A review of time–frequency representations, with application to sound/music analysis–resynthesis

PAUL MASRI, ANDREW BATEMAN and NISHAN CANAGARAJAH

Digital Music Research Group, University of Bristol, 5.01 Merchant Venturers Building, Woodland Road, Bristol BS8 1UB, UK
E-mail: Paul.Masri@bristol.ac.uk URL: http://www.fen.bris.ac.uk/elec/dmr/

Analysis–resynthesis (A–R) systems gain their flexibility for creative transformation of sound by representing sound as a set of musically useful features. The analysis process extracts these features from the time domain signal by means of a time–frequency representation (TFR). The TFR provides an intermediate representation of sound that must make the features accessible and measurable to the rest of the analysis. Until very recently, the short-time Fourier transform (STFT) has been the obvious choice for time–frequency representation, despite its limitations in terms of resolution. Recent and ongoing developments are providing several alternative schemes that allow for a more considered choice of TFR. This paper reviews these contemporary approaches in comparison with the more classical ones and with reference to their applicability, merits and shortcomings for application to sound analysis. (Where they have been successfully applied, details are provided.) The techniques reviewed include linear, bilinear and higher-order spectra, nonparametric and parametric methods and some sound-model-specific TFRs.

1. INTRODUCTION

To the analysis–resynthesis (A–R) system user, the analysis process is the key to creative transformation. It unlocks the time-domain waveform and makes available the properties of a sound as a set of modifiable parameters. The challenge to the A–R system designer is to discover how to extract that information from a time-domain sample.

The time-domain waveform encapsulates all the detail of a sound, but as a one-dimensional signal. The role of a time–frequency representation (TFR) is to expand that signal into two dimensions, by separating out the spectrum so that frequency components become distinguishable. It is the location, strength and evolution of these components that make up the feature set of most A–R systems. Therefore it is important that the TFR be capable of providing sufficient detail for effective extraction and accurate synthesis.

The short-time Fourier transform (STFT) is the time-varying form of the Fourier transform, composed of FFTs of successive time windows. It remains the primary method of time–frequency representation, but it is by no means the only method. With the A–R system quality being dependent on the ability of the TFR to render sufficient detail, the designer must consider which TFR is most appropriate to the application.

The importance of the TFR to the analysis–resynthesis of sound is discussed in an associated paper (Masri, Bateman and Canagarajah 1997) which also deals with how the designer should make this choice. In this paper, the contemporary approaches to time–frequency representation are reviewed. With the ongoing development of new TFRs and the variants and extensions to existing ones, this review cannot be exhaustive. Instead the aim is to present a broad synopsis of available techniques, particularly where there has been some application to musical signals. These approaches are presented according to four categories: Fourier-based transforms, higher-order nonparametric distributions, parametric estimators and model-informed representations. The first includes the 'standard' TFR of A–R systems to date, the short-time Fourier transform (STFT), as well as the wavelet transform (WT) and the bounded-Q Fourier transform (BQFT), which are strongly related to the Fourier transform (FT). The higher-order nonparametric distributions include the Cohen class of bilinear and higher-order spectra, the most famous of which is the Wigner–Ville (W–V) distribution. Those in the parametric class employ time-series modelling as opposed to energy modelling and they too include linear and higher-order possibilities. Finally, the model-informed class includes some approaches that elicit more information from the STFT by applying foreknowledge about the types of signals to expect. In so doing, although they use the STFT, they are more powerful than the STFT alone.

The STFT is described in greatest detail of all the TFRs (in the section following) – how it is constructed, what assumptions accompany it and why, and how these have impacted existing models. Its merits and limitations for sound representation are also presented, in order to provide a clear picture of what is required of a TFR in an A–R system. The other techniques are summarised with a brief explanation of their construction, but without delving into the deeper mathematics, for easy understanding of

Organised Sound **2**(3): 193–205 © 1997 Cambridge University Press. Printed in the United Kingdom.

https://doi.org/10.1017/S1355771898009042 Published online by Cambridge University Press

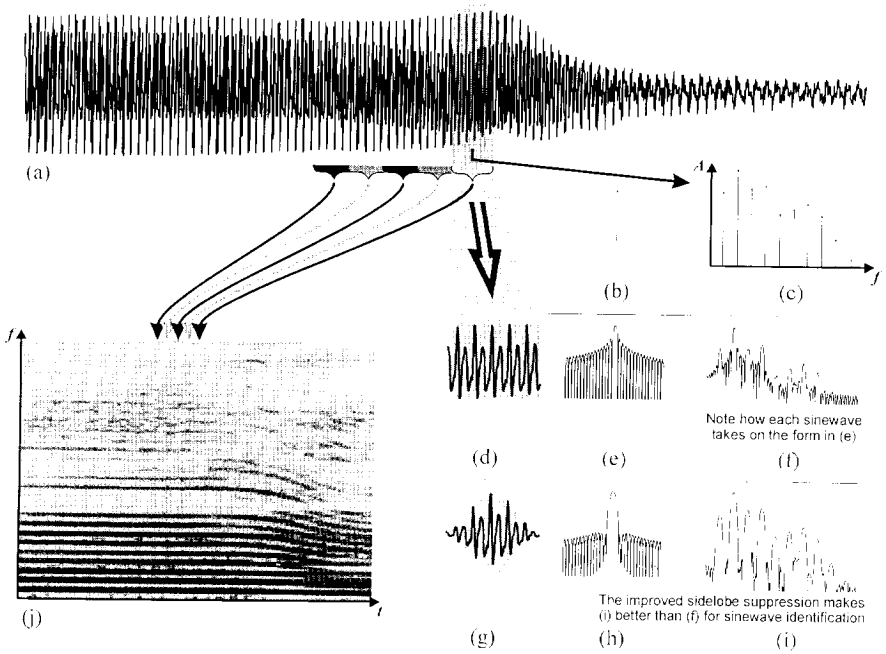


Figure 1. Construction of the STFT: (a) sampled female singing, with analysis window segmentation indicated; (b) delta function: the theoretical spectrum of a sinewave; (c) theoretical instantaneous spectrum (at the centre of the analysis window); (d) waveform segment within the analysis window; (e) spectrum of a time-limited sinewave (rectangular window); (f) Fourier transform of (d); (g) waveform segment with Hamming window function; (h) spectrum of a Hamming windowed sinewave: note the mainlobe width and sidelobe suppression compared to (e); (i) Fourier transform of (g); and (j) STFT of (a), using the Hamming window and overlapping time frames.

the approach and straightforward assessment of their application to different types of sound. (Particular detail is paid to those techniques which have been implemented recently for sound analysis.) For those who wish to pursue a particular technique for themselves, the necessary references are provided.

2. FOURIER-BASED TRANSFORMS

2.1. Short time Fourier transform (STFT)

2.1.1. Construction of the STFT

The STFT is composed of a series of time frames, where each frame is calculated as a fast Fourier transform (FFT) of a different portion of the signal (see figure 1(a) and (j)). A window function picks out the section of the signal to be represented, in the time domain. For a window length of N samples, the FFT yields values for energy localised in N frequency bands (often called *bins*). Since the sound signal is a band-limited, real signal, useful information is contained in the first $(N/2+1)$ bins, covering the range from DC to the Nyquist frequency (i.e. $[0, F_s/2]$, where F_s is the sampling rate).

Since the number of samples in the time domain window is the same as the number of frequency bands in the frequency domain, the time resolution is inversely proportional to the frequency resolution. The longer the time frame, the more points there are in the frequency domain and hence the better the frequency resolution, but at the cost of poorer temporal resolution.

