Multitaper time-frequency reassignment for nonstationary spectrum estimation and chirp enhancement

Jun Xiao, Patrick Flandrin

Multitaper Time-Frequency Reassignment for Nonstationary Spectrum Estimation and Chirp Enhancement

Jun Xiao and Patrick Flandrin, *Fellow, IEEE*

Abstract

A method is proposed for obtaining time-frequency distributions of chirp signals embedded in nonstationary noise, with the twofold objective of a sharp localization for the chirp components and a reduced level of statistical fluctuations for the noise. The technique consists in combining time-frequency reassignment with multitapering, and two variations are proposed. The first one, primarily aimed at nonstationary spectrum estimation, is based on sums of estimates with different tapers, whereas the second one makes use of differences between the same estimates for a sake of chirp enhancement. The principle of the technique is outlined, its implementation based on Hermite functions is justified and discussed, and some examples are provided for supporting the efficiency of the approach, both qualitatively and quantitatively.

Index Terms

Multitapers, time-frequency, reassignment, chirps.

I. Introduction

In nonstationary contexts, it is well-known [9] that Fourier-based methods of (time-varying) spectrum estimation are classically faced with intrinsic limitations and different kinds of trade-offs: (*i*) from a statistical point of view, the usual bias-variance trade-off inherent to any estimation procedure is amplified when analyzing nonstationary stochastic processes, since time-averaging—aimed at reducing variance—introduces some bias not only in the frequency direction but also in time; (*ii*) from a geometrical perspective, windowing—aimed at guaranteeing a form of local stationarity—ends up with a different kind of trade-off related to the time-frequency (TF) localization in the case of chirp-like signals. Such difficulties have been recognized long ago, and numerous studies have tried to address the problem. As far as localization is concerned, Wigner-based approaches have been developed and shown to outperform windowed (Fourier or wavelet-based) methods, at least in the case of noise-free single chirps [9]. In

J. Xiao and P. Flandrin are with the Physics Department (UMR 5672 CNRS), Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07 France. Tel.: +33 (0)4 72 72 81 60; Fax: +33 (0)4 72 72 80 80; E-mail: {Jun.Xiao,flandrin}@ens-lyon.fr. This work was supported in part by the ENS-ECNU Doctoral Program. Part of this paper was first presented at the 14th European Signal Proc. Conf. EUSIPCO-06 (Florence, I) [17].

more realistic situations of multi-chirps, a dramatic improvement over both Fourier and Wigner-based methods has come from the use of the *reassignment* technique [3], with an efficiency that is however limited to the cases where the signal-to-noise ratio is high enough. Turning to the estimation issue in a statistical sense, different attempts have been made to take advantage of the idea of *multitapering*, pioneered by D.J. Thomson in a stationary setting [14], and thanks to which an improved statistical stability can be obtained without a time-averaging step. Numerous extensions of multitaper techniques to nonstationary situations have been proposed, see, e.g., [6], [11] and, for a more comprehensive covering of the topics and of the literature, [5], [16] and references therein. When extended in a direct way, the "classical" method of multitapering suffers however still from the TF localization trade-off mentioned above. Some attempts have been made to circumvent this limitation, in particular by identifying chirp-like components and excising them prior applying the multitaper machinery [5]. The purpose of this paper is to avoid such a complication and to rather combine multitapering (for a sake of variance reduction) with reassignment (for localization).

More precisely, the paper is organized as follows. In Section II, the issue of nonstationary spectrum estimation is briefly addressed, and the key concepts of reassignment and multitapering are recalled. A first combination of both ideas is discussed in Section III-A, whose aim is to improve statistical stability while preserving TF localization. This is achieved by summing estimates based on different tapers, the rationale being that such estimates tend to behave as well-localized, weakly correlated, surrogate data whose sum combines coherently chirp components and incoherently noise contributions. A companion perspective is then envisioned in Section III-B, which exploits differences between estimates, the idea being in this case to get rid of noise by masking those regions where different tapers lead to significantly different estimates. Quantitative performance evaluations are provided for supporting the efficiency of the approach, and possibilities and limitations of both variations are illustrated in Section III-C by a number of numerical experiments.

The algorithms related to this paper and the specific procedures used for producing most of the figures can be freely obtained as Matlab routines [1]. These must be explicitly considered as part of the paper, since they not only allow the readers to reproduce the claimed numerical results in a spirit of "reproducible research" (see, e.g., [2]), but also to make their own variations on their production.

II. NONSTATIONARY SPECTRUM ESTIMATION

Defining a time-varying "spectrum" for a nonstationary process $\{x(t), t \in \mathbb{R}\}$ is a question that has no unique answer[1]. Among the various possibilities stands first the *Wigner-Ville Spectrum* (WVS), whose definition reads:

$$\mathbf{W}_x(t,f) = \int_{-\infty}^{+\infty} \mathbb{E}\left\{x\left(t+\frac{\tau}{2}\right)x^*\left(t-\frac{\tau}{2}\right)\right\} e^{-i2\pi f\tau}\, d\tau,$$

where t and f refer to time and frequency, respectively, and $\mathbb{E}\{.\}$ stands for the expectation operator. This definition, though not unique, presents the advantage of extending the usual concept of *Power Spectrum Density* (PSD) and

[1]The brief overview in this Section summarizes classical results that are exposed, e.g., in [9], where further references and details can be found.

making it time-dependent in a rather natural way. Without entering into much details, it is worth recalling that the WVS reduces to the PSD at every instant if the analyzed process happens to be stationary. Moreover, it satisfies the important property of preserving time and frequency supports, and its marginal distributions are directly related to meaningful quantities (variance in time, Loève's distribution function in frequency).

If we now introduce the (non random) quantity

$$W_x(t,f) = \int_{-\infty}^{+\infty} x\left(t+\frac{\tau}{2}\right)\, x^*\left(t-\frac{\tau}{2}\right)\, e^{-i2\pi f\tau}\, d\tau,$$

which is referred to as the *Wigner-Ville Distribution* (WVD), it can be shown that, under mild conditions, the WVS of a process is nothing but the ensemble average of the WVDs of its possible realizations:

$$\mathbf{W}_x(t,f) = \mathbb{E}\left\{W_x(t,f)\right\}. \tag{1}$$

Given one observed realization of a nonstationary process, estimating the WVS amounts to find a substitute for the unattainable ensemble average entering eq.(1). One standard way is to assume for $x(t)$ a form of local stationarity in both time and frequency, i.e., some locally slow evolution of the WVS in the two directions. Such an assumption paves the road for a replacement of the ensemble average at a given TF location by a local smoothing over a neighbouring domain. This idea can be formalized by introducing a TF smoothing kernel $\Pi(t,f)$ and defining as a WVS estimator the quantity:

$$\hat{\mathbf{W}}_x(t,f) = \iint_{-\infty}^{+\infty} W_x(s,\xi)\, \Pi(s-t,\xi-f)\, ds\, d\xi. \tag{2}$$

This expression, which coincides with *Cohen's class* $C_x(t,f;\Pi)$ for the observed realization, offers a unified setting in which the two trade-offs mentioned previously (regarding fluctuations and localization) appear clearly. If we consider for instance the toy example of a linear chirp embedded in broadband noise, the fluctuations of the WVD due to noise will be smoothed out provided that $\Pi(t,f)$ is low-pass. However, the WVD of the linear chirp (which has the unique property of being perfectly localized along the instantaneous frequency) will be smoothed out too. A way out is however possible by reconsidering the apparently contradictory issues of fluctuations reduction and localization at the light of the two refinements offered by reassignment and multitapering.

