Adaptive Suppression of Wigner Interference-Terms Using Shift-Invariant Wavelet Packet Decompositions

Israel Cohen*, Shalom Raz and David Malah

Department of Electrical Engineering, Technion — Israel Institute of Technology, Technion City, Haifa 32000, Israel

June 8, 1998

Abstract

The Wigner distribution (WD) possesses a number of desirable mathematical properties relevant to time-frequency analysis. However, the presence of interference terms renders the WD of multi-component signals extremely difficult to interpret. In this work, we propose adaptive suppression of interference terms using the *Shift-Invariant Wavelet Packet Decomposition*. A prescribed signal is expanded on its best basis and transformed into the Wigner domain. Subsequently, the interference terms are eliminated by adaptively thresholding the cross WD of interactive basis functions, according to their amplitudes and distance in an idealized time-frequency plane. We define a distance measure that weighs the Euclidean distance with the local distribution of the signal. The amplitude and distance thresholds control the cross-term interference, the useful properties of the distribution, and the computational complexity. The properties of the resultant *modified Wigner distribution* (MWD) are investigated, and its surpassing performance, in eliminating interference terms while still retaining high energy resolution, is compared with that of other existing approaches. It is shown that the proposed MWD is directly applicable to resolving multicomponent signals. Each component is determined as a partial sum of basis-functions over a certain equivalence class in the time-frequency plane.

Keywords: Shift-invariance; Best-basis; Wavelets; Wavelet-packets; Wigner distribution; Time-frequency; Interference terms

*Corresponding author. Tel.: 972 4 879 5033; fax: 972 4 879 5315; e-mail: cisrael@shoshan.technion.ac.il.

1 Introduction

The Wigner distribution (WD) has long been of special interest, because it possesses a number of desirable mathematical properties [4, 13], including maximal autocomponent concentration in the time-frequency plane. However, practical applications of the WD are restricted due to the presence of interference terms. These terms render the WD of multicomponent signals extremely difficult to interpret.

Several methods, developed to reduce noise and cross-components at the expense of reduced time-frequency energy concentration, employ smoothing kernels or windowing techniques [3, 24, 25, 22]. Unfortunately, the specific choice of kernel dramatically affects the appearance and quality of the resulting time-frequency representation. Consequently, adaptive representations [25, 1, 15] often exhibit performance far surpassing that of fixed-kernel representations. However, such methods are either computationally expensive or have a very limited adaptation range. Another approach striving for cross-term suppression with minimal resolution loss [34, 29] uses the Gabor expansion to decompose the WD. Interference terms are readily identified as cross WD of distinct basis functions. Here, a major drawback is the dependence of the performance on the choice of the Gabor window. An appropriate window selection depends on the data and may vary for different components of the same signal. Furthermore, distinct basis functions which are "close" in the time-frequency plane are often related to the same signal component. Accordingly, their cross-terms are not interpretable as interference terms, but rather may have a significant effect on the time-frequency resolution. Qian and Chen [30] proposed to decompose the WD into a series of Gabor expansions, where the order of the expansion is defined by the maximum degree of oscillation. They showed that such harmonic terms contribute minimally to the useful properties, but are directly responsible for the appearance of interference terms. In this case, the manipulation of cross-terms is equivalent to including cross-terms of Gabor functions whose *Manhattan distance* is smaller than a certain threshold. However, the order of the expansion has to be determined adaptively and generally depends on the local distribution of the signal. In [33], the signal is decomposed into frequency bands, and the Wigner distributions of all the subbands are superimposed. This attenuates interferences between

subbands, but still suffers interferences within the subbands. Therefore it is merely suitable for signals that possess a single component in each subband. Moreover, the exclusion of beneficial cross-terms, which join neighboring basis-functions, invariably degrades the energy concentration and may artificially split a given signal component into several frequency-bands.

In this paper, we propose an adaptive suppression of interference terms using the *Shift-Invariant Wavelet Packet Decomposition* (SIWPD) [6, 9]. A prescribed signal is expanded on its best SIWPD basis, and subsequently transformed into the Wigner domain. The interference terms are controlled by adaptively thresholding the cross WD of interactive basis functions according to their distance and amplitudes in an idealized time-frequency plane (an abstract representation where each basis-function is associated with a rectangular time-frequency tile, *e.g.*, [35]). When the distance-threshold is set to zero, the *modified* Wigner distribution (MWD) precludes any cross-terms, so essentially there is no interference terms but the energy concentration of the individual components is unacceptably low. When the amplitude-threshold is set to zero and the distance-threshold goes to infinity, the MWD converges to the conventional WD. By adjusting the distance and amplitude thresholds, one can effectively balance the cross-term interference, the useful properties of the distribution, and the computational complexity.

The distance measure in the idealized plane is related to a degree of adjacency by weighing the Euclidean time-frequency distance with the self distribution of the basis-functions. Since the basis-functions are adapted to the signal's local distribution, the thresholding of the cross-terms is also adapted to the local distribution of the signal. This dispenses with the need for local adjustments of the associated distance-threshold.

We note that the MWD constitutes an effective tool for resolving multicomponent signals. By defining equivalence classes in the time-frequency plane, we show that a prescribed component of a multicomponent signal can be determined as a partial sum of basis-functions. The signal components are well delineated in the time-frequency plane, and can be recovered from the energy distribution to within a constant phase factor.

This paper is organized as follows. In Section 2, we review the Wigner distribution, the origin

of interference terms and the relation to Cohen's class of distributions. In Section 3, we define the extended library of wavelet packets and demonstrate the shift-invariant properties of the SIWPD. Section 4 introduces the MWD. We present adaptive decompositions of the WD and show that the interference terms can be eliminated by thresholding the cross-terms according to a degree of adjacency in the idealized time-frequency plane. The general properties of the MWD are presented in Section 5. Inversion and uniqueness of the MWD are the subjects of Section 6.

2 The Wigner Distribution

Let $R_g(t,\tau)$ be the instantaneous auto-correlation of a complex signal $g(t)$, defined as

$$R_g(t,\tau) = g(t+\tau/2)g^*(t-\tau/2) \tag{1}$$

where g^* denotes the complex conjugate of g. The Wigner distribution of $g(t)$ is then defined as the Fourier transform (FT) of $R_g(t,\tau)$ with respect to the lag variable τ [36]:

$$W_g(t,\omega) = \int R_g(t,\tau)e^{-j\omega\tau}\,d\tau = \int g(t+\tau/2)g^*(t-\tau/2)e^{-j\omega\tau}\,d\tau\,, \tag{2}$$

or equivalently as

$$W_g(t,\omega) = \frac{1}{2\pi}\int G(\omega+\xi/2)G^*(\omega-\xi/2)e^{j\xi t}\,d\xi\,, \tag{3}$$

where $G(\omega)$ is the Fourier transform of $g(t)$ (the range of integrals is from $-\infty$ to $+\infty$ unless otherwise stated). The WD satisfies a large number of desirable mathematical properties [4, 13]. In particular, the WD is always real-valued, it preserves time and frequency shifts and satisfies the marginal properties:

$$\tfrac{1}{2\pi}\textstyle\int W_g(t,\omega)\,d\omega = |g(t)|^2 \tag{4}$$

$$\textstyle\int W_g(t,\omega)\,dt = |G(\omega)|^2\,. \tag{5}$$

One major drawback of the WD is the interference terms between signal components. Suppose that a given signal consists of two components,

$$g(t) = g_1(t) + g_2(t) \tag{6}$$

Then, by substituting this into (2) we have

$$W_g(t,\omega) = W_{g_1}(t,\omega) + W_{g_2}(t,\omega) + 2\,\mathrm{Re}\{W_{g_1,g_2}(t,\omega)\} \tag{7}$$

where

$$W_{g_1,g_2}(t,\omega) = \int g_1(t+\tau/2)g_2^*(t-\tau/2)e^{-j\omega\tau}\,d\tau \tag{8}$$

is the cross WD of $g_1(t)$ and $g_2(t)$. This shows that the WD of the sum of two signals is not the sum of their respective WDs, but has the additional term $2\,\mathrm{Re}\{W_{g_1,g_2}(t,\omega)\}$. This term is often called the interference term or cross term and it is often said to give rise to artifacts. However, one has to be cautious with the interpretations these words evoke, because any signal can be broken up into an arbitrary number of parts and the so-called cross terms are therefore not generally unique and do not characterize anything but our own division of a signal into parts [5]. There exists a natural decomposition where beneficial cross terms, which enhance the energy concentration, are distinguished from the undesirable interference terms, which obscure the time-frequency representation. This issue is addressed in Sections 4 and 6.

The WD, as well as the Choi-williams [3] and cone-kernel distributions [39] are members of a more general class of distributions, called *Cohen's class* [11]. Each member of this class is given by

$$C_g(t,\omega;\phi) = \frac{1}{2\pi}\iiint e^{j(-\theta t-\omega\tau+\theta u)}\phi(\theta,\tau)g(u+\tau/2)g^*(u-\tau/2)\,du\,d\theta\,d\tau \tag{9}$$

$$= \iint W_g(u,\xi)\Phi(t-u,\omega-\xi)\,du\,d\xi \tag{10}$$

where $\phi(\theta,\tau)$ is the kernel of the distribution, and $\Phi(t,\omega)$ is the 2-D Fourier transform of $\phi(\theta,\tau)$. Different kernels produce different distributions obeying different properties. For example, $\phi(\theta,\tau) = 1$, $e^{j\theta|\tau|/2}$, $e^{-\theta^2\tau^2/\sigma}$ and $w(\tau)\,|\tau|\,\sin(\alpha\theta\tau)/\alpha\theta\tau$ correspond to the Wigner, Page, Choi-Williams and Cone-kernel distributions, respectively [22]. The spectrogram, the squared magnitude of the short-time Fourier transform, is also a member of Cohen's class, since it can be obtained as a 2-D convolution of the WD's of the signal and the window.

The interference terms associated with the WD are highly oscillatory, whereas the auto terms are relatively smooth. Therefore, the reduced-interference distributions are designed to attenuate

the interference terms by smoothing the WD with a low-pass kernel [24, 37]. Unfortunately, this procedure invariably entails a loss of time-frequency concentration. Accordingly, high energy concentration and effective suppression of interference terms cannot be achieved simultaneously by merely smoothing the Wigner distribution.

3 The Extended Library of Wavelet Packets

Overcomplete libraries of waveforms that span redundantly the signal space encourage adaptive signal representations. Instead of representing a prescribed signal on a fixed basis, it is often useful to choose a suitable basis that facilitates a desired application, such as compression, identification, classification or noise removal (denoising) [27, 32, 35]. Of particular interest are the libraries of wavelet packet bases, which consist of translations and dilations of wavelet packets, and libraries of local trigonometric bases, comprising sines and cosines multiplied by smooth window functions [14, 35]. The basis functions are localized in the time-frequency plane, and organized in a binary tree structure where efficient search algorithms for the best basis can be implemented.

A serious drawback of the wavelet packet decomposition (WPD) and local cosine decomposition (LCD) [14] is the lack of shift-invariance. The expansion, as well as the information cost measuring its suitability for a particular application, may be significantly influenced by the alignment of the input signal with respect to the basis functions. Furthermore, the time-frequency tilings, produced by the best-basis expansions, do not generally conform to standard time-frequency energy distributions [9]. Hence we employ modified versions which induce shift-invariance, lower information cost and improved time-frequency resolution [6, 8, 7].

Let us specifically consider the *shift invariant wavelet packet decomposition* (SIWPD) [6, 9]. The library of bases is extended by introducing an additional degree of freedom that adjusts the time-localization of the basis functions. This degree of freedom is practically incorporated into the search algorithm as an adaptive even-odd down-sampling. That is, following the low-pass and high-pass filtering, when expanding a parent-node, we retain either all the odd samples or all the even samples, according to the choice which minimizes the cost function.

