Gaussian Noise Time-Varying Power Spectrum Estimation with Minimal Statistics

Julien Huillery, Fabien Millioz, Nadine Martin

HAL Id: hal-01062165
https://hal.science/hal-01062165

Submitted on 9 Sep 2014

Gaussian Noise Time-Varying Power Spectrum Estimation with Minimal Statistics

Julien Huillery, Fabien Millioz and Nadine Martin, *Member, IEEE*

Abstract—Given the spectrogram of an unknown signal embedded in a Gaussian noise, the Minimal Statistics Maximum Likelihood (MiniSMaL) estimator of the noise time-varying power spectrum is presented and a method to tune one of its parameter is studied. The objective of the minimal statistics approach is to separate the signal of interest from the noise in order to estimate properly the probabilistic properties of the latter. Considering an initial time-frequency estimation neighborhood, the strategy relies on the selection of a minimal subset containing the time-frequency coefficients with the smallest values. Estimators of the noise are then sought from this minimal subset. In this work the case of a spectrogram constructed from a finite-length discrete-time noisy signal is presented. This study extends previous works on minimal statistics on two aspects: first, the Maximum Likelihood estimate of the noise is formulated according to a clear analysis of the probability distribution of the time-frequency coefficients. Second, the choice of an optimal minimal subset is investigated. The signal versus noise discrimination property of the spectral kurtosis is used to select a minimal subset which ensures a fair trade-off between the bias and the variance of the estimator. The resulting performances are discussed and compared with those of other methods through numerical simulations on synthetic signals. The use of the MiniSMaL estimator in a time-frequency detection procedure is finally illustrated on a real-world signal.

Index Terms—Noise estimation, Time-varying power spectrum, Minimal statistics, Maximum likelihood, Spectral kurtosis, Spectrogram.

I. INTRODUCTION

Noise power estimation is one of the fundamental questions in signal processing. As some noise is generally present in every sensed signal, estimation is necessary in order to build up rigorous signal detection or filtering techniques. Among others, an interesting review with audio applications can be found in [1]. We cite here three approaches which have received a particular claim for the estimation of a non-stationary noise power spectral density. All the following strategies are based on the Short Time Fourier Transform (STFT) of the noisy signal. The STFT belongs to the class of the linear time-frequency representations [2] and the present work is also rooted in this class of transforms.

J. Huillery is with the École Centrale de Lyon, Ampère laboratory, CNRS, Écully F-69134, France (e-mail:julien.huillery@ec-lyon.fr).

F. Millioz is with the University Lyon 1, Lyon Neuroscience Research Center, Villeurbanne F-69000, France (e-mail:fabien.millioz@univ-lyon1.fr).

N. Martin is with the University of Grenoble Alpes, CNRS, GIPSA-lab, Grenoble F-38000, France (e-mail:nadine.martin@gipsa-lab.grenoble-inp.fr).

Manuscript received ...

The first approach is based upon a recursive averaging of the STFT coefficients [3], [4], [5]. The averaging is applied along the time axis for each frequency bin. The central question concerning this method is the adaptation of the filter parameters to the time variations of the signal in one hand and to those of the noise in the other hand. The estimation of the Signal to Noise Ratio (SNR) and the probability of presence of the signal are typically used to work out this adaptation. The advantages of this method are twofold. Technically first, the averaging reduces the variance of the estimator and benefits from a strong methodological background. Second, the causal FIR structure of the filter enables a real-time implementation which can be required by some applications.

The second approach that has been proposed for non-stationary noise estimation is to identify the noise-only coefficients within a time-frequency neighborhood centered on the time-frequency bin of interest. In that case, a simple statistic (generally the arithmetic mean) can be applied to estimate the noise power [6], [7], [8]. The approach typically consists in the iteration of a two-step procedure, namely a signal rejection step based on a binary hypothesis test followed by a noise estimation step over the set of non-rejected time-frequency coefficients. As some signal remains present within the set of non-rejected coefficients, this approach estimates the noise from above. The iterations are stopped when a statistical criterion indicates that the set of non-rejected time-frequency coefficients contains only noise coefficients. Several criterion have been proposed: a threshold on the convergence of the likelihood [6]; a threshold on the skewness [7]; a threshold on the kurtosis [8]. The drawback of this approach is to be quite slow as the iterative procedure needs to be apply at every time-frequency bin, and is strongly dependent on the quality of the stopping criterion.

A third approach, initiated by R. Martin [9], [10], [11] and described in the next section, can be named the *minimal statistics* approach. As with the previous approach, the philosophy remains to isolate a subset of noise-only coefficients from which an appropriate estimator can be derived. However, the minimal statistics approach keeps only the smallest time-frequency coefficients which originate most probably from the noise alone. The aim of this paper is to provide further developments concerning the use of the minimal statistics approach for the problem of non-white and non-stationary Gaussian noise estimation. The results presented are twofold. First, the Maximum Likelihood (ML) estimate of the noise is presented and evaluated in the context of minimal statistics. A simple procedure to compute this estimate is also given. To our point of view, this estimator represents a suitable benchmark
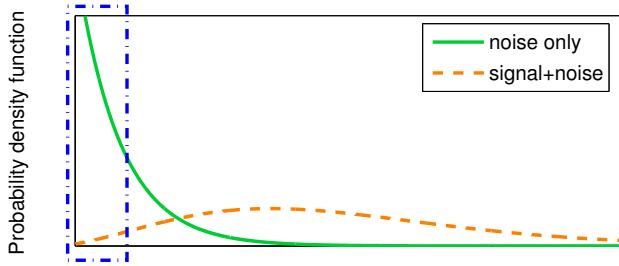
Figure 1. Probability density function of a spectrogram coefficient under the 'noise only' (plain) and 'signal+noise' (dashed) hypotheses. As pointed out within the dot-dashed square, the smallest values are much more likely to represent the noise-only contribution than a signal+noise contribution. This motivates the minimal statistics approach for noise estimation.

against which other kind of estimators could be compared. As a second contribution, we investigate a procedure to fix an important parameter involved in the minimal statistics estimation approach, namely the threshold which isolates the minimal subset. In this paper, the spectral kurtosis is studied as a potential indicator for the determination of this threshold.

The paper is organized as follows. Section II outlines the minimal statistics approach. Considering the STFT and the spectrogram, the section III describes the probability distributions of the time-frequency coefficients as well as their truncated versions. Section IV presents the Minimal Statistics Maximum Likelihood estimator of the noise time-varying power spectrum. The bias and variance of this estimator is studied in regard to the presence of signal within the estimation sample. The determination of a fair minimal subset is the focus of section V. A procedure based on the spectral kurtosis of the STFT coefficients is proposed and discussed. Section VI illustrates the validity of the method on synthetic signals and provides a comparison with some of the above-mentioned estimators. An application of the proposed estimator in a time-frequency detection procedure is also proposed and discussed. Finally, section VII summarizes the main results, draws some conclusions and suggests future studies.

II. MINIMAL STATISTICS ESTIMATION

For the purpose of illustration, the probability density functions (pdf) of two kinds of spectrogram coefficients are plotted in figure 1. The plain curve is the pdf of a central chi-square variable which describes the distribution of a noise-only spectrogram coefficient. The dashed curve is the pdf of a non-central chi-square variable describing the distribution of a signal+noise spectrogram coefficient. This figure emphasizes the fact that if one keeps only the smallest values within a given neighborhood of time-frequency coefficients, the selected subset will contain noise-only coefficients in a much higher proportion than signal+noise coefficients. The minimal statistics approach relies on the possibility to infer the characteristics of the noise from this kind of minimal values subset.

