



Article submitted to journal

Subject Areas:

applied mathematics, signal analysis

Keywords:

Synchrosqueezing transform,
reassignment method, multitaper,
conceFT, instantaneous frequency

Author for correspondence:

Hau-tieng Wu

e-mail: hauwu@math.toronto.edu

ConceFT: Concentration of
Frequency and Time via a
multitapered
synchrosqueezed transform

Ingrid Daubechies¹, Yi (Grace) Wang²,
Hau-tieng Wu³

¹Mathematics, Duke University

²Mathematics, Syracuse University

³Mathematics, University of Toronto

A new method is proposed to determine the time-frequency content of time-dependent signals consisting of multiple oscillatory components, with time-varying amplitudes and instantaneous frequencies. Numerical experiments as well as a theoretical analysis are presented to assess its effectiveness.

1. Introduction

Oscillatory signals occur in a wide range of fields, including geophysics, biology, medicine, finance and social dynamics. They often consist of several different oscillatory components, the nature, time-varying behavior and interaction of which reflect properties of the underlying system. In general, we want to assess the number, strength and rate of oscillation of the different components constituting the signal, to separate noise from signal, and to isolate individual components; efficient and robust extraction of this information from an observed signal will help us better describe and quantify the underlying dynamics that govern the system. For each of the quantities of interest listed, we thus want an estimator that is consistent, that has (ideally) small variance and that produces results robust to different types of noise.

If the observed signal f can be written as a finite sum of so-called harmonic components, i.e. $f(t) = \sum_\ell a_\ell \cos(2\pi\xi_\ell t + \delta_\ell)$, where $a_\ell > 0$ (respectively $\xi_\ell > 0$) represents the strength or amplitude (respectively frequency) of the ℓ -th component, then one can recover the a_ℓ and ξ_ℓ from time-samples of $f(t)$ via the Fourier transform \hat{f} of f , defined by $\hat{f}(\xi) := \int f(t) e^{-i2\pi\xi t} dt$. (If the ξ_ℓ are all integer multiples of a common $1/t_0$, then the integral can be taken over an interval of length t_0 ; when this is not the case, one can resort to integrals over long time intervals and average. Typically only discrete samples $f(t_n)$, $n \in \mathbb{Z}$, are known, rather than the continuous time course $f(t)$, $t \in \mathbb{R}$, and the integrals are estimated by quadrature.) However, oscillatory signals of interest often have more complex behavior. We shall be interested in particular in signals that are still the combination of “elementary” oscillations, but in which both the amplitude and the frequency of the components are no longer constant; they can be written as

$$f(t) = \sum_{k=1}^K A_k(t) \cos(2\pi\varphi_k(t)), \quad (1.1)$$

where $K \in \mathbb{N}$, $A_k(t) > 0$ and $\varphi'_k(t) > 0$ for all k , but $A_k(t)$ and $\varphi'_k(t)$ are not constants. One can compute the Fourier transform \hat{f} of such signals, and recover f from \hat{f} (this can be validly done for a much wider class of functions), but it is now less straightforward to determine the time-varying amplitudes $A_k(t)$ and the so-called “instantaneous frequencies” $\varphi'_k(t)$ from \hat{f} . Although the time-local behavior of the oscillations, and their deviation from perfect periodicity, cannot be captured by the Fourier transform in an easily “readable” way, an accurate description of this instantaneous behavior is nevertheless important in many applications, both to understand the system producing the signal and to predict its future behavior. Examples in the medical field include studies of the circadian [1,2] and cortical rhythms [3], or of heart-rate [4–6] and respiratory variability [7–9], all widely studied to understand physiology and predict clinical outcomes.

The last 50 years have seen many approaches, in applied harmonic analysis and signal processing, to develop useful analysis tools for signals of this type; this is the domain of *time-frequency (TF) analysis*. Several algorithms and associated theories have been developed (see, e.g., the overview [10]) and widely applied; well known examples include the short time Fourier transform (STFT), continuous wavelet transform (CWT) and Wigner-Ville distribution (WVD). The main idea is often to “localize” a portion of the signal in time, and then “measure” the oscillatory behavior of this portion. For example, given a function $f \in L^2$, the *windowed* or *short time Fourier transform* (STFT) associated with a window function $h(t)$ can be defined as:

$$V_f^{(h)}(t, \eta) := \int f(s) h(t-s) e^{-i2\pi\eta(t-s)} ds$$

where $t \in \mathbb{R}$ is the time, $\eta \in \mathbb{R}^+$ is the frequency, h is the window function chosen by the user – a commonly used choice is the Gaussian function with kernel bandwidth $\sigma > 0$, i.e. $h(t) = (2\pi\sigma)^{-1/2} e^{-t^2/\sigma^2}$. (The overall phase factor $e^{-i2\pi\eta t}$ is not always present in the STFT, leading to the name *modified short time Fourier transform* (mSTFT) for this particular form in [11].)

Other, more specialized methods, targeting in particular signals of type (1.1), include the empirical mode decomposition [12], ensemble empirical mode decomposition [13], the sparsity

approach [14], iterative convolution-filtering [15,16], the approximation approach [17], non-local mean approach [18], time-varying autoregression and moving average approach [19] as well as the synchrosqueezing transforms introduced and studied by some of us [11,20–23].

All TF methods that target reasonably large classes of functions (as opposed to functions with such specific models that complete characterization requires only fitting a small number of parameters) must face the Heisenberg uncertainty principle, limiting how accurately oscillatory information can be captured over short time intervals; for toy signals specially designed to have precise TF properties (e.g., chirps), this typically expresses itself by a “blurring” or “smearing out” of their TF representation, regardless of the analysis tool used. *Reassignment methods* [24–26], introduced in 1978 and recently attracting more attention again, were proposed to analyze and possibly counter this. Their main idea is to analyze the local behavior in the TF plane of portions of the representation, and determine nearby possible TF concentration candidates that best explain it; each small portion is then reallocated to its “right” place in the TF plane, to obtain a more concentrated TF representation that, one hopes, gives a faithful and precise rendering of the TF properties of the signal. Reassignment methods can be applied to very general TF representations [10,25]; they can be adaptive as well [27]. It has been argued recently [18] that reassignment methods can be viewed as analogs of non-local mean techniques commonly applied in image processing; this provides an intuitive explanation for their robustness to noise.

The synchrosqueezing transform (SST) can be viewed as a special reassignment method [25–27]. In SST, the STFT or CWT coefficients are reassigned *only* in the frequency “direction” [21, 23,28]; this preserves causality, making it possible to reconstruct each component with real-time implementation [29]. The STFT-based SST of f is defined as

$$S_f^{(h)}(t, \xi) := \lim_{\alpha \rightarrow 0} \int V_f^{(h)}(t, \eta) g_\alpha(|\xi - \omega_f^{(h)}(t, \eta)|) d\eta,$$

where g_α is an “approximate δ -function” (i.e. g is smooth and has fast decay, with $\int g(x)dx = 1$, so that $g_\alpha(t) := \frac{1}{\alpha}g(\frac{t}{\alpha})$ tends weakly to the delta measure δ as $\alpha \rightarrow 0$), and with $\omega_f^{(h)}$ defined by

$$\omega_f^{(h)}(t, \eta) := \frac{-i\partial_t V_f^{(h)}(t, \eta)}{2\pi V_f^{(h)}(t, \eta)} \text{ if } V_f^{(h)}(t, \eta) \neq 0, \text{ and } \omega_f^{(h)}(t, \eta) := -\infty \text{ otherwise.}$$

The SST for CWT is defined similarly; see [21,30], or Section 2. SST was proposed originally for sound signals [20,31]; its theoretical properties have been studied extensively [21,28–30,32–34], including its stability to different types of noise [30,35]. Several variations of the SST have been proposed [33,36–39]; in particular, the SST-approach can also be used for other TF representations, such as the wave packet transform [37], and it can be extended to two-dimensional signals (such as images) [40,41]. In addition, its practical usefulness has been demonstrated in a wide range of fields, including medicine [6,7,9,42–46], mechanics [39,47,48], finance [49,50], geography [51–53], denoising [33], atomic physics [54–56] and image analysis [57,58].

The SST approach can extract the instantaneous frequency and reconstruct the constitutional oscillatory components of a signal of type (1.1) in the presence of noise [30,35]. However, its performance suffers when SNR gets low: as the noise level increases, and even before it completely obscures the main concentration in the TF plane of the signal, spurious concentration areas appear elsewhere in the TF plane, caused by correlations introduced by the overcomplete STFT or CWT analysis tool. The effect of these misleading perturbations, which downgrade the quality of the results, can be countered, to some extent, by *multi-tapering*.

Multi-tapering is a technique originally proposed to reduce the variance and hence stabilize power spectrum estimation in the spectral analysis of stationary signals [59–61]. Sampling the signal during only a finite interval leads to artifacts, traditionally reduced by tapering; an unfortunate side effect of tapering is to diminish the impact of samples at the extremes of the time interval. Thomson [59] showed that one can nevertheless exploit optimally the information provided by the samples at the extremities, by using several orthonormal functions as tapers: the average of the corresponding power spectra is a good estimator with reduced variance. This

technique has since been applied widely [60]. Multi-tapering was later extended to non-stationary TF analysis by combining it with reassignment [62–64]: a more robust “combined” reassigned TF representation is obtained by picking orthonormal “windows” (used to isolate portions of the TF representation when working with a reassignment method), and averaging the reassigned TF representations determined by each of the individual windows. Heuristically, the concentration for a “true” constituting component of the signal will be in similar locations in the TF plane for each of the individual reassigned TF representations, whereas the spurious concentrations, artifacts of correlations between noise and the windowing function, tend to not be co-located and have a diminished impact when averaged. In the SST context a similar multi-taper idea was used by one of us in a study of the anesthesia depth [6,46], in which J different window functions h_j , $j = 1, \dots, J$ are considered, and the multi-taper SST (MTSST) is computed as follows:

$$\text{MS}_f(t, \xi) := \frac{1}{J} \sum_{j=1}^J S_f^{(h_j)}(t, \xi).$$

Using multiple tapers reduces artifacts, and the MTSST remains “readable” at higher noise levels than a “simple” SST [6,46]. To optimally suppress noise artifacts it is tempting to consider increasingly larger J . However, the area in the TF plane over which the signal TF information is “smeared out” also increases (linearly) with J , and a balance needs to be observed; in the multi-taper reassignment method of [62], for instance, 6 Hermite functions were used (i.e. $J = 6$.)