A. Reassignment

The smoothing kernel in (2) is *a priori* arbitrary, and a possibility is to choose as $\Pi(t,f)$ the WVD of some signal $h(t)$ supposed to be well localized in both time and frequency, a property that carries over to its WVD. Doing so, it is easy to show that $C_x(t,f;W_h) = S_x^{(h)}(t,f)$, where $S_x^{(h)}(t,f)$ is nothing but the *spectrogram* of $x(t)$ with window $h(t)$, a TF distribution that is usually rather expressed as:

$$S_x^{(h)}(t,f) = \left|F_x^{(h)}(t,f)\right|^2, \tag{3}$$

where

$$F_x^{(h)}(t,f) = \int_{-\infty}^{+\infty} x(s)\, h(s-t)\, e^{-i2\pi f s}\, ds$$

stands for the *Short-Time Fourier Transform* (STFT).

A spectrogram appears therefore as an estimator for the WVS, with the well-known TF localization trade-off attached to this type of distribution: the shorter the window $h(t)$, the better the time localization, but the poorer the frequency localization, and *vice versa*. In this respect however, the alternative interpretation of the spectrogram as a smoothed WVD (according to eq.(2)) rather than as a squared STFT (according to eq.(3)) gives the clue for improving upon its localization limitations. Indeed, if we recall that the WVD of a linear chirp perfectly localizes on a TF line, the spreading of any corresponding spectrogram just comes from the fact that, when centering the analysis window at some TF point that does not belong to this line, a non-zero contribution is nevertheless obtained as long as the line passes through the local TF window (whose joint support cannot be made arbitrarily small). Reasoning by a mechanical analogy identifying energy with mass, the situation is as if a whole distribution of mass within a domain (here, the TF window) would be replaced by one single number (the total mass) assigned to the *geometrical* center of the domain. Such an assignment is clearly not well adapted to situations where the distribution is not uniform over the domain, a much more meaningful assignment being the *center of mass* within the domain. This is precisely the essence of the "reassignment" technique [3], [10], which consists in (*i*) evaluating for each TF location, not only the integrated signal WVD within the TF domain of the window (in other words, the spectrogram value at this point), but also the center of mass of the signal WVD, and (*ii*) reassigning the spectrogram value to this location. In the idealized case where one single linear chirp intersects the TF window, it is clear that the center of mass necessarily belongs to the line along which the WVD is localized, thus guaranteeing a perfect localization of the spectrogram after its reassignment.

Previous studies [3], [10] have shown that an efficient evaluation of the local centers of mass $(\hat{t}_{t,f}, \hat{f}_{t,f})$ can be made implicitly, according to

$$\left\{ \begin{array}{lcl} \hat{t}_{t,f} & = & t + \mathrm{Re}\{F_x^{(\mathcal{T}h)}(t,f)/F_x^{(h)}(t,f)\} \\ \hat{f}_{t,f} & = & f - \mathrm{Im}\{F_x^{(\mathcal{D}h)}(t,f)/F_x^{(h)}(t,f)\} \end{array} \right.$$

where the two additional windows needed in the computation are defined from the mother window $h(t)$ as $(\mathcal{T}h)(t) = t\,h(t)$ and $(\mathcal{D}h)(t) = (dh/dt)(t)$. Given the field of all above centroids, the reassigned spectrogram $RS_x^{(h)}(t,f)$ attached to the conventional spectrogram $S_x^{(h)}(t,f)$ follows as:

$$RS_x^{(h)}(t,f) = \iint_{-\infty}^{+\infty} S_x^{(h)}(s,\xi)\,\delta(t-\hat{t}_{s,\xi})\delta(f-\hat{f}_{s,\xi})ds d\xi.$$

B. *Multitapering*

In the case of stationary processes, the spectral characterization is fully described by means of the PSD $\mathbf{S}_x(f)$, which could be thought of as:

$$\mathbf{S}_x(f) = \lim_{T\to\infty} \mathbb{E}\left\{ \frac{1}{T} \left| \int_{-T/2}^{+T/2} x(t)\,e^{-i2\pi ft}\,dt \right|^2 \right\}.$$

In practice, the above quantity is unattainable when only one realization of finite duration is given. The Squared Fourier Transform (SFT) of a single observation is a crude, non consistent estimator, whose variance is of the order of the squared PSD [13]. Since an improvement can only come from averaging (almost) uncorrelated estimations, an ergodic argument suggests to chop the observation into (almost) disjoint blocks and average their SFTs (a procedure sometimes referred to as the Welch method of averaged periodograms). Adopting the notation of spectrograms, it turns out that the corresponding (Welch) estimator can be written as:

$$\hat{\mathbf{S}}_{x,W}(f) = \frac{1}{K}\sum_{k=1}^{K} S_x^{(h)}(t_k, f),$$

where the spacing $\Delta = t_{k+1} - t_k$ between adjacent t_k's is of the order of the window length. Assuming that this spacing ensures an approximate decorrelation between blocks, one can expect that the variance is inversely proportional to the number of blocks K (i.e., roughly T/Δ for an observation of duration T). Although variance can be decreased this way by increasing K, the finite duration constraint necessarily leads to shorten Δ, increasing in turn the bias in frequency since a window of duration Δ has a frequency resolution of the order of $1/\Delta$.

In order to circumvent this trade-off, D.J. Thomson suggested [14] to still average SFTs stemming from (almost) uncorrelated sequences in order to reduce variance, but to construct such sequences by using for each of them the whole data set so as to not sacrifice bias. The way to achieve this program consists in projecting the observation on a family of basis functions $\{h_k(t), k \in \mathbb{N}\}$ that are orthonormal over the observation interval $(-T/2, +T/2)$. This results in a (Thomson) estimator that can be written as:

$$\hat{\mathbf{S}}_{x,T}(f) = \frac{1}{K}\sum_{k=1}^{K} S_x^{(h_k)}(0, f), \tag{4}$$

with all windows $h_k(t)$ extending over $(-T/2, +T/2)$.

Assuming that the spectrum is flat over a given bandwidth B associated with the expected frequency resolution, the basis can be obtained as the family of orthonormal functions (on the given time interval) that maximize their energy in the given frequency band. The solution to this problem is given by the family of functions known as the *Prolate Spheroidal Wave Functions* or, in a discrete-time setting, as the *Discrete Prolate Spheroidal Sequences* (DPSS) [13].

Extending the approach to nonstationary situations is appealing [11], [16]. The main reason is that the inconsistency (and large variance) of a PSD estimator based on a crude SFT directly carries over to spectrograms considered as WVS estimators. The traditional way out would be to smooth over time and frequency, but at the expense of further increasing bias. In this respect, resorting to multitapers allows for a variance reduction with a bias that only sticks to the common length of the different windows. This is certainly an improvement as compared to (smoothed) spectrograms with respect to statistical efficiency, but the question of TF resolution still remains not improved. Wedding multitapering with reassignment is therefore proposed as the key for such an improvement.

III. Multitaper Time-Frequency Reassignment

A. Nonstationary Spectrum Estimation

1) Principle and implementation: The direct application of multitapering to nonstationary processes consists in making the estimator (4) time-dependent according to [11], [16]:

$$\hat{\mathbf{S}}_{x,T}(f) \rightarrow S_{x,K}(t,f) = \frac{1}{K}\sum_{k=1}^{K} S_x^{(h_k)}(t,f),$$

with the functions $\{h_k(t), k \in \mathbb{N}\}$ given again as a family of short-time windows. What we propose here is to adopt the same strategy, but applied to reassigned spectrograms, i.e., to consider as a WVS estimator the quantity:

$$RS_{x,K}(t,f) = \frac{1}{K}\sum_{k=1}^{K} RS_x^{(h_k)}(t,f). \tag{5}$$

The rationale for this approach can be justified in a twofold way: (*i*) as far as chirp components are concerned, reassignment increases localization in a way that can be made independent of the window, thus permitting (5) to act as a *coherent* averaging; (*ii*) in noise regions on the contrary, the same windows lead to uncorrelated surrogate data whose TF distributions are different, (5) acting in this case as a form of *incoherent* averaging tending to smooth the estimate.