In the initial works of Martin based on a smoothed periodogram [9], the smallest coefficient within a sliding time window of predefined length was sought at every frequency bin. According to the probability distribution of this minimum, a multiplicative gain was applied to recover the noise power. However, in spite of the recursive smoothing, the variance of the minimum and hence of the estimator is high. So as to reduce the variance, it has been proposed to replace the minimum by the qth quantile observed within the same time window [12], or within a time-frequency neighborhood [13]. The choice of the quantile remains however *ad hoc* and the application of a multiplicative gain is unspecified. In [14], the Best Linear Unbiased estimator of the noise power is derived for a minimal subset containing the Z smallest values within a time-frequency neighborhood. But again, the choice of the size Z of the minimal subset is arbitrary.

In the context of a non-white and non-stationary noise, the minimal statistics approach requires three key elements to be considered:

i) How to determine the initial time-frequency estimation neighborhood? To estimate the noise statistics, the initial set of time-frequency coefficients must contain noise-only coefficients. If trivial, this requirement imposes to the estimation neighborhood to be large enough to ensure the presence of noise-only coefficients, even in the case of long-duration and wide-band signals. On the other hand, as required by any statistical inference process, the noise must be identically distributed within this initial neighborhood. In the presence of possibly non-white and non-stationary noise, this second constraint requires to minimize the size of the initial time-frequency neighborhood. As fundamental as can be this question, no procedure has yet been proposed to adapt this initial neighborhood to any given signal. This question is out of the scope of this paper and we consider the initial neighborhood is given. Finally, the probabilistic independence of the time-frequency coefficients within the estimation neighborhood is of great interest as it simplifies the mathematical derivation of the estimator. In time-frequency signal analysis, this independence can be controlled via the parameters involved in the construction of the time-frequency representation and will also be assumed satisfied[1]. The case of correlated samples has been considered in [11].

ii) How to determine the threshold which selects a minimal subset with noise-only coefficients? This threshold (or quantile) should isolate the low valued noise-only coefficients from the higher valued signal+noise coefficients. This threshold controls the trade-off between the bias and the variance of the estimator. With a low threshold, the minimal subset will be small and will only contain noise-only coefficients, assuring unbiasedness but a high variance. On the other hand, a minimal subset obtained with a high threshold will contain a large number of coefficients but will also probably include some signal+noise coefficients. This will result in a low variance but possibly biased estimation of the noise. To this regard, the threshold assuring an unbiased estimation with the smallest variance corresponds to the smallest signal+noise coefficient. Except for the original choice of the minimum value [9], no

---

[1]For the STFT and the spectrogram, a factor of zero-padding less than 2 and a shift of the analysis window greater than its half-length result in reasonably independent coefficients in time and frequency.

procedure has been proposed concerning the determination of this threshold. One of the two contributions of the present work, initiated in [15], is to investigate the possibility to rely on the spectral kurtosis of the time-frequency coefficients to determine a suitable threshold.

iii) Given a minimal subset, how to estimate the noise? Given a subset of minimal values, the mathematical expression of the estimator must be formulated. This formulation must take into account the truncation induced by the thresholding and should be robust to a possibly small number of selected coefficients. Many formulations are possible according to the kind of estimation criterion. In this paper, the performances of the Maximum Likelihood (ML) estimator are investigated.

III. PROBABILITY DISTRIBUTIONS IN THE TIME-FREQUENCY DOMAIN

Based on time domain assumptions, the random behavior of the signal in the time-frequency domain is described when considering the STFT and the spectrogram.

A. Time-domain signal

We consider a discrete-time real-valued observation $x[n]$ composed of the sum of a signal of interest $d[n]$ and a noise $w[n]$, that is

$$x[n] = d[n] + w[n]. \tag{1}$$

The signal and the noise are assumed to be independent. The random perturbation $w[n]$ is assumed to be Gaussian, centered, with the non-stationary autocovariance function $\Gamma_w[n,l] = \mathbb{E}\{w[n]w[n+l]\}$, where $\mathbb{E}\{.\}$ stands for the expectation. Note that for a white and stationary Gaussian noise with variance σ^2, one has $\Gamma_w[n,l] = \sigma^2\delta[l]$, with $\delta[.]$ the discrete Dirac function.

B. Short-Time Fourier Transform and the truncated case

Given an analysis window $h[n]$ centered about $n = 0$, the STFT of the signal x, denoted $F_x[n,k]$, can be written as

$$
\begin{aligned}
F_x[n,k] &= \sum_{l=-\frac{L}{2}}^{\frac{L}{2}} h[l]x[n+l]\exp\left(-i2\pi k\frac{l}{K}\right), \quad (2)\\
&= F_d[n,k] + F_w[n,k],
\end{aligned}
$$

with L the length of the analysis window and K the number of frequency bins. Except for the frequency indexes k close to 0, $K/2$ and K (i.e. for the normalized frequencies k/K close to 0, 0.5 and 1), and under the condition that the analysis window smoothly reaches the value 0 at its boundaries, the noise-only STFT coefficients $F_w[n,k]$ are distributed as a circular complex Gaussian variable [16], [17], [18], [19]. The circularity means that the real and imaginary parts are independent and have the same variance. The probability density function (pdf) of the noise-only STFT complex coefficients $F_w[n,k] = \Re(F_w[n,k]) + i\Im(F_w[n,k])$, with $\Re$ and $\Im$ the real and imaginary parts, is

$$
\begin{aligned}
&p_F(\Re(F_w[n,k]), \Im(F_w[n,k])) = \\
&\frac{1}{\pi\gamma_w[n,k]}\exp\left(-\frac{\Re(F_w[n,k])^2 + \Im(F_w[n,k])^2}{\gamma_w[n,k]}\right). \quad (3)
\end{aligned}
$$

where the scale coefficient $\gamma_w[n,k]$ is the time-varying power spectrum of the noise defined as the second order moment of its STFT, i.e.

$$\gamma_w[n,k] = \mathbb{E}\left\{F_w[n,k]F_w^*[n,k]\right\}. \tag{4}$$

The estimation of $\gamma_w[n,k]$ is the focus of the present work. Note that for a white and stationary Gaussian noise with variance σ^2, one has $\gamma_w[n,k] = \sigma^2$ if the analysis window is taken such that $\sum_n w^2[n] = 1$. Substituting the polar variables $(\rho[n,k],\theta[n,k])$ for Cartesian variables $(\Re(F_w[n,k]),\Im(F_w[n,k]))$ in the equation (3) and integrating over $\theta[n,k]$ as suggested by the circularity of F, the pdf of the modulus of a noise-only STFT coefficient can be written as [15]

$$p_F(\rho[n,k]) = \frac{2\rho[n,k]}{\gamma_w[n,k]}\exp\left(-\frac{\rho[n,k]^2}{\gamma_w[n,k]}\right). \tag{5}$$

The minimal statistics approach relies on the selection of the smallest time-frequency coefficients by the application of a threshold $\sqrt{u}$ on their modulus $\rho[n,k]$ or u on their squared modulus as it will be the case with the spectrogram. As a consequence, the selected time-frequency coefficients are distributed as a truncated circular complex Gaussian variable, denoted T_u, whose pdf is, for $0 \le \rho[n,k] \le \sqrt{u}$,

$$
\begin{aligned}
&p_{T_u}(\rho[n,k]) = \\
&\frac{2}{1-\exp\left(-\frac{u}{\gamma_w[n,k]}\right)}\frac{\rho[n,k]}{\gamma_w[n,k]}\exp\left(-\frac{\rho[n,k]^2}{\gamma_w[n,k]}\right). \quad (6)
\end{aligned}
$$

The kurtosis (fourth-order cumulant) of this distribution will be used in the section V for the determination of a suitable threshold u.