In this paper, we introduce a new approach to obtain better concentrated time-frequency representations, which we call *ConceFT*, for **C**oncentration in **F**requency and **T**ime. It is based on STFT- or CWT-based SST, but the approach could be applied to yet other TF decomposition tools. The ConceFT algorithm will be defined precisely below, in Section 2. Like MTSST, ConceFT starts from a multi-layered time-frequency representation, but instead of averaging the SST results obtained from STFT or CWT for orthonormal windows, which can be viewed as elements in a vector space of time-frequency functions, it considers many different projections in this same vector space, and averages the corresponding SSTs; for more details, see Section 2. Section 3 studies the theoretical properties of ConceFT, and explains how it can provide reliable results under challenging SNR conditions; finally, in Section 4, we provide several numerical results.

To conclude this introduction, we illustrate ConceFT on a simulated signal, in which the clean signal $s(t)$ is composed of two oscillatory components: $s(t) = s_1(t) + s_2(t)$, where $s_1(t) = A_1(t) \cos(2\pi\varphi_1(t))\chi_{[3,12]}(t)$, and $s_2(t) = A_2(t) \cos(2\pi\varphi_2(t))\chi_{[0,8]}(t)$ (here χ stands for the indicator function, $\chi_{[a,b]}(t) = 1$ if $a \leq t \leq b$, $\chi_{[a,b]}(t) = 0$ otherwise); $A_i(t) > 0$ and $\varphi'_i(t) > 0$ for $i = 1, 2$. This signal is sampled at rate 100Hz, from $t = 0$ to $t = 12$ seconds. To these signal samples we add independent realizations of a fat-tailed noise ξ , which is identically-t4-Student-distributed with variance 2.036. The left panels in Figure 1 show the three constituents of the total (noisy) signal $Y(t) = s_1(t) + s_2(t) + \xi(t)$; note that each of s_1 and s_2 “lives” during only part of the full time observation interval; the fat-tailed nature of the noise causes the bursty behavior evident in the plot of $\xi(t)$. The individual plots of the s_i show the amplitude modulations $A_i(t)$ of the s_i ; Figure 1 also graphs $\varphi'_i(t) > 0$ for $i = 1, 2$. In addition, Figure 1 shows the time course of both the clean signal $s(t)$ and the noisy signal $Y(t)$, at the same scale; their signal-to-noise ratio is -0.85 , computed as $20 \log_{10} \left(\frac{\text{std}(s(t))}{\text{std}(\xi(t))} \right)$, where std stands for standard deviation. Figure 2 shows several SST-based results for this (quite challenging) example. For the clean signal s , the “mono-SST” (STFT-based, with a Gaussian window) performs quite well, with only some artifacts at the onset and cessation of the s_i ; many structured artifacts are visible in the mono-SST of the noisy signal Y . Both MTSST and ConceFT remove the onset and cessation artifacts for the clean s (shown only for ConceFT in the figure, but similar for MTSST); the improvement is much more marked for the noisy signal Y : the spurious “bubbles” are suppressed to some extent in the MTSST-based representation (using 2 orthonormal windows: the same Gaussian and the next higher-order Hermite function); a more dramatic improvement is seen in the ConceFT-representation corresponding to the same vector space of windows.

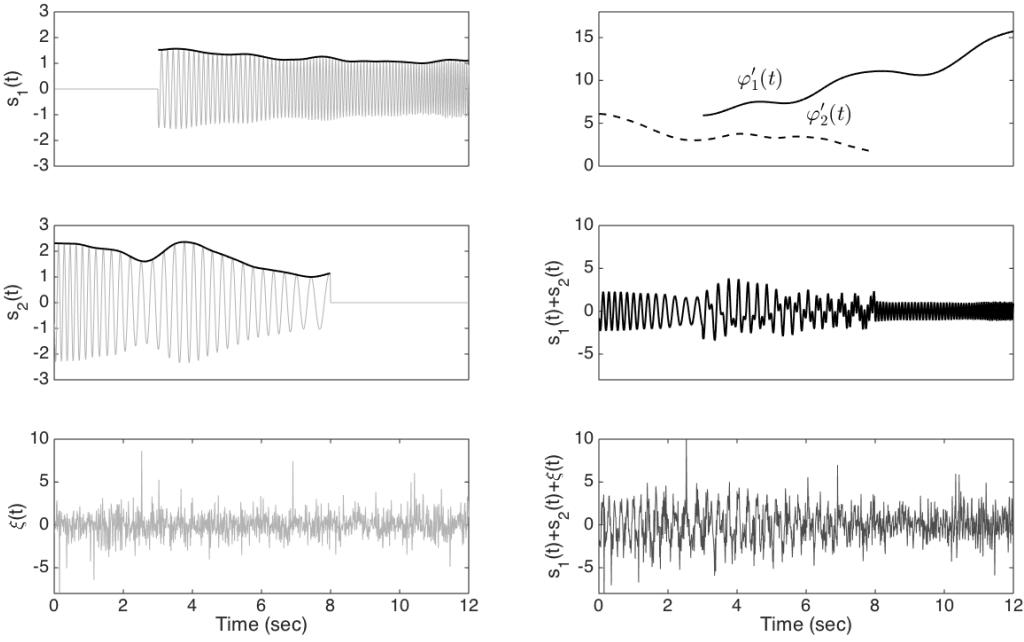


Figure 1. Left panels: the three constituents of the noisy signal $Y(t)$: oscillatory components $s_1(t)$ (top), and $s_2(t)$ (middle), and the bursty iid t4-Student noise $\xi(t)$ (bottom). Note that $s_1(t) \neq 0$ only for $t > 3$, $s_2(t) \neq 0$ only for $t < 8$ sec.; their respective amplitudes $A_i(t)$ are plotted as envelopes for each. Right panels: plots of $\varphi'_1(t)$ (solid) and $\varphi'_2(t)$ (dashed) in top panel; the clean signal $s = s_1 + s_2$ (middle) and noisy signal, $Y(t) = s(t) + \xi(t)$ (bottom), plotted with the same scale.

2. The ConceFT algorithm

We start by quickly reviewing SST. In the introduction, we defined STFT-based SST, discussed in more detail in [23,28]; to show that the situation is very similar for CWT-based SST, we discuss that case here; see [21,30] for details. We start with the wavelet ψ with respect to which the CWT will be computed, which must necessarily have mean zero; that is, $\int \psi(t) dt = 0$; let's also pick it to be a Schwartz function. Then the *Continuous Wavelet Transform* $W_f^{(\psi)}(a, b)$ of a tempered distribution f , with the variables a, b standing for scale and time location, is defined as the inner product of f with $\psi^{(a,b)}(t) = |a|^{-1/2} \psi((t-b)/a)$. Even if the Fourier transform \hat{f} is very concentrated around some frequency ω_0 , the magnitude $|W_f^{(\psi)}(a, b)|$ of the CWT will be spread out over a range of scales a , corresponding to a neighborhood of ω_0 . However, the phase information of $W_f^{(\psi)}$ will still hold a ‘‘fingerprint’’ of ω_0 on that whole neighborhood, in that $W_f^{(\psi)}(a, b)$ will show oscillatory behavior in b , with frequency ω_0 , for a range of different a . This is the motivation for the synchrosqueezing transform, which shifts the CWT coefficients ‘‘back’’, according to certain reassignment rules determined by the phase information. More concretely, we set a *threshold* $\Gamma > 0$, and then define

$$\Omega_f^{(\psi, \Gamma)}(a, b) := \begin{cases} \frac{-i\partial_b W_f^{(\psi)}(a, b)}{2\pi W_f^{(\psi)}(a, b)} & \text{when } |W_f^{(\psi)}(a, b)| > \Gamma \\ -\infty & \text{when } |W_f^{(\psi)}(a, b)| \leq \Gamma. \end{cases}$$

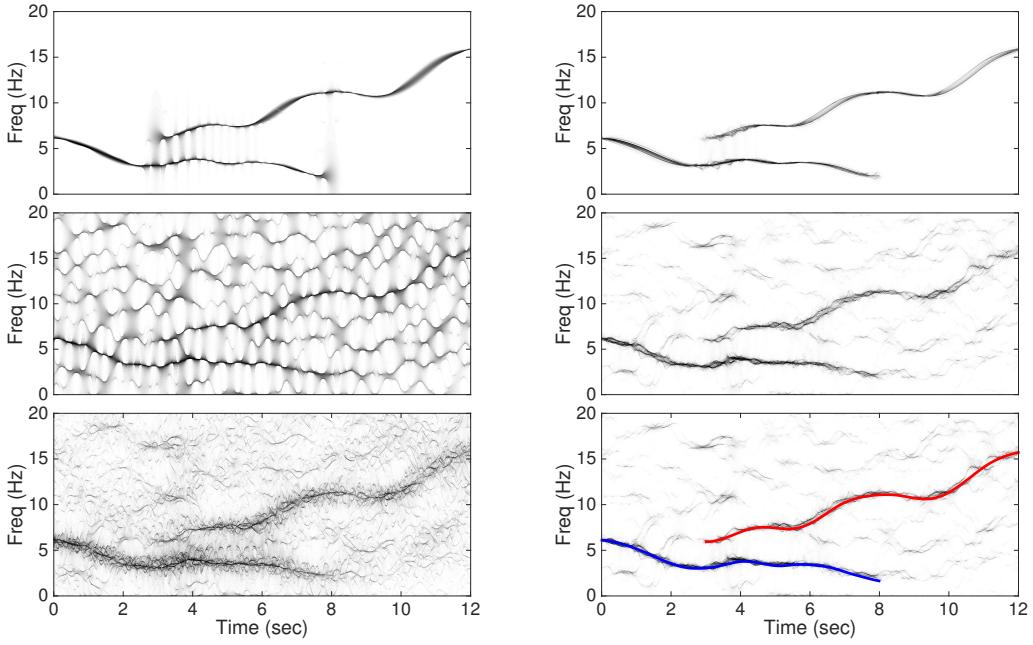


Figure 2. Top left: STFT-based synchrosqueezing transform (SST) of the clean signal $s(t)$ for a Gaussian window h ; middle left: STFT-based SST of the noisy signal $Y(t)$, with the same window. Bottom left: multi-taper SST of $Y(t)$, choosing the Gaussian and the next 5 Hermite functions as windows; this result is closer to the STFT-based SST of s . Top right: ConceFT of $s(t)$ based on the same two Hermite functions; middle right: ConceFT of $Y(t)$ based on the two Hermite functions; bottom right: same as middle right, with plots of $\varphi'_1(t)$ and $\varphi'_2(t)$ superimposed.

where ∂_b denotes the partial derivative with respect to b . Here the hard thresholding parameter Γ can be adjusted for best reduction of the numerical error and noise influence. The CWT-based synchrosqueezed transform (or CWT-based SST) is then defined by using $\Omega_f^{(\psi, \Gamma)}(a, b)$ as a reassignment guide, to move the CWT coefficient $W_f^{(\psi)}(a, b)$ at scale a to a new frequency slot:

$$S_f^{(\psi, \Gamma, \alpha)}(b, \xi) := \int_{\{a : |W_f^{(\psi)}(a, b)| > \Gamma\}} W_f^{(\psi)}(a, b) \frac{1}{\alpha} g\left(\frac{|\xi - \Omega_f^{(\psi, \Gamma)}(a, b)|}{\alpha}\right) a^{-3/2} da,$$

where $0 < \alpha \ll 1$ is chosen by the user, g is a smooth function so that $\frac{1}{\alpha} g(\frac{\cdot}{\alpha}) \rightarrow \delta$ in the weak sense as $\alpha \rightarrow 0$, and the factor $a^{-3/2}$ is introduced to ensure that the integral of $S_f^{(\psi, \Gamma, \alpha)}(b, \xi)$ over ξ yields a close approximation to the original $f(b)$. For more details, we refer the reader to [21,30].