The shape of the time-domain window is as important as its length. The default window function is rectangular, isolating a signal segment without modification (see figure 1(d)–(f)). If the signal being represented in this window contains sinusoidal elements whose periods are not exact fractions of the window length, spectral leakage occurs, significantly reducing the readability of the representation. Spectral leakage occurs as a result of a discontinuity between the signal at either end of the window. This can be solved by using a modified window shape.

If the window is tapered to zero at either end, this problem is eliminated, but such modification is at the expense of frequency resolution. The design choice involves a compromise between the width of the mainlobe and the prominence of sidelobes. (Ideally the mainlobe should be very thin to enable separate detection of closely spaced sinusoids and the sidelobes should be very low, so as not to interfere with neighbouring elements and to reduce the chance of being misinterpreted as low-energy mainlobes.) The most popular window function is the Hamming window, which has a smooth, bell-shaped curve in the

https://doi.org/10.1017/S1355771898009042 Published online by Cambridge University Press

time domain and provides a good trade-off between mainlobe width and sidelobe suppression in the frequency domain (see figure 1(g)–(i)).

2.1.2. The STFT as a filterbank

Note that the filterbank model of the STFT (in which the frequency bin outputs are provided by a bank of band-pass filters) is identical to the FFT frame-based model. If linear filters are used with a frequency profile equivalent to the spectrum of the window function used in FFT frames, then the response time of the filter (cf. its slew rate) will match the time-domain form of the window function. Therefore, the time–frequency resolution trade-off exists regardless of the model.

2.1.3. Zeropadding for extra frequency resolution

In sinusoidal-based models for analysis–resynthesis (see section 2.5 below), the purpose of the TFR is to enable detection, location and measurement of the partials, the sinusoids belonging to the harmonic structure. Multiplication of the time-domain signal with the window function causes convolution of the partial sinusoids with the window function's spectrum in the frequency domain, so the partials appear as copies of the window function's spectrum, where the maximum of each mainlobe is centred on the associated partial's frequency.

In order that the partials be successfully detected, there must be some separation between the mainlobe peaks. Therefore, frequency resolution takes priority to the detriment of the inversely related time resolution. However, time resolution must not be degraded so far that the approximation to stationarity over the duration of the window becomes poor, otherwise detection will be impaired as the peaks become distorted. A popular solution is to make the windows as long as possible without significantly impairing detection and then to apply zeropadding prior to the FFT calculation. This increases the number of samples in the FFT without using more signal samples.

Although zeropadding does increase the apparent frequency resolution, it does not actually provide extra spectral detail, as shown in figure 2. That is, if the windowed signal segment were analysed with the FT, under the assumption that the signal is eternally zero before and after the window, a continuous spectrum would result, whose peaks are as wide as the mainlobe of the window function (see figure 2(b)). This is equivalent to applying infinite zeropadding. For practical use, the signal is usually padded with zeros up to eight times the window length; this provides extra data points that sample the curve of the continuous spectrum at closer intervals than the unpadded FFT (see figure 2(c)). In summary, zeropadding can be useful in eliciting more detail of the spectral curve (particularly for *locating* the mainlobe maxima), but it does not provide better resolution of partials *per se*, and so it does not enhance *detection* (compare figure 2(c) and (d)).

2.1.4. Merits and shortcomings of the STFT for sound A–R

The limited time–frequency resolution of the STFT is the prime motivator in the search for an alternative TFR, for sound A–R. This shortcoming places several limits on the kinds of signals that can be analysed, including those with detailed or highly dynamic partial trajectories, percussive elements or other abrupt changes to spectral composition, multiple simultaneous pitches whose harmonic structures overlap with closely spaced partials, and so on. However, it is useful to note the merits of the STFT which may have been taken for granted, but which become significant in comparing it with the alternative techniques:

- The STFT is a linear function of the signal under analysis. Therefore if the signal contains additive elements, they will appear additively in the spectrum – i.e. without generation of inter-element products.
- The STFT is invertible, so all the information about the signal, stationary or otherwise, must be contained within the representation.
- The constituents of the harmonic structure can be considered as separate sinusoids. The STFT represents these in a simple way that facilitates easy extraction as magnitude peaks (under favourable conditions).
- The STFT is nonparametric, so it does not place any interpretation on the image. Although this is not necessarily an advantage, it does guarantee that the TFR contains a complete picture of the signal.

Figures 3 and 4, at the end of this article, demonstrate these properties of the STFT for two challenging but realistic sound examples. (These figures compare a few sample TFRs from those that are reviewed for each sound example.)

2.2. Wavelet transform (WT)

The window-length of the STFT is fixed over the entire frequency range, such that the time–frequency 'cells' are uniformly spaced in both time and frequency. The wavelet transform (WT) (Kronland-Martinet 1988, Rioul and Vetterli 1991) takes the

https://doi.org/10.1017/S1355771898009042 Published online by Cambridge University Press

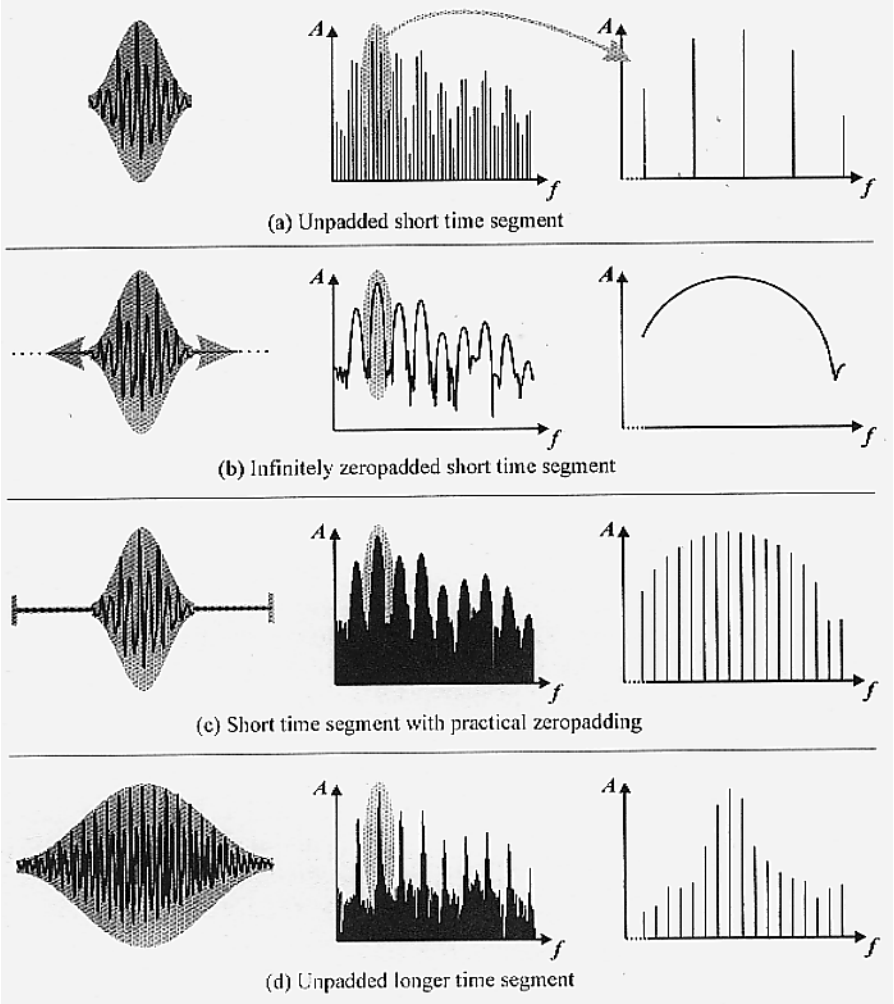


Figure 2. Why zeropadding is no substitute for a longer window.

alternative approach of making the window-length inversely proportional to frequency, so that each window contains the same number of periods of the centre frequency.