Let $\{\psi_n(t) : n \in \mathbb{Z}_+\}$ be a wavelet packet family [14] generated by

$$\psi_{2n}(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} h_k \psi_n(2t-k) \tag{11}$$

$$\psi_{2n+1}(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} g_k \psi_n(2t-k) \tag{12}$$

where $g_k = (-1)^k h_{1-k}$, and $\psi_0(t) \equiv \varphi(t)$ is an orthonormal scaling function, satisfying

$$\langle \varphi(t-p), \varphi(t-q) \rangle = \delta_{p,q}, \quad p, q \in \mathbb{Z}\,. \tag{13}$$

The extended library of wavelet packets is defined as the collection of all the orthonormal bases which are subsets of

$$\left\{ B_{\ell,n,m} \ : \ 0 \le \ell \le L,\, 0 \le n, m < 2^{L-\ell} \right\}, \tag{14}$$

where L denotes the finest resolution level, and

$$B_{\ell,n,m} \equiv \left\{ 2^{\ell/2} \psi_n \left[2^{\ell}(t - 2^{-L} m) - k \right] \ : \ 0 \le k < 2^{\ell} \right\}. \tag{15}$$

Although this library is larger than the standard wavelet packet library by a square power, it still retains a tree configuration facilitating fast search algorithms [6]. The additional parameter m provides a crucial degree of freedom, required for adjusting the time location of basis functions. When an analyzed signal is translated in time by $\tau = q \cdot 2^{-L}$ ($q \in \mathbb{Z}$), a new best-basis is selected whose elements are also translated by τ compared to the former best-basis. Thus the expansion coefficients remain unchanged, and the corresponding representation is time-shifted by the same period.

The relative advantages of SIWPD over WPD are as follows [9]: 1) Shift-invariance; 2) Lower information cost; 3) Improved time-frequency resolution; 4) A more stable information cost across a prescribed data set; 5) Controllable computational complexity (down to $O(N log_2 N)$) at the expense of the information cost. To demonstrate the shift-invariant properties of the SIWPD and its enhanced time-frequency representation, we compare the expansions of signals $g(t)$ (Fig. 1) and $g(t - 2^{-6})$. These signals contain $2^7 = 128$ samples, and are identical to within 2 time-shifted samples. For definiteness, we choose C_{12} to serve as the scaling function (C_{12} corresponds to 12-tap

coiflet filters [17, page 261] [18]) and the Shannon entropy as the information cost function, defined by $\mathcal{M}(\{x_i\}) = -\sum_{i:x_i\neq 0} x_i^2 \log x_i^2$ (*e.g.*, [14]). Figs. 2 and 3 display the best-basis expansions under the WPD and the SIWPD algorithms, respectively. The sensitivity of WPD to temporal shifts is obvious, while the best-basis SIWPD representation is indeed shift-invariant and characterized by a lower entropy and improved time-frequency resolution.

4 Adaptive Decomposition of the Wigner Distribution and Elimination of Interference Terms

In this section, we present adaptive decompositions of the WD using overcomplete libraries of orthonormal bases. The Wigner domain interference terms are controlled adaptively by thresholding the cross WD of interactive basis functions according to their degree of adjacency in the idealized time-frequency plane. In particular, we demonstrate the superiority of the modified distribution by employing the shift-invariant wavelet packet decomposition.

Let $\mathcal{B}$ denote an overcomplete library of orthonormal bases, and let

$$g(t) = \sum_{\lambda\in\mathbf{N}} c_\lambda \varphi_\lambda(t)\,, \quad \{\varphi_\lambda\}_{\lambda\in\mathbf{N}} \in \mathcal{B} \tag{16}$$

be the best-basis expansion of the signal g. Then by inserting (16) into (2), the Wigner distribution of g can be written as

$$W_g(t,\omega) = \sum_{\lambda,\lambda'\in\mathbf{N}} c_\lambda c^*_{\lambda'} W_{\varphi_\lambda,\varphi_{\lambda'}}(t,\omega) \tag{17}$$

$$= \sum_{\lambda\in\mathbf{N}} |c_\lambda|^2 W_{\varphi_\lambda}(t,\omega) + 2\sum_{\lambda>\lambda'} \mathrm{Re}\{c_\lambda c^*_{\lambda'} W_{\varphi_\lambda,\varphi_{\lambda'}}(t,\omega)\}\,. \tag{18}$$

Equation (18) partitions the traditional WD into two subsets. The superposition of the auto WD of the basis-functions, represents the auto-terms. The second summation, comprising cross WD of basis-functions, represents the cross-terms. Cross terms associated with the Wigner distribution, and other bilinear distributions, should not be always interpreted as interference terms. Any signal

can be sub-divided in an infinite number of ways, each generating different cross terms. Therefore, we need to distinguish between generally undesirable interference-terms and beneficial cross-terms that primarily enhance useful time-frequency features.

The cross WD of distinct basis-functions is oscillating and centered in the midway of the corresponding auto-terms [20, 21]. The oscillation rate is proportional to the distance between the auto-terms. On the other hand, useful properties such as the time marginal, frequency marginal, energy concentration and the instantaneous frequency property [13], are achieved by *averaging* the Wigner distribution. Therefore the overall contribution of each cross-term component is *inversely* proportional to the distance between the corresponding basis-functions in the time-frequency plane [30, 31].

A useful distance measure between pairs of basis-functions is obtainable in the idealized time-frequency plane [10]. Recall that in the idealized plane, each basis-function is symbolized by a rectangular cell (tile) whose area is associated with Heisenberg's uncertainty principle, and its shade is proportional to the corresponding squared coefficient [35]. We define the distance between a pair of basis-functions by

$$d(\varphi_\lambda, \varphi_{\lambda'}) = \left[\frac{(\bar{t}_\lambda - \bar{t}_{\lambda'})^2}{\Delta t_\lambda \Delta t_{\lambda'}} + \frac{(\bar{\omega}_\lambda - \bar{\omega}_{\lambda'})^2}{\Delta \omega_\lambda \Delta \omega_{\lambda'}} \right]^{1/2} \tag{19}$$

where $(\bar{t}_\lambda, \bar{\omega}_\lambda)$ is the position of the cell associated with φ_λ; Δt_λ and $\Delta \omega_\lambda$ denote the time and frequency widths (uncertainties), respectively. Similar notations apply to $\varphi_{\lambda'}$.

Since the best basis tends to represent the signal using a relatively small number of significant expansion coefficients, the summations in (18) can be restricted to basis-functions whose coefficients are above a prescribed cutoff, and to pairs that are "close" (sufficiently small values of $d(\varphi_\lambda, \varphi_{\lambda'})$). The modified Wigner distribution (MWD) is then given by

$$T_g(t,\omega) = \sum_{\lambda \in \Lambda} |c_\lambda|^2 W_{\varphi_\lambda}(t,\omega) + 2 \sum_{\{\lambda,\lambda'\} \in \Gamma} \mathrm{Re}\{c_\lambda c_{\lambda'}^* W_{\varphi_\lambda, \varphi_{\lambda'}}(t,\omega)\} \tag{20}$$

where

$$\Lambda = \{\lambda \,|\, |c_\lambda| \geq \varepsilon M\}, \quad M \equiv \max_\lambda\{|c_\lambda|\} \tag{21}$$

$$\Gamma = \{\{\lambda, \lambda'\} \,|\, 0 < d(\varphi_\lambda, \varphi_{\lambda'}) \leq D,\, |c_\lambda c_{\lambda'}| \geq \varepsilon^2 M^2\}\,. \tag{22}$$

ε and D denote thresholds of relative amplitude and time-frequency distance, respectively. When $D = 0$, the MWD precludes any cross-terms, so essentially there are no interference terms but the energy concentration of the individual components is generally low. As D goes to infinity and ε goes to zero, the MWD converges to the conventional WD. By adjusting the distance and amplitude thresholds, one can effectively balance the cross-term interference, the useful properties of the distribution, and the computational complexity.

Here, rather that the usual Euclidean distance ($\sqrt{(\bar{t}_\lambda - \bar{t}_{\lambda'})^2 + (\bar{\omega}_\lambda - \bar{\omega}_{\lambda'})^2}$) or the Manhattan distance ($|\bar{t}_\lambda - \bar{t}_{\lambda'}| + |\bar{\omega}_\lambda - \bar{\omega}_{\lambda'}|$) [30], we use the measure defined in (19), which weighs the time-frequency distance with the self distribution of the basis elements. Since the basis elements are selected to best match the signal's local distribution, such a distance measure implicitly characterizes the signal itself. Accordingly, the thresholding of the cross-terms is also adapted to the local distribution of the signal, dispensing with the need for local adjustments of the associated distance-threshold.

The extended library of wavelet packets includes basis-functions of the form

$$\psi_{\ell,n,m,k}(t) = 2^{\ell/2}\psi_n\left[2^\ell(t - 2^{-L}m) - k\right] \tag{23}$$

where ℓ is the resolution-level index ($0 \leq \ell \leq L$), n is the frequency index ($0 \leq n < 2^{L-\ell}$), m is the shift index ($0 \leq m < 2^{L-\ell}$) and k is the position index ($0 \leq k < 2^\ell$). Each basis-function is symbolically associated with a rectangular tile in the time-frequency plane which is positioned about

$$\bar{t} = 2^{-\ell}k + 2^{-L}m + (2^{L-\ell} - 1)C_h + (C_h - C_g)R(n)\,, \tag{24}$$

$$\bar{f} = 2^{\ell-L}[GC^{-1}(n) + 0.5]\,, \tag{25}$$

where

$$C_h \triangleq \frac{1}{\|h\|^2} \sum_{k \in \mathbb{Z}} k|h_k|^2 \,, \quad C_g \triangleq \frac{1}{\|g\|^2} \sum_{k \in \mathbb{Z}} k|g_k|^2 \,, \tag{26}$$

are respectively the energy centers of the low-pass and high-pass quadrature filters [35, 19], $R(n)$ is an integer obtained by bit reversal of n in a $L - \ell$ bits binary representation, and GC^{-1} is the inverse Gray code permutation. The width and height of the tile are given by

$$\Delta t = 2^{-\ell} \,, \quad \Delta f = 2^{\ell - L} \,. \tag{27}$$

For a given signal, the SIWPD yields the best expansion in the extended library with respect to an additive cost function. It is demonstrated below that it would be advantageous to search for the best *orthonormal* basis using an *extended* library of wavelet packets, rather than using computationally expensive algorithms for searching optimal (not necessarily orthonormal) expansions in a *conventional* wavelet packet library. The extended library provides flexibility in expanding the signal, while the orthonormality contributes to a manageable complexity of the search procedure.

For example, Fig. 4 depicts the Wigner distribution and spectrogram for $g(t)$. The signal $g(t)$ (Fig. 1) comprises a short pulse, a tone and a component with nonlinear frequency modulation. The spectrogram has no interference terms, at the expense of comparatively poor energy concentration. The optimal expansions of $g(t)$ obtained by the Method of Frames (minimum l^2 norm) [16], Matching Pursuit [26], Basis Pursuit (minimum l^1 norm) [2] and WPD are illustrated in Fig. 5. While these algorithms use the conventional library of wavelet packets and fail to represent the signal efficiently, the SIWPD (Fig. 5(f)) facilitates an efficient representation by a small number of coefficients. Furthermore, its computational complexity ($\sim 3,580$ multiplications) is significantly lower than those associated with the Matching Pursuit ($\sim 44,800$ multiplications) and the Basis Pursuit ($\sim 331,500$ multiplications).

Fig. 6 illustrates the MWD for $g(t)$, using various distance-thresholds. When $D = 0$, there are no interference terms, but the energy concentration of individual components is insufficient. $D = 2$ leads to improved energy concentration, yet, no significant interference terms are present. As D gets larger, the interference between components becomes visible and the MWD converges to the

conventional WD (*cf.* Fig. 4(a)). An acceptable compromise is usually found between $D = 1.5$ and $D = 2.5$.