Although both the CWT $W_f^{(\psi)}$ and its derived SST $S_f^{(\psi, \Gamma, \alpha)}$ depend on the choice of the reference wavelet ψ , this is much less pronounced for the SST; CWT-based SST corresponding to different reference wavelets lead to different but very similar TF representations. (Theoretical reasons for this can be found in [21,30].) In particular, the dominant components in the TF representations are very similar. Moreover, even when the signal is contaminated by noise, these dominant components in the TF representations are not significantly disturbed [30]. However, the distribution of artifacts across the TF representation, induced by the noise, as seen in e.g. the middle left panel of Figure 2, vary from one reference wavelet to another; this can be intuitively explained by observing that the CWT is essentially a convolution with (scaled versions of) the reference wavelet, so that the wavelet transforms of i.i.d. noise based on different orthogonal reference wavelets are *independent*. These observations lead to the idea of a multi-taper SST algorithm [6,46]. In brief, given J orthonormal reference wavelets $\psi_j, j = 1, \dots, J$, one determines

the reassignment rules $\Omega_f^{(\psi_j, \Gamma)}(a, b)$, as well as the corresponding $S_f^{(\psi, \Gamma, \alpha)}(b, \xi)$, and then defines the MTSST by

$$\text{MS}_f^{\Gamma, \alpha}(b, \xi) := \frac{1}{J} \sum_{j=1}^J S_f^{(\psi_j, \Gamma, \alpha)}(b, \xi).$$

This suggests that averaging over a large number of orthonormal reference wavelets would smooth out completely the TF artifacts induced by the noise, as originally discussed for the reassignment method [62]. However, in order for reassignment to make sense, the reference function, whether it is the window h for STFT or the wavelet ψ for CWT, must itself be fairly well concentrated in time and frequency, so that inner products with modulated window functions or scaled wavelets do not mix up different components and behaviors of the signal. On the other hand, there is a limit to how many orthonormal functions can be “mostly” supported in a concentrated region in the TF-plane – a rule of thumb generalizing the Nyquist sampling density states that for a region \mathcal{R} in the TF-plane, one can find only $\text{Area}(\mathcal{R})/(2\pi)$ orthonormal functions that are mostly concentrated on \mathcal{R} [65]. This severely limits how many different orthonormal ψ_j can be used in practice in MTSST.

ConceFT uses the different TF “views” provided by the CWT transforms $W_f^{(\psi)}$ in a different way, exploiting the *non-linearity* of the SST operation. For each choice of ψ , the collection of CWT $W_f^{(\psi)}$, where f ranges over the class of signals of interest, span a subspace of \mathcal{F} , where \mathcal{F} is a set of all reasonably smooth functions of the two variables a, b . Different orthonormal ψ_j generate different subspaces in \mathcal{F} ; combined, they generate a larger subspace, in which one can define an infinite number of “sections”, each corresponding to the collection of CWT generated by one reference wavelet. Each linear combination of the ψ_j defines such a CWT-space, in which one can carry out the corresponding SST operation. For $\psi = \sum_{j=1}^J r_j \psi_j$, where $r_j \in \mathbb{R}$, one has $W_f^{(\psi)} = \sum_{j=1}^J r_j W_f^{(\psi_j)}$; because synchrosqueezing is a highly nonlinear operation, the corresponding $S_f^{(\psi, \Gamma, \alpha)}$ are however not linear combinations of the $S_f^{(\psi_j, \Gamma, \alpha)}$. In practice, the artificial concentrations in the TF-plane, triggered by fortuitous correlations between the noise and the (overcomplete) $\psi^{(a,b)}$, occur at locations sufficiently different, for different choices of the vector $\mathbf{r} = (r_1, \dots, r_J)$, that averaging over many choices of \mathbf{r} successfully suppresses noise artifacts.

More precisely, the CWT-based ConceFT algorithm proceeds as follows:

- Take J orthonormal reference wavelets, ψ_1, \dots, ψ_J , in the Schwartz space, with good concentration in the TF-plane.
- Pick N random vectors \mathbf{r}_n , $n = 1, \dots, N$, of unit norm, in \mathbb{R}^J ; that is, uniformly select N samples in S^{J-1} .
- For each n between 1 and N , define $\psi_{[n]} := \sum_{j=1}^J (\mathbf{r}_n)_j \psi_j$, and $W_f^{(\psi_{[n]})} = \sum_{j=1}^J (\mathbf{r}_n)_j W_f^{(\psi_j)}$.
- Select the threshold $\Gamma > 0$ and the approximation parameter $\alpha > 0$, and evaluate, for each n between 1 and N , the corresponding CWT-based SST of f by computing the reassignment rule $\Omega_f^{(\psi_{[n]}, \Gamma)}(a, b)$, and hence $S_f^{(\psi_{[n]}, \Gamma, \alpha)}(b, \xi)$, as defined above, with the minor adjustment that when the expression $(\mathbf{r}_n)_j W_f^{(\psi_j)}(a, b)$ in the reassignment rule denominator has a negative real part, we switch to the vector $-\mathbf{r}_n$.
- The final ConceFT representation of f is then the average

$$C_f^{\Gamma, \alpha}(b, \xi) := \frac{1}{N} \sum_{n=1}^N S_f^{(\psi_{[n]}, \Gamma, \alpha)}(b, \xi). \quad (2.1)$$

In practice, J could be as small as 2, while N could be chosen as large as the user wishes.

The square of the magnitude of $C_f^{\Gamma,\alpha}(b, \xi)$,

$$\tilde{S}_f(b, \xi) := |C_f^{\Gamma,\alpha}(b, \xi)|^2,$$

can be of interest in its own right, as an estimated *time-varying Power Spectrum* (tvPS) of f . STFT-based ConceFT representations are defined entirely analogously, based on the STFT-reassignment rule given in Section 1. (See the *Electronic Supplementary Materials* (or ESM) for the sketch of an alternate way in which one could extend multi-taper CWT, not pursued in this paper, however.)

3. Theoretical Results

In this section, we list and explain theoretical results about CWT-based ConceFT. The detailed mathematical computations and proofs can be found in the ESM. Entirely similar results hold for STFT-based ConceFT; since they are established by the same arguments, we skip those details. We start by recalling the structure of our signal space, as introduced in [21,30]. We emphasize that this is, to a large extent, a purely phenomenological model, constructed so as to reflect the fairly (but not exactly) periodic nature of many signals of interest, in particular (but not only) those of a physiological origin (see the discussion in [7]).

A single-component or *intrinsic-mode type* (IMT) function has the following form:

$$F(t) = A(t) \cos(2\pi\varphi(t)),$$

where the amplitude modulation $A(t)$ and the phase function $\varphi(t)$ are both reasonably smooth; in addition, both $A(t)$ and the derivative $\varphi'(t)$ (or the “instantaneous frequency”) are strictly positive at all time as well as bounded; finally, we assume that A and φ' vary in time at rates that are slow compared to the instantaneous frequency of F itself. For the precise mathematical formulation of these conditions we refer to the ESM; this precise formulation invokes a few parameters, one of which, ϵ , bounds the ratio of the rate of change of A and φ' . This parameter will play a role in our estimates below. [Although we are assuming the signal to be real-valued here, all this can easily be adapted to the complex case by replacing the cosine with the corresponding complex exponential; the discussion in the remainder of this section can be adapted similarly.] We also consider signals that contain several IMT components, that is, functions of the type

$$G(t) = \sum_{\ell=1}^L F_\ell(t) = \sum_{\ell=1}^L A_\ell(t) \cos(2\pi\varphi_\ell(t)), \quad (3.1)$$

where each F_ℓ is an IMT function, and we assume in addition that the instantaneous frequencies $\varphi'_\ell(t)$ are ordered (higher ℓ corresponding to larger φ'_ℓ) and well-separated,

$$\varphi'_{\ell+1}(t) - \varphi'_\ell(t) > d(\varphi'_{\ell+1}(t) + \varphi'_\ell(t)) \quad (3.2)$$

for all $\ell = 1, \dots, L-1$, for some d with $0 < d < 1$. We denote by \mathcal{A} the set of all such functions G ; it provides a flexible *adaptive harmonic model* space for a wide class of signals of interest. (Strictly speaking, they are not “truly” harmonic, if harmonicity is interpreted – as it often is – as “having components with frequencies that are integer multiples of a fundamental frequency”.)

Next, we turn to the noise model for which we prove our main theoretical result. For the purposes of this theoretical discussion, we use a simple additive Gaussian white noise (even though, as illustrated by the figures in the introduction, the approach works for much more challenging noise models as well!). That is, we consider our noisy signals to be of the form

$$Y(t) = G(t) + \sigma\Phi(t) = \sum_{\ell=1}^L F_\ell(t) + \sigma\Phi(t) = \sum_{\ell=1}^L A_\ell(t) \cos(2\pi\varphi_\ell(t)) + \sigma\Phi(t), \quad (3.3)$$

where $G = \sum_{\ell=1}^L F_\ell$ is in \mathcal{A} , Φ is a Gaussian white noise with standard deviation 1 and $\sigma > 0$ is the noise level. Note that in general Y is a generalized random process, since by definition G is a tempered distribution. We could extend this, introducing also the trend and a more

general noise model as in [30], the wave-shape function used in [32], or the generalized IMT functions that model oscillatory signals with fast varying instantaneous frequency of [66]. None of these generalizations would significantly affect the mathematical analysis, but to simplify the discussion, we restrict ourselves to the model (3.3).