In explaining how the FFT is created, it is common to think of the signal being windowed and then convolved with sinewaves at the frequencies of the bins, and indeed this is how it is implemented. Mathematically, the same outcome results from convolving the signal with windowed versions of the sinewaves. This latter explanation is more appropriate for the WT, where the customary terminology describes the windowed sinewave as the *analysing wavelet*. Since every wavelet contains the same number of periods of the sinewave, each wavelet is a dilated or time-scaled version of the others, where the dilation or scale depends on the frequency.

In the filterbank model of the WT, the filters have constant Q (unlike the STFT filters which have constant bandwidth), and for equal coverage of the spectrum they are therefore logarithmically spaced.

The same mathematical laws that govern resolution for the STFT apply to the WT (discussed in section 2.4 below), so the narrower higher-frequency wavelets have a wider bandwidth, thereby preserving the total duration-bandwidth coverage of each 'cell'. The logarithmic frequency spacing of cells appears to be more in tune with the way in which we perceive sounds. That is, the pitch scale is logarithmically related to frequency (inasmuch as the perceptual property of pitch can be related to the physical property of frequency), so the WT provides a constant profile for all pitches.

This fact makes the STFT more useful for sound A–R than the WT, because the STFT provides the linear frequency spacing that is essential for distinguishing the equally spaced harmonics of a pitched sound. This is evidenced by the sound examples in figures 3 and 4. However, in applications that require pitch extraction or the accurate location or characterisation of sudden changes, the WT outperforms the STFT.

2.3. Bounded-Q Fourier transform (BQFT)

The bounded-Q Fourier transform (BQFT) (Chafe, Jaffe, Kashima, Mont-Reynaud and Smith 1985) is

https://doi.org/10.1017/S1355771898009042 Published online by Cambridge University Press

a hybrid of the WT and the FFT, developed as a compromise that provides *locally linear* spacing. The spectrum is divided into broad bands such that higher-frequency bands are calculated to favour good time resolution, as with the WT. However, within each band the frequency spacing is linear, as with the STFT. The BQFT was implemented in a sinusoidal model by the authors (Masri and Bateman 1994) with the aim of providing an analysis scale that automatically adapts to the pitch of the analysed instruments – higher-pitched sounds appear in the upper bands and their harmonics are more widely spaced.

The following advantages of the BQFT over the STFT were observed:

- Automatic TFR scaling, because each pitch is analysed at ideal frequency spacing for the first few (and often the most important) partials.
- The automation enables analysis of multiple simultaneous pitches (chords or multiple instruments), where each pitch is analysed optimally (for the first few partials).
- The initial transients of percussive pitched instruments (e.g. the onset of piano notes) and some noisiness are captured within the sinusoidal analysis; in fact many tonal instruments do not require a stochastic aspect to the model for faithful reproduction when using the BQFT.

The following disadvantages were also observed:

- Above the first few partials, the lack of frequency resolution leads to erroneous detection. Upon resynthesis this is audible, although very quiet, as high-frequency susurrus artefacts (similar to a radio with bad reception or a person making the sound 'srsrsrsr').
- Some sounds, particularly the human singing voice, have a wide bandwidth with significant energy in the higher partials, so enunciation is compromised by loss of the higher harmonics; the resultant sound is perceived as 'closed' after resynthesis, due to the effective low-pass filtering.
- Although transients and noisiness are captured by the sinusoidal representation, their synthesis is phase dependent; therefore distortion can result upon transformation.

2.4. Being certain about the uncertainty principle

The STFT, WT and BQFT are all limited to the same fixed resolution (in terms of the time–bandwidth product) by the same mathematical law. This law is often quoted as the Heisenberg uncertainty principle, and almost any technical paper on the subject of time–frequency analysis will mention it at some point (e.g. Claasen and Mecklenbrauker 1984, Fineberg and Mammone 1992). The law is encapsulated in the inequality

$$\Delta t \Delta \omega \geqslant \tfrac{1}{2}, \tag{1}$$

where Δt and $\Delta \omega$ are the effective duration and effective bandwidth (in angular frequency) of a signal. The Fourier transform and its variants achieve the minimum energy localisation permitted by equation (1), i.e. $\Delta t \Delta \omega = \frac{1}{2}$.

In actual fact, to call the inequality of equation (1) the Heisenberg uncertainty principle is a misnomer, because this inequality relates to a limit for representation resolution, not a statistical uncertainty (Cohen 1989). However, both derive from the Cauchy–Schwarz inequality, a property of every linear time-invariant system; it results from making linear approximations to an essentially nonlinear world (Kosko 1994: 107–14), or in the case of the STFT, a linear phase function approximation to signals that have essentially nonlinear phase functions. Throughout the rest of this paper, it is referred to as the linear resolution principle (LRP).

From its equation, the STFT can be seen to be linear, time-invariant:

$$\mathrm{STFT}(t, \omega) = \int s(u) w(u - t)\, \mathrm{e}^{-\mathrm{j}\omega u}\, \mathrm{d}u \tag{2}$$

where t is time, u is the offset (centre of transform time-window), ω is the (angular) frequency, $s(t)$ is the time-domain signal under analysis, $w(t)$ is the window function, and $\mathrm{STFT}(t, \omega)$ is the short-time Fourier transform. The equation is linear because the integral includes a single, linear function of the signal $s(t)$. It is time-invariant because a time-shifted version of $s(t)$ would result in time shifting of the time–frequency representation and no other effect. Time-invariance is a desirable property because the results of the transform are identical for signals with the same content, regardless of when they occur. Linearity means that there is no conflict between components, a positive attribute (as will become evident), but it does limit the usability of the transform.

In the representation, the STFT is optimal for signals with components of linearly increasing phase. That is, for signals containing linear phase (stationary frequency) components, the representation is a delta function – a single, narrow spike – that takes up the minimum space on the time–frequency map – a single STFT cell. (This assumes that the default rectangular window has been used and that there is no spectral leakage.) Hence the familiar assumption or rule that the signal must be stationary, for optimal use of the STFT.

2.5. Application of Fourier-based TFRs to sound A–R

As expounded in the previous sections, when the STFT is applied to A–R, there is the assumption that the sound signal is quasi-periodic or quasi-stationary. This forces the system to conform fairly closely to the

https://doi.org/10.1017/S1355771898009042 Published online by Cambridge University Press

assumption of signal stationarity within each FFT frame. In the interpretation of the TFR in the sinusoidal model (McAulay and Quatieri 1986) and the deterministic-plus-stochastic (DPS) model (also known as spectral modelling synthesis (Serra 1989, Serra and Smith 1989)), there is the expectation that peaks in each FFT (magnitude response) correspond to sinusoidal components in the sound signal, whose instantaneous frequency, amplitude and phase can be measured individually at the maxima of the associated peaks. This assumption and expectation limit the application of the sinusoidal and DPS models to gradually evolving sounds. If rapid changes are presented to the system, their representation in the model domain will be degraded, often resulting in poor synthesis quality.