In stationary spectrum estimation, multitapers are chosen as DPSSs because the data is of finite duration and estimation concerns frequency only. In the nonstationary case, there is no *a priori* reason to dissymmetrize time and frequency by choosing tapers that would be perfectly localized in the time domain rather than in frequency. Indeed, it makes much more sense to fully exploit the two degrees of freedom offered by the TF plane and, as suggested in [5], to rather pick up those functions that maximally concentrate in TF domains with elliptic symmetry. As shown in [7] in the context of coherent states and in [8] within a Wigner framework, those functions are the *Hermite* functions (HF), whose definition is given by

$$h_k(t) = \left((t-\mathcal{D})^k g\right)(t)/\sqrt{\pi^{1/2}2^k k!},$$

with $g(t) = \exp\{-t^2/2\}$. In practice, HFs can be computed recursively, according to

$$h_k(t) = g(t)\, H_k(t)/\sqrt{\pi^{1/2}2^k k!},$$

where the $\{H_k(t), k \in \mathbb{N}\}$ stand for the Hermite polynomials that obey the recursion:

$$H_k(t) = 2\,t\,H_{k-1}(t) - 2\,(k-2)\,H_{k-2}(t), k \geq 2$$

with the initialization $H_0(t) = 1$ and $H_1(t) = 2\,t$.

Not only the HFs are orthonormal, but they also guarantee a perfect localization of the corresponding reassigned spectrograms in the case of a linear chirp, whatever k. This can be easily understood by noting that the WVD of a HF (which is basically a 2D Laguerre function) has elliptic symmetry [5], [8]. Recalling that the WVD is covariant with respect to dilations and rotations, it is enough to check that reassignment ends up with a perfect localization in the case of a pure tone, which can be done by an elementary calculation.

In the context of reassignment, HFs offer one further advantage, as compared to DPSSs. In the standard implementation of spectrogram reassignment, only the mother window $h(t)$ has to be given and the two additional windows $(\mathcal{T}h)(t)$ and $(\mathcal{D}h)(t)$ that are needed are evaluated numerically [10]. This may cause difficulties when differentiating tapers whose order k is large, since they are highly oscillating. This problem can be circumvented when using HFs since their successive derivatives also obey a recursion that can be explicitly plugged in the algorithm, namely

$$(\mathcal{D}h_k)(t) = (\mathcal{T}h_k)(t) - \sqrt{2(k+1)}\, h_{k+1}(t).$$

The effectiveness of this implementation of multitaper reassignment is illustrated in Fig. 1 which evaluates the spreading of cumulative Hermite estimates in the case of a linear chirp and of a white Gaussian noise (wGn). The measure used for this evaluation is a Rényi entropy of order p, defined as [4]

$$R_p(P) = \frac{1}{1-p}\, \log_2 \sum_{n=1}^{N} \sum_{m=1}^{N} (P[n,m])^p\, ; p > 0,$$

for any normalized discrete TF distribution $P[n,m]$ with N points in time and N frequency bins.

In the case where the distribution is perfectly localized on, say, the diagonal of the plane (the situation expected to happen when reassigning the spectrogram of a linear chirp), we have ideally $P_{chirp}[n,m] = \delta_{n,m}/N$ and $R_p(P_{chirp}) = \log_2 N$. This situation contrasts with that of a wGn whose distribution is expected to spread uniformly all over the plane as $P_{noise}[n,m] = 1/N^2$, thus leading to $R_p(P_{noise}) = 2\, \log_2 N$. In the multitaper spectrogram case (without reassignment), Fig. 1 evidences that the spreading can approach the theoretical prediction in the noise case by increasing the maximum taper order K, but at the expense of increasing at the same time the spreading in the chirp case, with a minimum value (for a taper length matched to the chirp rate) that remains significantly large. This contrasts with multitaper reassigned spectrograms which compare favourably with the theoretical prediction for chirps for all Hermite tapers over a wide range of orders and lengths. Comparison is less favourable in the noise case, but it is continuously improved when increasing K, without degrading the corresponding localization for chirps.

2) Performance evaluation: In order to go beyond the numerical check described above, a first example is concerned with the idealized situation of a bandpass filtered wGn within a time-limited support. Although not strictly attainable (because of the uncertainty relation), the model $M_D(t,f)$ for the WVS of such an observation is the indicator function of a rectangle of area D within the TF plane[2]. Fig. 2 illustrates what happens in such a case by comparing the WVD and a sample (reassigned) spectrogram with the corresponding multitaper estimates based on $K = 10$ Hermite functions. The two effects of reduced fluctuations and support preservation are clearly visible, and ensemble averages (based on 10 independent estimates) are also provided for supporting the effectiveness of the approach and its improved convergence rate as compared to an empirical estimate of the WVS.

[2]As suggested by one of the reviewers, the evaluation could be carried out with the admissible model of a Gaussian modulated Gaussian bandpass filtered white Gaussian noise. This model is detailed in [1], together with the associated Matlab codes, leading to results which proved to be similar to those reported here.

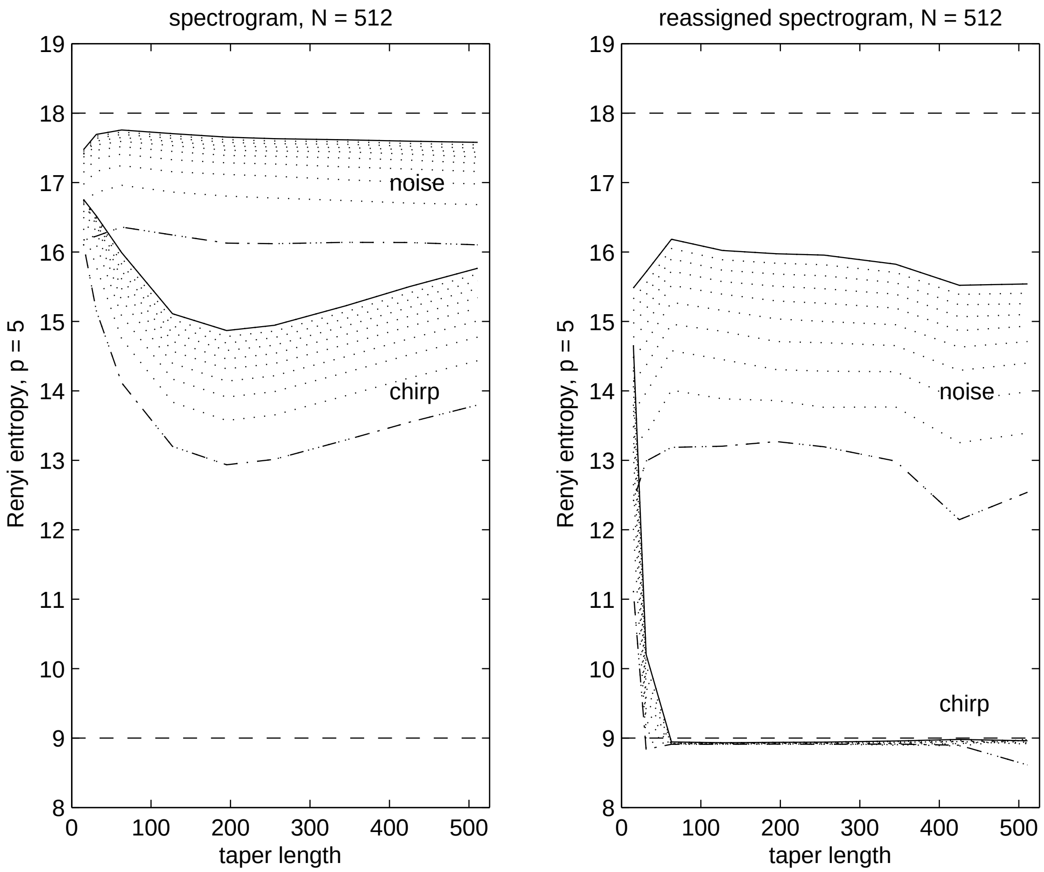

Fig. 1. Spreading of Hermite multitaper estimates. Both diagrams (left: spectrogram; right: reassigned spectrogram) display the Rényi entropies R_5 of Hermite multitaper TF estimates, as a function of the taper length, and parameterized by the maximum Hermite order K, from $K = 1$ (dashed dotted line) to $K = 10$ (full line). A linear chirp signal and a sample of wGn (512 points each) are considered, and the theoretical predictions corresponding to a perfect localization and a uniform spreading ($R_5 = 9$ and 18, respectively, see text) are superimposed as dotted lines.