Finally, we describe the wavelets ψ_1, \dots, ψ_J with respect to which we compute the CWT of Y . For the sake of convenience of the theoretical analysis, we assume that they are smooth functions with fast decay, and that their Fourier transforms $\widehat{\psi}_j$ are all real functions with compact support, $\text{supp } \widehat{\psi}_j \subset [1 - \Delta_j, 1 + \Delta_j]$, where $0 < \Delta_j < 1$. We also assume that the ψ_1, \dots, ψ_J form an orthonormal set, that is, $\int \psi_i(x) \overline{\psi_j(x)} dx = \delta_{i,j}$, where $\delta_{i,j}$ is the Kronecker delta. To build appropriate linear combinations of the ψ_j , we define, for any unit-norm vector $\mathbf{r} = (r_1, \dots, r_J)$ in \mathbb{R}^J , the corresponding combination as $\psi^{[\mathbf{r}]} := \sum_{j=1}^J r_j \psi_j$. It is convenient to characterize intervals for the scale a such that the support of $\widehat{\psi}_j^{(a,b)}$ overlaps $\varphi'_\ell(b)$, where $\psi_j^{(a,b)}(t) := \frac{1}{\sqrt{a}} \psi_j\left(\frac{t-b}{a}\right)$; we thus introduce the notation $Z_\ell^{(j)}(b) = [(1 - \Delta_j)/\varphi'_\ell(b), (1 + \Delta_j)/\varphi'_\ell(b)]$. It then follows from the definition of the CWT as the inner product between the signal and scaled, translated versions of the wavelets that (see [21,30])

$$W_{F_\ell}^{(\psi_j)}(a, b) = \begin{cases} e^{i2\pi\varphi_\ell(b)} Q_{j,\ell}(a, b) + \epsilon_j(a, b) & \text{when } a \in Z_\ell^{(j)}(b) \\ \epsilon_j(a, b) & \text{otherwise,} \end{cases}$$

where

$$Q_{j,\ell}(a, b) = A_\ell(b) \sqrt{a} \overline{\widehat{\psi}_j(a \varphi'_\ell(b))} \in \mathbb{R} \quad (3.4)$$

and ϵ_j is of order ϵ for all $j = 1, \dots, J$. Here ϵ_j depends on the first three absolute moments of ψ_j and ψ'_j and the model parameters. It follows that the wavelet transform of Y , with respect to ψ_j , is given by

$$W_Y^{(\psi_j)}(a, b) = \sum_{\ell=1}^L e^{i2\pi\varphi_\ell(b)} Q_{j,\ell}(a, b) \chi_{Z_\ell^{(j)}(b)} + \epsilon_j(a, b) + \sigma \Phi(\psi_j^{(a,b)}),$$

where $\chi_{Z_\ell^{(j)}(b)}$ is the indicator function of the set $Z_\ell^{(j)}(b)$; note that the $\epsilon_j(a, b)$ -term, again of order ϵ , need not be the same as before. As shorthand notations, we will use bold symbols to regroup quantities indexed by $j = 1, \dots, J$ into one J -dimensional vector, e.g. $\boldsymbol{\epsilon}(a, b) = [\epsilon_1(a, b), \dots, \epsilon_J(a, b)]^\top$ (which has norm of order ϵ), $\boldsymbol{\Phi}(a, b) = [\Phi(\psi_1^{(a,b)}), \dots, \Phi(\psi_J^{(a,b)})]^\top$ (a complex Gaussian random vector [67], with mean $[0, \dots, 0]^\top \in \mathbb{R}^J$, and covariance as well as relation matrix equal to $I_{J \times J}$ – see ESM), $\mathbf{W}_Y^\psi(a, b) = [W_Y^{(\psi_1)}(a, b), \dots, W_Y^{(\psi_J)}(a, b)]^\top$, and $\mathbf{Q}_\ell(a, b) := [Q_{1,\ell}(a, b), \dots, Q_{J,\ell}(a, b)]^\top$. Finally, $W_Y^{(\psi^{[\mathbf{r}]})}(a, b) := \mathbf{r}^\top \mathbf{W}_Y^\psi(a, b)$ or, more explicitly,

$$W_Y^{(\psi^{[\mathbf{r}]})}(a, b) = \sum_{\ell=1}^L \sum_{j=1}^J r_j e^{i2\pi\varphi_j(b)} Q_{j,\ell}(a, b) \chi_{Z_\ell^{(j)}}(a, b) + \mathbf{r}^\top [\boldsymbol{\epsilon}(a, b) + \sigma \boldsymbol{\Phi}(a, b)]. \quad (3.5)$$

Under the general assumptions for our model,

$$-i\partial_b W_Y^{(\psi^{[\mathbf{r}]})}(a, b) = 2\pi \left(\sum_{j=1}^J \sum_{\ell=1}^L r_j \varphi'_j(b) e^{i2\pi\varphi_j(b)} Q_{j,\ell}(a, b) \chi_{Z_\ell^{(j)}}(a, b) + \mathbf{r}^\top [\tilde{\boldsymbol{\epsilon}}(a, b) + \sigma \tilde{\boldsymbol{\Phi}}(a, b)] \right),$$

where $\tilde{\boldsymbol{\epsilon}}(a, b)$ is again a J -dimensional vector of order ϵ , and $\tilde{\boldsymbol{\Phi}}(a, b)$ is again a complex Gaussian random vector. The scalar products $\mathbf{r}^\top \boldsymbol{\Phi}(a, b)$ and $\mathbf{r}^\top \tilde{\boldsymbol{\Phi}}(a, b)$ are independent complex Gaussian random variables, with mean 0 and variance $\|\mathbf{r}\|^2$ and $\sum_{j=1}^J r_j^2 \|\widehat{\psi}'_j\|^2 / (2\pi a)^2$, respectively. (See ESM.) Set now $Z_\ell(b) = \cap_{j=1}^J Z_\ell^{(j)}(b)$. Then it follows that for $a \in Z_\ell(b)$, we get the following

reassignment for the CWT $W_Y^{(\psi[r])}$:

$$\omega_Y^{(\psi[r])}(a, b) = \frac{-i\partial_b W_Y^{(\psi[r])}(a, b)}{2\pi W_Y^{(\psi[r])}(a, b)} = \frac{\mathbf{r}^\top [\varphi'_\ell(b)e^{i2\pi\varphi_\ell(b)}\mathbf{Q}_\ell(a, b) + \tilde{\epsilon}(a, b) + \sigma\tilde{\Phi}(a, b)]}{\mathbf{r}^\top [e^{i2\pi\varphi_\ell(b)}\mathbf{Q}_\ell(a, b) + \epsilon(a, b) + \sigma\Phi(a, b)]},$$

which is a ratio random variable of two dependent complex Gaussian random variables with non-zero means. Next, we consider, for each *fixed* realization of the random noise, the unit-norm vector $\mathbf{r} \in \mathbb{R}^J$ as a random vector, picked uniformly from $S_\kappa = \{\mathbf{r} \in S^{J-1}; |\mathbf{r}^\top \mathbf{W}_Y^\psi(a, b)| > 2\kappa\} \subset S^{J-1}$. Restricting the choice of \mathbf{r} to the subset of S^{J-1} for which the inner product of \mathbf{r} and $\mathbf{W}_Y^\psi(a, b)$ has absolute value larger than 2κ reflects the threshold used in the SST algorithm (see Section 2); restricting \mathbf{r} so that the inner product has a positive part means that we sample \mathbf{r} from a half sphere rather than the whole sphere. (See ESM for more details.)

Assuming that the bound on the noise is such that $\|\epsilon + \sigma\Phi\|_2^2 < \kappa$, then the expectation of $\omega_Y^{(\psi[r])}(a, b)$ over S_κ is given by

$$\mathbb{E}_{\mathbf{r}} \omega_Y^{(\psi[r])}(a, b) = \varphi'_\ell(b) + e^{-i2\pi\varphi_\ell(b)} \mathfrak{p}_{\mathbf{Q}_\ell(a, b)}(\mathbf{V}_\ell(a, b)) + E_1,$$

where $\mathbf{V}_\ell(a, b) := \tilde{\epsilon}(a, b) + \sigma\tilde{\Phi}(a, b) - \varphi'_\ell(b)[\epsilon(a, b) + \sigma\Phi(a, b)]$, \mathfrak{p}_v denotes “taking the component along” a vector v , that is, $\mathfrak{p}_v(u) = \frac{v^\top u}{\|v\|}$, and E_1 is bounded by

$$|E_1| \leq \frac{1}{2} \left(\left[1 - \frac{c}{J-1} \right] |\mathfrak{p}_{\mathbf{Q}_\ell(a, b)}(\mathbf{V}_\ell(a, b))|^2 + c \frac{\|\mathbf{V}_\ell(a, b)\|^2}{J-1} \right)^{1/2}.$$

Furthermore the variance is bounded by

$$\text{Var}_{\mathbf{r}} \omega_Y^{(\psi[r])}(a, b) \leq \frac{5}{2} \left(\left[1 - \frac{c}{J-1} \right] |\mathfrak{p}_{\mathbf{Q}_\ell(a, b)}(\mathbf{V}_\ell(a, b))|^2 + c \frac{\|\mathbf{V}_\ell(a, b)\|^2}{J-1} \right).$$

A detailed derivation, and an explicit expression for the constant c is given in the ESM; as J becomes large, we have $c \approx 2\sqrt{2}/[\kappa\sqrt{\pi J}]$. The vector $\mathfrak{p}_{\mathbf{Q}_\ell(a, b)}(\mathbf{V}_\ell(a, b)) = \mathfrak{p}_{\mathbf{Q}_\ell(a, b)}(\tilde{\epsilon}(a, b) + \sigma\tilde{\Phi}(a, b) - \varphi'_\ell(b)[\epsilon(a, b) + \sigma\Phi(a, b)])$, which occurs in several of these estimates, is a vector that, with high probability (with respect to the random noise process), has a smaller norm for larger J , because vectors that are unrelated (as is the case for $\mathbf{Q}_\ell(a, b)$ and $\mathbf{V}_\ell(a, b)$) have a higher chance of being close to orthogonal in higher dimensions. The other terms in $\mathbb{E}_{\mathbf{r}} \omega_Y^{(\psi[r])}(a, b) - \varphi'_\ell(b)$ and in the variance $\text{Var}_{\mathbf{r}} \omega_Y^{(\psi[r])}(a, b)$ all have a factor $J-1$ in the denominator. Our theoretical analysis thus proves that ConceFT, using a larger dimensional space of TF representations, and subsequently averaging over the SST corresponding to random vectors in this larger dimensional space, leads to sharper estimates of the instantaneous frequencies for signals in \mathcal{A} that are corrupted by noise. Even when J is not large, our bounds show that ConceFT leads to a reduction in potential deviation of the tvPS from the itvPS.