Fig. 7(a) shows the MWD for $g(t)$, obtained via the SIWPD with thresholds $D = 2$ and $\varepsilon = 0.1$. Figs. 7(b), (c), (d), (e) and (f) describe, respectively, the WD, the Smoothed pseudo Wigner distribution, the Choi-Williams distribution, the cone-kernel distribution and the reduced interference distribution [22]. Clearly, the SIWPD based MWD achieves high time-frequency resolution, and is superior in eliminating interference terms.

The particular basis, selected for representing a prescribed signal, plays an important role in the MWD. As long as the "best" basis elements are localized in time-frequency and reasonably matched to the local distribution of the signal, each signal component is characteristically represented by a few significant elements. Thus, by restricting the cross-terms to neighboring basis-functions, we eliminate interference terms between distinct components, and even within components having a nonlinear frequency modulation. On the other hand, whenever the signal is arbitrarily decomposed into elements that have no relation to the actual signal distribution, the performance of the MWD may deteriorate. The SIWPD constitutes an efficient algorithm for selecting the most suitable basis. Similarly to standard WPD, the SIWPD library is generated by a single "mother-wavelet" [9]. Although the library is flexible and versatile enough to describe various local features of the signal, the choice of the mother-wavelet may affect the eventual performance.

The signal $g(t)$, depicted in Fig. 1, can be represented by seven basis-function, belonging to the extended wavelet packet library with C_{12} as the mother-wavelet (*cf.* Fig. 3(a)). If the SIWPD utilizes decomposition filters that correspond to a different mother-wavelet, then the entropy of the representation is expected to be higher and correspondingly the performance of the MWD will deteriorate. Figs. 8 and 9 illustrate best-basis expansions and MWDs for $g(t)$, obtained by the SIWPD with D_6 and S_9 as mother-wavelets (D_6 corresponds to 6-tap Daubechies least asymmetric wavelet filters, and S_9 corresponds to 9-tap Daubechies minimum phase wavelet filters [17, pp. 195,198]). A comparison with Figs. 5(f) and 6(b) shows that despite variations in the time-frequency tilings, the MWD managed to delineate the components of the signal and effectively

eliminate the interference terms.

Fig. 10 illustrates the best-basis expansion and MWD for $g(t)$, obtained using an extended library of local trigonometric bases and a corresponding best-basis search algorithm, namely the *shift-invariant adapted-polarity local trigonometric decomposition* (SIAP-LTD) [7, 8]. Here, the basis-functions fail to represent the signal efficiently. We may compare the entropy (= 2.81) with that obtained with the SIWPD (1.88 with C_{12}, 2.09 with D_6, and 2.32 with S_9). The reduced performance of the SIAP-LTD for this particular signal stems from the fact that short pulses, expanded on local trigonometric bases, require a large number of decomposition levels [8]. This entails a steeper rising cutoff function, and consequently basis-functions which are less localized in frequency [35]. Notice that the "visual quality" of the MWD is well correlated with the entropy attained by the best basis expansion. Lower entropy generally yields "better" (well delineated components, high resolution and concentration) MWD. It appears that "entropy" can serve as a reasonable measure for a quantitative comparison between MWDs.

5 General Properties

In this section we investigate the MWD in more detail.

Realness: The MWD is always real, even if the signal or the basis functions are complex.

$$T_g^* = T_g \tag{28}$$

This property is a direct consequence of the realness of the Wigner distribution.

Shift-Invariance: Shifting a signal by $\tau = k \cdot 2^{-J} \quad (k, J \in \mathbb{Z})$, where J is finest resolution level of the best-basis decomposition, entails an identical shift of the MWD, *i.e.*,

$$\text{if} \quad \tilde{g}(t) = g(t-\tau) \quad \text{then} \quad T_{\tilde{g}}(t,\omega) = T_g(t-\tau,\omega)\,. \tag{29}$$

This property follows from the shift-invariance property of the best-basis decomposition. To see this, let $g(t) = \sum_\lambda c_\lambda \varphi_\lambda(t)$ be the best-basis expansion of g, and let $\tilde{g}(t) = g(t-\tau)$, $\tau = k \cdot 2^{-J}$.

Then, using the shift-invariance of the best-basis decomposition, we have

$$\tilde{g}(t) = \sum_{\lambda} \tilde{c}_\lambda \tilde{\varphi}_\lambda(t) = \sum_{\lambda} c_\lambda \varphi_\lambda(t-\tau) \tag{30}$$

is the best-basis expansion of $\tilde{g}$, *i.e.*, the best-basis for $\tilde{g}$ is identical to within a time-shift τ to the best-basis for g, and the corresponding expansion coefficients are the same. The MWD of g and $\tilde{g}$ are given by

$$T_g(t,\omega) = \sum_{\lambda\in\Lambda} |c_\lambda|^2 W_{\varphi_\lambda}(t,\omega) + 2 \sum_{\{\lambda,\lambda'\}\in\Gamma} \mathrm{Re}\{c_\lambda c^*_{\lambda'} W_{\varphi_\lambda,\varphi_{\lambda'}}(t,\omega)\}\,, \tag{31}$$

$$T_{\tilde{g}}(t,\omega) = \sum_{\lambda\in\tilde{\Lambda}} |c_\lambda|^2 W_{\varphi_\lambda}(t-\tau,\omega) + 2 \sum_{\{\lambda,\lambda'\}\in\tilde{\Gamma}} \mathrm{Re}\{c_\lambda c^*_{\lambda'} W_{\varphi_\lambda,\varphi_{\lambda'}}(t-\tau,\omega)\}\,, \tag{32}$$

where we used the shift-invariance property,

$$W_{\tilde{\varphi}_\lambda,\tilde{\varphi}_{\lambda'}}(t,\omega) = W_{\varphi_\lambda,\varphi_{\lambda'}}(t-\tau,\omega)\,, \qquad W_{\tilde{\varphi}_\lambda}(t,\omega) = W_{\varphi_\lambda}(t-\tau,\omega)\,. \tag{33}$$

Now, since the expansion coefficients of g and $\tilde{g}$ are identical ($\tilde{c}_\lambda = c_\lambda$), and the time-frequency distance between pairs of basis-functions remains unchanged ($d(\tilde{\varphi}_\lambda,\tilde{\varphi}_{\lambda'}) = d(\varphi_\lambda,\varphi_{\lambda'})$), the sets $\tilde{\Lambda}$ and $\tilde{\Gamma}$ are identical to Λ and Γ, respectively. It is therefore concluded that $T_{\tilde{g}}(t,\omega) = T_g(t-\tau,\omega)$.

Symmetry in Frequency: Real signals have symmetrical spectra. For symmetric spectra, the Wigner distribution is symmetric in the frequency domain,

$$W_g(t,-\omega) = W_g(t,\omega), \quad W_{g,s}(t,-\omega) = W_{s,g}(t,\omega)\,. \tag{34}$$

Thus, for real signals and real basis-functions, the MWD retains the same symmetries, *i.e.*,

$$T_g(t,-\omega) = T_g(t,\omega)\,. \tag{35}$$

Symmetry in Time: For symmetrical signals, the Wigner distribution is symmetric in the time domain,

$$W_g(-t,\omega) = W_g(t,\omega), \quad W_{g,s}(-t,\omega) = W_{s,g}(t,\omega)\,. \tag{36}$$

However, the MWD is not necessarily symmetric, since the best-basis decomposition is generally asymmetric. Still, confining ourselves to symmetric basis-functions (entailing either biorthogonal or complex-valued basis-functions [17]) and restricting $\mathcal{B}$, the library of bases, to those bases satisfying

$$\{\varphi_\lambda\}_{\lambda\in\mathbf{N}} \in \mathcal{B} \quad \Longrightarrow \quad \{\varphi_\lambda(t)\}_{\lambda\in\mathbf{N}} = \{\varphi_\lambda(-t)\}_{\lambda\in\mathbf{N}} \,,$$

the best-basis decomposition becomes symmetric, rather than shift-invariant. In that case, the MWD is symmetric in time,

$$\begin{aligned} T_g(-t,\omega) &= \sum_{k\in\Lambda} |c_k|^2 W_{\varphi_k}(-t,\omega) + 2 \sum_{\{k,\ell\}\in\Gamma} \mathrm{Re}\{c_k c_\ell^* W_{\varphi_k,\varphi_\ell}(-t,\omega)\} \\ &= \sum_{k'\in\Lambda} |c_{k'}|^2 W_{\varphi_{k'}}(t,\omega) + 2 \sum_{\{k',\ell'\}\in\Gamma} \mathrm{Re}\{c_{k'} c_{\ell'}^* W_{\varphi_{k'},\varphi_{\ell'}}(t,\omega)\} \\ &= T_g(t,\omega) \,. \end{aligned}$$

Total Energy: Integrating the general form of the MWD with respect to time and frequency shows that the total energy is bounded by the energy of the signal:

$$\frac{1}{2\pi}\int dt \int d\omega \; T_g(t,w) = \sum_{\lambda\in\Lambda} |c_\lambda|^2 \le \sum_{\lambda} |c_\lambda|^2 = \|g\|^2 \tag{37}$$

where we have used

$$\frac{1}{2\pi}\int dt \int d\omega \; W_{\varphi_k,\varphi_\ell}(t,w) = \langle \varphi_k, \varphi_\ell \rangle = \delta_{k,\ell} \,.$$

Observe that the difference between the total energy and the energy of the signal essentially stems from the smallest expansion coefficients. In fact, if we set the amplitude threshold (ε) to zero, the set of indices Λ runs over all the basis-functions, and thus the total energy equals the energy of the signal.

Positivity: The interpretation of the conventional WD as a pointwise time-frequency energy density is generally restricted by the uncertainty principle and by the fact that the WD may locally assume negative values [28, 12, 23]. However, the nonnegativity and interference terms are closely related, and in many cases the suppression of interference terms accompanies reduction of negative

values in magnitude [24]. Thus, reduction of the interference terms associated with the WD, entails comparable attenuation of its negative values.

6 Inversion and Uniqueness

In this section we show that the components that comprise a given signal can be recovered from the MWD, to within an arbitrary constant phase factor and to within the errors caused by neglecting low weight basis constituents.

6.1 Equivalence Classes in the Time-Frequency Plane

A multicomponent signal is one that has well delineated regions in the time-frequency plane. Examples of multicomponent signals are illustrated in Fig. 11. One of the advantages of the MWD is its capability to resolve a multicomponent signal into disjoint time-frequency regions.

Definition 1 *Let* $X = \Lambda \cup \{\lambda \mid \{\lambda, \lambda'\} \in \Gamma$ *for some* $\lambda' \in \Lambda\}$ *be the indices set of the significant basis functions, i.e., the basis functions which contribute to the MWD. A pair of indices* $k, \ell \in X$ *are said to be equivalent, denoted by* $k \sim \ell$*, if* $k \equiv \ell$ *or alternatively there exists a finite series* $\{\lambda_i\}_{i=1}^{N}$ *such that* $\{\lambda_i, \lambda_{i+1}\} \in \Gamma$ *for* $i = 1, 2, \ldots, N-1$ *and* $\{k, \lambda_1\}, \{\ell, \lambda_N\} \in \Gamma$.

Clearly, $\sim$ is an equivalence relation on X, since it is reflexive ($k \sim k$ for all $k \in X$) symmetric ($k \sim \ell$ implies $\ell \sim k$) and transitive ($k \sim \ell$ and $\ell \sim m$ imply $k \sim m$). The equivalence relation means that the corresponding basis-functions are linked in the time-frequency plane by a series of consecutive adjacent basis-functions.