Fig. 2 gives a qualitative account of the behaviour of the method, that can be supplemented by the quantitative measure

$$E(K) = \frac{1}{\|M_D\|_1} \iint_{-\infty}^{+\infty} \left| \hat{\mathbf{W}}_{x,K}(t,f) - M_D(t,f) \right| \, dt\, df, \tag{6}$$

where $\hat{\mathbf{W}}_{x,K}(t,f)$ stands for the WVS estimate ($S_{x,K}(t,f)$ or $RS_{x,K}(t,f)$), the L_1-norm being here chosen so as to put emphasis on localization in the estimates.

Fig. 3 presents results with different domains, all rectangular and centered in the analyzed TF region, but with different areas D. In the pure wGn case where the model support extends over the whole plane (in this case, $D = 256$), we observe for both spectrograms and reassigned spectrograms that the error measure behaves asymptotically as $E(K) \propto K^{-1/2}$ when using K tapers. In the spectrogram case, this can be justified since, for each taper, the

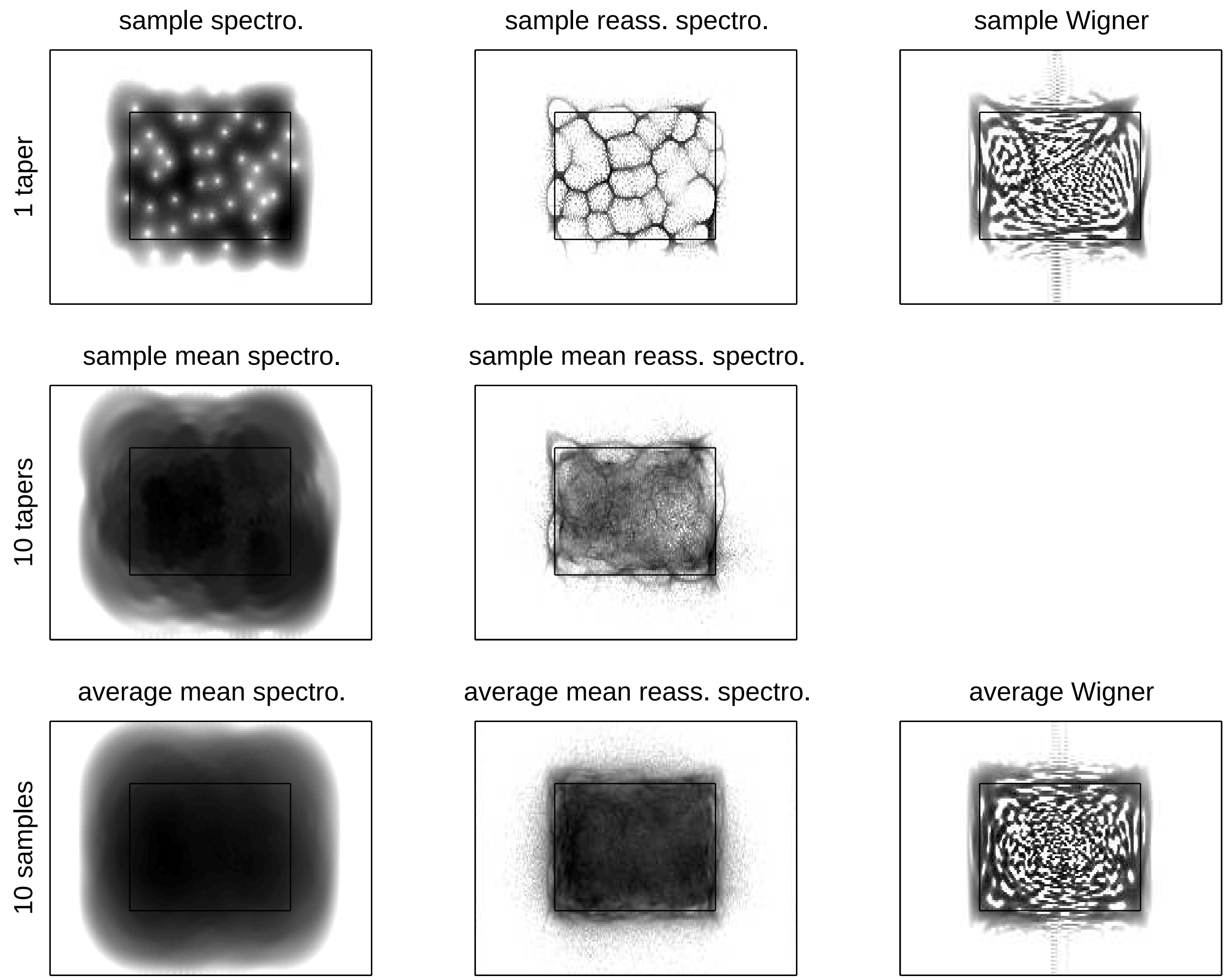

Fig. 2. Comparison of noise WVS estimates. Each diagram represents a WVS estimate in the case of a wGn limited in time and frequency within the superimposed rectangular domain. The first row consists of a spectrogram, its reassigned version and the WVD, based on one realization. The corresponding multitaper estimates (10 Hermite functions) are given in the middle row, whereas the bottom row displays ensemble averages of such estimates (10 independent realizations), together with the empirical WVS estimate on the same data set. In each diagram, time is horizontal, frequency vertical and the energy is coded logarithmically with a dynamic range of 30 dB.

values are known to be χ^2 distributed with 2 degrees of freedom [13]. It follows from the orthogonality of the tapers that the sum of the K first Hermite spectrograms is also χ^2 distributed, but with $2K$ degrees of freedom. Given a χ^2 distribution with ν degrees of freedom:

$$p_\nu(y) = y^{\nu/2-1}\, e^{-y/2} 2^{-\nu/2} / \Gamma(\nu/2); y \geq 0,$$

its absolute mean deviation (the quantity on which the performance measure (6) is based) can be evaluated by using