The detailed estimates given in section ESM-3 are derived under the restrictive conditions listed at the start of this section for the signals and the wavelets used. However, as noted above, these conditions can be relaxed significantly (at the price of more intricate estimates). In practice, we observe similar behavior in our numerical examples even for more complex situations; in particular, the method can handle noise models that are much more challenging, as illustrated in the next section as well as by Figure 2 in section 1.

4. Numerical Experiments

In this Section, we demonstrate the results of the ConceFT algorithm on examples; we also discuss different choices for some of the different parameters involved. The ConceFT Matlab code and the codes leading to the figures in this paper could be found in <https://sites.google.com/site/hautiengwu/home/download>.

The first choice to be made, when applying CWT- or STFT-based ConceFT, concerns the family of orthonormal reference functions (wavelets or window functions) for the underlying wavelet or windowed Fourier transform. In both cases, we pick a family of eigenfunctions for a time-frequency localized operator designed for the CWT or STFT framework; as shown in [68,69] these can provide “optimal” localization within a restricted region of time-frequency space, where the size of the region depends solely on the number of functions used. More precisely, we use orthonormal Hermite functions for the STFT case [62,68] (see also Figure 2), and Morse wavelets for the CWT case [69,70]. In both cases, the shape of the localization domain in TF plane is not completely fixed, but can be adjusted by varying some parameters; for details, see ESM. Once the family of orthonormal reference functions is fixed, we need to decide how many ψ_j , $j = 1, \dots, J$ we pick; this corresponds to choosing the size of the corresponding domain of concentration in the TF plane. Flexibility in the choices of shape and size of the TF localization domain make it possible to adapt ConceFT, to some extent, to the family of signals under consideration. Finally, ConceFT also depends on the number N of random projections chosen (see sections 2 and 3). In principle, the larger N , the closer the results are to the expected value of the random process, and the more we expect accidental correlations between reference function and the noise to cancel out in regions of the TF plane where the signal does not reside; in practice, increasing N beyond a certain value does not appreciably improve the results. In what follows, we explore these different choices for the CWT case, on a simple family of challenging examples, with noise of different types (white Gaussian, Poisson and ARMA), and of different strengths. Results for the STFT case are similar; we will come back to them briefly below in subsection (f) as well as in the ESM.

For our test data, we restrict ourselves to simulated signals only, so as to be able to quantify the deviation from the “ground truth”, usually not available in real-life applications (ConceFT results on concrete signal, will appear elsewhere [71]). On the other hand, we want to avoid parametric models, so as to be sufficiently general. Accordingly, we generate a class \mathcal{C} of non-stationary data via a random process described below in subsection (a); each realization provides us not only with a (simulated) clean signal, but also with the exact “ground truth” for the time-dependent instantaneous frequency and amplitude of the components of that signal. The same subsection also describes in detail three different noise models (white Gaussian, Poisson and ARMA(1,1)) for which the approach is tested. After applying ConceFT to signals in \mathcal{C} , we want to compare the ConceFT results with the optimal, ground truth TF representation; to quantify their (dis)similarity, we use an Optimal Transport (OT) distance, as described in subsection (b). In subsection (c), we discuss how choices of the parameters and of the number of orthogonal Morse wavelets impact the ConceFT results, for this family of examples; subsection (d) illustrates the effect of the number N of random projections. Finally, in subsection (e), we explore the effect of different noise levels, for each of the three noise types we consider.

(a) Data simulation

To generate a typical multi-component signal, we use smoothed Brownian path realizations to model the non-constant amplitudes and the instantaneous frequencies of the components; more precisely, if W is the standard Brownian motion defined on $[0, \infty)$, then we define the *smoothed Brownian motion with bandwidth $h > 0$* as $\Phi_h := W * K_h$, where K_h is the Gaussian function with standard deviation $h > 0$ and $*$ denotes the convolution operator. Given $T > 0$ and parameters $\zeta_1, \dots, \zeta_6 > 0$, we then define the following family of random processes on $[0, T]$:

$$\Psi_{[\zeta_1, \dots, \zeta_6]}(t) := \zeta_1 + \zeta_2 t + \zeta_3 \frac{\Phi_{\zeta_4}(t)}{\|\Phi_{\zeta_4}\|_{L^\infty[0,T]}} + \zeta_5 \int_0^t \frac{\Phi_{\zeta_6}(s)}{\|\Phi_{\zeta_6}\|_{L^\infty[0,T]}} ds.$$

For the amplitude $A(t)$ of each IMT, we set $\zeta_2 = \zeta_5 = 0$; every realization then varies smoothly between ζ_1 and $\zeta_1 + \zeta_3$. In the examples shown below and in the ESM, the signal consists of two components (i.e. $L = 2$) on $[0, 60]$; their two amplitudes are independent realizations of $\Psi_{[2, 0, 1, 200, 0, 0]}(t)$. To simulate a phase function, we set $\zeta_1 = \zeta_3 = 0$; $\Psi_{[0, \zeta_2, 0, 0, \zeta_5, \zeta_6]}(t)$ is then, appropriately, a monotonically increasing process. In the examples we consider, we take for $\varphi_1(t)$

a realization of $\Psi_{[0,10,0,0,6,400]}(t)$ for $t \in [0, 60]$, and for $\varphi_2(t)$ a realization of $\Psi_{[0,2\pi,0,0,2,300]}(t)$. Finally, we also constrain each component to “live” on only part of the interval, by setting

$$s(t) = A_1(t) \cos(2\pi\varphi_1(t))\chi_{[18, 60]}(t) + A_2(t) \cos(2\pi\varphi_2(t))\chi_{[0, 36]}(t) =: s_1(t) + s_2(t),$$

where $\chi_{[\tau_1, \tau_2]}$ is the indicator function of $[\tau_1, \tau_2]$; that is, $\chi_{[\tau_1, \tau_2]}(t) = 1$ if $\tau_1 \leq t \leq \tau_2$, $\chi_{[\tau_1, \tau_2]}(t) = 0$ otherwise. We shall denote the resulting class of two-component signals by \mathcal{C} . In our examples, signals in \mathcal{C} are sampled uniformly at rate 160Hz, corresponding to 9600 samples. Figure 3 plots $s(t)$ for one example $s \in \mathcal{C}$, as well as the instantaneous frequencies (IFs) of its two components, all restricted to the subinterval $[15, 40] \subset [0, 60]$.

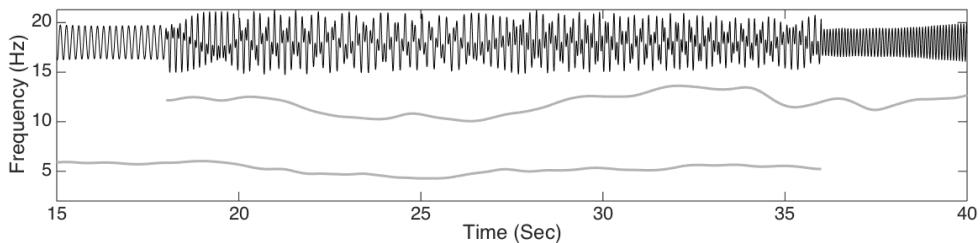


Figure 3. The signal s (in black) and the corresponding instantaneous frequencies (in gray) of the two components, restricted to the time interval $[15, 40]$.

Note that the signal s should not be viewed as a random process itself – we use the random processes $\Psi_{[\zeta_1, \dots, \zeta_6]}$ as a means to generate signals consisting of several components for which the amplitudes and instantaneous frequencies are not easily expressed analytically, but we will not consider or compute expectations with respect to these processes – once $s \in \mathcal{C}$ is generated, we consider it fixed when we apply ConceFT to it. (In further subsections, we shall encounter other elements of \mathcal{C} .)

To study the performance of ConceFT in the presence of noise, we add noise to $s(t)$, setting $Y(t_k) = s(t_k) + \sigma\xi(t_k)$, where t_k is the k -th sampling time and ξ is a stationary random process. We shall consider three different noise models; in each case we set the value of σ so that the signal to noise ratio (SNR),

$$\text{SNR} := 20 \log \frac{\text{std}(s)}{\sigma \text{std}(\xi)},$$

equals 0 dB. The three noise models we consider are Gaussian white noise, an auto-regressive-and-moving-average (ARMA) noise and Poisson noise. For the ARMA case, we consider an ARMA(1, 1) model determined by autoregression polynomial $a(z) = 0.5z + 1$ and moving averaging polynomial $b(z) = -0.5z + 1$; for the innovation process we use independent and identically distributed Student t_4 random variables. [Note that this ARMA(1, 1) noise is not white, because of the time dependence; in addition, the Student t_4 random variable has a “fat-tailed” distribution, resulting in possibly spiky realizations.] For the Poisson noise, we pick the $\xi(t_k)$ to be independent and identically sampled from the Poisson distribution with parameter $\lambda = 1$. Figure 4 plots a realization of $Y(t) = s(t) + \sigma\xi(t)$ for each of these three noise processes, restricted to the subinterval $[15, 40]$.