Denote by

$$\Lambda_k = \{\lambda \in X \mid \lambda \sim k\} \tag{38}$$

the equivalence class for $k \in X$. Then, for any $k, \ell \in X$ either $\Lambda_k = \Lambda_\ell$ or $\Lambda_k \cap \Lambda_\ell = \emptyset$. Hence, $\{\Lambda_k | k \in X\}$ forms a partition of X, and each equivalence class can be related to a single component

of the signal. The number of components which comprise the signal g is determined by the number of distinct equivalence classes in X.

For example, refer to the multicomponent signal $s(t)$, depicted in Fig. 12. Its best-basis decomposition (Fig. 13) shows that it can be expressed as the sum of six basis-functions: $s(t) = \sum_{k=1}^{6} c_k \varphi_k$. In this case, with an appropriate distance-threshold ($D = 2$), we obtain

$$\Lambda = \{1,2,3,4,5,6\} = X\,,$$
$$\Gamma = \{\{1,2\}\ \{2,3\}\ \{4,5\}\ \{5,6\}\}\,.$$

Thus there are two distinct equivalence classes on X,

$$\Lambda_1 = \Lambda_2 = \Lambda_3 = \{1,2,3\} \equiv \Lambda_I\,,$$
$$\Lambda_4 = \Lambda_5 = \Lambda_6 = \{4,5,6\} \equiv \Lambda_{II}\,.$$

Accordingly, we presume that the signal consists of two components:

$$s = s_I + s_{II}$$

where

$$s_I = \sum_{k \in \Lambda_I} c_k \varphi_k\,, \quad s_{II} = \sum_{k \in \Lambda_{II}} c_k \varphi_k\,.$$

These components, depicted in Fig. 14, are associated with the two well delineated time-frequency regions in the MWD domain (Fig. 15(a)).

6.2 Recovering the Components of a Multicomponent Signal

The components of a multicomponent signal are given by the partial sums of basis-functions with respect to equivalence classes. They can also be recovered from the MWD to within an arbitrary constant phase factor in each signal component, and to within errors generated by neglecting small basis constituents (small auto-terms, small cross-terms, as well as interference terms that correspond to distant basis functions).

Lemma 1 *Let $\{\varphi_k\}_{k\in\mathbb{N}}$ be the best basis for $g(t)$, and let $W_{k,\ell} \equiv W_{\varphi_k,\varphi_\ell}$ be the cross Wigner distribution of pairs of basis-functions. Then the set $\{W_{k,\ell}\}_{k,\ell\in\mathbb{N}}$ is an orthonormal basis for $L_2(\mathbb{R}^2)$, and the expansion coefficients for the MWD are given by*

$$c_{k,\ell} = \langle T_g, W_{k,\ell}\rangle = \begin{cases} c_k c_\ell^*, & \text{if } \ k=\ell\in\Lambda \ \text{ or } \ \{k,\ell\}\in\Gamma\,, \\ 0, & \text{otherwise }\,, \end{cases} \tag{39}$$

where

$$\langle T_g, W_{k,\ell}\rangle \stackrel{\triangle}{=} \frac{1}{2\pi}\iint T_g(t,\omega) W_{k,\ell}^*(t,\omega)\, dt\, d\omega$$

Proof: We first need to show that the system $\{W_{k,\ell}\}_{k,\ell\in\mathbb{N}}$ is orthonormal and complete in $L_2(\mathbb{R}^2)$. Orthonormality is given by

$$\begin{aligned} \langle W_{k,\ell}, W_{m,n}\rangle &= \frac{1}{2\pi}\int dt \int d\omega \int d\tau \int d\tau'\ \varphi_k(t+\frac{\tau}{2})\varphi_\ell^*(t-\frac{\tau}{2})\varphi_m^*(t+\frac{\tau'}{2})\varphi_n(t-\frac{\tau'}{2})e^{-j\omega(\tau-\tau')} \\ &= \int dt \int d\tau\ \varphi_k(t+\frac{\tau}{2})\varphi_\ell^*(t-\frac{\tau}{2})\varphi_m^*(t+\frac{\tau}{2})\varphi_n(t-\frac{\tau}{2}) \\ &= \langle\varphi_k,\varphi_m\rangle\,\langle\varphi_n,\varphi_\ell\rangle = \delta_{k,m}\delta_{\ell,n}\,, \end{aligned}$$

and completeness is satisfied by

$$\begin{aligned} &(2\pi)^{-1}\sum_{k,\ell\in\mathbb{N}} W_{k,\ell}(t,\omega)W_{k,\ell}^*(t',\omega') \\ &= \frac{1}{2\pi}\sum_{k,\ell\in\mathbb{N}}\int d\tau\int d\tau'\ \varphi_k(t+\frac{\tau}{2})\varphi_\ell^*(t-\frac{\tau}{2})e^{-j\omega\tau}\varphi_k^*(t'+\frac{\tau'}{2})\varphi_\ell(t'-\frac{\tau'}{2})e^{+j\omega'\tau'} \\ &= \frac{1}{2\pi}\int d\tau\int d\tau'\ \delta(t-t'-\frac{\tau}{2}+\frac{\tau'}{2})\delta(t-t'+\frac{\tau}{2}-\frac{\tau'}{2})e^{-j\omega\tau+j\omega'\tau'} \\ &= \delta(t-t')\cdot\frac{1}{2\pi}\int e^{j\tau(\omega'-\omega)}d\tau = \delta(t-t')\delta(\omega-\omega')\,. \end{aligned}$$

Now, the MWD can be expressed in the following form:

$$T_g = \sum_{k\in\Lambda}|c_k|^2 W_{k,k} + 2\sum_{\{k,\ell\}\in\Gamma}\mathrm{Re}\{c_k c_\ell^* W_{k,\ell}\} = \sum_{k,\ell\in\mathbb{N}} c_{k,\ell}W_{k,\ell}\,. \tag{40}$$

Therefore, by the uniqueness of the expansion, the relation in Eq. (39) holds. □

Let $k \in \Lambda$, and let Λ_k be its equivalence class. Then for any $\ell \in \Lambda_k$ there exists a finite series $\{\lambda_i\}_{i=1}^N$ such that $\{\lambda_i, \lambda_{i+1}\} \in \Gamma$ for $i = 1, \ldots, N-1$ and $\{k, \lambda_1\}, \{\ell, \lambda_N\} \in \Gamma$. By Eq. (39) we have

$$|c_k|^2 = \langle T_g, W_{k,k} \rangle \ , \tag{41}$$

$$c_k c_{\lambda_1}^* = \langle T_g, W_{k,\lambda_1} \rangle \ , \tag{42}$$

$$c_{\lambda_i} c_{\lambda_{i+1}}^* = \langle T_g, W_{\lambda_i,\lambda_{i+1}} \rangle \ , \qquad i = 1, \ldots, N-1 \ , \tag{43}$$

$$c_{\lambda_N} c_\ell^* = \langle T_g, W_{\lambda_N,\ell} \rangle \ , \tag{44}$$

which shows that c_ℓ has a recursive relation to c_k, and c_k can be recovered from the MWD up to a phase factor. Accordingly, each component of the signal can also be recovered up to an arbitrary constant phase factor by

$$s_k = \sum_{\ell \in \Lambda_k} c_\ell \varphi_\ell \ . \tag{45}$$

The constant phase factor in each component of the signal clearly drops out when we calculate the MWD (as it does for the WD). Therefore, it cannot be recovered. Summation of distinct signal components generally yields a different signal that has the same MWD. For example, we observed that the signal s in Fig. 12 consists of two components, $s = s_I + s_{II}$. The difference of these components, generates another signal $\tilde{s} = s_{II} - s_I$ (*cf.* Fig. 16), which has the same MWD as s. In some applications, such as pattern recognition, it is actually desirable that signals consisting of the same components will be identified, irrespective of their relative phase. The MWD provides an efficient technique for doing so.

7 Summary

The main issue investigated in this paper is that of adaptive decompositions of the Wigner distribution and suppression of interference terms, leading to a newly defined modified Wigner space. A prescribed signal is expanded on its best basis using the SIWPD, and subsequently transformed into the Wigner domain. The resulting distribution is modified by restricting the auto-terms

and cross-terms to basis-functions whose normalized coefficients are larger in magnitude than a certain amplitude-threshold ε, and to pairs whose time-frequency distance is smaller than a specified critical distance D. We have shown that the distance and amplitude thresholds control the cross-term interference, the useful properties of the distribution, and the computational complexity. A smaller distance-threshold better eliminates the interference terms, but tends to lower the energy concentration. A larger distance-threshold improves the time-frequency resolution at the expense of retaining additional interference terms. When the amplitude-threshold is set to zero and the distance-threshold goes to infinity, the MWD converges to the conventional WD. Appropriate threshold values ($D \approx 2$, $\varepsilon \approx 0.1$) combine high resolution, high concentration and suppressed cross-term interference at a manageable computational complexity.

We have compared alternative libraries, showing that interference terms between distinct components can be efficiently eliminated, as long as the localization properties of basis elements aptly resemble that of the signal. The visual quality of the MWD is well correlated with the entropy attained by the best basis expansion, facilitating a quantitative comparison between energy distributions. The MWD is thus effective for resolving multicomponent signals. The signal components are determined as partial sums of basis-functions over certain equivalence classes in the time-frequency plane.

The proposed methodology is extendable to other distributions (*e.g.*, the Cohen class) and other "best-basis" decompositions. However, the properties of the resulting modified forms clearly depend on the particular distribution, library of bases and best-basis search algorithm which are employed.

References

[1] R. G. Baraniuk and D. J. Jones, "A signal-dependent time-frequency representation: Fast algorithm for optimal kernel design", IEEE Trans. on Signal Processing, Vol. 42, No. 1, Jan. 1994, pp. 134–146.

[2] S. Chen and D. L. Donoho, "Atomic decomposition by basis pursuit", Technical Report, Dept.of Statistics, Stanford Univ., Feb. 1996 (http://playfair.stanford.edu /reports/chen_s).

[3] H. I. Choi and W. J. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels", IEEE Trans. on Acoust., Speech and Signal Processing, Vol. 37, No. 6, June 1989, pp. 862–871.

[4] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "The Wigner distribution — a tool for time-frequency signal analysis. Part I. Continuous-time signals", Philips J. Res., Vol. 35, 1980, pp. 217–250.

[5] A. Cohen and J. Kovačević, "Wavelets: The mathematical background", Proc. IEEE, Vol. 84, No. 4, 1996, pp. 514–522.

[6] I. Cohen, S. Raz and D. Malah, "Shift invariant wavelet packet bases", Proc. of the 20th IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-95, Detroit, Michigan, 8–12 May 1995, pp. 1081–1084.

[7] I. Cohen, S. Raz, D. Malah and I. Schnitzer, "Best-basis algorithm for orthonormal shift-invariant trigonometric decomposition", Proc. of the 7th IEEE Digital Signal Processing Workshop, DSPWS'96, Loen, Norway, 1–4 Sep. 1996, 1–4 Sep. 1996, pp. 401–404.

[8] I. Cohen, S. Raz and D. Malah, "Orthonormal shift-invariant adaptive local trigonometric decomposition", Signal Processing, Vol. 57, No. 1, Feb. 1997, pp. 43–64.

[9] I. Cohen, S. Raz and D. Malah, "Orthonormal shift-invariant wavelet packet decomposition and representation", Signal Processing, Vol. 57, No. 3, Mar. 1997, pp. 251–270.

[10] I. Cohen, S. Raz and D. Malah, "Eliminating interference terms in the Wigner distribution using extended libraries of bases", Proc. of the 22th IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-97, Munich, Germany, 20–24 Apr. 1997, pp. 2133–2136.

[11] L. Cohen, "Generalized phase-space distribution functions", J. Math. Phys., Vol. 7, 1966, pp. 781–786.

[12] L. Cohen and T. Posch, "Positive time-frequency distribution functions", IEEE Trans. on Acoust., Speech and Signal Processing, Vol. 33, 1985, pp. 31–38.