C. Spectrogram and the truncated case

The spectrogram is the squared modulus of the STFT,

$$S_x[n,k] = F_x[n,k]F_x^*[n,k]. \tag{7}$$

When the STFT is a circular complex Gaussian variable, the spectrogram is distributed as a chi-square variable with 2 degrees of freedom [20], [19]. Consequently, the noise-only coefficients have a central chi-square distribution with 2 degrees of freedom (this also corresponds to an exponential distribution), with the scale parameter $\gamma_w[n,k]$ defined in (4). The corresponding pdf is given by

$$p_S(s) = \frac{1}{\gamma_w[n,k]}\exp\left(-\frac{s}{\gamma_w[n,k]}\right). \tag{8}$$

Taking into account the truncation threshold u used in the minimal statistics approach, the pdf of the right-truncated exponential distribution is, for $0 \le s \le u$,

$$
\begin{aligned}
p_{S_u}(s) &= \frac{1}{C(u)}.p_S(s) \quad (9)\\
&= \frac{1}{1-\exp\left(-\frac{u}{\gamma_w[n,k]}\right)}.\frac{1}{\gamma_w[n,k]}\exp\left(-\frac{s}{\gamma_w[n,k]}\right),
\end{aligned}
$$

where $C(.)$ stands for the cumulative distribution function of an exponential variable with scale parameter γ_w [21]. Based on this pdf, the Minimal Statistics ML estimate of the scale parameter γ_w is presented in section IV.

IV. MINIMAL STATISTICS MAXIMUM LIKELIHOOD ESTIMATION OF THE NOISE

This section presents and evaluates the Minimal Statistics Maximum Likelihood (MiniSMaL) estimate of the noise time-varying power spectrum $\gamma_w[n,k]$ defined in (4). We emphasize that this quantity provides a mean time-frequency energetic description of a random signal w. However, other quantities pursuing the same objective can easily be defined considering other quadratic or bilinear time-frequency distributions. The Wigner-Ville spectrum defined in [22] states the case of the Wigner-Ville distribution. The meaningful quantity to be estimated is induced by the initial choice of a time-frequency representation. Because of its intuitive meaning and its practical usefulness, the definition given in (4) has been coined with the terminology *physical spectrum* in [23].

A. Formulation

Consider a sample $\left\{S_{(i)}\right\}_{i=1:Z}$ of Z independent and identically distributed (*i.i.d*) spectrogram coefficients whose pdf is given in (10). According to the truncation threshold u and the scale parameter γ_w, the log-likelihood function is

$$L\left(\left\{S_{(i)}\right\}_{i=1:Z}, u, \gamma_w\right) = $$
$$-Z\left[\ln \gamma_w + \ln\left(1 - \exp\left(-\frac{u}{\gamma_w}\right)\right) + \frac{m_u}{\gamma_w}\right], \quad (10)$$

where m_u denotes the empirical mean of the minimal subset $\left\{S_{(i)}\right\}_{i=1:Z}$

$$m_u = \frac{1}{Z}\sum_{i=1}^{Z} S_{(i)}. \quad (11)$$

Through derivative with respect to γ_w, the ML estimator $\hat{\gamma}_{w\text{ML}}$ of the noise is given as the solution of the equation

$$\hat{\gamma}_{w\text{ML}} - m_u - \frac{u}{\exp\left(\frac{u}{\hat{\gamma}_{w\text{ML}}}\right) - 1} = 0. \quad (12)$$

This equation is solved numerically with the Newton-Raphson method. The equations relative to this method are given in the Appendix A. We emphasize that one problematic situation appears as the sample mean m_u becomes greater than the half-threshold $u/2$. For these cases, the ML equation (12) does not admit any solution. This situation may happen when the number Z of coefficients is very small (order of 10) and their empirical distribution looses the monotonic decreasing nature of the exponential distribution (which is satisfied when $m_u < u/2$). An inspection of the mean m_u is thus required before solving the ML equation. If the condition is not met, one may increase the threshold u. If the threshold is already high, one may also set the estimated value $\hat{\gamma}_w$ to zero. This last possibility preserves the variance of the estimator and ensures a probability of existence equal to 1 [21].

B. Estimator performances

The ML estimator (12) has the nice asymptotic (as Z and u approach infinity) properties of gaussianity, unbiasness, consistency and achieves the Cramer-Rao Lower Bound (CRLB)

[24], [25]. Its asymptotic variance is given by [25]

$$var_\infty\left\{\hat{\gamma}_{w\text{ML}}\right\} = \frac{\gamma_w^2}{Z}\left[1 - \frac{u^2}{\gamma_w^2}\frac{\exp\left(u/\gamma_w\right)}{\left(\exp\left(u/\gamma_w\right) - 1\right)^2}\right]^{-1}. \quad (13)$$

This indicates that the performance of the estimator rapidly degrades as the sample size Z gets smaller, but most important, as the normalized threshold u_n, defined as

$$u_n = \frac{u}{\gamma_w}, \quad (14)$$

gets also smaller. Indeed, the second term in the right hand side of the equation (13) (including the exponent -1) increases from 1 to $+\infty$ as the normalized threshold (14) decreases from $+\infty$ to 0. This term is of order 1, 10 and 10^3 as u_n takes the value 10, 1 and 0.1. However, these large sample optimal properties may not be true in the case of finite and possibly small values of Z and u_n. Typically, we are looking for an estimation which could handle a sample size Z as small as 100 and a normalized threshold u_n between 0.5 and 1. The "MiniSMaL estimator" $\hat{\gamma}_{w\text{MiniSMaL}}$ refers to the ML estimator defined in (12) applied in this kind of practical context. Figure 2-a) displays two distribution curves of the estimates obtained with samples of size $Z \sim 40$ (circled green) and $Z \sim 250$ (crossed red). The estimation samples (*i.e.* the minimal subsets) are constructed by applying a normalized threshold u_n with value 0.5 and 1 on an exponentially distributed random sample of size $N = 100$ and $N = 400$ respectively[2]. In our context of estimating the power of a non-white and non-stationary noise in the time-frequency plane, these two curves are designed to account for a pessimistic situation ($N = 100$, $u_n = 0.5$) as opposed to an optimistic one ($N = 400$, $u_n = 1$). The associated probability distribution are clearly non-Gaussian, especially in the pessimistic case, indicating that the asymptotic properties are far from being verified. For both cases the normalized bias, defined for any given noise estimate $\hat{\gamma}_w$ as

$$B_n = \mathbb{E}\left\{\frac{\hat{\gamma}_w - \gamma_w}{\gamma_w}\right\}, \quad (15)$$

is found to be of the order of 10^{-2} which we consider satisfactory in our context. For a closer inspection, we consider the normalized variance of the MiniSMaL estimator computed as $var\left\{\hat{\gamma}_{w\text{MiniSMaL}}\right\}/\gamma_w^2$ where $var\left\{\hat{\gamma}_{w\text{MiniSMaL}}\right\}$ is the empirical variance of the MiniSMaL estimator observed over a set of experiments with finite Z and u. The result is displayed in figure 2-b) as a function of the sample size Z considering $u_n = 0.5$ (dashed green lines) and $u_n = 1$ (plain blue lines). For comparison, the normalized asymptotic variance defined in (13) is also plotted. Considering the case $u_n = 1$, the normalized variance is found superior to 1 and is 8 times greater than the asymptotic prediction when $Z = 50$. Safer performances are obtained with $Z > 150$ and the asymptotic CRLB is nearly reached when $Z = 250$ (the normalized variance of the MiniSMaL estimator takes the value 0.086 as compared to 0.05). The situation is worst for a smaller threshold $u_n = 0.5$ as the normalized variance is found to