(b) Performance evaluation

To evaluate the performance of ConceFT, we propose comparing the time-varying power spectrum or tvPS (defined at the end of Section 2) of the results of the ConceFT analysis of Y with the *ideal time-varying power spectrum* (itvPS) of our simulated signal s , which can easily be

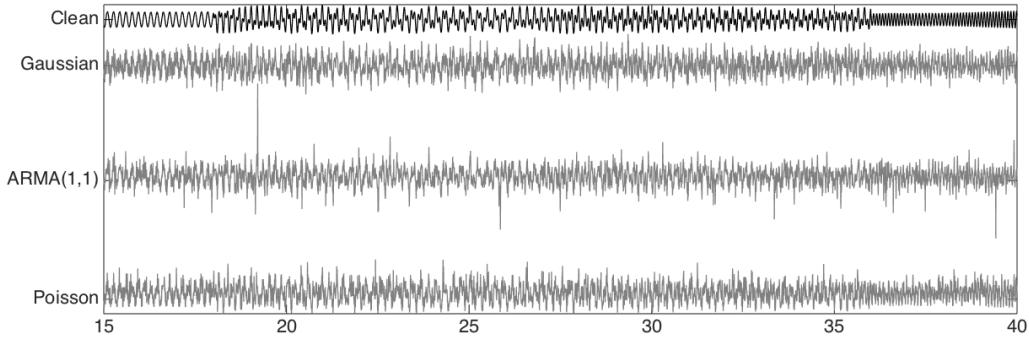


Figure 4. The restrictions to $[15, 40]$ of the clean signal s (top) and of the noisy signal $Y = s + \sigma\xi$, where the added noise is Gaussian, ARMA(1,1), or Poisson noise (2nd row to bottom, in order); in each case σ is picked so that the noisy signal has 0 dB SNR (signal to noise ratio). All signals are plotted at the same scale.

defined explicitly (because our construction was designed accordingly) as follows:

$$S_s(t, \omega) := \sum_{k=1}^2 A_k^2(t) \delta_{\varphi'_k(t)(\omega)}.$$

In order to quantify the (dis)similarity between the ConceFT-estimated tvPS \tilde{S}_Y and the itvPS S_s , we use the *Optimal Transport* distance (also called the Earth Mover distance). Because the principle of ConceFT is to “reassign” content in the TF plane, keeping the time-variable fixed (see Section 2), we also keep t fixed for the OT-distance. That is, we interpret, at each time t , $\tilde{S}_Y(t, \omega)$ and $S_s(t, \omega)$ as (probability) distributions in ω and compute the OT-distance between them, which essentially measures how much one distribution needs to be “deformed” in order to coincide with the other; this is repeated for all t , and the average of the t -dependent individual OT-distances then indicates the quality of the estimator \tilde{S}_Y for S_s .

The precise definition of (the discretized version of) the OT-distance we use is given in the ESM; Figure 5 displays 4 examples in which two delta-measures localized on curves in the TF-plane (similar to the itvPS defined above) lie at similar OT-distances of each other – although in each example the distance indicates a different type of “distortion”. Together, these examples give an intuitive understanding of the way in which OT distances capture the difference between the TF distributions of interest to us here.

(c) Parameter selection

As described in [70], generalizing the construction in [69], orthonormal families of Morse wavelets can be defined for different values of two parameters, β and γ ; different choices correspond to different shapes of the domain in the TF-plane on which they are mostly localized (see ESM). Once the values of β and γ are chosen, determining the family of ψ_j , one also needs to select J , the total number of orthonormal reference wavelets used in the ConceFT method. For signals in \mathcal{C} (see (a)), we explored systematically a range of (β, γ) pairs, as well as different values of J , to find the choice that, under different types of noise, with SNR of 0 dB, gave rise to the smallest OT-based distance (as described above) between the itvPS and the ConceFT-estimated tvPS. Surprisingly, the optimal choice depended very little on the type of noise; the optimal values we found are $\beta = 30$, $\gamma = 9$ and $J = 2$. (Detailed results are given in the ESM.)

(d) Effect of the number of random projections

The ConceFT Algorithm averages the SST results computed with N randomly picked reference wavelets (or windows, for STFT) from the linear span of the ψ_j , $j = 1, \dots, J$. It is expected that

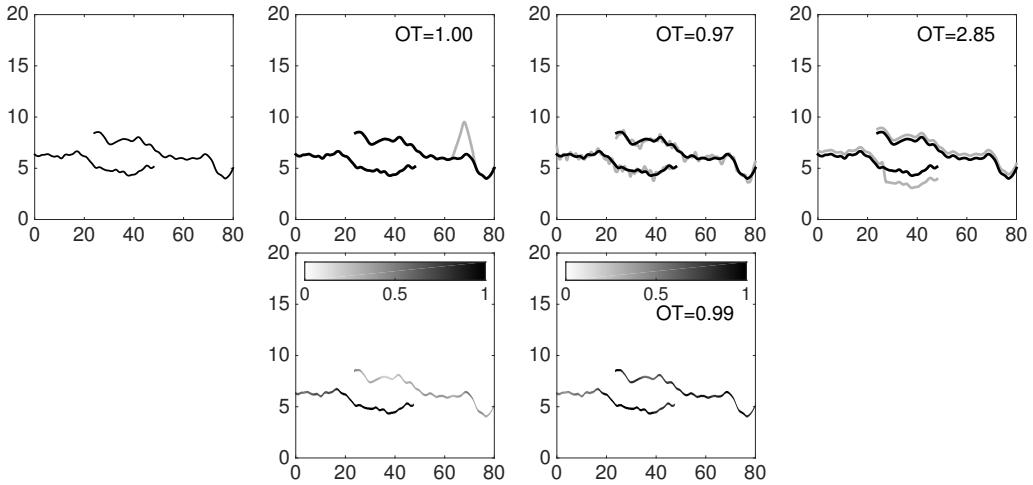


Figure 5. Top row: left: the TF-localization of the ideal time varying power spectrum (itvPS) of a 2-component simulated signal s_α (not showing the amplitude modulation (AM)). All other itvPS shown in the top row are for signals s_β , s_γ and s_δ that have a fairly small different OT-distance with respect to s_α ; the two components have the same time-dependent amplitudes as for s_α , but the instantaneous frequency curves have been moved (in order from left to right) by a narrow bump (left), a random dither (middle) and a shift (right). Bottom row: illustration of amplitude change; left: the original itvPS with the AM values indicated by the gray color. Bottom right: an itvPS example with the same IF but different AM. In all the figures, the x-axis is time and the y axis is the frequency. The image for each “deformed” itvPS indicates its OT-distance to the original itvPS (shown in the leftmost image).

the concentration in the TF-plane observed with ConceFT kicks in only when N is sufficiently large; on the other hand, the larger N , the more expensive the computation. To explore the trade-off, we applied ConceFT to the three noisy versions of the signal $s \in \mathcal{C}$ (see subsection (a)), with N ranging from 1 to 200. In all cases, the ConceFT algorithm uses the optimal parameters as described in subsection (c), i.e. it uses the first 2 Morse wavelets with parameters $\beta = 30$, $\gamma = 9$. In this simulation, each ConceFT computation was repeated 300 times and the mean and standard deviation of the OT distances of the ConceFT-tvPS to the itvPS were computed. Figure 6 plots the results. For each of the three noise types, the graph of the average OT-distance shows an “elbow” shape, i.e. regime in which the decrease is faster, as N increases, followed by one in which the decrease is less marked. The elbow is located around $N = 20$; the standard deviation is also quite small for this N . We accordingly decided to set $N = 20$ in our further experiments.

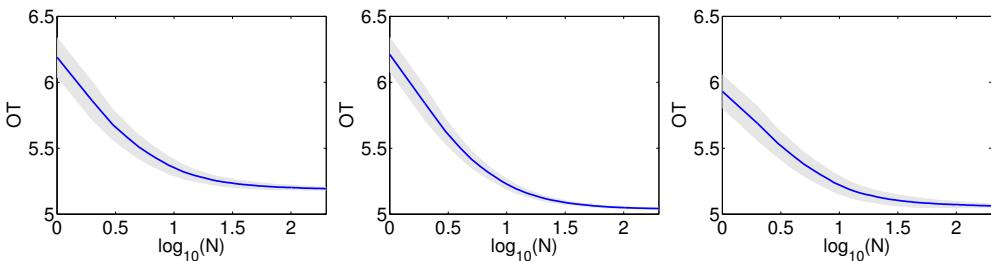


Figure 6. The OT distance as a function of the number N of random projections. The shaded band indicates the standard deviation of the OT distance at the corresponding number of random projections. The left column is for the first example and the right column for the second. From left to right, the noise types are Gaussian, ARMA(1,1) and Poisson respectively. For all three experiments, $\beta = 30$, $\gamma = 9$, and the first two Morse wavelets are used.

(e) ConceFT results for noisy signals

We now show the result of using ConceFT with the calibrated parameter choices. We illustrate the performance of ConceFT on signals of the simulation class \mathcal{C} (see subsection (a)), for a range of SNR, as well as on deterministic signals.

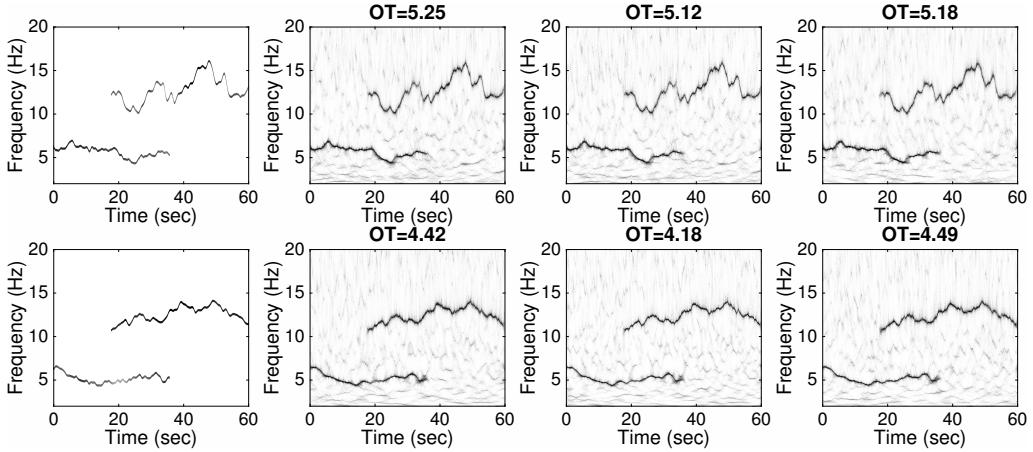


Figure 7. First row: results for the signal s ; second row: results for a new example s^* . Left to right: ideal time-varying TF power spectrum (itvPS) for the clean signal, followed by results of ConceFT with Morse wavelets after (in order) Gaussian, ARMA(1,1) or Poisson noise was added, with SNR of 0 dB. Clearly, even for a signal-to-noise ratio is as low as 0 dB, the results approximate the truth with high precision. For each of the tvPS panels, the header gives the OT distance to the corresponding itvPS.