Since the WT and BQFT do not escape the LRP but merely alter the preference toward time or frequency resolution, they do not ultimately offer any improvement over the resolution of the STFT. That is not to say their properties are similar, though. The way in which the time–frequency resolution trade-off is mapped over the spectrum affects the suitability of each for different applications. The linear frequency spacing of the STFT means that it remains the preferred method for determination of partial trajectories. However, the time-scale properties of the WT make it the preferred method for detection of transient events and audio compression.

3. NONPARAMETRIC DISTRIBUTIONS

3.1. The Wigner distribution and the bilinear form

The Wigner distribution (Wigner 1932) was first introduced in 1932 in the context of studying quantum statistical mechanics (Cohen 1989, Jeong and Williams 1992). In 1948, Ville applied it in the context of signal processing (Ville 1948). The Wigner–Ville distribution is

$$(3)\quad \mathrm{WV}(t,\omega)=\frac{1}{2\pi}\int s^{*}\left(t-\frac{1}{2}\tau\right)\mathrm{e}^{-\mathrm{j}\omega\tau}s\left(t+\frac{1}{2}\tau\right)\mathrm{d}\tau,$$

where t is time, τ is the delay (time-lag), ω is the (angular) frequency, $s(t)$ is the time-domain signal under analysis, $s^{*}(t)$ is its complex conjugate, and $\mathrm{WV}(t,\omega)$ is the Wigner–Ville time–frequency distribution. The Wigner–Ville distribution is bilinear, time-invariant because it has two linear functions of the signal $s(t)$. As a result it is able to provide an optimal representation for quadratic phase (linear frequency modulation) signals, of which stationary frequency is one case. In terms of the representation quality, this means that nonstationary signals exhibit reduced distortion and the time–frequency resolution is better than that prescribed by the linear resolution principle.

3.2. The Cohen class and higher-order spectra

In 1966, Leon Cohen developed an equation that could be used to describe all bilinear time–frequency distributions (Cohen 1966, 1989), including the spectrogram and the Wigner–Ville distribution:

$$(4)\quad P(t,\omega)=\frac{1}{4\pi^{2}}\iiint \mathrm{e}^{-\mathrm{j}\theta t-\mathrm{j}\tau\omega+\mathrm{j}\theta u}\times\Phi(\theta,\tau)s^{*}\left(u-\frac{1}{2}\tau\right)s\left(u+\frac{1}{2}\tau\right)\mathrm{d}u\,\mathrm{d}\tau\,\mathrm{d}\theta,$$

where t is the time, u is the offset (centre of transform time-window), τ is the delay (time-lag), ω is the angular frequency, θ is the Doppler (frequency-) shift, $s(t)$ is the signal to be represented, $\Phi(\theta,\tau)$ is the 'kernel' (see next section), and $P(t,\omega)$ is the representation domain. From this equation, the spectrogram, which is the magnitude-squared function of the STFT and is therefore bilinear in $s(t)$, can be shown to be linearly related to the Wigner–Ville distribution. (In fact the spectrogram is a smoothed version of the Wigner–Ville distribution, an alternative explanation for the time–frequency averaging that causes reduced resolution.)

Cohen's generalised form for representing all bilinear time–frequency distributions, known as the Cohen class, enabled similar extensions to higher order spectra. Initially, the Wigner distribution was 'upgraded' to the Wigner bispectrum (Gerr 1988), the third order (trilinear) Wigner distribution, and later the generalised form for higher-order spectra was developed (Boashash and O'Shea 1991, Boashash and Frazer 1992). Called the generalised Wigner–Ville distribution (GWVD), it enables design of distributions to optimally represent signals whose frequency trajectory could be described by any arbitrary polynomial.

The process by which a signal is represented is shown to be a mapping followed by a FFT. The mapping converts the frequency modulation (up to the targeted order) onto the linear phase plane (zero frequency modulation), where it can be analysed by Fourier techniques. This can be understood in terms of autocorrelation. The spectrogram can be generated as the magnitude FFT of the autocorrelation of a signal. In the autocorrelation, a time-shifted version of the signal is correlated with the original signal, generating peaks at the signal's periodicities. General bilinear distributions can be considered to work in a similar way, except the correlation is between a time-shifted, frequency-shifted version of the signal and the original signal. Therefore, peaks arise where locally periodic elements have changed with linear frequency modulation.

A bilinear time-invariant transform can achieve a better effective resolution than a linear time-invariant

https://doi.org/10.1017/S1355771898009042 Published online by Cambridge University Press

one, not because it increases frequency resolution as such, but because it is more tolerant of nonstationarities than the linear transform. In practice, this means that the time window (defined by the kernel in equation (4)) can be longer than that for the STFT without losing resolution, and therefore the frequency resolution can be finer.

Just as the bilinear achieves a better resolution than the linear, so the higher-order spectra achieve better resolution than the bilinear, the accuracy increasing with every order (for an arbitrarily nonlinear frequency trajectory). The process can be extended to an arbitrary order, but at each increase in order there is an increase in system complexity and computation. Perhaps more importantly for applications to sound analysis, there are also extra interactions between the additive spectral components that can obscure the 'real' spectrum.

3.3. Cross-terms and modifications to the kernel

For a monocomponent signal (e.g. a nonstationary sinusoid), the bilinear and higher-order spectra have been proven capable of extracting detail from highly complex frequency trajectories. However, in the Fourier series-based model of sound, a sound signal contains many spectral components that are additively combined – a multicomponent signal. In the bilinear and higher-order spectra, as the signal is mapped to stationary frequencies (by the autocorrelation function), these components interact with one another. This situation is examined for the bilinear case.

The expansion of the mathematical expression $(x+y)^2$ gives a term in each of the variables, x^2 and y^2, and a term in both, $2xy$. In an analogous way, the expansion of $s^*(u-\frac{1}{2}\tau)s(u+\frac{1}{2}\tau)$ yields terms in each of the spectral components (auto-terms) and terms in each combination pair (cross-terms). In the time–frequency representation of the signal, the auto-terms appear, as expected, at the time and frequency locations of the actual signal components, whereas the cross-terms appear halfway between the corresponding components. This has two undesirable effects: there appears to be energy in the signal at times and frequencies where there is none, and some cross-terms lie on auto-terms, obscuring or distorting them.

Although the expression $s^*(u-\frac{1}{2}\tau)s(u+\frac{1}{2}\tau)$ yields the cross-terms, it is the kernel, $\Phi(\theta, \tau)$, that determines the representation. Through modifications to the kernel, many authors have sought to suppress the cross-terms. Perhaps the most widely recognised solution is the exponential kernel developed by Choi and Williams (1989), known commonly as the Choi–Williams distribution. From these studies it was recognised that just as there is a trade-off between time and frequency resolution in the STFT, so too there is a trade-off between cross-term suppression and auto-term resolution. In the case of the spectrogram, the cross-terms are maximally suppressed, and the auto-term resolution is the worst. (Similar kernel modifications have been applied to higher-order spectra to reduce the cross-terms interference (Fonollosa and Nikias 1992).)

3.4. Application of the Cohen class of TFR to sound A–R

A few studies have compared the various distributions for application to multicomponent nonstationary signal representation. Jones and Parks have compared the spectrogram with the Wigner and Choi–Williams distributions (Jones and Parks 1992). They conclude that for nonstationary, multicomponent signals the spectrogram outperforms both of the others (if the window-length is chosen to match the signal).