[13] L. Cohen, "Time-frequency distributions — a review", Proc. IEEE, Vol. 77, No. 7, July 1989, pp. 941–981.

[14] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection", IEEE Trans. Inform. Theory, Vol. 38, No. 2, Mar. 1992, pp. 713–718.

[15] R. N. Czerwinski and D. J. Jones, "Adaptive cone-kernel time-frequency analysis", IEEE Trans. on Signal Processing, Vol. 43, No. 7, July 1995, pp. 1715–1719.

[16] I. Daubechies, "Time-frequency localization operators: a geometric phase space approach", IEEE Trans. Inform. Theory, Vol. 34, 1988, pp. 605-612.

[17] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM Press,Philadelphia, Pennsylvania, 1992

[18] I. Daubechies, "Orthonormal bases of compactly supported wavelets, II. Variations on a theme", SIAM J. Math. Anal., Vol. 24, No. 2, 1993, pp. 499–519.

[19] N. Hess-Nielsen and M. V. Wickerhauser, "Wavelets and time-frequency analysis", Proc. IEEE, Vol. 84, No. 4, 1996, pp. 523–540.

[20] F. Hlawatsch, "Interference terms in the Wigner distribution", Proc. Int. Conf. on Digital Signal Processing, Florence, Italy, Sept. 5–8,1984, pp. 363–367.

[21] F. Hlawatsch and P. Flandrin, "The interference structure of the Wigner distribution and related time-frequency signal representations", in: W. Mecklenbrāuker, ed., *The Wigner Distribution - Theory and Applications in Signal Processing*, North Holland, Elsevier Science, 1992.

[22] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations", IEEE SP Magazine, Apr. 1992, pp. 21–67.

[23] A. J. E. M. Janssen and T. A. C. M. Claasen, "On positivity of time-frequency distributions", IEEE Trans. on Acoust., Speech and Signal Processing, Vol. 33, No. 4, Aug. 1985, pp. 1029–1032.

[24] J. Jeong and W. J. Williams, "Kernel design for reduced interference distributions", IEEE Trans. on Signal Processing, Vol. 40, No. 2, Feb. 1992, pp. 402–412.

[25] D. L. Jones and T. W. Parks, "A high resolution data-adaptive time-frequency representation", IEEE Trans. on Acoust., Speech and Signal Processing, Vol. 38, No. 12, Dec. 1990, pp. 2127–2135.

[26] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries", IEEE Trans. on Signal Processing, Vol. 41, No. 12, Dec. 1993, pp. 3397–3415.

[27] Y. Meyer, *Wavelets: Algorithms and Applications*, SIAM, Philadelphia, 1993.

[28] M. Mugur-Schachter, "A study of Wigner's theorem on joint probabilities", Found. Phys., Vol. 9, 1979, pp. 389–404.

[29] S. Qian and J. M. Morris, "Wigner distribution decomposition and cross-terms deleted representation", Signal Processing, Vol. 27, No. 2, May 1992, pp. 125–144.

[30] S. Qian and D. Chen, "Decomposition of the Wigner-Ville distribution and time-frequency distribution series", IEEE Trans. on Signal Processing, Vol. 42, No. 10, Oct. 1994, pp. 2836–2842.

[31] S. Qian and D. Chen, *Joint Time-Frequency Analysis: Methods and Applications*, Prentice-Hall Inc., 1996.

[32] N. Saito, *Local Feature Extraction and Its Applications Using a Library of Bases*, Ph.D. Dissertation, Yale Univ., New Haven, Dec. 1994.

[33] M. Wang, A. K. Chan and C. K. Chui, "Wigner-Ville distribution decomposition via wavelet packet transform", Proc. of the 3rd IEEE-SP Int. Symposium on Time-Frequency and Time-Scale Analysis, Paris, France, 18-21 June 1996, pp. 413–416.

[34] J. Wexler and S. Raz, "On minimizing the cross-terms of the Wigner distribution", Technical Report, EE PUB No. 809, Technion - Israel Institute of Technology, Haifa, Israel, Nov. 1991.

[35] M. V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*, AK Peters, Ltd, Wellesley, Massachusetts, 1994.

[36] E. Wigner, "On the quantum correction for thermodynamic equilibrium", Phys. Rev., Vol. 40, 1932, pp. 749–759.

[37] W. J. Williams, "Reduced interference distributions: biological applications and interpretations", Proc. IEEE, Vol. 84, No. 9, Sep. 1996, pp. 1264–1280.

[38] P. M. Woodward, *Probability and Information Theory with Applications to Radar*, Pergamon, London, 1953.

[39] Y. Zhao, L. E. Atlas and R. J. Marks, "The use of cone-shaped kernels for generalized time-frequency representations of nonstationary signals", IEEE Trans. on Acoust., Speech and Signal Processing, Vol. 38, No. 7, July 1990, pp. 1084–1091.

Figure Captions

Fig. 1: Test signal $g(t)$ consisting of a short pulse, a tone and a nonlinear chirp.

Fig. 2: Effect of a temporal shift on the time-frequency representation using the WPD with 12-tap coiflet filters: (a) $g(t)$ in its best basis; Entropy= 3.07. (b) $g(t - 2^{-6})$ in its best basis; Entropy= 2.94.

Fig. 3: Time-frequency representation using the SIWPD with 12-tap coiflet filters: (a) $g(t)$ in its best basis; Entropy= 1.88. (b) $g(t - 2^{-6})$ in its best basis; Entropy= 1.88. Compared with the WPD (Fig. 2), beneficial properties are shift-invariance and lower information cost.

Fig. 4: Contour plots for the signal $g(t)$: (a) Wigner distribution; (b) Spectrogram. Compared with the WD, the spectrogram does not have undesirable interference terms but the energy concentration is poor.

Fig. 5: Time-frequency tilings for the signal $g(t)$, using the library of wavelet packet bases (generated by 12-tap coiflet filters) and various best-basis methods: (a) Method of Frames (minimum l^2 norm). (b) Matching Pursuit. (c) Basis Pursuit (minimum l^1 norm). (d) Wavelet Packet Decomposition (minimum l^1 norm). (e) Wavelet Packet Decomposition (minimum Shannon entropy). (f) Shift-Invariant Wavelet Packet Decomposition (minimum Shannon entropy).

Fig. 6: The modified Wigner distribution for the signal $g(t)$, combined with the SIWPD and various distance-thresholds: (a) $D = 0$; (b) $D = 2$; (c) $D = 3$; (d) $D = 5$. For $D = 0$, the energy concentration is not sufficient. For $D = 2$, the energy concentration is improved by

cross-terms within components. As D gets larger, the interference between components becomes visible and the modified Wigner distribution converges to the conventional WD (*cf.* Fig. 4). A good compromise has been found for $1.5 \leq D \leq 2.5$.

Fig. 7: Mesh plots for the signal $g(t)$: (a) The modified Wigner distribution combined with the SIWPD and distance-threshold $D = 2$; (b) Wigner distribution; (c) Smoothed pseudo Wigner distribution; (d) Choi-Williams distribution; (e) Cone-kernel distribution; (f) Reduced interference distribution. The modified Wigner distribution yields an *adaptive* distribution where high resolution, high concentration, and suppressed interference terms are attainable.

Fig. 8: Time-frequency representation for the signal $g(t)$, using the SIWPD with 6-tap Daubechies least asymmetric wavelet filters: (a) The best-basis tiling; entropy= 2.09. (b) The modified Wigner distribution ($D = 2$, $\varepsilon = 0.1$).

Fig. 9: Time-frequency representation for the signal $g(t)$, using the SIWPD with 9-tap Daubechies minimum phase wavelet filters: (a) The best-basis tiling; entropy= 2.32. (b) The modified Wigner distribution ($D = 2$, $\varepsilon = 0.1$).

Fig. 10: Time-frequency representation for the signal $g(t)$, using the SIAP-LTD: (a) The best-basis tiling; entropy= 2.81. (b) The modified Wigner distribution.

Fig. 11: Examples of multicomponent signals: (a) Superposition of two linear chirps. (b) Superposition of two nonlinear chirps. Neither the time representation nor the energy spectral density indicate whether the signals are multicomponent. The joint time-frequency representations, however, show that the signals are well delineated into regions.

Fig. 12: A multicomponent signal $s(t)$.

Fig. 13: The best-basis decomposition of $s(t)$.

Fig. 14: The components of the signal s. (a) The component s_I associated with the equivalence class Λ_I. (b) The component s_{II} associated with the equivalence class Λ_{II}.

Fig. 15: Contour plots for the signal $s(t)$: (a) Modified Wigner distribution; (b) Wigner distribution.

Fig. 16: The signals $\tilde{s} = -s_I + s_{II}$ (bold line) and $s = s_I + s_{II}$ (light line) are different. However, since they consist of the same components, they have the same modified Wigner distribution.

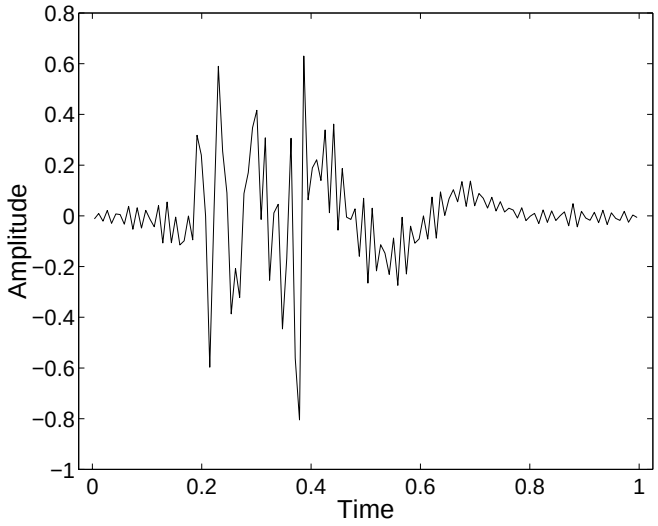
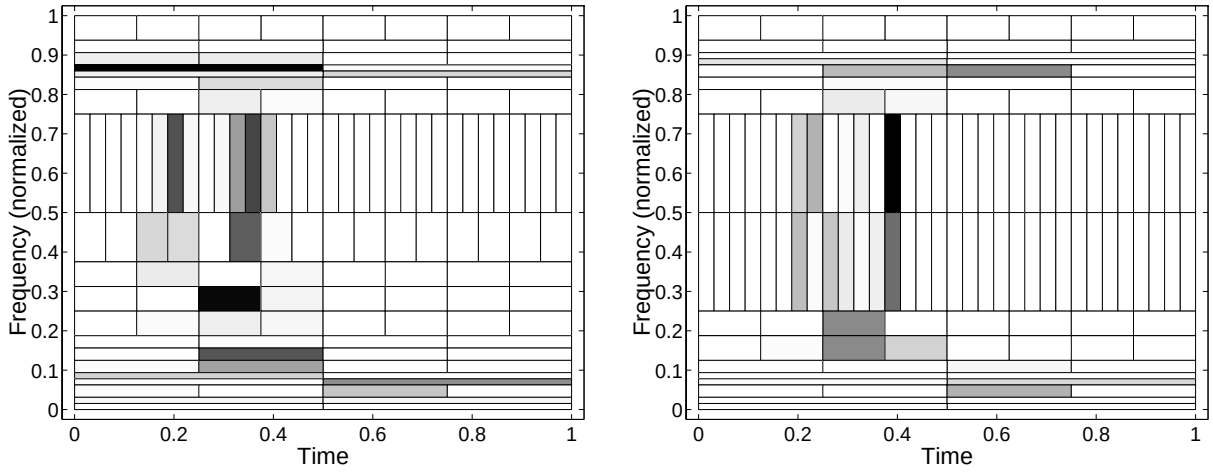
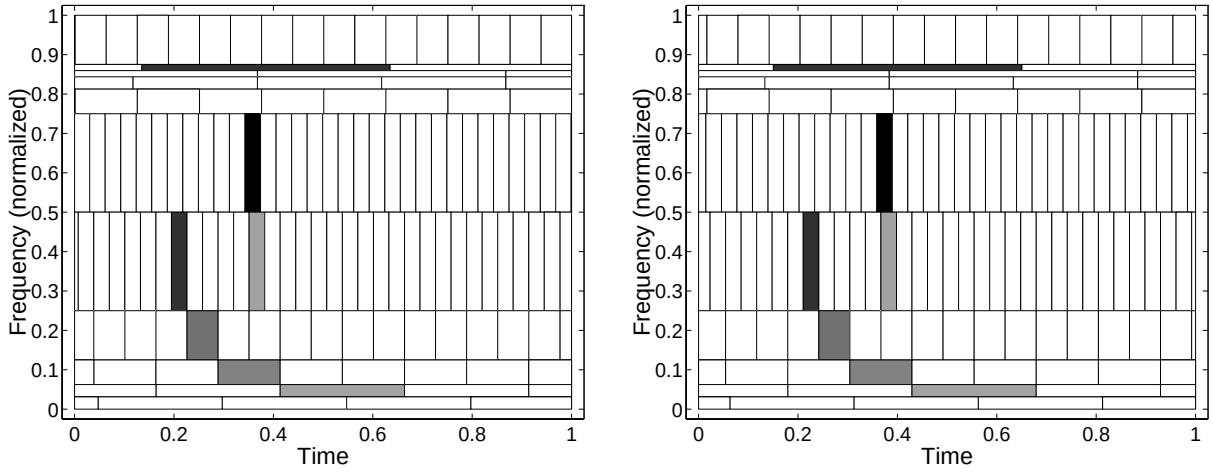