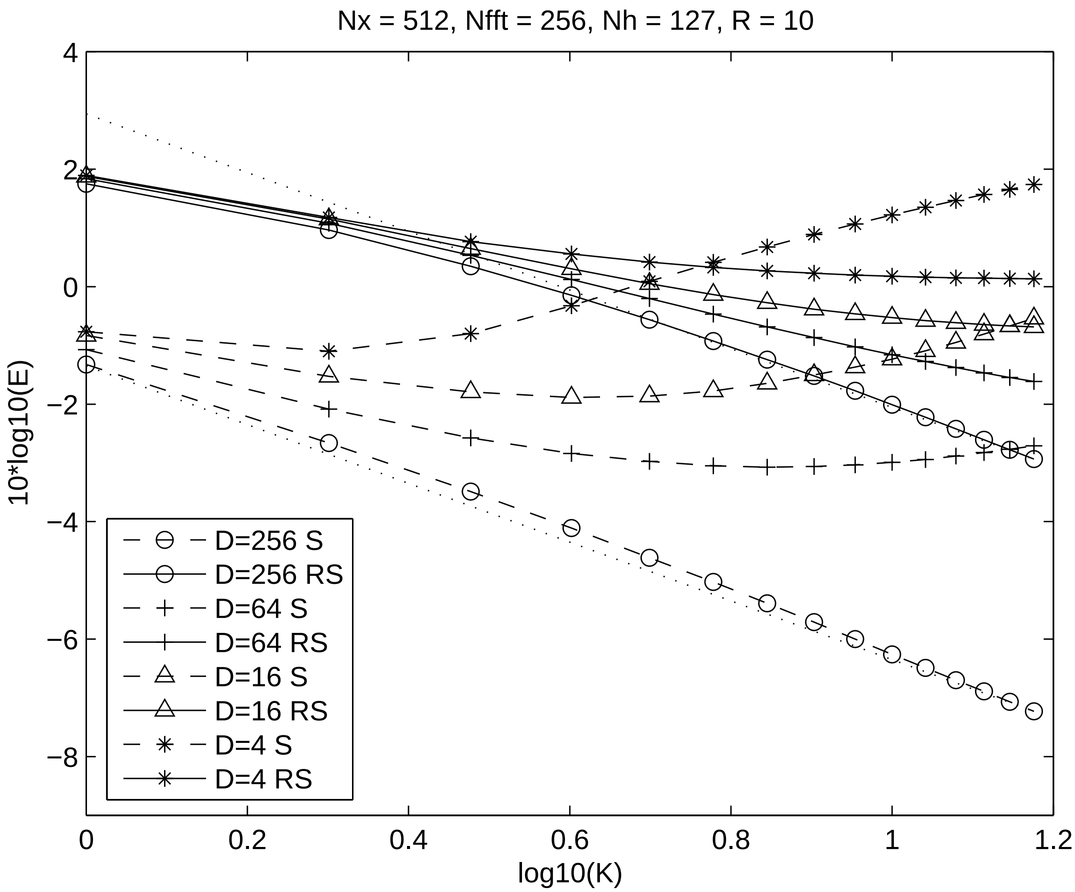

Fig. 3. Error measures in WVS multitaper estimates. The figure plots, as a function of the number of tapers K, the error measure (6) attached to multitaper (reassigned) spectrograms when the model is a wGn limited in time and frequency over a rectangular domain of area D. The simulations have been conducted (with up to $K = 15$ Hermite tapers, each of length 127) on the basis of 10 independent realizations of 512 data points each, with 256 frequency bins over the whole frequency range $[0, 1/2)$. In the pure wGn situation ($D = 256$), asymptotic decays in $K^{-1/2}$ (see text) have been superimposed as dotted lines.

properties of the incomplete Gamma function. A simple calculation ends up with the result:

$$\int_0^\infty |y-\nu|\, p_\nu(y)\, dy = 4\,(\nu/2)^{\nu/2}\, e^{-\nu/2}/\Gamma(\nu/2).$$

and, for large ν's, we can apply Stirling's formula from which we get the asymptotic behavior:

$$\lim_{\nu\to\infty} \frac{1}{2}\sqrt{\frac{\pi}{\nu}} \int_0^\infty |y-\nu|\, p_\nu(y)\, dy = 1.$$

Specifying this result to the case where $\nu = 2K$, we therefore justify that the sum of K spectrograms based on orthogonal tapers has an absolute mean deviation which varies as $K^{1/2}$ for large K's, leading to the $K^{-1/2}$ behavior for the mean $S_{x,K}(t,f)$. Although no proof is available yet, the experiments reported in Fig. 3 evidence a similar behaviour for multitaper reassigned spectrograms $RS_{x,K}(t,f)$, but with a higher level of fluctuations. However, when the area of the domain D is reduced, the situation evolves quite differently for the two estimates: on the one hand, $S_{x,K}$ is smoother that $RS_{x,K}$; on the other hand, $RS_{x,K}$ is essentially confined to the non-zero support of the model, whereas $S_{x,K}$ spreads outside this domain. The criterion (6) can therefore be viewed as a measure of a trade-off between a “bias” term (due mostly to the existence of non-zero values outside the model domain) and a “variance” term (related to non-constant values within the domain). In the case of pure wGn, no bias in the aforementioned sense enters the measure and $E(K)$ decreases when K is increased. When D becomes smaller, the variance reduction is balanced by a bias term increasing with K, since the higher the order of an Hermite taper, the larger its TF support. This analysis applies to the spectrogram but not to the reassigned spectrogram, since squeezing smeared spectra concentrates the error on the fluctuations term which globally decreases always the same way. This is illustrated in Fig. 3, evidencing eventually crossings indicating that multitaper spectrograms may be outperformed by their reassigned counterparts when localized components are to be analyzed.

3) Variations: While preserving the basic idea, different variations upon the method proposed above can be considered. Indeed, the key point in the method is to combine different estimates, and the arithmetic means used in the “classical” multitaper approach as well as in eq.(5) appears as one possibility only amongst others. Comprehensive approaches have been reported [12] on such generalizations in the general context of optimally combining different TF distributions, given some criterion to minimize. What has been shown is that arithmetic averaging naturally results from the requirement that the combined distribution be at a minimum L_2-distance of all distributions to be combined. Changing the chosen distance ends up with different ways of averaging and, e.g., replacing the L_2-distance by a Kullback-Leibler divergence leads to a geometrical mean instead of the arithmetic one.

Given N positive numbers $\{X_n, n = 1, \dots N\}$, their arithmetic and geometric means, defined respectively as

$$A(X_1, \dots X_N) = \frac{1}{N} \sum_{n=1}^{N} X_n ; G(X_1, \dots X_N) = \left(\prod_{n=1}^{N} X_n \right)^{1/N}$$

are clearly such that

$$\log G(X_1, \dots X_N) = A(\log X_1, \dots \log X_N),$$

evidencing that geometrically averaging quantities essentially amounts to arithmetically averaging their logarithms. In the context of multitaper spectrum estimation, this corresponds to applying the machinery to log-spectra, a possibility that has been explored in [15] in a stationary context, and that can be adapted to nonstationary situations. Given the fact that reassigned spectrograms have a high variability, crude geometric means could tend to favor the small (or even zero) values that might appear in one individual estimate only. A possible improvement is to first jacknifing estimates based on arithmetic means prior their geometric averaging. In other words, an alternative to the procedure previously described is

$$\widetilde{RS}_{x,K}(t,f) = \exp \left\{ \frac{1}{K} \sum_{k=1}^{K} \log \widetilde{RS}_x^{(h_k)}(t,f) \right\},$$

with

$$\widetilde{RS}_x^{(h_k)}(t,f) = \frac{1}{K-1} \sum_{m=1, m \neq k}^{K} RS_x^{(h_m)}(t,f).$$

This variation is proposed as an option in the Matlab routines available at [1], but it has been observed to lead to results (not reported here) that were similar, or only incrementally improved. It was however worth mentioning it since the use of log-distributions that underlies the idea of geometrical averaging proved useful in the companion approach described in the following Section.

B. Chirp Enhancement

1) Principle and implementation: The previous Section, aimed at reducing variability in nonstationary spectrum estimation while preserving localization of chirp components, was based on an idea of averaging, i.e., of an *additive* combination of estimates obtained from different tapers. Making use of similar arguments now based on *differences*, a companion perspective can be envisioned for a sake of chirp enhancement, i.e., of displaying at best localized chirp components in the TF plane while getting rid of noise. The rationale for this new approach can be justified by the following twofold argument: (*i*) as far as signal components are concerned and as mentioned above, reassignment increases localization whatever the window: reassigned spectrograms obtained from different tapers are thus expected to be similar, so that their differences take on small values; (*ii*) in noise regions on the contrary, the same set of windows lead to weakly correlated surrogate data whose TF distributions are significantly different, leading in this case to large values for their differences.