---

[2]The parameter N indicates the size of the initial estimation time-frequency neighborhood from which the minimal subset of the Z smallest coefficient is constructed.
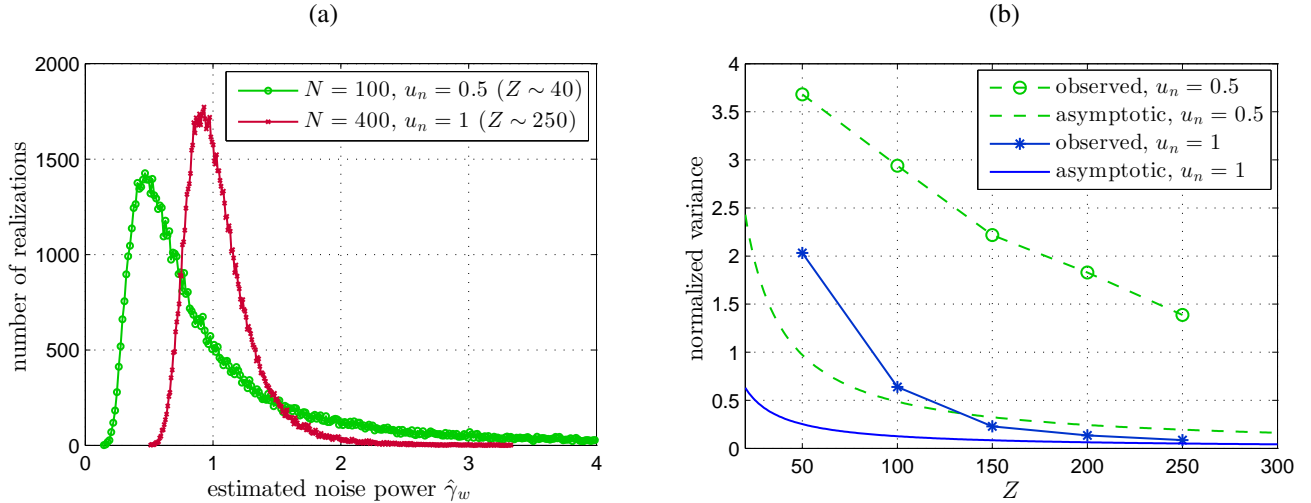
(a)

(b)

Figure 2. Performances of the MiniSMaL estimator. (a) Two distribution curves, one pessimistic (circled green) and one optimistic (crossed red), of the estimates $\hat{\gamma}w_{\text{MiniSMaL}}$ obtained with a normalized threshold u_n (14) applied on an initial set of N samples. Z represents the approximate number of samples kept for the estimation. Histograms were constructed from 10^5 experiments. (b) Empirical normalized variance $var\{\hat{\gamma}w_{\text{MiniSMaL}}\}$ of the MiniSMaL estimator as a function of Z for $u_n = 0.5$ (circled dashed green) and $u_n = 1$ (crossed blue), and comparison with the asymptotic expression (13). Empirical variances are evaluated from 10^4 experiments.

be always greater than one and far from the asymptotic in the range of Z considered in figure 2-b). For $Z = 250$, we observe a normalized variance of 1.4 (the asymptotic prediction gives 0.2). Not reported in the figure, the performances of the MiniSMaL estimator with $u_n = 0.5$ are comparable with the asymptotic predictions for a sample size Z of the order of 10^3.

The second point that strongly degrades the performance of the noise estimator is the presence of signal coefficients among the minimal subset. This situation appears when the truncation threshold is too high. So as to examine the impact of this situation, two models are considered for the signal $d[n]$ (see equation (1)): a deterministic one and a stochastic one. In both cases, the performances of the MiniSMaL estimator in terms of the normalized bias defined in (15) and the normalized standard deviation defined as

$$Std_n = \sqrt{\mathbb{E}\left\{\left(\frac{\hat{\gamma}_w - \gamma_w}{\mathbb{E}\{\hat{\gamma}_w\}}\right)^2\right\}}, \tag{16}$$

are evaluated as a function of two parameters : the signal to noise ratio and the proportion P of signal within the initial estimation neighborhood. Considering an estimation neighborhood containing N time-frequency coefficients among which S coefficients belong to the time-frequency region of support of the signal d, the proportion of signal is defined as $P = S/N$.

Case 1 : deterministic signal $d[n]$. When the signal of interest $d[n]$ in (1) is a deterministic signal, noted $d_1[n]$, we model the STFT coefficients $F_{d_1}[n, k]$ as

$$F_{d_1}[n, k] = A[n, k]e^{i\phi}, \tag{17}$$

with $A[n, k]$ a given (deterministic) amplitude and ϕ a random phase uniformly distributed in the interval $[-\pi, \pi]$. The corresponding spectrogram coefficients are given by

$$S_{d_1}[n, k] = A^2[n, k]. \tag{18}$$

This model aims to describe the general class of amplitude and frequency modulated signals. As those signals are narrowband over a short period of time, it can be assumed that they correspond to rather small values of P. In this case, the signal to noise ratio is defined as

$$\text{SNR}_1[n, k] = \frac{A^2[n, k]}{\gamma_w[n, k]}. \tag{19}$$

Case 2 : stochastic signal $d[n]$. In the case where $d[n]$ in (1) is a stochastic signal, noted $d_2[n]$, assumed to be Gaussian and centered, its spectrogram coefficients are distributed as a central chi-square variable with 2 degrees of freedom,

$$S_{d_2}[n, k] \sim \chi^2(2, \gamma_d[n, k], 0). \tag{20}$$

The scale parameter γ_d of this variable is the physical spectrum (as defined in (4)) of the signal $d_2[n]$. As compared with the first model, this stochastic model is more adapted to represent the class of broad-band signals with a wide time-frequency region of support and consequently large P. The appropriate signal to noise ratio is defined for this case as

$$\text{SNR}_2[n, k] = \frac{\gamma_d[n, k]}{\gamma_w[n, k]}. \tag{21}$$

Note that the SNRs defined in (19) and (21) correspond to local time-frequency SNRs and are meaningful only over the time-frequency region of support of the signal d.