Equation (4) facilitates development of an infinite number of bilinear time–frequency distributions, through kernel design. However, only a small proportion satisfy certain 'desirable' properties for an energy distribution. These include, for example, that the energy should never be negative – it can be with the Wigner distribution. The possibility for violation of such properties demonstrates a mismatch between the mathematics and the real world – only when the properties are satisfied can a physical interpretation truly be considered valid.

Nevertheless, new distributions continue to be developed through kernel design, with some promising results. Recent publications indicate that these techniques are beginning to be applied to sound A–R. Pielemeier and Wakefield have developed a kernel which achieves good resolution without cross-terms interference for sounds that do not vary rapidly (Pielemeier and Wakefield 1992, 1996), and this has recently been applied to new work into synthesis of the piano (Guevara and Wakefield 1996). Also, work has begun on applying the smoothed Wigner transform for A–R, with the aim of timbre morphing (Hope and Furlong 1997). The Pielemeier–Wakefield 'modal' transform is compared with other TFRs in figures 3 and 4.

Higher-order statistics, which are closely related to the Cohen class and the higher-order spectra, but whose mathematical foundation differs, are also beginning to be applied for sound representation. The techniques also use autocorrelation of the time-domain signal (in the calculation of the higher-order cumulants) and take the Fourier transform of the result to generate the 'polyspectra' (Mendel 1991). Shlomo Dubnov began using the bispectrum, a third-order statistic, to analyse microvariations of partials

https://doi.org/10.1017/S1355771898009042 Published online by Cambridge University Press

during the steady-state portion of sounds, which he was able to simulate with frequency modulation jitter (Dubnov, Tishby and Cohen 1996). In collaboration with Xavier Rodet, this work has been extended to note the phase (and hence frequency) coherence between partials of different instruments (Dubnov and Rodet 1997). Although this research is at the investigative stage, it shows great promise for application to A–R, where it could help to reveal greater detail in the partial trajectories, evidence of which has been observed from the widening of peaks in the magnitude STFT (Masri 1996, section 2.3.4), but the detail of which has not been measurable using the STFT.

An informative introduction to the Cohen class, its history, derivation and properties, as well as some personal insights, can be found in Cohen (1989). The mathematical rigours are also presented in Boashash (1990) with graphs comparing distributions and consideration of multicomponent analysis, with application to underwater sound analysis. A comprehensive, but significantly mathematical, tutorial on higher-order statistics and polyspectra can be found in Mendel (1991).

4. PARAMETRIC ESTIMATORS

The spectral estimators considered so far have all been nonparametric techniques: they generate their spectral representations through mapping functions and distributions that do not alter or interpret the data in any way, but simply change the form of (literally trans-form) the data. As such, they do not bias the outcome of an estimate; all interpretation is left to the subsequent analysis algorithms.

There also exist parametric techniques for spectral estimation, which in contrast incorporate the designer's implicit model into the TFR, in effect preemptively interpreting the results. The expectations about the signal's content decide the form of the TFR. Generation of the representation is then a matter of finding the set of parameters that best fit the signal data. Since foreknowledge about the signal is built in, less information is required of the signal, so as a general rule less signal data is required for parametric techniques than nonparametric TFRs. Of course, this presumes that the foreknowledge is correct, or at worst a good approximation!

4.1. Autoregression and moving average

The most popular parametric models are autoregression (AR), moving average (MA), and a combination of the two (ARMA). In brief, ARMA is based on a z-plane pole-zero model of the spectrum: AR models the poles of the signal spectrum (in the z-plane), which define narrow peaks, and MA models the zeros, which define localised minima. The model appears as the assumption that the signal has been generated by feeding a white noise input to a system with a transfer function of the form

$$H(z) = \frac{B(z)}{A(z)}, \tag{5}$$

where

$$A(z) = \sum_{k=0}^{p} a[k] z^{-k}, \tag{6}$$

and

$$B(z) = \sum_{k=0}^{q} b[k] z^{-k}, \tag{7}$$

where $H(z)$ is the system transfer function, $A(z)$ is the z-transform of the AR branch (for p poles), and $B(z)$ is the z-transform of the MA branch (for q zeros).

AR performs best (i.e. provides the most concise parameter set) where the resonances of a system or the harmonics of a signal are of interest, whereas MA is best at describing broadband noise spectra. Where there are sinusoids embedded in white noise, ARMA is the best choice (because AR degrades in the presence of noise). In practice, the most widely used technique is AR, because parameter fitting amounts to solving a set of linear equations, whereas MA and ARMA both require solutions to a set of (usually highly complex) nonlinear equations (Kay 1988: 153).

4.1.1. Model order

The order of a parametric model (i.e. the number of poles and/or zeros) defines the level of detail of the spectral estimate. If the model order is too small, then a highly smoothed spectral estimate results. If it is too high, then extra spurious peaks can appear. The need to find a suitable order adds a complication to the process. Several techniques have been developed, relying on some measure of closeness of the model to the actual spectrum at each iteration (Makhoul 1975, Marple 1987: 229, Kay 1988: 234).

4.1.2. Further reading

For the interested reader, Kay (1988) and Marple (1987) provide a general grounding in linear parametric modelling techniques. Kay approaches the subject from a largely theoretical basis, comparing and contrasting different implementation techniques for statistical performance and computational efficiency. Marple's book is geared more toward applications, providing an easier route to implementation. Makhoul (1975) provides a more readable introduction to AR and ARMA, from the perspective of linear prediction, in terms of system function rather than mathematical rigour.

https://doi.org/10.1017/S1355771898009042 Published online by Cambridge University Press

4.2. Higher-order parametric modelling

Just as the above linear parametric techniques can improve over the time–frequency resolution of the FFT, so it is also possible to develop higher-order parametric techniques that improve over the higher-order spectra. The key to parametric model design is once again the kernel. The Cohen class of distributions can be made parametric by making the kernel a function of the signal (Cohen 1989). Similarly, the higher-order generalised Wigner–Ville distributions could be made parametric by including the signal as a parameter of the kernel.

Shimon Peleg and Benjamin Friedlander have recently developed an iterative method for estimating the polynomial phase functions (up to an arbitrary order) of a multicomponent signal (Peleg and Friedlander 1996). The coefficients of the polynomial phase function are the parameters of the model. These are estimated for each signal component in turn. Once a component has been completely specified, it is subtracted from the signal, to enable estimation of the next signal component. The drawback of the approach is the computational load. For a specific signal segment to be estimated, an FFT calculation is required for each polynomial order of each signal component (and more for their deletion from the signal). For a modest harmonic sound signal, this easily equates to hundreds of FFTs per analysis segment (which might be the duration of a single STFT frame). Furthermore, the estimates are affected by cross-terms interference, so Peleg and Friedlander recommend a second iteration of the whole process to improve the estimates.

4.3. Application of parametric TFRs for sound A–R

The linear parametric techniques, described earlier, are best known to computer musicians in the form of linear predictive coding (LPC) (Markel and Gray 1976). This is the basis of the well-established speech model, in which voiced sounds are synthesised by filtering excitation functions through the model filter of the speech spectrum. In its simplest form (i.e. without codebooks), voiced sounds are synthesised by exciting the system with a pulse train at the pitch period; for unvoiced sounds, white noise excitation is used. In this application, the model order is deliberately low, so that the spectral envelope (i.e. the formant structure) will be modelled instead of the individual partials.