What we propose therefore is to consider such differences between the estimates based on the different tapers. As for "sums", "differences" can be considered in a generalized sense and, in accordance with the usual way of displaying spectra in dB units (see figures), *logarithms* of TF estimates will be used. In fact, evaluating differences between log-distributions just amounts to evaluating ratios between the distributions themselves, leading to define as a measure of average difference between estimates based on successive tapers the quantity:

$$RSD_{x,K}(t,f) = \frac{1}{K-1} \sum_{k=1}^{K-1} \frac{RS_x^{(h_{k+1})}(t,f)}{RS_x^{(h_k)}(t,f)}. \tag{7}$$

It is this quantity that is proposed to be thresholded in order to distinguish between the ("coherent") signal components and the ("incoherent") noise regions. More precisely, the idea is (*i*) to define a binary mask function taking on zero values for those TF points where the criterion (7) differs significantly from unity, and (*ii*) to multiply this mask with the average reassigned spectrogram $RS_{x,K}(t,f)$, so as to end up with a masked distribution $RSM_{x,K}(t,f)$ expected to preserve the chirp components while erasing most of the noise. In practice however, we cannot expect to completely get rid of noise by averaging a limited number of difference distributions. Therefore, a threshold is proposed to be first applied to the average reassigned spectrogram $RS_{x,K}(t,f)$ so as to get a pre-denoised version $RSP_{x,K}(t,f)$. Proceeding from this point as above, this pre-processing leads, after masking, to a final, combined estimate $RSF_{x,K}(t,f)$. Interestingly, the binary nature of the two masking functions involved in the two thresholding operations detailed above allows to write

$$RSF_{x,K}(t,f) = \sqrt{RSP_{x,K}(t,f).RSM_{x,K}(t,f)},$$

thus making of the combined estimate $RSF_{x,K}(t,f)$ the *geometric mean* of the two distributions based on the (thresholded) *sums* and *differences* of the individual taper estimates.

A key question in the approach is of course the choice of the thresholds. Since no theory is available yet for the probability distribution function of reassigned spectrograms (and, *a fortiori*, of their ratios), a pragmatic approach is to set the threshold values on the basis of numerical experiments. The nominal values given in [1] and used in the examples reported here have been determined this way (for details about this determination, see [1]), and they proved to guarantee fairly robust estimations in a large variety of experimental setups. The pre-denoising threshold T_p—applied to $RS_{x,K}(t,f)$—was chosen according to a given ("detection") probability P_p for rejecting the null hypothesis of wGn. Similarly, the masking threshold T_d—applied to $RSD_{x,K}(t,f)$—was chosen so as to guarantee some prescribed ("false alarm") probability P_d for the rectified quantity $|RSD_{x,K}(t,f)-1|+1$ in the wGn case. In practice, T_p (resp., T_d) turned out to essentially depend on $(1-P_p)K$ (resp., P_dK), with typical values $T_p \approx 7$ times the empirical mean of $RS_{x,K}(t,f)$ for $P_p \approx 0.95$ and $T_d \approx 1.5$ for $P_d \approx 0.05$ when K ranges in between 4 and 8.

2) Performance evaluation: In order to test the relevance of the proposed methodology, one of the simplest cases to consider is concerned with a pure tone (signal with constant frequency and amplitude) embedded in a background of zero-mean wGn. Fig. 4 illustrates what happens in such a case with SNR $=10$ dB. The two effects of reduced noise and support preservation of the localized signal component thanks to using differences are clearly visible, as well as the impact of pre-denoising for a sake of a further noise reduction.

Fig. 4 gives a qualitative account of the behaviour of the method, which can be supplemented by the quantitative measure

$$C(\mathrm{SNR}) = \frac{1}{\|\hat{\mathbf{W}}_{x,K}\|_1} \int_{-\infty}^{+\infty} \hat{\mathbf{W}}_{x,K}(t,f_0)\,dt, \tag{8}$$

where $\hat{\mathbf{W}}_{x,K}(t,f)$ stands for the WVS estimate ($RS_{x,K}(t,f)$, $RSP_{x,K}(t,f)$, $RSM_{x,K}(t,f)$ or $RSF_{x,K}(t,f)$), and f_0 is the actual frequency of the tone. The criterion (8), which can be interpreted as a contrast measure, is illustrated in Fig. 5 for different SNRs ranging from -40 to $+40$ dB. Comparing the criterion for the 4 different estimates, two regimes are observed: for SNR <0 dB, the method of pre-denoising ($\hat{\mathbf{W}}_{x,K}(t,f) = RSP_{x,K}(t,f)$) is more efficient than the one based on differences only ($\hat{\mathbf{W}}_{x,K}(t,f) = RSM_{x,K}(t,f)$), whereas the latter outperforms the former when SNR >0 dB. This situation is uniformly improved when combining the two estimates according to $\hat{\mathbf{W}}_{x,K}(t,f) = RSF_{x,K}(t,f)$.

C. Examples

Only schematic examples have been considered so far, and this Section is devoted to slightly more elaborated ones. In this respect and for a sake of comparison, we first consider in Fig. 6 the case already discussed in [5] and [6], with both a (nonlinear) chirp component and a (bandpass) time-varying noise. Concerning spectrum estimation, the effectiveness of the approach is clearly supported by this example which evidences the good trade-off achieved between TF localization along the chirp and smoothness within the (time-varying) frequency band of the noise. As far as chirp enhancement is concerned, the final combined estimate identifies in a very localized way the frequency trajectory while erasing most of the noise.