Figures 3 and 4 show the values of the normalized bias and the normalised standard deviation as a function of the normalized threshold u_n, in case 1 (figure 3) and in case 2 (figure 4). The estimations have been carried over 2000 noise runs. For both cases, the sequence $P = 0, 10, 20, 50$ and 80% have been chosen for the signal proportion P (left columns) while the variations over the SNR lie between -3 and 9 dB (right columns). Similar general results are observed in regard to the deterministic or stochastic nature of the signal. In both cases and for every values of SNR and signal proportion
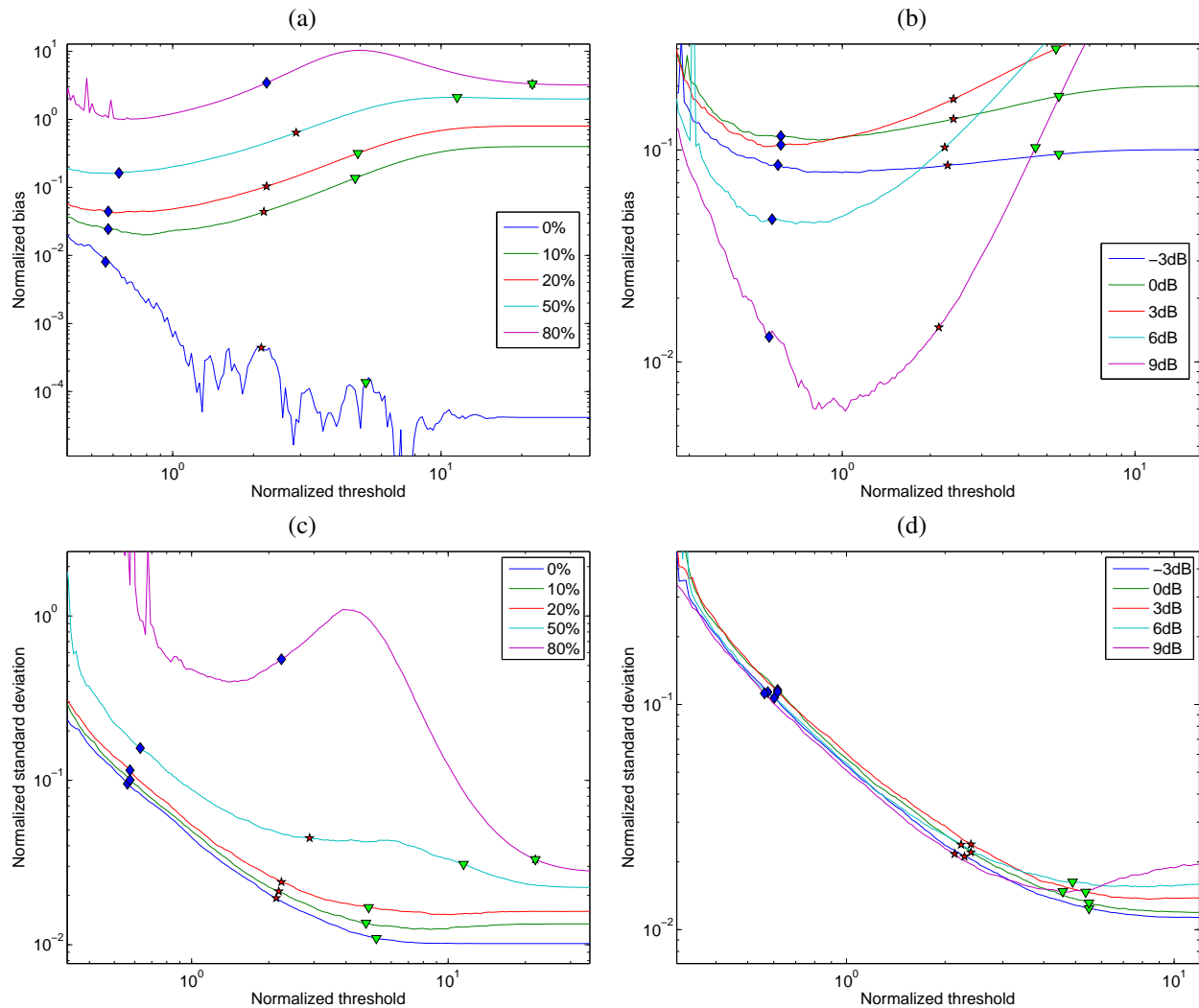
Figure 3. Performances of the MiniSMaL noise estimator in presence of a deterministic signal. Left column: (a) absolute value of the normalized bias and (c) normalized standard deviation for $SNR = 4$dB and for different signal proportion P. Right column: (b) absolute value of the normalized bias and (d) normalized standard deviation for a signal proportion $P = 20\%$ and for different SNR. The results are obtained over 2000 noise runs. For each plot, the markers identify the operating points obtained with a spectral kurtosis κ equal to -0.6 (Blue diamonds), -0.4 (red stars) and -0.1 (green triangles). These latter points are described in section V.

P (figures 3 and 4, plots -a) and -c)), the bias decreases as the threshold increases, admits a minimum, goes up and finally stabilizes as all the coefficients become smaller than the threshold, and hence are selected within the minimal subset. We emphasize that this minimum observed in the curve of the bias appears when the threshold is placed around the value of the smallest signal+noise coefficient. In this situation, the largest possible number of noise-only coefficient is selected within the minimal subset. For higher thresholds, some signal coefficients are selected and the estimator becomes more biased. We note that the value of the minimum bias can be far from zero, indicating a rather biased estimation as soon as the initial signal proportion P becomes too high. As an example, the normalized bias is more than 10% for a signal proportion P higher than 50%. On the other side, a higher SNR leads to a smaller bias, as the distinction between signal and noise is more effective. Interestingly, the minimum bias is always observed for a threshold u close to the noise power γ_w

(*i.e.* $u_n \simeq 1$). Concerning the variance of the ML estimator, we observe a logical reduction with increasing threshold (and hence increasing sample size Z), see figures 3 and 4, plots -b) and -d). Note that the variations over the SNR have a limited impact on the estimator variance.

From these observations, we conclude that the minimal statistics ML estimator is highly sensitive to the presence of signal within the minimal subset. In those unfavorable contexts, a strong bias is observed. In between, it appears that a minimum bias is reached for some specific value of the truncation threshold u. We suggest that the interest of the minimal statistics approach is linked with the ability to pick out this specific threshold. The choice of this threshold hence appears as a crucial key element if one wants to use the minimal statistics approach. In the next section, the spectral kurtosis is studied as a potential tool to work out this specific thresholding task. We also refer to [15] for an initial presentation of this work.
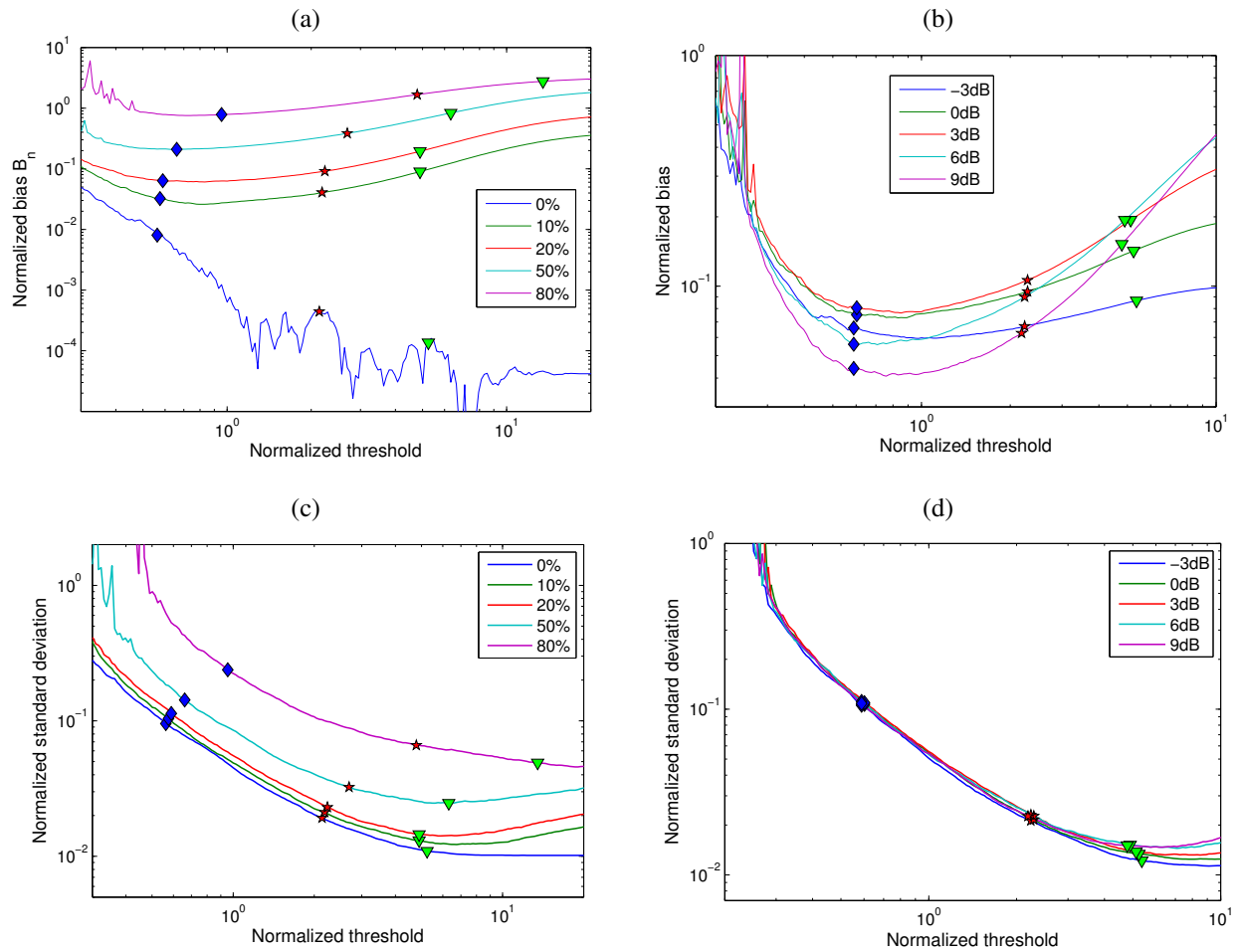
Figure 4. Performances of the MiniSMaL noise estimator in presence of a stochastic signal. Left column: (a) absolute value of the normalized bias and (c) normalized standard deviation for $SNR = 4$dB and for different signal proportion P. Right column: (b) absolute value of the normalized bias and (d) normalized standard deviation for a signal proportion $P = 20\%$ and for different SNR. The results are obtained over 2000 noise runs. For each plot, the markers identify the operating points obtained with a spectral kurtosis κ equal to -0.6 (Blue diamonds), -0.4 (red stars) and -0.1 (green triangles). These latter points are described in section V.

