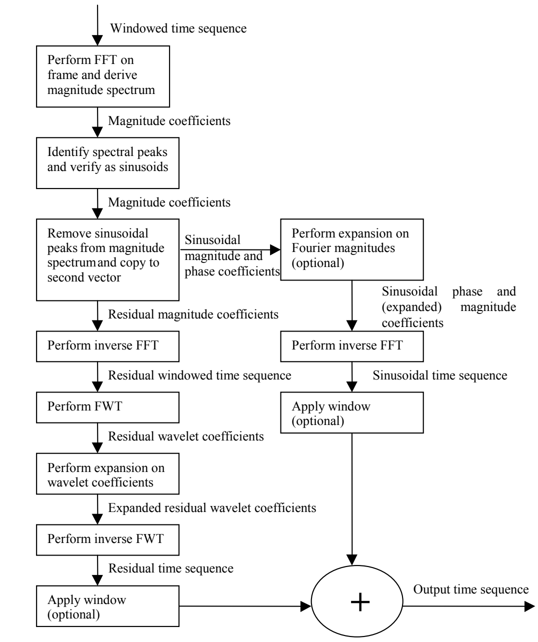

Figure 3*: Combined analysis algorithm.*

signal after Fourier analysis and apply wavelet analysis to the residual. Figure 3 illustrates the combined analysis algorithm.

We use a combination of peak identification and derived frequency verification (described in [3]) for frame by frame sinusoidal extraction. Using a 1024 point window for signals sampled at 44.1 kHz we search for local maxima. A bin is considered to contain a local maximum if its magnitude is higher than its four surrounding bins. The frequency for that bin is derived by comparing the bin magnitude from the DFT of the signal to the magnitude of the same bin from the DFT of the first derivative of the signal as in (1).

$$f_{peak} = \sin^{-1}\left(\frac{DFT'[m_{peak}]}{DFT[m_{peak}]} \times \frac{1}{2F_s}\right) \quad (1)$$

where f_{peak} is the frequency of the sinusoidal function representing that peak, F_s is the sampling frequency and $DFT[m_{peak}]$ and $DFT'[m_{peak}]$ are the DFT magnitudes of the bin containing that peak for the signal and its derivative respectively.

If this peak in the DFT spectrum is caused by a single stationary sinusoid (or similar function) then the derived frequency for that peak should be within half a bin width each side of the centre frequency of that bin as in (2).

$$lower_{peak} \le f_{peak} \le upper_{peak} \quad (2)$$

where

$$lower_{peak} = (peak - 0.5) \times (F_s / N)$$
$$upper_{peak} = (peak + 0.5) \times (F_s / N) \quad (3)$$

where F_s is the sampling frequency, N is the size of the DFT frame and peak refers to the index of the bin in which the peak resides. If (2) is not the case then the peak is rejected as being a single sinusoid.

If a peak is identified as a single sinusoid its magnitude is corrected by applying the frequency deviation (D) of the derived frequency from the centre frequency of that bin (normalised to the width of one bin) to the calculated power spectrum for the windowing function (Hann in this case) as in (4).

$$DFT[m_{sinusoid}] = \frac{\sin \pi D}{\pi D(1-D^2)} \times DFT[m_{peak}], 0 < D \leq 0.5 \quad (4)$$

We can also estimate what contribution to the surrounding bins this sinusoidal function will have made from (5) where D_{side} is the difference between centre frequency of the surrounding bin and the derived frequency of the sinusoid.

$$DFT[m_{side}] = \frac{\pi D_{side}(1-D_{side}^2)}{\sin \pi D_{side}} \times DFT[m_{sinusoid}], 0 < D_{side} \quad (5)$$

Since we are concerned with real-time operation the number of corrected bins for each peak is limited to 5 (the peak bin and the four surrounding bins). Future versions of *WaveThresh* will offer the user control over how many bins the algorithm takes account of when considering the contribution of a sinusoid so that the tradeoff between computational intensity and accuracy can be tuned to taste and\or the processing capabilities of the host platform.

When a sinusoidal peak has been identified and its magnitude corrected it is copied to a sinusoidal magnitude array along with estimated magnitudes for surrounding bins. Magnitudes in the sinusoidal array are subtracted from magnitudes in the original array (which, at the end of the sinusoidal identification and extraction process, becomes the residual array).