As a warm-up, we start with the signal s seen before. The top row of Figure 7 plots the tvPS \tilde{S}_Y of the three noisy versions of s next to the itvPS S_s . To compress the dynamical range of the tvPS plots, we carry out the following procedure. We first normalize the discretized version $\tilde{S}_Y \in \mathbb{R}^{m \times n}$ of \tilde{S}_Y (where m and n stand for the number of discrete frequencies and the number of time samples, respectively) by multiplying it by a constant so that the mean energy of all entries is $\theta > 0$; that is, $\frac{1}{nm} \sum_{k=1}^m \sum_{l=1}^n (\tilde{S}_Y)_{k,l} = \theta$. We then plot a gray-scale visualization of $\mathbf{R} \in \mathbb{R}^{m \times n}$ rather than the (normalized) $\tilde{S}_Y \in \mathbb{R}^{m \times n}$ itself, where $R_{k,l} := \log(1 + \min\{\tilde{S}_{k,l}, q\})$, $k = 1, \dots, m$, $l = 1, \dots, n$ and q is a (very high) cut-off to downplay the effect of far-off outliers. We choose q to be the same for all three tvPS, so that comparable gray levels on the different tvPS panels indicate comparable values of \mathbf{R} (see Section 4f in the ESM for a more extensive discussion of choosing q and gray-scale plotting of tvPS). In the figures, we choose $\theta = 5$ and $q = 5.718$, which is the minimal value of different 99.8% quantiles of all tvPSs. The second row of Figure 7 gives the results for s^* , a signal of the simulation class \mathcal{C} that was not used (unlike s) to calibrate the parameter N of ConceFT. The results are similarly highly accurate.

Next, we study the effect on the ConceFT performance of the noise level, as quantified by SNR. To this end, we revisit the analysis of the signal s^* (and s in the ESM). For each signal, each type of noise (Gaussian, ARMA(1,1) or Poisson) and each SNR considered (SNR = x dB, where $x \in \{-7, -6, \dots, 6, 7\}$), we considered 20 independent realizations of the noise process; for each of the resulting noisy signals we carried out the ConceFT analysis and computed the OT-distance of the tvPS to the itvPS of the clean signal; we then computed the mean and the standard deviation for each. The results are shown in Figure 8. The same figure also compares the ConceFT results with those of simple SST (using either the first Morse wavelet with parameters $\beta = 30$, $\gamma = 9$ as reference wavelet, or one random linear combination of the two first Morse wavelets) and of multi-taper SST (denoted as orgMT), using the same ψ_j as ConceFT. For each of these alternate methods,

we likewise computed the mean OT-distance of the tvPS to the itvPS for 20 noise realizations. It is striking that the ConceFT method outperforms the other methods in all cases.

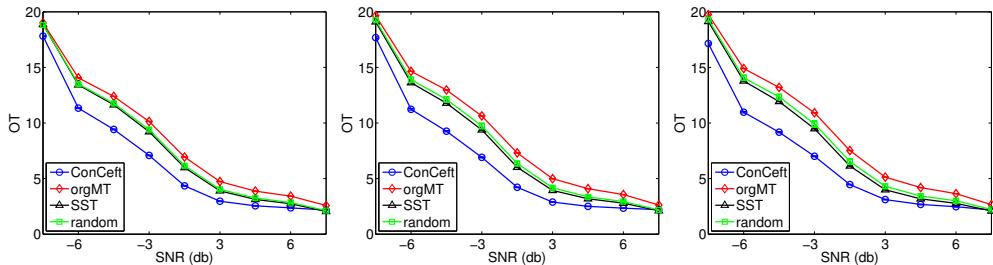


Figure 8. OT distance of ConceFT results against signal to noise ratio (SNR) of the signal $s^*(t)$, and comparison with standard SST and standard multi-taper SST (see text). Noise type (left to right): Gaussian, ARMA(1,1), and Poisson. The standard deviation is smaller, at the scale of this figure, than the height of the markers, and has not been plotted.

Finally, to address possible concerns that the randomness in the generation and plots of $\varphi'(t)$ and $A(t)$ somehow ‘help’ ConceFT in these estimations, we show in Figure 9 the results for yet another signal, which (in contrast to s and s^*) is completely deterministic; it consists of 3 components, each given by an explicit, analytic formula (again for $t \in [0, 60]$):

$$s^\circ(t) = \chi_{[10,48]}(t) \left(1 + 0.3 \cos(\pi(t-10)/20)^2 \right) \cos(\pi/3 + 5t + t^2/50) \\ + \left(0.4 + 0.9 \sin(\pi t/60)^2 \right) \cos(12t + \sin(\pi t/6)) + 1.2 \chi_{[15,60]}(t) \cos(17t + (t-35)^3/800).$$

Figure 9 shows that the results are of a quality similar to those in Figure 7.

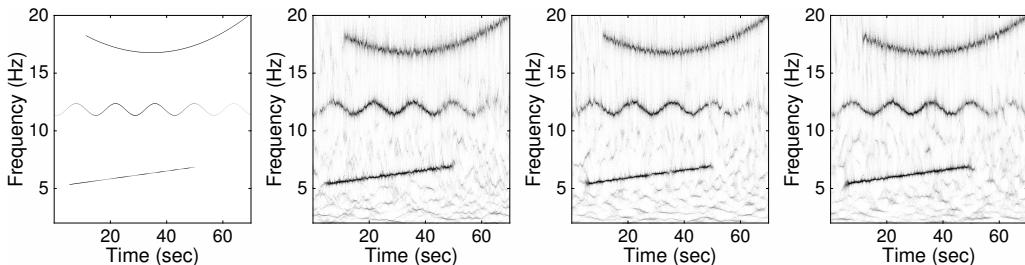


Figure 9. Results for the three-component deterministic signal s° . Left: ideal time-varying TF power spectrum (itvPS) for the clean signal, followed by results of ConceFT with Morse wavelets after (in order) Gaussian, ARMA(1,1) or Poisson noise was added, with SNR of 0 dB.

(f) ConceFT with STFT

As described earlier, the ConceFT approach can be carried out for STFT-based SST as well as for CWT-based SST. Figure 2 in section 1 already showed the results of STFT-ConceFT on one example. Other examples are shown in the ESM, together with values of the OT-distance of the STFT-ConceFT estimated tvPS to the itvPS. In these experiments, as in Figure 2, the reference windows are chosen to be Hermite functions; 20 random projections are used to compute the ConceFT averages. Although STFT-based ConceFT achieves better OT-distance with respect to the ground truth than STFT-based multi-taper SST, and also achieves a better reduction of ‘background noise’ (i.e. the structures in zones away from itvPS concentration, due to fortuitous

correlations between the noise and the overcomplete frame of TF reference functions; see the description in section 1), the performance of STFT-based ConceFT is not quite as impressive, on the class \mathcal{C} , as CWT-based ConceFT. We provide some discussion in the ESM.

5. Conclusion

We consider signals that are the linear combination of a small number of “intrinsic-mode functions”, each of which can be reasonably viewed as an oscillatory function with well-defined but time-varying amplitude and “instantaneous frequency”. We have introduced a new approach, called ConceFT, to determine the time-frequency representation of such signals, combining multi-taper estimation ideas with averaging over random projections. Theoretical analysis shows that this leads to improved estimation of the time-varying characteristics of the signals of interest; numerical results confirm the theoretical promise, even when the signals are corrupted by significant and challenging noise.

We also introduced two tools to evaluate the effectiveness of this method (or other similar methods), which may be of interest in their own right to others’ works in the field. On the one hand, we introduced a class of explicit, easy to construct signals that have time-varying characteristics that can be characterized explicitly, even though the signals themselves are not given by explicit formulas; the explicit time-varying amplitude and instantaneous frequency give a “ground truth” with which estimations can be compared. On the other hand, we introduced a distance between time-frequency representations that can be useful in comparing results obtained by different methods, by computing for each the distance to the “ground truth” time-frequency representation.

References

1. Golombek D, Rosenstein R. 2010 Physiology of circadian entrainment. *Physiol. Rev.* **90**, 1063–1102.
2. Takeda N, Maemura K. 2011 Circadian clock and cardiovascular disease. *J. Cardiol.* **57**, 249–256.
3. Wang X. 2010 Neurophysiological and computational principles of cortical rhythms in cognition. *Physiol. Rev.* **90**, 1195–1268.
4. Ahmad S, Tejuja A, Newman K, R Z, Seely A. 2009 Clinical review: a review and analysis of heart rate variability and the diagnosis and prognosis of infection. *Crit. Care.* **13**, 232.
5. Lewis GF, Furman Sa, McCool MF, Porges SW. 2012 Statistical strategies to quantify respiratory sinus arrhythmia: Are commonly used metrics equivalent? *Biol. Psychol.* **89**, 349–364.
6. Lin YT, Wu HT, Tsao J, Yien HW, Hseu SS. 2014 Time-varying spectral analysis revealing differential effects of sevoflurane anaesthesia: non-rhythmic-to-rhythmic ratio. *Acta Anaesthesiologica Scandinavica* **58**, 157–167.
7. Wu HT, Hseu SS, Bien MY, Kou YR, Daubechies I. 2013 Evaluating physiological dynamics via synchrosqueezing: Prediction of ventilator weaning. *IEEE Trans. Biomed. Eng.* **61**, 736–744.
8. Seely AJE, Bravi A, Herry C, Green G, Longtin A, Ramsay T, Fergusson D, McIntyre L, Kubelik D, Maziak DE, Ferguson N, Brown SM, Mehta S, Martin C, Rubenfeld G, Jacono FJ, Clifford G, Fazekas A, Marshall J. 2014 Do heart and respiratory rate variability improve prediction of extubation outcomes in critically ill patients? *Crit. Care* **18**, R65.
9. Baudin F, Wu HT, Bordessoule A, Beck J, Jouvet P, Frasch M, Emeriaud G. 2014 Impact of ventilatory modes on the breathing variability in mechanically ventilated infants. *Frontiers in Pediatrics, section Neonatology* **2**.
10. Flandrin P. 1999 *Time-frequency/time-scale analysis*, volume 10 of *Wavelet Analysis and its Applications*. Academic Press Inc.