Figure 1: Test signal $g(t)$ consisting of a short pulse, a tone and a nonlinear chirp.



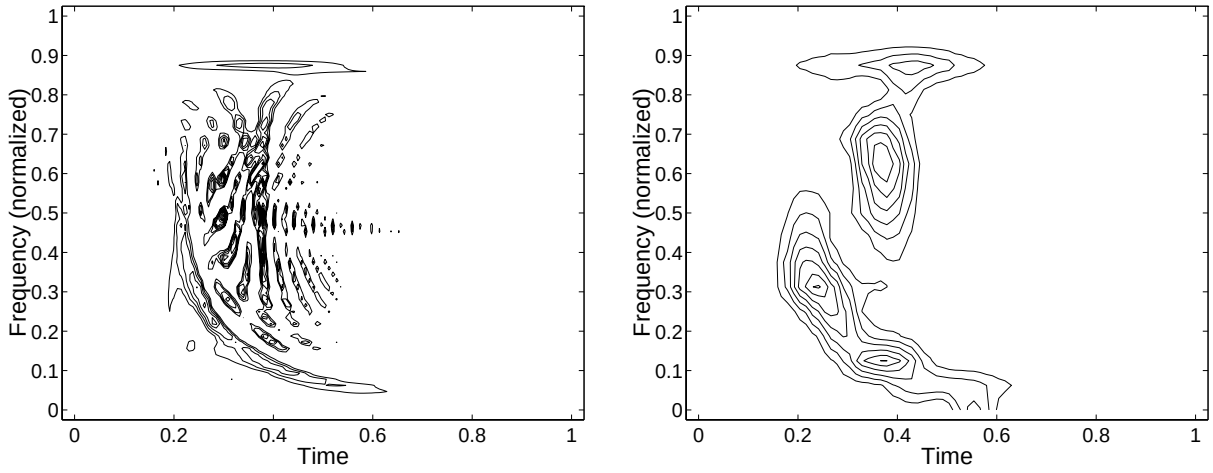
(a) (b)

Figure 2: Effect of a temporal shift on the time-frequency representation using the WPD with 12-tap coiflet filters: (a) $g(t)$ in its best basis; Entropy= 3.07. (b) $g(t - 2^{-6})$ in its best basis; Entropy= 2.94.



(a) (b)

Figure 3: Time-frequency representation using the SIWPD with 12-tap coiflet filters: (a) $g(t)$ in its best basis; Entropy= 1.88. (b) $g(t-2^{-6})$ in its best basis; Entropy= 1.88. Compared with the WPD (Fig. 2), beneficial properties are shift-invariance and lower information cost.



(a) (b)

Figure 4: Contour plots for the signal $g(t)$: (a) Wigner distribution; (b) Spectrogram. Compared with the WD, the spectrogram does not have undesirable interference terms but the energy concentration is poor.

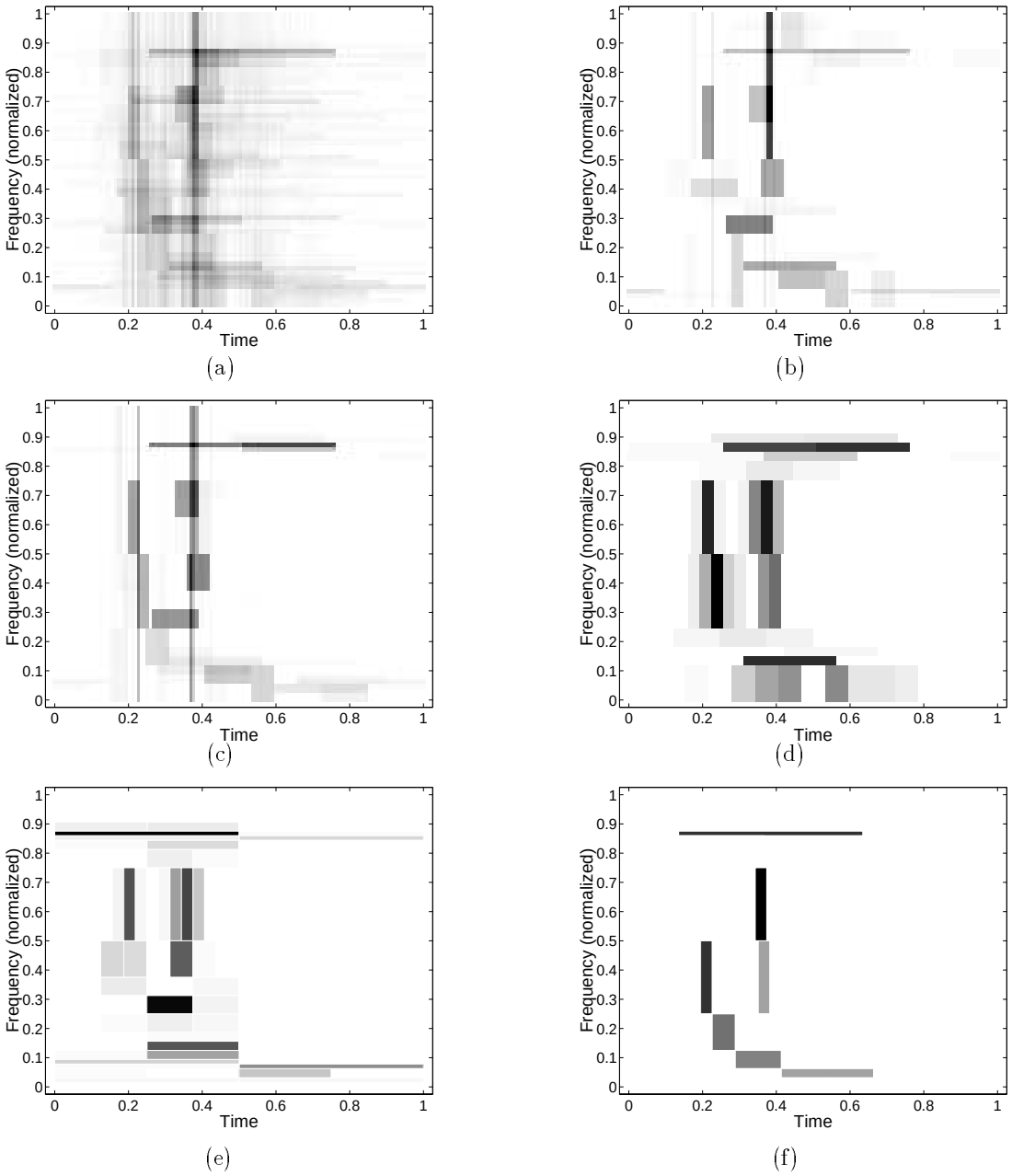


Figure 5: Time-frequency tilings for the signal $g(t)$, using the library of wavelet packet bases (generated by 12-tap coiflet filters) and various best-basis methods: (a) Method of Frames (minimum l^2 norm). (b) Matching Pursuit. (c) Basis Pursuit (minimum l^1 norm). (d) Wavelet Packet Decomposition (minimum l^1 norm). (e) Wavelet Packet Decomposition (minimum Shannon entropy). (f) Shift-Invariant Wavelet Packet Decomposition (minimum Shannon entropy).

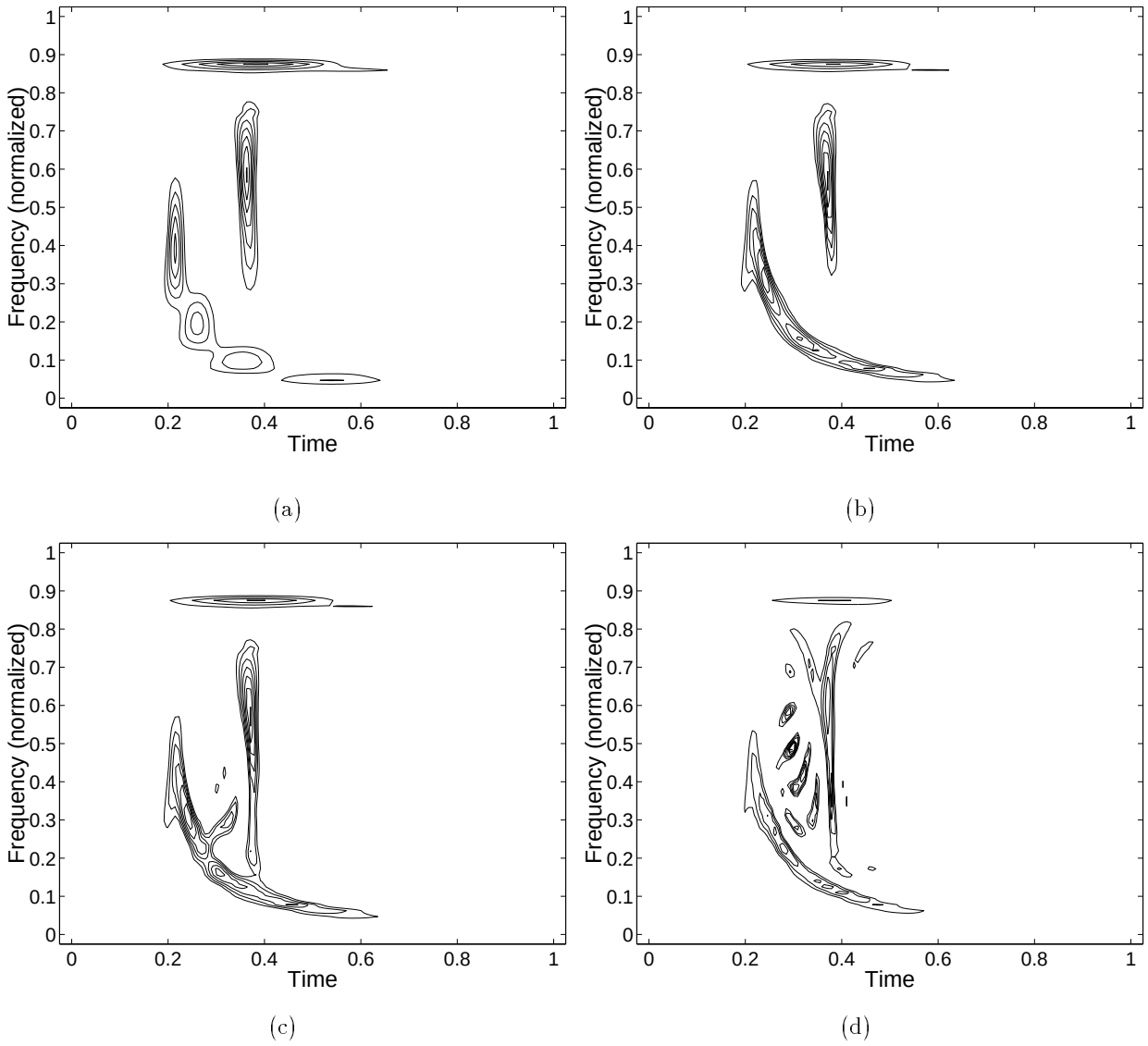


Figure 6: The modified Wigner distribution for the signal $g(t)$, combined with the SIWPD and various distance-thresholds: (a) $D = 0$; (b) $D = 2$; (c) $D = 3$; (d) $D = 5$. For $D = 0$, the energy concentration is not sufficient. For $D = 2$, the energy concentration is improved by cross-terms within components. As D gets larger, the interference between components becomes visible and the modified Wigner distribution converges to the conventional WD (*cf.* Fig. 4). A good compromise has been found for $1.5 \leq D \leq 2.5$.