V. MINIMAL SUBSET DETERMINATION WITH SPECTRAL KURTOSIS

In this section, the spectral kurtosis of the STFT coefficients is considered as a potential tool to determine the minimal statistics threshold u. As noted in section III-B, the truncated STFT of a Gaussian noise has a truncated circular complex Gaussian distribution. The spectral kurtosis of this kind of random variable is first presented. Second, in the context of the minimal statistics estimation, the influence of signal on this spectral kurtosis is investigated. Finally, the use of the spectral kurtosis in the minimal statistics approach is discussed.
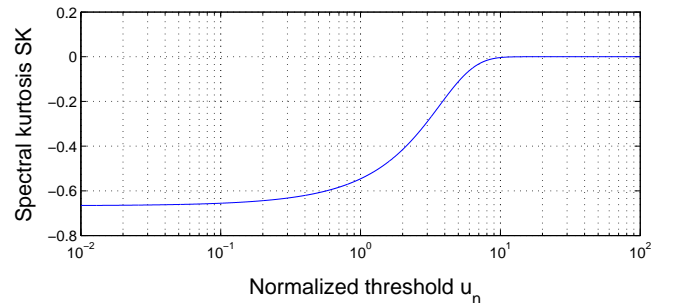

Figure 5. Evolution of the spectral kurtosis κ of a $\sqrt{u}$-truncated circular complex Gaussian variable with covariance matrix $\gamma_w/2 \times I_2$, with respect to the normalized threshold $u_n = u/\gamma_w$.

A. Spectral kurtosis: definition

The spectral kurtosis is the definition of kurtosis applied to complex random variables [26]. The spectral kurtosis of a $\sqrt{u}$-truncated[3] circular complex Gaussian variable T_u, whose

pdf is given in (6), is given by [15]

$$
\begin{aligned}
\mathrm{SK}\left\{T_u\right\}(u_n) &= \frac{\mathbb{E}\left\{T_u^2 T_u^{*2}\right\}}{\mathbb{E}\left\{T_u T_u^*\right\}^2} - 2 \qquad (22) \\
&= \frac{\left[\exp\left(\frac{u}{\gamma_w}\right) - 1\right]\left(2\frac{u}{\gamma_w} - \frac{u^2}{\gamma_w^2}\right) - 2\frac{u^2}{\gamma_w^2}}{\left[\exp\left(\frac{u}{\gamma_w}\right) - 1 - \frac{u}{\gamma_w}\right]^2}.
\end{aligned}
$$

---

[3]A threshold u applied on the spectrogram corresponds to a threshold $\sqrt{u}$ applied on the modulus of the STFT.

u_n		SK	SK		u_n
0.1	$\rightarrow$	-0.655	-0.6	$\rightarrow$	0.57
1	$\rightarrow$	-0.546	-0.4	$\rightarrow$	2.12
10	$\rightarrow$	-0.0036	-0.1	$\rightarrow$	5.19

Table I

TABLE OF CORRESPONDENCE BETWEEN SOME SPECIFIC VALUES OF THE SPECTRAL KURTOSIS OF A TRUNCATED CIRCULAR GAUSSIAN RANDOM VARIABLE AND THE ASSOCIATED NORMALIZED THRESHOLDS u_n.

We highlight that the spectral kurtosis is not directly dependent on γ_w but rather on the normalized threshold u_n defined in (14). The monotonic increasing evolution of the spectral kurtosis as a function of u_n is depicted in figure 5. When the threshold increases, the truncated complex Gaussian variable tends to be a full complex Gaussian variable and its spectral kurtosis becomes 0. In complement to the figure 5, table I gives some specific correspondences between the spectral kurtosis and the normalized threshold.

The monotonic behavior of the spectral kurtosis as a function of the normalized threshold u_n is of great interest in regard to the choice of a (non-normalized) minimal statistics threshold u. Indeed, for any threshold $\sqrt{u}$ applied on an initial set of noise-only STFT coefficients, the spectral kurtosis of the remaining subset provides, with a one to one relation, the ratio between the threshold u and the noise power γ_w. The determination of the minimal statistics threshold can thus be conducted with the following procedure: for any *a priori* choice of the normalized threshold, a unique spectral kurtosis exists and one can change the threshold applied to the time-frequency coefficients until the desired spectral kurtosis is observed. It has been observed in section IV-B that the bias of the MiniSMaL noise estimator is minimum when the normalized threshold u_n is close to 1. If no signal is present within the minimal subset, this threshold corresponds to a spectral kurtosis equal to -0.546 (see table I). In an ideal context, this value of kurtosis could thus be used for the determination of a suitable threshold. However, the presence of signal within the minimal subset modifies its kurtosis and has to be taken into account as it will be observed in the next paragraph.

Note: as proposed in [15], one could think about estimating the noise directly with the kurtosis as it provides the ratio between the known applied threshold u and the noise γ_w. However, the variance of the resulting estimator has been found to be higher than the variance of the ML estimator presented in section IV.