Within the last year, details of a new A–R model have been published that apply a second-order parametric approach to the principles of the deterministic-plus-stochastic (DPS) model. The model, created by Yinong Ding and Xiaoshu Qian has been named QUASAR (QUAdratic polynomial-phase Sinusoids And Residual) (Ding and Qian 1997a, b). In this, the deterministic aspect is modelled using a parametric estimator based on the technique pioneered by Peleg and Friedlander described in the previous section, but simplified for a maximum of second-order estimation. The residual is estimated by LPC. With this approach Ding and Qian claim to have solved the problem of the noisiness failing to fuse with the partials that was evident with the DPS model, in addition to enabling analysis and resynthesis of more complex partial trajectories. (The authors have not heard a demonstration to be able to make more informed comments.)

5. MODEL-INFORMED REPRESENTATIONS

The schemes in this section differ from those in previous sections in that they are not so much alternatives to the STFT as extensions to it. Since the application for these TFRs is the representation of sound, in particular the musical properties of sound, there are some assumptions that can be made that are specific to this class of signals. Application of this information is a form of parametrisation, and it can allow for extra detail to be elicited from the STFT.

5.1. Multiresolution Fourier transform

It has been established that the STFT can provide good time resolution (GTR) or good frequency resolution (GFR) but not both simultaneously. Using a GFR-STFT for sound representation, partials appear as well-separated ridges running horizontally (along the time line). With a GTR-STFT, the partials cannot be distinguished, but vertical lines (parallel to the frequency axis) indicate the note onset points. By intelligent combination of these representations, it should be possible to enhance the resolution in both time and frequency.

One method, termed the multiresolution Fourier transform, was developed for improving note onset detection, for automatic score transcription (Pearson and Wilson 1990). This uses the *min* function, so that energy peaks are represented only where they appear in both STFTs. Thus the note onsets appear well localised in time and frequency.

In application to sound representation, the results appear good where there are energy boundaries, such as percussive onsets or stable partials. Therefore the results should be good for note onset localisation. Where tracking of gradually evolving phenomena is required, there is no improvement: the broadband peaks in the GTR-TFR do not enable better frequency tracking of the partials. In fact, because the GTR-TFR varies across individual periods of stable

https://doi.org/10.1017/S1355771898009042 Published online by Cambridge University Press

waveforms, its minima break up the otherwise continuous ridges of the GFR-TFR, thereby degrading the combined representation.

5.2. Parametric modelling of the STFT

Extraction of partials from an STFT is done in two stages: the first stage is detection, the second is location. In order to detect partials from the STFT, it is necessary to have sufficient frequency resolution that the peaks are separated. A rule of thumb safe separation is four bins, which corresponds to a window-length of four periods of the fundamental frequency. Having established the presence and rough location of a partial, accurate location is a simpler task, because the shape of each partial peak in the frequency domain is largely predictable. This is especially true if the partial trajectory is a good approximation to local stationarity.

A new method was first published a year ago (Depalle and Tromp 1996) which makes an initial estimate of the number and location of partials from an FFT using a window-length of four periods of the waveform. Analysis then continues at a shorter window-length ($1\frac{1}{2}$ periods of the waveform), using a least-squares-error method to iteratively improve the accuracy of each of the estimates, taking each partial in turn. The method alternates between improving amplitude and phase values whilst fixing frequency, and improving frequency location whilst fixing amplitude and phase. The latest developments (Depalle and Helie 1997) include:

- a sidelobe-free window function that should drastically reduce the problem of peaks due to sidelobe artefacts (but will increase additive interference because of a higher noise floor), and
- an algorithm manager that refines estimates iteratively in batches. This estimates the most significant components and then removes them from the spectrum before the next iteration. Presumably this combats the increased additive effects of the new window.

5.3. Phase distortion analysis of the FFT

Despite the assumption of stationarity that accompanies the FFT, it is nevertheless an invertible transform:

$$\text{IFFT}\{\text{FFT}[s(t)]\} = s(t). \tag{8}$$

Therefore, the FFT fully describes the signal, regardless of whether the assumption is observed or violated. Ordinarily, the presence of nonstationarities in a signal are perceived as distortion in the spectral representation. However, that distortion is actually the source of information about the higher-order phase and amplitude functions. Fortunately, the majority of the 'distortion' is localised around the peak that represents a nonstationary component, so there is scope for estimating more detail of the component's trajectory by observation and classification of the peak shape.

A preliminary investigation by the authors revealed that the effect on phase across the mainlobe is not only straightforward, but that independent estimations of linear frequency modulation and exponential amplitude modulation (which appears linear on the decibel scale) can be made from just two phase measurements around each partial peak (Masri and Bateman 1995, Masri 1996). As a result, the STFT can be used to effectively gain improved frequency resolution, without needing to compromise time resolution. Furthermore, the success of this study suggests that, through several measurements of phase and amplitude, it should be possible to make a robust estimate of higher orders of frequency and amplitude modulation. (Work in this direction has progressed and new results will be published shortly.)

6. CONCLUSIONS

At the time that sound and music A–R models were first developed, the FFT algorithm had been newly developed. It made the implementation of A–R a feasible proposition on the computer systems of the day. With sound modelling in its infancy, the resolution afforded by the STFT was adequate.

With the accelerated development of the model over the past few years, development of the TFR has lagged behind, to the point of non-consideration. However, independently of sound model development, there has over the same period been considerable theoretical research into alternative, higher-resolution methods, the wealth of which is just becoming appreciable. Until recently, many of these methods could not be usefully applied to multicomponent signals (due to cross-terms distortion). Even Cohen, who unified the bilinear transforms in 1966, concluded (in his review paper of 1989):

> The enigma of these distributions is that they sometimes give very reasonable results and sometimes absurd ones.... The fact that these distributions cannot be used in a consistent manner is one of the main areas that needs much further theoretical development. (Cohen 1989)

It has been the aim within this and its related paper (Masri *et al*. 1997) to bring TFR design into the mainstream of A–R design, through consideration of design methodology, and to stimulate interest by showing what approaches are already available. The fact that within the past year or so there have been several publications that have begun to tap this mathematical reservoir to improve sound A–R, is indeed encouraging.

https://doi.org/10.1017/S1355771898009042 Published online by Cambridge University Press