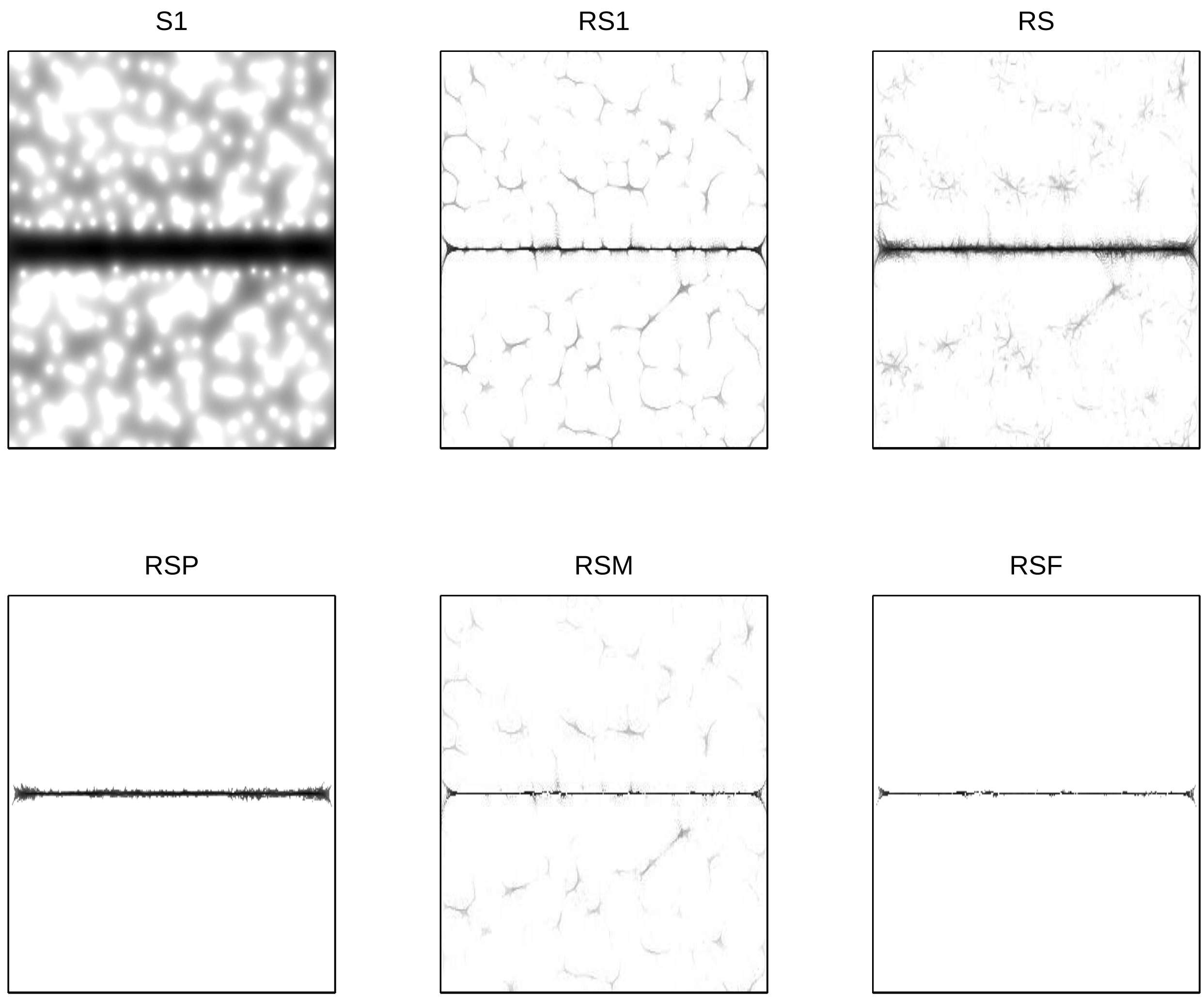

Fig. 4. Comparison of WVS estimates: pure tone in stationary noise. Each diagram represents a WVS estimate in the case of one realization of a pure tone embedded in wGn with SNR = 10 dB. The first row consists of a spectrogram (left), its reassigned version (middle) and the corresponding multitaper reassigned estimate (based on 4 Hermite functions), whereas the bottom row displays the pre-denoised reassigned spectrogram (left), the masked reassigned spectrogram (middle) and the final combined estimate (right). In each diagram, time is horizontal, frequency vertical and the energy is coded logarithmically with a dynamic range of 30 dB.

As a second example, we consider the case where two chirps (a linear one added to the nonlinear one of the previous example) are closely superimposed, with crossings of their instantaneous frequencies in the TF plane, some transient noise being added in a disjoint domain. The overall behaviour of the different estimates is similar to what has been observed in Fig. 6, with the noticeable additional benefit for the combined estimate of improving upon the suppression of the interference terms existing between the two closely spaced chirps.

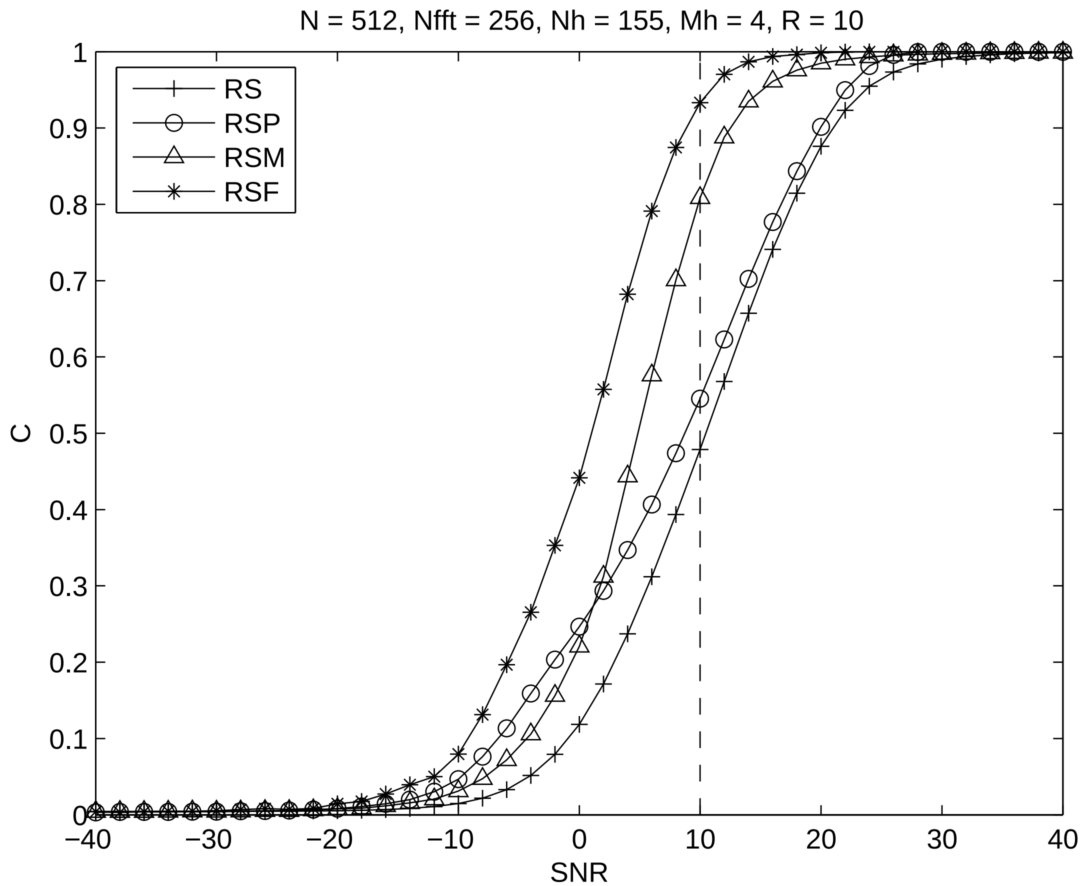

Fig. 5. Contrast in the case of a pure tone embedded in wGn. The figure plots, as a function of SNR, the measure (8) for each WVS estimate (see text). The simulations have been conducted (with 4 Hermite tapers, each of length 155), on the basis of 10 independent realizations of 512 data points each, with 256 frequency bins over the whole frequency range $[0, 1/2)$. The dotted vertical line corresponds to the SNR used in Fig. 4.

IV. Conclusion and perspectives

A novel approach, combining reassignment and multitapering, has been proposed for better estimating time-varying spectra with possibly localized components. The general principles of the approach have been outlined and illustrated, leaving room for many possible extensions. Some possible variations (related to the choice of the averaging type) have been mentioned but, in the specific context of reassignment, other possibilities are offered which are worth to be explored further such as, e.g., combining reassignment vector fields and not only reassigned distributions. More fundamentally, both the performance evaluation in the spectrum estimation context and the tuning of parameters (such as thresholds) for chirp enhancement call for a theoretical analysis of the statistics of reassigned spectrograms.