B. Influence of signal on the spectral kurtosis

The influence of the signal on the spectral kurtosis of the minimal subset is depicted by the figures 6, 7 and 8. Figure 6 shows the evolution, with respect to the normalized threshold u_n, of the spectral kurtosis of the truncated STFT for a Gaussian noise alone (plain thin line), a deterministic signal embedded in that noise (dashed line) and a random signal embedded in that noise (dot-dashed line). The signals are those defined in section IV-B. The SNR is fixed at 3dB and the proportion P of signal is equal to 10%. As expected, the presence of signal makes the spectral kurtosis diverge from
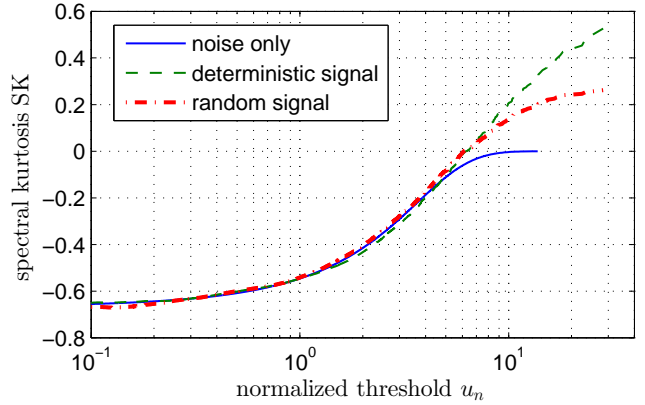

Figure 6. Spectral kurtosis SK of the truncated STFT for a Gaussian noise alone (plain blue line), a deterministic signal (dashed green line) and a random signal (dot-dashed red line) with respect to the normalized threshold u_n. The SNR is equal to 3 dB and the proportion of signal contained in the initial set of time-frequency coefficients is $P = 10\%$.
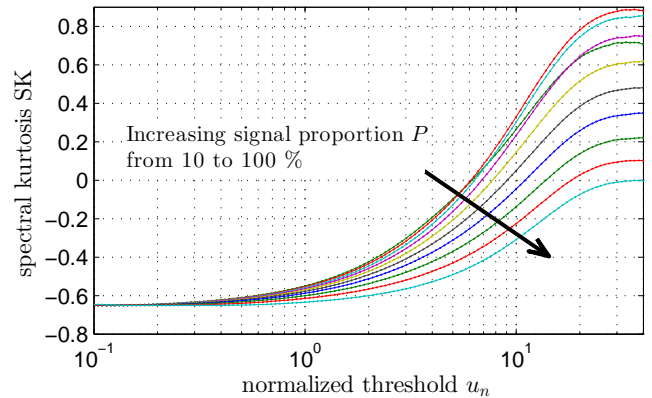

Figure 7. Spectral kurtosis SK of the truncated STFT for a random noisy signal with respect to the normalized threshold u_n. The SNR is equal to 4 dB and the proportion of signal contained in the initial set of time-frequency coefficients varies from $P = 10\%$ (upper line) to $P = 100\%$ with a step of 10%.

the noise-only case as the threshold gets higher. However, as the normalized threshold u_n remains less than 4, the spectral kurtosis of the three signals are very close. Note that the spectral kurtosis evaluation becomes unstable for very low thresholds, as only a small number of coefficients are selected.

Figure 7 shows the evolution, with respect to the normalized threshold u_n, of the spectral kurtosis of the truncated STFT of the noisy random signal as the proportion P of signal varies from 0 to 100%. The SNR is fixed at 4dB. When the proportion of signal increases, the spectral kurtosis of a signal+noise subset tends to diminish and becomes smaller than for a noise-only estimation subset. This divergence will affect the selected threshold.

To further study the influence of signal on the resulting threshold, we consider the random signal and the deterministic signal of section IV-B. For both signals, the influence of the SNR is investigated for a proportion of signal $P = 50\%$, and
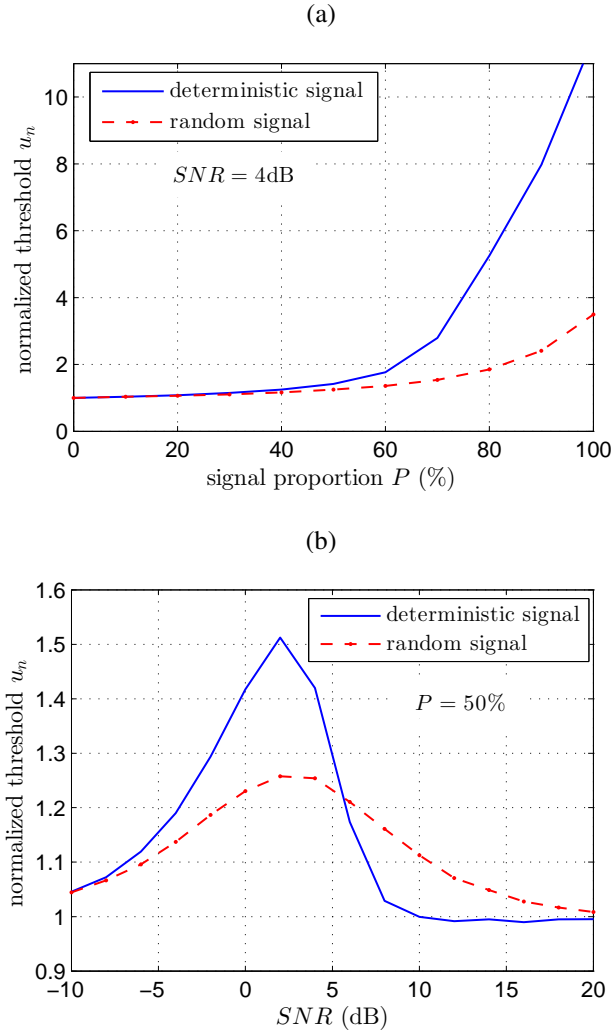
(a)

(b)

Figure 8. Influence of the signal parameters on the normalized threshold u_n calculated at the spectral kurtosis $\kappa = -0.546$. (a) u_n as a function of the proportion of signal P for $SNR = 4$dB and (b) u_n as a function of the SNR for $P = 50\%$. The curves display the mean result over 2000 noise runs.

the influence of the proportion of signal is investigated for $SNR = 4$dB. For each parameter, the threshold corresponding to SK $= -0.546$, that is the spectral kurtosis corresponding to a normalized threshold $u_n = 1$ in the noise-only case, is computed and averaged over 2000 realizations of noise. Figure 8 displays the results. The higher the proportion of signal, the higher the threshold. In other words, when the signal is sparse in the time-frequency domain, it has a small influence on the threshold. Moreover, the threshold keeps close to one for proportion of signal up to $P = 50\%$. On the other hand, the SNR does not have a monotonic influence on the threshold. The highest divergence appears around $SNR = 4$dB. Given a quite high proportion of signal $P = 50\%$, the obtained threshold in the worst SNR case does not exceed $u_n = 1.5$.

C. Discussion

In summary, the spectral kurtosis (23) is a monotonically increasing function of the truncation threshold u. As a consequence one can easily determine a suitable threshold to

use in the MiniSMaL estimator of section IV by imposing a value κ to the spectral kurtosis of the minimal subset. The determination of the truncation threshold only depends on the spectral kurtosis of the minimal subset, independently from the noise variance or the signal. The choice of a small κ leads to a low threshold, while a high κ leads to a high threshold, but in both cases bias and standard deviation keep reasonably low. The usual trade-off between bias and variance is made by choosing κ between -0.6 and -0.1.

On figures 3 and 4 are plotted the thresholds corresponding to a spectral kurtosis of -0.6 (blue diamonds), -0.4 (red stars) and -0.1 (green triangles). One can observe that a threshold corresponding to spectral kurtosis between -0.6 and -0.4 leads to good values of bias and variance. Note that the presence of signal tends to slightly increase the desired threshold, especially if the signal is not sparse.

VI. APPLICATION AND COMPARISON

This section illustrates the results of the MiniSMaL noise estimator with a minimal subset determined by the method proposed in the previous section. First a synthetic signal embedded in a white and stationary noise is analysed. All the parameters being under control, the estimation performance is calculated and compared with the Median Absolute Deviation (MAD) method [27]. Second, estimation of the noise in a non-white and non-stationary context is examined. Considering a synthetic signal and noise, the MiniSMaL estimator is compared to three other methods which belong respectively to the iterative, recursive and minimal statistics classes of estimators mentioned in the introduction. Third, considering a real-world acoustic dolphin signal, the noise estimate obtained with the MiniSMaL estimator is injected into a time-frequency detection procedure as possibly required for time-frequency filtering or classification.