The residual time sequence is then constructed from the inverse DFT (IDFT), the WT of this sequence is calculated and the wavelet coefficients are thresholded. In many applications it is not desirable to threshold the Fourier magnitudes of the sinusoidal components as even small magnitudes in bins surrounding a peak contribute to that sinusoid and are considered purely deterministic, however the option to threshold these magnitudes as well as the wavelet coefficients is made available to the user.

4. RESULTS

Figure 4 shows the separation of sinusoidal and residual signal components for a piano recording. Listening to these two signals we can hear that the process has been largely successful in separating the steady state, pitched portion of the piano notes from the non-pitched sound of the hammer initially hitting the strings. There are some pitched components just audible in the residual but these are very low in level compared to the non-pitched components in this signal and the components in the sinusoidal part.

We have added white noise to this recording and used thresholding to attempt to remove this noise as a single ended

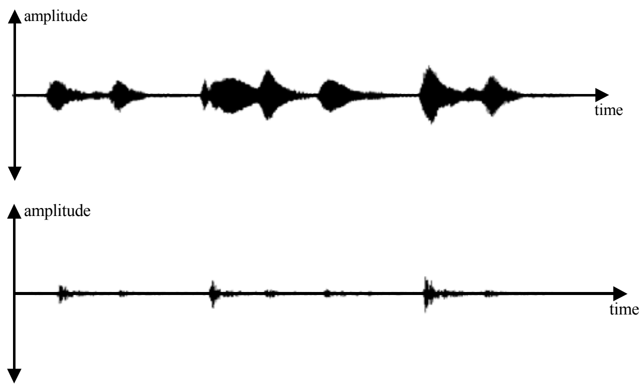

Figure 4*: The sinusoidal (top) and residual (bottom) parts of a signal after synthesis by IDFT*

process. It is interesting to note that some thresholding of the sinusoidal magnitudes is required to reduce 'birdy noise' in the sinusoidal signal (other noise reduction algorithms, e.g. [1], use knowledge of previous and successive frames to introduce hysteresis into the thresholding process but such a causal plus anti-causal approach is not suitable for real time operation). Thresholding of the wavelet coefficients can be adjusted and auditioned in real time to find the best tradeoff between removing the unwanted noise and leaving musical noise (hammer and pedal noise in this example) intact.

Figure 5 illustrates noise reduction of this type on a synthetic signal (short burst of broad band noise followed by a steady state sinusoidal signal) where the process can be more easily visualised. In order to make the results clear hard thresholding with high threshold levels has been applied. The waveform in (a) shows the original signal and (b) shows the same signal with white noise added at a relative level of –18 dB. The signal waveform after Fourier thresholding to remove the added noise is shown in (c). We can see that the noise has been removed whilst the overall shape of the signal waveform has been retained but the detail in the first part of the original signal has been lost. The signal waveform after thresholding with a Daubechies wavelet of order 10 is shown in (e). This wavelet offers a useful compromise between transient following and frequency localisation – there is preservation of the detail in the first part of the waveform and the fundamental period of the second half of the waveform. However the waveform shape in the second half has been distorted and this is clearly audible as a change of timbre in the steady state portion of the sound. The result of combined thresholding using Fourier bases for the deterministic part of the signal and Daubechies 10 for the residual is shown in (f). Some of detail of the first part of the signal is retained along with a waveform shape for the second part that is almost identical to that of the original signal with all of the additive noise removed. In order to remove all of the noise from the second portion of the signal some detail has been lost from the first but this tradeoff can be controlled by adjusting the threshold if desired.

5. APPLICATIONS

We have discussed the results of applying this algorithm to real time single ended noise reduction, however its greatest potential