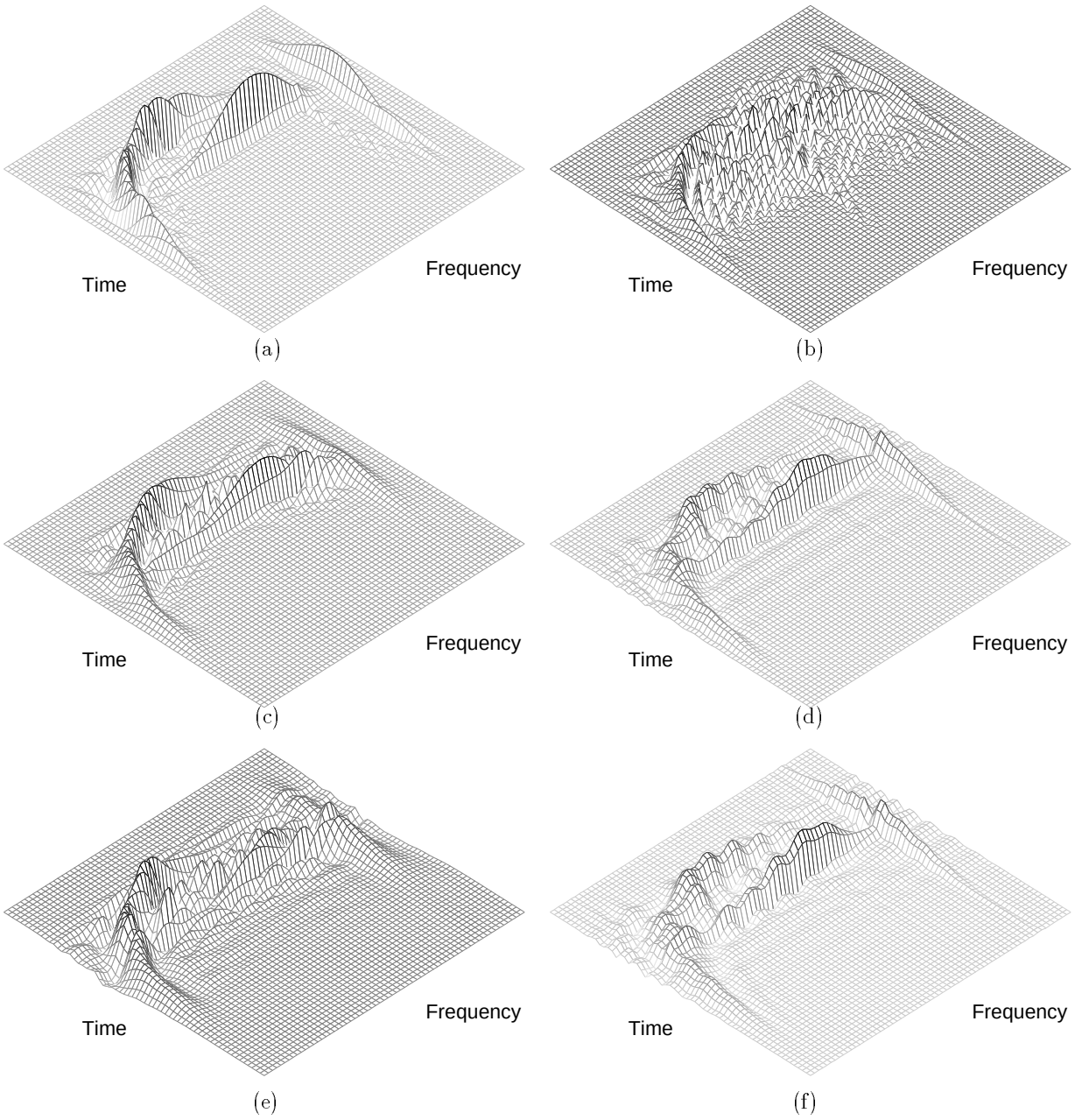
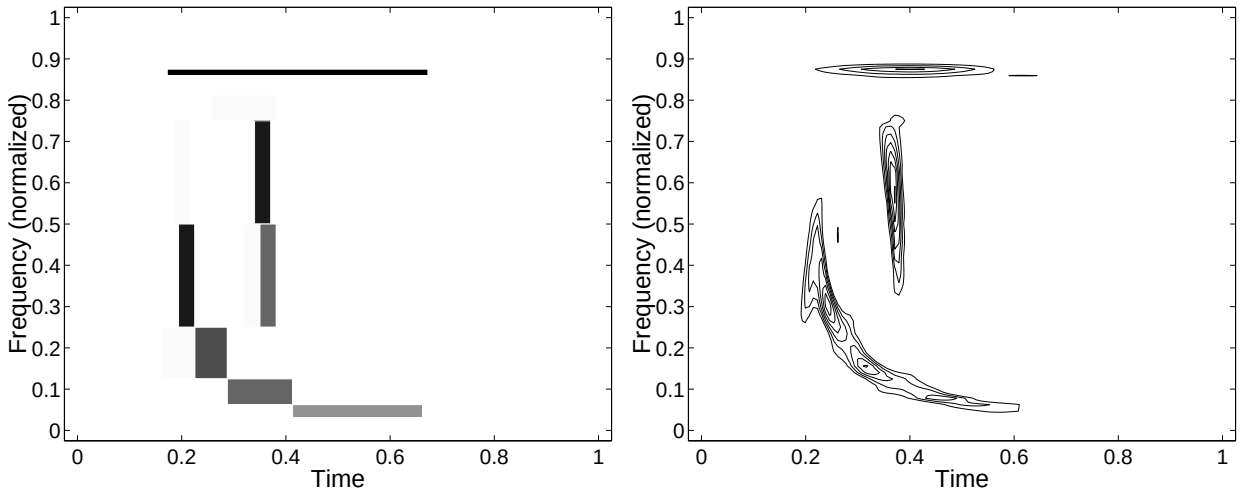
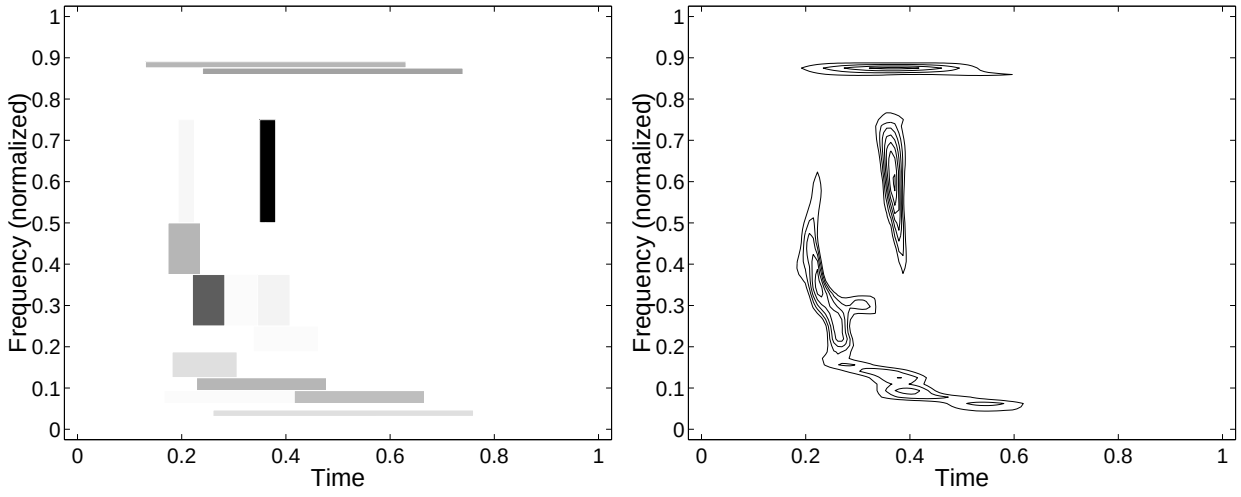


Figure 7: Mesh plots for the signal $g(t)$: (a) The modified Wigner distribution combined with the SIWPD and distance-threshold $D = 2$; (b) Wigner distribution; (c) Smoothed pseudo Wigner distribution; (d) Choi-Williams distribution; (e) Cone-kernel distribution; (f) Reduced interference distribution. The modified Wigner distribution yields an *adaptive* distribution where high resolution, high concentration, and suppressed interference terms are attainable.



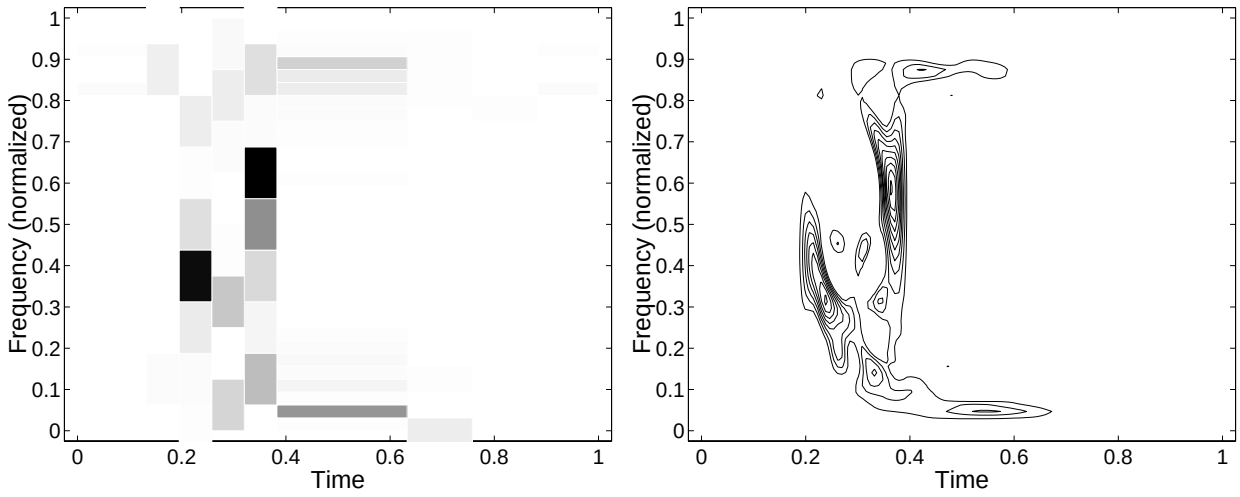
(a) (b)

Figure 8: Time-frequency representation for the signal $g(t)$, using the SIWPD with 6-tap Daubechies least asymmetric wavelet filters: (a) The best-basis tiling; entropy= 2.09. (b) The modified Wigner distribution ($D = 2$, $\varepsilon = 0.1$).