11. Thakur G, Wu HT. 2011 Synchrosqueezing-based Recovery of Instantaneous Frequency from Nonuniform Samples.
SIAM J. Math. Anal. **43**, 2078–2095.
12. Huang NE, Shen Z, Long SR, Wu M, Shih H, Zheng Q, Yen NC, Tung CC, Liu HH. 1998 The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis.
Proc. R. Soc. Lond. A **454**, 903–995.
13. Wu Z, Huang NE. 2009 Ensemble empirical mode decomposition: a noise-assisted data analysis method.
Adv. Adapt. Data Anal. **1**, 1–41.
14. Tavallali P, Hou T, Shi Z. 2014 Extraction of intrawave signals using the sparse time-frequency representation method.
Multiscale Modeling & Simulation **12**, 1458–1493.
15. Lin L, Wang Y, Zhou H. 2009 Iterative filtering as an alternative for empirical mode decomposition.
Adv. Adapt. Data Anal. **1**, 543–560.
16. Huang C, Wang Y, Yang L. 2009 Convergence of a convolution-filtering-based algorithm for empirical mode decomposition.
Adv. Adapt. Data Anal. **1**, 561–571.
17. Chui CK, Mhaskar H. 2015 Signal decomposition and analysis via extraction of frequencies.
Appl. Comput. Harmon. Anal. .
18. Galiano G, Velasco J. 2014 On a non-local spectrogram for denoising one-dimensional signals.
Applied Mathematics and Computation **244**, 1–13.
19. De Livera AM, Hyndman RJ, Snyder RD. 2011 Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing.
J. Am. Stat. Assoc. **106**, 1513–1527.
20. Daubechies I, Maes S. 1996 A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models.
Wavelets in Medicine and Biology pp. 527–546.
21. Daubechies I, Lu J, Wu HT. 2011 Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool.
Appl. Comput. Harmon. Anal. **30**, 243–261.
22. Wu HT, Flandrin P, Daubechies I. 2011 One or Two Frequencies? The Synchrosqueezing Answers.
Adv. Adapt. Data Anal. **3**, 29–39.
23. Wu HT. 2011 *Adaptive Analysis of Complex Data Sets*.
Ph.D. thesis, Princeton University.
[Https://sites.google.com/site/hautiengwu/home](https://sites.google.com/site/hautiengwu/home)
24. Kodera K, Gendrin R, Villedary C. 1978 Analysis of time-varying signals with small bt values.
IEEE Trans. Acoust., Speech, Signal Processing **26**, 64 – 76.
25. Auger F, Flandrin P. 1995 Improving the readability of time-frequency and time-scale representations by the reassignment method.
IEEE Trans. Signal Process. **43**, 1068 –1089.
26. Chassande-Mottin E, Auger F, Flandrin P. 2003 Time-frequency/time-scale reassignment.
In *Wavelets and signal processing*, Appl. Numer. Harmon. Anal., pp. 233–267. Boston, MA: Birkhäuser Boston.
27. Auger F, Chassande-Mottin E, Flandrin P. 2012 Making reassignment adjustable: The levenberg-marquardt approach.
In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pp. 3889–3892.
28. Thakur G. 2014 The synchrosqueezing transform for instantaneous spectral analysis.
In *Excursions in Harmonic Analysis vol. 3*. Springer.
29. Chui CK, Lin YT, Wu HT. 2015 Real-time dynamics acquisition from irregular samples – with application to anesthesia evaluation.
Analysis and Applications In Press.
30. Chen YC, Cheng MY, Wu HT. 2014 Nonparametric and adaptive modeling of dynamic seasonality and trend with heteroscedastic and dependent errors.
J. Roy. Stat. Soc. B **76**, 651–682.

31. Maes S. 1995 The synchrosqueezed representation yields a new reading of the wavelet transform.
In *Proceedings SPIE95 on OE/Aerospace Sensing and Dual Use Photonics*. Orlando, FL.
32. Wu HT. 2013 Instantaneous frequency and wave shape functions (I).
Appl. Comput. Harmon. Anal. **35**, 181–199.
33. Meignen S, Oberlin T, McLaughlin S. 2012 A new algorithm for multicomponent signals analysis based on synchrosqueezing: With an application to signal sampling and denoising.
IEEE Trans. Signal Process. **60**, 5787–5798.
34. Lin YT, Flandrin P, Wu HT. 2015 Interpolation-induced reflection artifact in the reassignment technique – with anesthesia example.
Submitted. ArXiv e-prints.
35. Thakur G, Brevdo E, Fuckar NS, Wu HT. 2013 The synchrosqueezing algorithm for time-varying spectral analysis: robustness properties and new paleoclimate applications.
Signal Processing **93**, 1079–1094.
36. Li C, Liang M. 2012 A generalized synchrosqueezing transform for enhancing signal time-frequency representation.
Signal Processing **92**, 2264 – 2274.
37. Yang H. 2015 Synchrosqueezed Wave Packet Transforms and Diffeomorphism Based Spectral Analysis for 1D General Mode Decompositions.
Appl. Comput. Harmon. Anal. **39**, 33–66.
38. Oberlin T, Meignen S, Perrier V. 2015 Second-order synchrosqueezing transform or invertible reassignment? towards ideal time-frequency representations.
IEEE Trans. Signal Process. **63**, 1335–1344.
39. Xi S, Cao H, Chen X, Zhang X, Jin X. 2015 A frequency-shift synchrosqueezing method for instantaneous speed estimation of rotating machinery.
ASME. J. Manuf. Sci. Eng. **137**, 031012–031012–11.
40. Yang H, Ying L. 2013 Synchrosqueezed wave packet transform for 2d mode decomposition.
SIAM Journal on Imaging Science **6**, 1979–2009.
41. Yang H, Ying L. 2014 Synchrosqueezed curvelet transform for 2d mode decomposition.
SIAM J. Math. Anal. **46**, 2052–2083.
42. Stankovski T, Duggento A, McClintock PVE, Stefanovska A. 2012 Inference of Time-Evolving Coupled Dynamical Systems in the Presence of Noise.
Physical Review Letters **109**, 1–5.
43. Iatsenko D, Bernjak A, Stankovski T, Shiogai Y, Owen-Lynch P, Clarkson PBM, McClintock PVE, Stefanovska A. 2013 Evolution of cardiorespiratory interactions with age Evolution of cardiorespiratory interactions with age.
Phil. Trans. R. Soc. A **371**, 1–18.
44. Wu HT, Chan YH, Lin YT, Yeh YH. 2014 Using synchrosqueezing transform to discover breathing dynamics from ecg signals.
Appl. Comput. Harmon. Anal. **36**, 354–359.
45. Wu HT, Talmon R, Lo YL. 2015 Assess sleep stage by modern signal processing techniques.
IEEE Transactions on Biomedical Engineering **62**, 1159–1168.
46. Lin YT. 2015 *The Modeling and Quantification of Rhythmic to Non-rhythmic Phenomenon in Electrocardiography during Anesthesia*.
Ph.D. thesis, National Taiwan University.
ArXiv 1502.02764
47. Li C, Liang M. 2012 Time-frequency signal analysis for gearbox fault diagnosis using a generalized synchrosqueezing transform.
Mechanical Systems and Signal Processing **26**, 205–217.
48. Feng Z, Chen X, Liang M. 2015 Iterative generalized synchrosqueezing transform for fault diagnosis of wind turbine planetary gearbox under nonstationary conditions.
Mechanical Systems and Signal Processing **52–53**, 360 – 375.
49. Guharay S, Thakur G, Goodman F, Rosen S, Houser D. 2013 Analysis of non-stationary dynamics in the financial system.
Economics Letters **121**, 454–457.
50. Vatter T, Wu HT, Chavez-Demoulin V, Yu B. 2013 Non-parametric estimation of intraday spot volatility: disentangling instantaneous trend and seasonality.
SSRN e-prints .2330159

51. Wang P, J G, Wang Z. 2014 Time-frequency analysis of seismic data using synchrosqueezing transform.
Geoscience and Remote Sensing Letters, IEEE **11**, 2042–2044.
52. Herrera RH, Han J, van der Baan M. 2014 Applications of the synchrosqueezing transform in seismic time-frequency analysis.
Geophysics **79**, V55–V64.
53. Tary JB, Herrera RH, Han J, van der Baan M. 2014 Spectral estimation—what is new? what is next?
Reviews of Geophysics pp. 723–749.
54. Li PC, Sheu YL, Laughlin C, , Chu SI. 2013 Role of laser-driven electron-multirescattering in resonance-enhanced below-threshold harmonic generation in he atoms.
Phy. Rev. A. **90**, 041401(R).
55. Sheu YL, Hsu LY, Wu HT, Li PC, Chu SI. 2014 A new time-frequency method to reveal quantum dynamics of atomic hydrogen in intense laser pulses: Synchrosqueezing transform.
AIP Advances **4**, 117138.
56. Li PC, Sheu YL, Laughlin C, Chu SI. 2015 Dynamical origin of near- and below-threshold harmonic generation of Cs in an intense mid-infrared laser field.
Nature Communication **6**.
57. Yang H, Lu J, Ying L. 2014 Crystal image analysis using 2d synchrosqueezed transforms.
Submitted
58. Yang H, Lu J, Brown WP, Daubechies I, Ying L. 2015 Quantitative canvas weave analysis using 2d synchrosqueezed transforms.
IEEE Signal Processing Magazine To appear.
59. Thomson DJ. 1982 Spectrum estimation and harmonic analysis.
Proceedings of the IEEE **70**, 1055–1096.
60. Percival DB, Walden AT. 1993 *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*.
Cambridge University Press.
61. Babadi B, Brown EN. 2014 A review of multitaper spectral analysis.
IEEE Trans. Biomed. Eng. **61**, 1555–1564.
62. Xiao J, Flandrin P. 2007 Multitaper Time-Frequency Reassignment for Nonstationary Spectrum Estimation and Chirp Enhancement.
IEEE Trans. Signal Process. **55**, 2851–2860.
63. Lin YT, Hseu SS, Yien HW, Tsao J. 2011 Analyzing autonomic activity in electrocardiography about general anesthesia by spectrogram with multitaper time-frequency reassignment.
IEEE-BMEI **2**, 628–632.
64. Orini M, Bailón R, Mainardi LT, Laguna P, Flandrin P. 2012 Characterization of dynamic interactions between cardiovascular signals by time-frequency coherence.
IEEE Trans. Biomed. Eng. **59**, 663–73.
65. Daubechies I. 1992 *Ten lectures on wavelets*.
SIAM.
66. Kowalski M, Meynard A, Wu HT. 2015 Convex Optimization approach to signals with fast varying instantaneous frequency.
ArXiv e-prints 1503.07591 .
67. Gallager R. 2008 Circularly-Symmetric Gaussian random vectors
.Http://www.rle.mit.edu/rgallager/documents/CircSymGauss.pdf
68. Daubechies I. 1988 Time-frequency localization operators: a geometric phase space approach.
IEEE Trans. Inform. Theory **34**, 605–612.
69. Daubechies I, Paul T. 1988 Time-frequency localization operators: a geometric phase space approach II. The use of dilations.
Inverse Problems **4**, 661–680.
70. Olhede SC, Walden AT. 2002 Generalized Morse Wavelets.
IEEE Trans. Signal Process. **50**, 2661–2671.
71. Lin YT, Wu HT. 2015 Application of ConceFT to heart rate variability analysis and analgesia analysis.
In preparation .