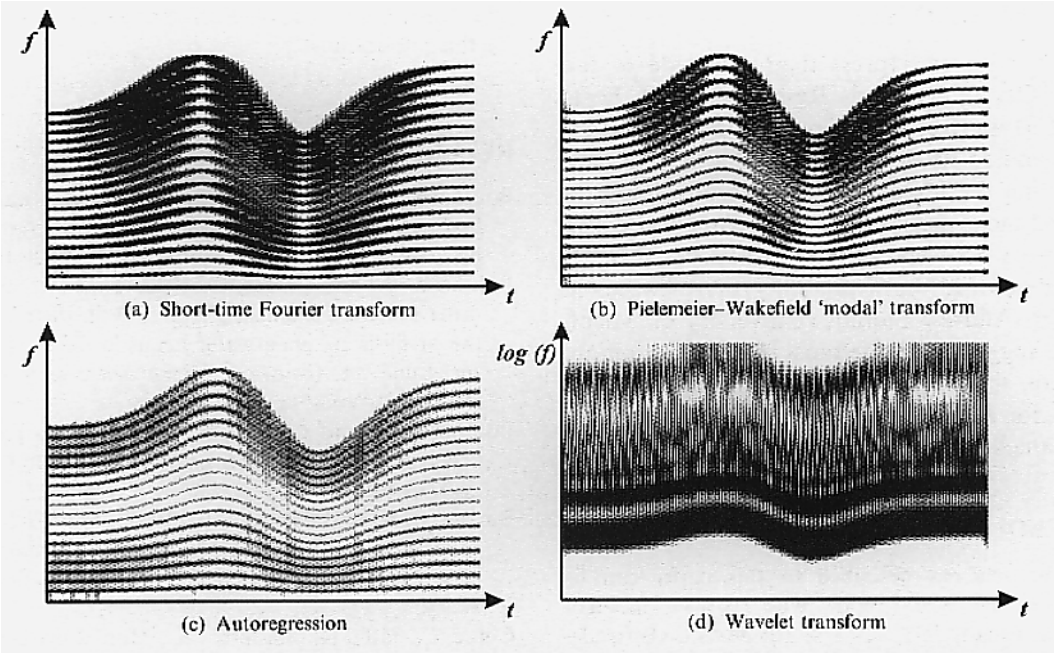


Figure 3. Representations of a synthetic waveform composed of harmonically related sinusoids of different amplitudes: (a) STFT: there is significant blurring where there is any significant nonstationarity; (b) modal: the blurring is much reduced; (c) AR: the trajectories are clear, except at their most dynamic where the top few partials show splitting; and (d) WT: at low frequencies the partials are clear, whilst at high frequencies the vertical banding indicates the peaks and troughs of the time-domain waveform.

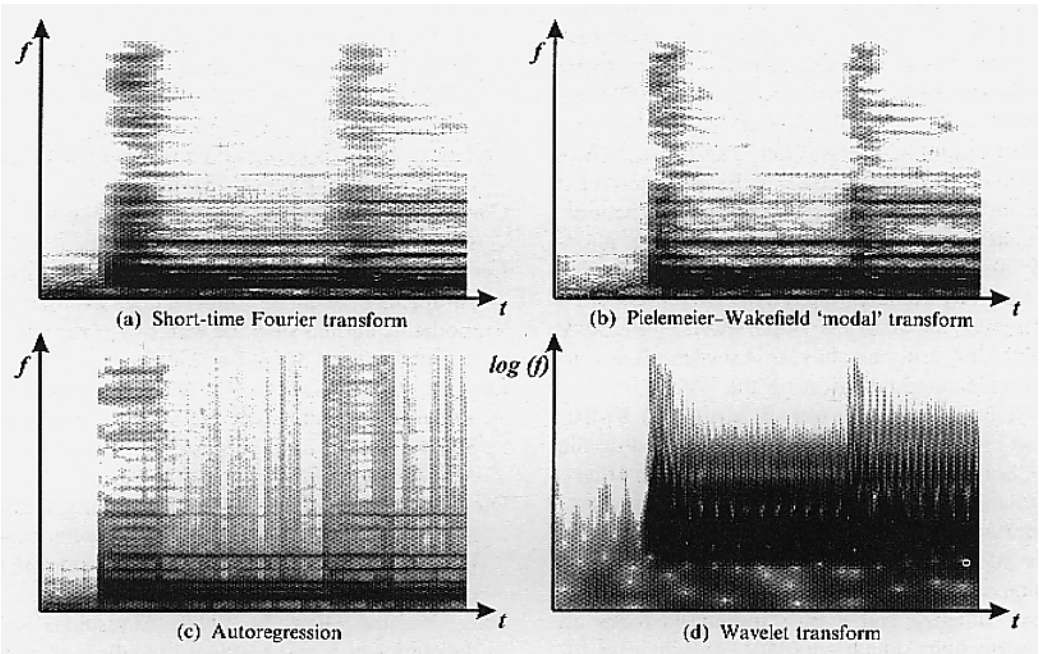


Figure 4. A recording (with hiss) of quiet followed by two guitar plucks in quick succession on the same note: (a) STFT: the louder partials are clear but the note onsets are blurred in time; (b) modal: the quieter partials are slightly better defined than the STFT and the note onsets are much better located; (c) AR: the presence of noise has greatly reduced the clarity in time and frequency compared to figure 3; and (d) WT: the dark vertical plumes give an accurate indication of note onset times, and the shape at the note onsets shows that there is a downward pitch slide onto the note in the first few milliseconds.

Additional Notes. In order to make each set of images comparable in figures 3 and 4, they were generated using the same time scale (except for the wavelet transform, which has a variable scale) and displayed with a greyscale indicating magnitude on a decibel scale. The WT was created using a complex Morlet wavelet (sinewave with Gaussian envelope) and the absolute magnitude is shown.

https://doi.org/10.1017/S1355771898009042 Published online by Cambridge University Press

ACKNOWLEDGEMENTS

The authors wish to express their gratitude to the Centre for Communications Research, which hosts the Digital Music Research Group at the University of Bristol, and to its director, Prof. Joe McGeehan, for supporting the research with generous access to facilities and the expertise of its staff. Thanks are also due to those who responded to e-mail and Usenet questions regarding the availability of software, in particular to Andrew Sterian (University of Michigan) whose suggestions were most helpful. On a more general note, thanks go to those who publish their algorithms for access over the Internet; it is a valuable and valued resource!

WWW RESOURCES

Many of the schemes described in this paper can be obtained from the World Wide Web (WWW) in algorithmic form, usually in C, C++ or for MATLAB (trademark of The MathWorks, Inc.). A small handful of these are presented here.

The FFT is so widely available that all programming languages and environments that support signal processing have an FFT function built into their libraries. This can be extended to the STFT by following the steps in section 2.1. Nevertheless, if you want to implement your own FFT, take a look at Burrus and Parks (1985) or Press, Teukolsky, Vetterling and Flannery (1992), which has an associated Web site at http://cfatab.harvard.edu/nr/. Alternatively there is a C library with documentation at http://theory.lcs.mit.edu/~fftw/ (untested by the authors).

For implementation of the wavelet transform, MATLAB has a Wavelet Toolbox which includes a graphical user interface and a variety of built-in graphing functions. Alternatively, there are also many Web sites with information about and software to implement the WT. The most memorable is probably http://www.wavelets.com/. Our favourite page, however, is http://www.amara.com/current/wavelet.html and this has an excellent index to wavelet transform-related resources on the Web.

Jeffrey O'Neill has created a set of MATLAB M-files that implement members of the Cohen class, including the Wigner and Choi–Williams distributions, amongst others. This can be downloaded by FTP from ftp://ftp.mathworks.com/pub/contrib/v5/signal/DiscreteTFDs/. For those in the UK, this site is mirrored at ftp://unix.hensa.ac.uk/mirrors/matlab/contrib/v5/signal/DiscreteTFDs/.

In the same category, but with a specialised focus on sound, http://www.eecs.umich.edu/asp/ contains easy-to-use MATLAB software that implements the Pielemeier–Wakefield 'modal' time–frequency distribution.

The WWW is constantly changing, so these links can only be verified at the time of writing. However, many search engines exist that should be able to locate these or similar resources. Usenet groups also exist for discussion on these subjects. Try comp.dsp or comp.soft-sys.matlab, for example.

All of the above sites (and more) can be accessed via the Digital Music Research Group Web pages. Visit http://www.fen.bris.ac.uk/elec/dmr/ and follow the links for this publication.

REFERENCES

Boashash, B. 1990. Time frequency signal analysis. In S. Haykin (ed.), *Advances in Spectral Analysis*, Vol. 1, pp. 418–517. Englewood Cliffs, NJ: Prentice Hall.

Boashash, B., and Frazer, G. 1992. Time-varying higher-order spectra, generalised Wigner–Ville distribution and the analysis of underwater acoustic data. *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing* (*ICASSP*), Vol. 5, pp. 193–6.