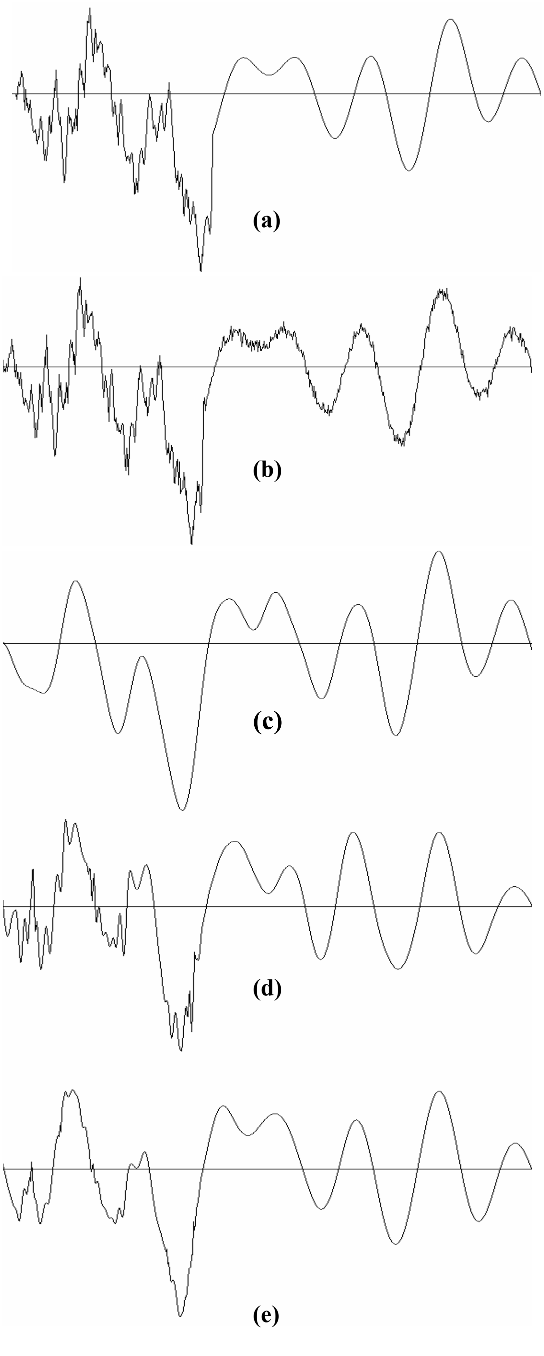

Figure 5*: waveform plots (amplitude – y axis versus time – x axis) of (a) test signal, (b) test signal plus noise, (c) Fourier thresholded signal, (d) wavelet thresholded signal and (e) combined thesholded signal*

may lie in creative sound transformation. We have produced audio examples where the distortion is progressively removed from an electric guitar sound whilst the steady state part is left unaltered by increasing the wavelet threshold over time. Wavelet thresholding of stochastic type sounds has been shown to offer convincing transformations which retain a high degree of source bonding (for example a heavy rain storm can be transformed into a light shower) [4]. With combined thresholding such transformations can be applied to deterministic–plus–stochastic type sounds [5]. We also have found *WaveThresh* to be useful in mixing situations where combinations of compression or expansion and equalisation might traditionally be applied.

6. CONCLUSIONS

WaveThresh a real time audio transformation process offering Fourier, wavelet and combined thesholding has been presented and discussed. This algorithm has been tested with both real and synthetic sounds containing combined narrow and broad band signal components. Sounds can be spectrally expanded using the combined algorithm with less distortion of sinusoidal components and greater preservation of transient detail. Waveform displays of transient followed by steady state sounds processed in this way demonstrate this improvement.

Audio examples demonstrating useful creative transformation and broad band denoising have been produced. These can be found, along with the plug-in (available for download free of charge) and accompanying documentation, on our web page at http://www.elec.york.ac.uk/ME/

Future work will focus on improving the sinusoidal extraction method. An improved method would take better account of intra frame non stationarity of sinusoidal components and fused sinusoids in ensembles of instruments or sounds when selecting components for inclusion in the deterministic part of the signal.

7. REFERENCES

[1] Maher, R. and Ramparu, P. "Methods for Reducing Audible Artifacts in a Wavelet-Based Broad Band Denoising System", in *Journal of the Audio Engineering Society 46[3]*, pp. 178-190, 1998.

[2] Lippe, C and Settel, Z., "Real-time Frequency-Domain Digital Signal Processing on the Desktop", in *Proceedings of the International Computer Music Conference*, 1998

[3] Desainte-Catherine, M. and Hanna, P. "High Precision Fourier Analysis of Sounds Using Signal Derivatives", in *Journal of the Audio Engineering Society 48[7/8]*, pp.654-667, 2000.

[4] Caudell, T. and Miner, N. "Using Wavelets to Synthesise Stochastic-based Sounds for Immersive Virtual Envornments", in *Proceedings of the International Conference for Auditory Display (ICAD) Conference*, 1997

[5] 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)*, pp. 12-24, 1990