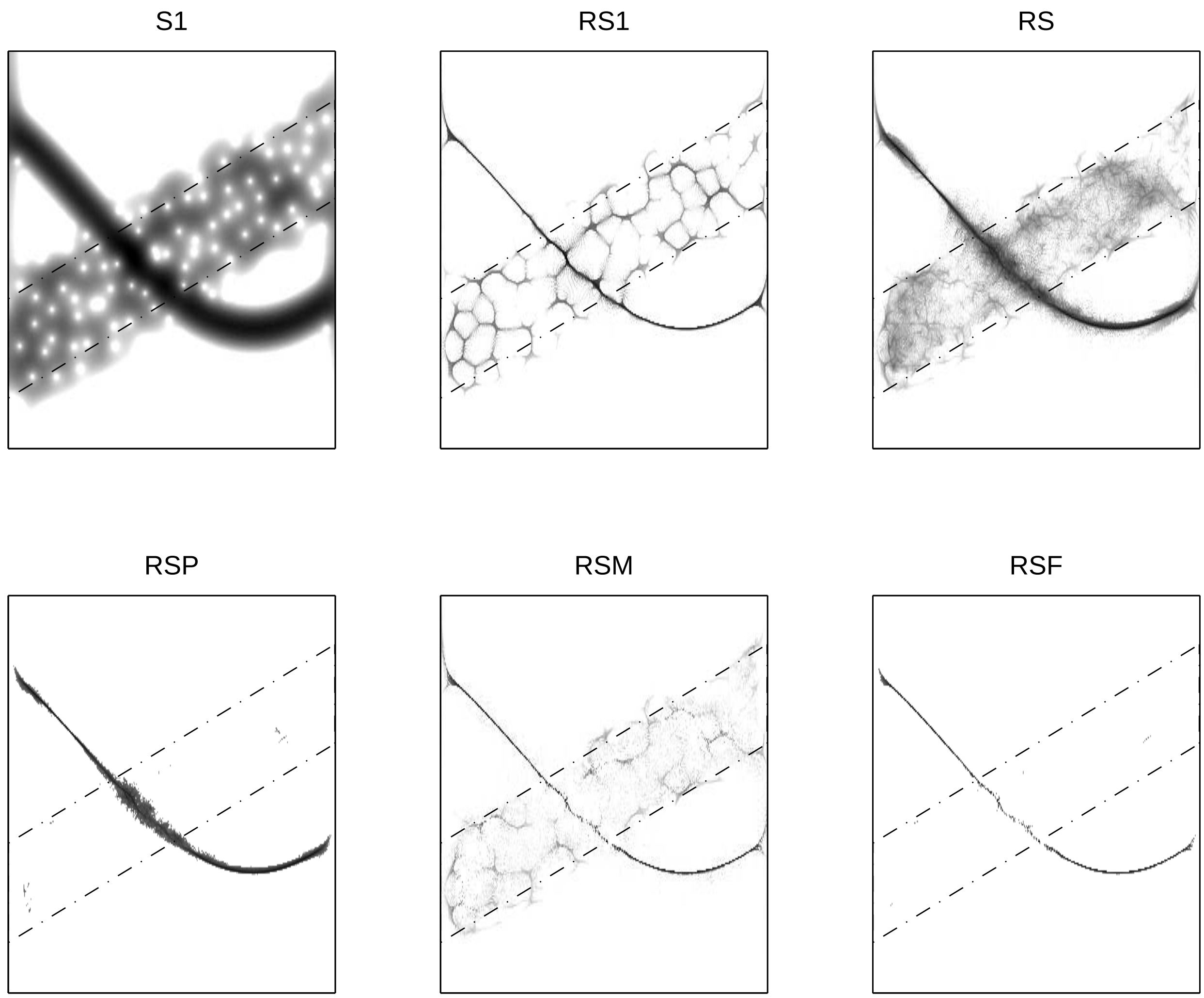

Fig. 6. Comparison of WVS estimates: one chirp in time-varying noise. Each diagram represents a WVS estimate in the case of one realization of a nonlinear chirp signal embedded in time-varying Gaussian noise with SNR = 5 dB. Same display as in Fig. 4, the limits of the noise TF domain being superimposed as dot-dash lines.

REFERENCES

[1] http://perso.ens-lyon.fr/patrick.flandrin/multitfr.html

[2] http://www-stat.stanford.edu/ donoho/Reports/1995/wavelab.pdf

[3] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Trans. Signal Proc.*, vol. SP-43(5), pp. 1068–1089, 1995.

[4] R.G. Baraniuk, P. Flandrin, A.J.E.M. Janssen, O. Michel, "Measuring time-frequency information content using the Rnyi entropies," *IEEE Trans. on Info. Th.*, Vol. 47(4), pp. 1391–1409, 2001.

[5] M. Bayram and R.G. Baraniuk, "Multiple window time-varying spectrum estimation," in *Nonlinear and Nonstationary Signal Processing* (W.J. Fitzgerald et al., eds.), pp. 292–316, Cambridge Univ. Press, 2000.

[6] F. Çakrak and P.J. Loughlin, "Multiple window time-varying spectral analysis," *IEEE Trans. Signal Proc.*, vol. SP-49(2), pp. 448–453, 2001.

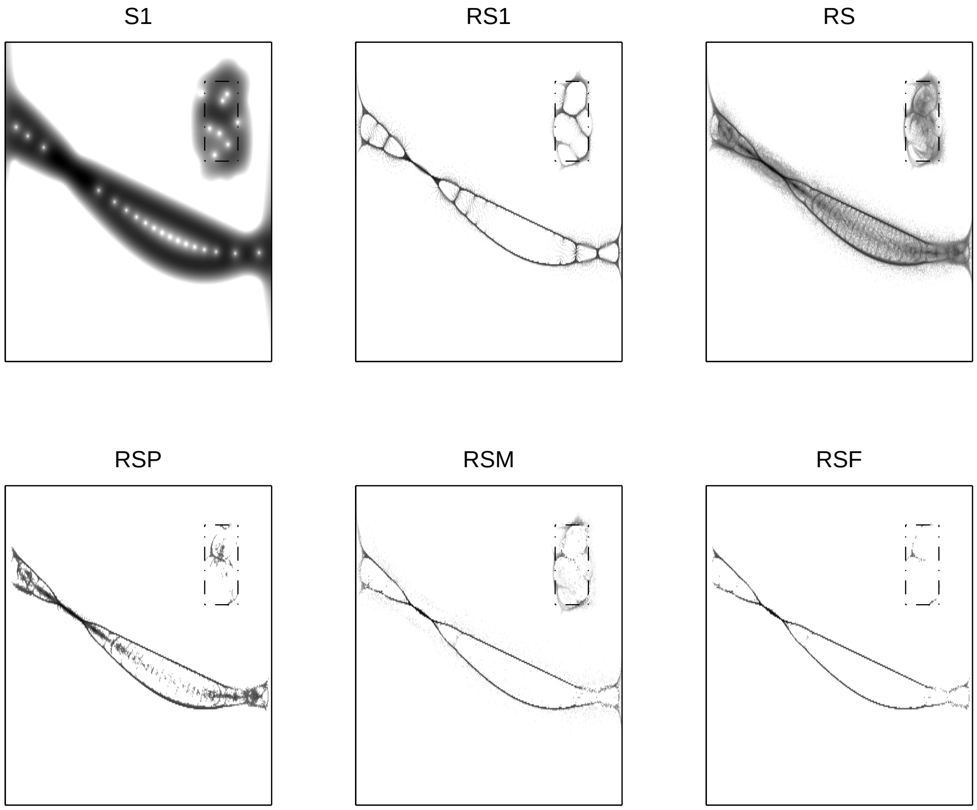

Fig. 7. Comparison of WVS estimates: two chirps and transient noise. Each diagram represents a WVS estimate in the case of one realization of a two component chirp signal superimposed to a transient Gaussian noise with SNR = 10 dB. Same display as in Fig. 6.

[7] I. Daubechies, "Time-frequency localization operators: A geometric phase space approach," *IEEE Trans. Info. Th.*, vol. 34(4), pp. 605-612, 1988.

[8] P. Flandrin, "Maximum signal energy concentration in a time-frequency domain," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Proc. — ICASSP'88*, (New York, NY), pp. 2176–2179, 1988.

[9] P. Flandrin, *Time-Frequency/Time-Scale Analysis*, Academic Press, 1999.

[10] P. Flandrin, F. Auger and E. Chassande-Mottin, "Time-Frequency Reassignment — From Principles to Algorithms," in *Applications in Time-Frequency Signal Processing* (A. Papandreou-Suppappola, ed.), Chap. 5, pp. 179-203, CRC Press, 2003.

[11] G. Frazer and B. Boashash, "Multiple window spectrogram and time-frequency distributions," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Proc. — ICASSP'94*, (Adelaide, AU), Vol. IV, pp. 293–296, 1994.

[12] P.J. Loughlin, J. Pitton and B. Hannaford "Approximating time-frequency density functions via optimal combinations of spectrograms," *IEEE Signal Proc. Lett.*, vol. 1(12), pp. 199–202, 1994.

[13] D.B. Percival and A.T. Walden, *Spectral Analysis for the Physical Sciences*, Cambridge Univ. Press, 1993.

[14] D.J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, pp. 1055–1096, 1982.

[15] D.J. Thomson, "Jacknifing multiple-window spectra," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Proc. — ICASSP'94*, (Adelaide, AU), Vol. 6, pp. VI.73–VI.76, 1994.

[16] D.J. Thomson, “Multitaper analysis of nonstationary and nonlinear time series data,” in *Nonlinear and Nonstationary Signal Processing* (W.J. Fitzgerald et al., eds.), pp. 317–394, Cambridge Univ. Press, 2000.

[17] J. Xiao and P. Flandrin, “Multitaper time-frequency reassignment,” in *Proc. 14th European Signal Proc. Conf. — EUSIPCO-06*, (Florence, I), 2006.