Boashash, B., and O'Shea, P. 1991. Time-varying higher order spectra. *Int. Conf. on Digital Signal Processing*, pp. 759–62.

Burrus, C. S., and Parks, T. W. 1985. *DFT/FFT and Convolution Algorithms – Theory and Implementation*. No. 1 in series *Topics in Signal Processing*. New York: John Wiley.

Chafe, C., Jaffe, D., Kashima, K., Mont-Reynaud, B., and Smith, J. O. 1985. Techniques for note identification in polyphonic music. *Proc. Int. Computer Music Conf.* (*ICMC*), pp. 399–405.

Choi, H. L., and Williams, W. J. 1989. Improved time–frequency representation of multicomponent signals using exponential kernels. *IEEE Trans. on Acoustics, Speech and Signal Processing* **37**(6): 862–71.

Claasen, T. A. C. M., and Mecklenbrauker, W. F. G. 1984. On the time–frequency discrimination of energy distributions: Can they look sharper than Heisenberg? *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing* (*ICASSP*), pp. 41.B.7.1–4.

Cohen, L. 1966. Generalised phase-space distribution functions. *J. Math. Phys.* **7**: 781–6.

Cohen, L. 1989. Time–frequency distributions – A review. *Proc of the IEEE*, **77**(7): 941–81.

Depalle, P., and Helie, T. 1997. Extraction of spectral peak parameters using a Short-Time Fourier Transform modeling and no sidelobe windows. *Proc. Int. Computer Music Conf.* (*ICMC*), pp. 256–9.

Depalle. P., and Tromp, L. 1996. An improved additive analysis method using parametric modelling of the Short-Time Fourier Transform. *Proc. Int. Computer Music Conf.* (*ICMC*), pp. 297–300.

Ding, Y., and Qian, X. 1997a. Estimating sinusoidal parameters of musical tones based on global waveform fitting. *IEEE Workshop on Multimedia Signal Processing* (awaiting publication).

Ding, Y., and Qian, X. 1997b. Sinusoidal and residual decomposition and residual modeling of musical tones using the QUASAR signal model. *Proc. Int. Computer Music Conf.* (*ICMC*), pp. 35–42.

Dubnov, S., and Rodet, X. 1997. Statistical modeling of sound aperiodicities. *Proc. Int. Computer Music Conf.* (*ICMC*), pp. 43–50.

Dubnov, S., Tishby, N., and Cohen, D. 1996. Influence of frequency modulating jitter on higher order moments of sound residual with applications to synthesis and classification. *Proc. Int. Computer Music Conf.* (*ICMC*), pp. 378–85.

https://doi.org/10.1017/S1355771898009042 Published online by Cambridge University Press

Fineberg, A. B., and Mammone, R. J. 1992. A method for instantaneous frequency tracking of multiple narrowband signals. *Signal Processing* **29**: 29–44.

Fonollosa, J. R., and Nikias, C. L. 1992. Analysis of transient signals using higher-order time–frequency distributions. *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing* (*ICASSP*), Vol. 5, pp. 197–200.

Gerr, N. L. 1988. Introducing a third order Wigner distribution. *Proc. of the IEEE* **76**: 290–2.

Guevara, R. C. L., and Wakefield, G. H. 1996. A modal distribution approach to piano analysis and synthesis. *Proc. Int. Computer Music Conf. (ICMC),* pp. 350–1.

Hope, C. J., and Furlong, D. J. 1997. Time–frequency distributions for timbre morphing: the Wigner distribution versus the STFT. *Proc. Brazilian Symp. on Computer Music*, pp. 99–110.

Jeong, J., and Williams, W. J. 1992. Kernel design for reduced interference distributions. *IEEE Trans. on Signal Processing* **40**(2): 402–12.

Jones, D. L., and Parks, T. W. 1992. A resolution comparison of several time–frequency representations. *IEEE Trans. on Signal Processing* **40**(2): 413–20.

Kay, S. M. 1988. *Modern Spectral Estimation: Theory & Application*. Englewood Cliffs, NJ: Prentice-Hall.

Kosko, B. 1994. *Fuzzy Thinking – The New Science of Fuzzy Logic*. London: Flamingo, HarperCollins.

Kronland-Martinet, R. 1988. The wavelet transform for analysis, synthesis and processing of speech and music sounds. *Computer Music Journal* **12**(4): 11–20.

Makhoul, J. 1975. Linear prediction: a tutorial review. *Proc. of the IEEE* **63**(4): 561–80.

Markel, J. D., and Gray, A. H. 1976. *Linear Prediction of Speech*. New York: Springer-Verlag.

Marple, S. L. 1987. *Digital Spectral Analysis with Applications*. Englewood Cliffs, NJ: Prentice-Hall.

Masri, P. 1996. *Computer Modelling of Sound for Transformation and Synthesis of Musical Signals*. PhD Thesis, University of Bristol.

Masri, P., and Bateman, A. 1994. Partial domain synthesis of music. In *Digital Signal Processing (UK)*, Vol. 1 (organised by New Electronics).

Masri, P., and Bateman, A. 1995. Identification of non-stationary audio signals using the FFT, with application to analysis-based synthesis of sound. *IEE Colloquium on Audio Engineering*, pp. 11.1–6.

Masri, P., Bateman, A., and Canagarajah, C. N. 1997. The importance of the time–frequency representation for sound/music analysis–resynthesis. *Organised Sound*, **2**(3): 207–14.

McAulay, R. J., and Quatieri, T. F. 1986. Speech analysis/synthesis based on a sinusoidal representation. *IEEE Trans. on Acoustics, Speech and Signal Processing* **34**(4): 744–54.

Mendel, J. M. 1991. Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. *Proc. of the IEEE* **79**(3): 277–305.

Pearson, E. R. S., and Wilson, R. G. 1990. Musical event detection from audio signals within a multiresolution framework. *Proc. Int. Computer Music Conf.* (*ICMC*), pp. 156–8.

Peleg, S., and Friedlander, B. 1996. Multicomponent signal analysis using the polynomial-phase transform. *IEEE Trans. on Aerospace and Electronic Systems* **32**: 378–87.

Pielemeier, W. J., and Wakefield, G. H. 1992. Time–frequency and time–scale analysis for musical transcription. *Proc. IEEE-SP Int. Symp. on Time–Frequency and Time–Scale Analysis*, pp. 421–4.

Pielemeier, W. J., and Wakefield, G. H. 1996. A high resolution time–frequency representation for musical instrument signals. *Journal of the Acoustical Society of America* **99**(4): 2,382–96.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. 1992. *Numerical Recipes in C: The Art of Scientific Computing*, 2nd edn. Cambridge: Cambridge University Press (published with corrections in 1995).

Rioul, O., and Vetterli, M. 1991. Wavelets and signal processing. *IEEE Signal Processing Magazine*, October issue, pp. 14–38.

Serra, X. 1989. *A System for Sound Analysis/Transformation/Synthesis Based on a Deterministic plus Stochastic Decomposition*. PhD Thesis, Stanford University.

Serra, X., and Smith, J. O. 1989. Spectral modeling synthesis. *Proc. Int. Computer Music Conf. (ICMC)*, pp. 281–4.

Ville, J. 1948. Theorie et applications de la notion de signal analytique. *Cables Transmission* **2A**: 61–74.

Wigner, E. P. 1932. On the quantum correction for thermodynamic equilibrium. *Physics Review* **40**: 749–59.

https://doi.org/10.1017/S1355771898009042 Published online by Cambridge University Press