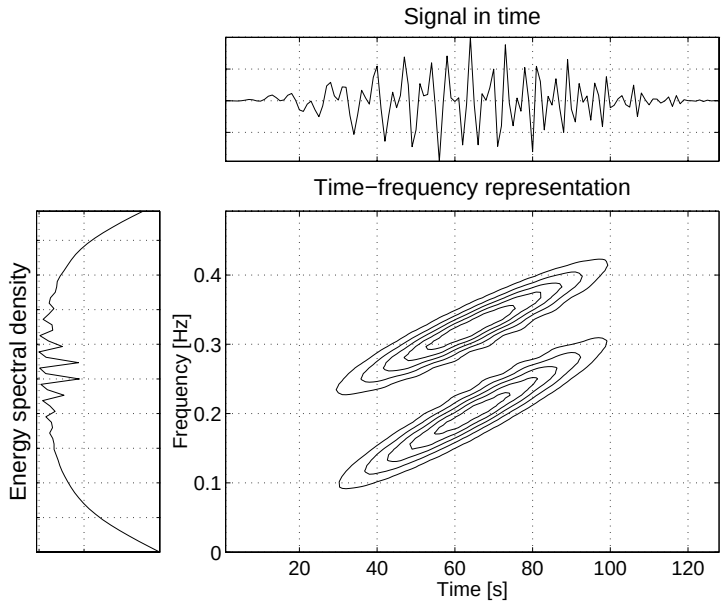
(a) (b)

Figure 9: Time-frequency representation for the signal $g(t)$, using the SIWPD with 9-tap Daubechies minimum phase wavelet filters: (a) The best-basis tiling; entropy= 2.32. (b) The modified Wigner distribution ($D = 2$, $\varepsilon = 0.1$).

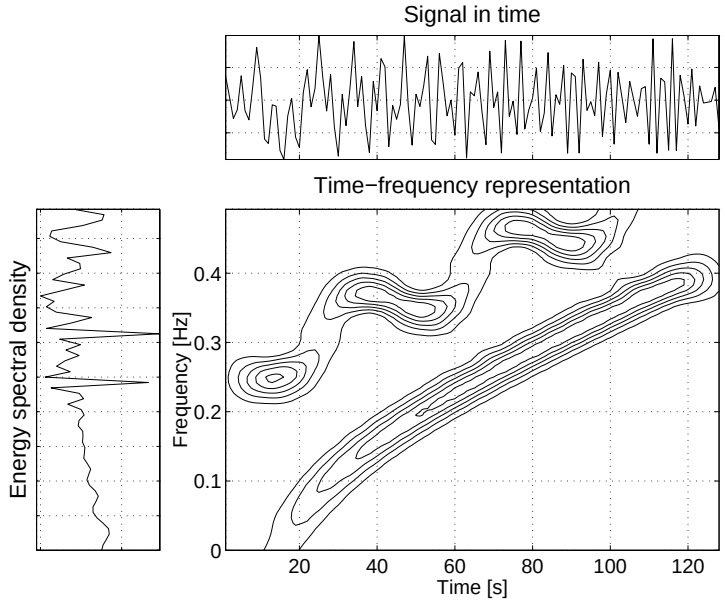


(a) (b)

Figure 10: Time-frequency representation for the signal $g(t)$, using the SIAP-LTD: (a) The best-basis tiling; entropy= 2.81. (b) The modified Wigner distribution.



(a)



(b)

Figure 11: Examples of multicomponent signals: (a) Superposition of two linear chirps. (b) Superposition of two nonlinear chirps. Neither the time representation nor the energy spectral density indicate whether the signals are multicomponent. The joint time-frequency representations, however, show that the signals are well delineated into regions.

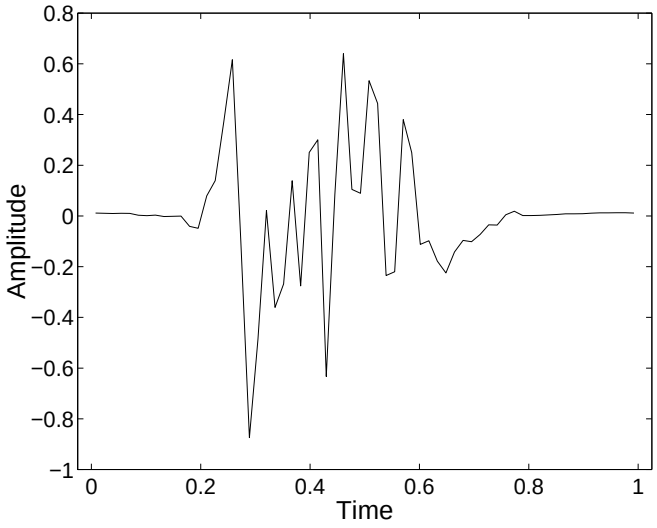


Figure 12: A multicomponent signal $s(t)$.

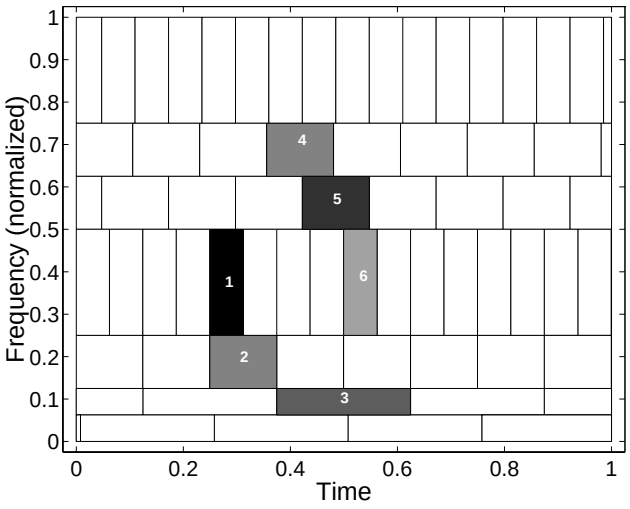
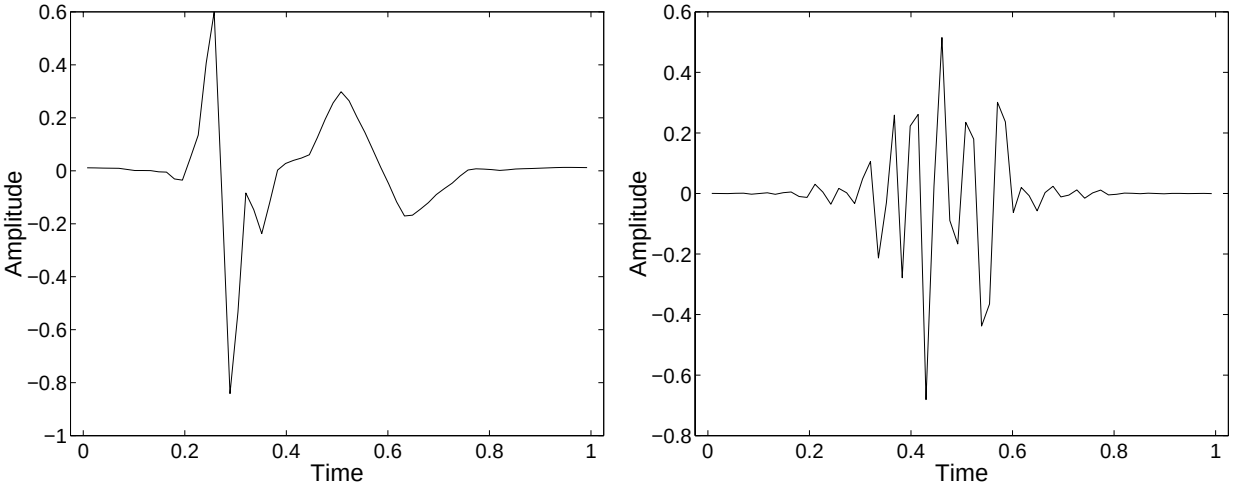
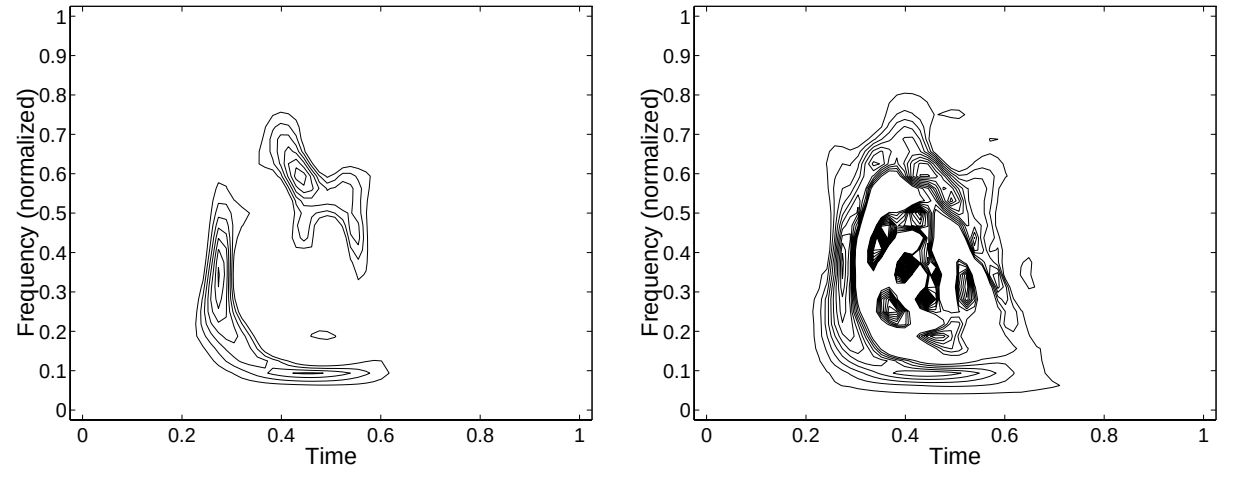


Figure 13: The best-basis decomposition of $s(t)$.



(a) (b)

Figure 14: The components of the signal s. (a) The component s_I associated with the equivalence class Λ_I. (b) The component s_{II} associated with the equivalence class Λ_{II}.



(a) (b)

Figure 15: Contour plots for the signal $s(t)$: (a) Modified Wigner distribution; (b) Wigner distribution.

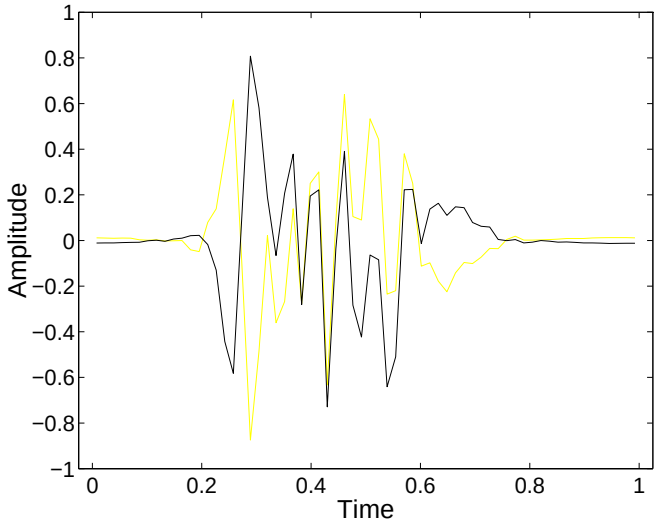


Figure 16: The signals $\tilde{s} = -s_I + s_{II}$ (bold line) and $s = s_I + s_{II}$ (light line) are different. However, since they consist of the same components, they have the same modified Wigner distribution.