A. Synthetic signals : white and stationary noise

We consider a 8000-sample synthetic signal made of three non-stationary signals: a linear chirp with an amplitude of 0.5, a sinusoidal frequency modulation with an amplitude of 1, and a Gaussian signal with a variance σ_s^2 equal to 32, limited in both time and frequency. Figure 9-a) shows the spectrogram of this signal embedded in an additive white and stationary Gaussian noise of variance $\sigma^2 = 0.5$. The spectrogram is constructed with a 127-sample Hanning window, an overlap of 62 samples and 512 frequency bins.

We performed noise estimations based on 1000 realizations of the additive noise with variance ranging from 0.125 to 256. The minimal subset truncation threshold was determined with a spectral kurtosis fixed to $\kappa = -0.4$. For comparison, we also considered the Median Absolute Deviation (MAD) estimator of the noise variance defined over a set $\{X_i\}$ of points as [27]:

$$\text{MaD} = \text{median}_i \left(|X_i - \text{median}_j(X_j)| \right). \quad (23)$$

The MAD is a consistent estimator of the standard deviation of the random variable X when multiplied by a constant which

(a)   (b)
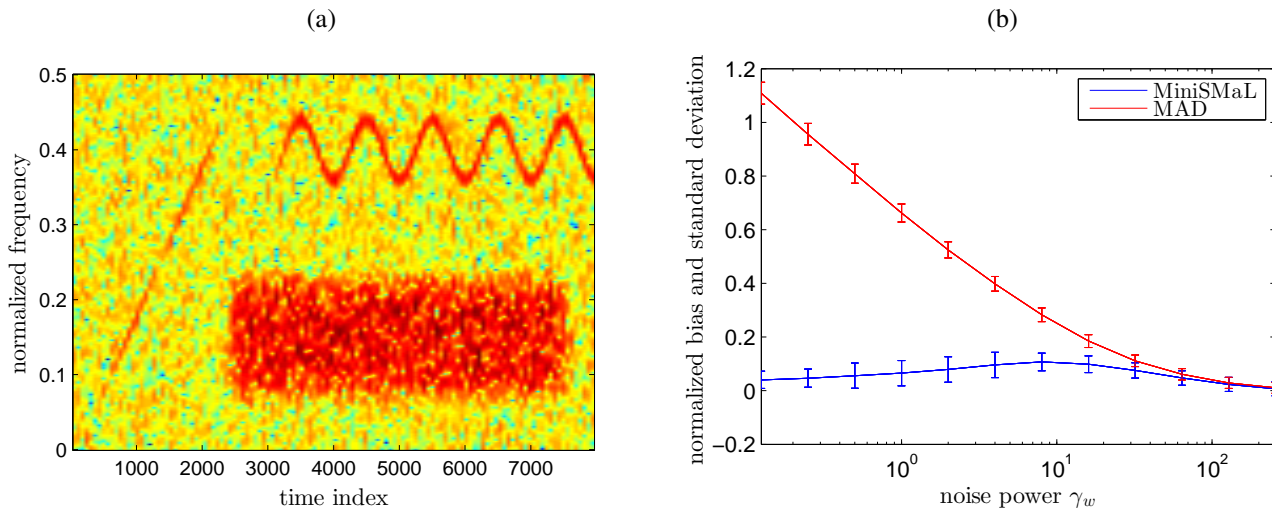

Figure 9. (a) Spectrogram of the synthetic signal given an additive white and stationary noise of variance $\sigma^2 = 0.5$. (b) Normalized bias of the MiniSMaL (plain blue line) and MAD (dashed red line) estimators as a function of the noise variance σ^2. Vertical bars indicate the normalized standard deviations. Performances are calculated over 1000 realizations of the noise.

depends on the distribution of X. For a Gaussian distribution, the estimate of the standard deviation is obtained as

$$\hat{\sigma} \approx 1.4826 \text{ MaD}. \tag{24}$$

Following the properties of the median, this noise estimator is known to be robust to outliers. This property is of interest in the case of a signal embedded in noise, as the signal may be considered as outliers of the noise. As the STFT coefficients are complex-valued, the MAD noise power spectrum estimate $\hat{\gamma}_{w_{\text{MAD}}}$ is computed as

$$\hat{\gamma}_{w_{\text{MAD}}} = \hat{\sigma}_{r_{\text{MAD}}}^2 + \hat{\sigma}_{i_{\text{MAD}}}^2, \tag{25}$$

where $\hat{\sigma}_{r_{\text{MAD}}}$ and $\hat{\sigma}_{i_{\text{MAD}}}$ are the MAD estimates of the standard deviations of the real and imaginary parts of the noise STFT.

Figure 9-b) displays the normalized bias B_n (15) of the MiniSMaL and MAD estimators. The error bars correspond to the normalized standard deviations Std_n defined in (16). For this signal, the MiniSMaL estimator is much more efficient than the MAD estimator. For instance considering the high SNRs (small noise variance), the MAD estimator is highly biased while the bias of the MiniSMaL remains stable. As the noise variance increases, the signal becomes negligible compared to the noise and both estimators tend to be identically unbiased. Note however that the MiniSMaL estimator has a higher standard deviation, as the minimal subset strategy decreases the size of the estimation sample. For the range of variances considered, the MiniSMaL estimator leads to a normalized bias always below 11%.

B. Synthetic signals : non-white and non-stationary noise

We consider now the case of a non-white and non-stationary additive noise whose spectrogram is displayed in figure 10-a). Note that for the convenience of the 3-D visualization, the time and frequency axes are in an unusual reversed direction. This noise is a 25000-sample signal obtained by applying

a time-varying multiplicative gain to the output of a linear time-invariant IIR filter excited by a white and stationary Gaussian noise with variance equal to 1. The signal of interest embedded in this noise is made of frequency modulated signals arranged so as to form the "chi^2" logo in the time-frequency plane, see figure 10-b). Spectrograms are constructed with a Hanning window of length 256, an overlap of 128 samples and 512 frequency bins which yield to a 194×256 time-frequency image. The sampling frequency is 5000 Hz.

The MiniSMaL estimator requires an initial estimation neighborhood within which the minimal subset will be selected. As mentioned in section II, this time-frequency neighborhood should be as large as possible while containing identically distributed noise coefficients. In the previous section the noise was white and stationary: the whole spectrogram was considered as the time-frequency neighborhood. For a non-white and non-stationary noise, the initial estimation neighborhood should be of limited size in both time and frequency. The estimation neighborhood used in the present application spans over 15 frequency bins and 15 time bins, which represents a total of $N = 225$ time-frequency coefficients. The minimal subset was selected according to a spectral kurtosis fixed at $\kappa = -0.4$. According to those parameters, the minimal subset has an average size of $Z \sim 130$ coefficients.

Figures 11 and 12 display the results obtained with the MiniSMaL estimator with comparison to three other estimators :

- The *RecAv* estimator was proposed in [5]. It belongs to the class of algorithms which use a recursive averaging of the STFT coefficients.
- The *IterDesc* estimator was proposed in [6]. It belongs to the second class of iterative estimation/rejection algorithms mentioned in the introduction.
- The third estimator used for comparison is the initial Minimal Statistics estimator proposed in [9]. We coin the name *MiniStatRecAv* as it uses the statistic of the smallest
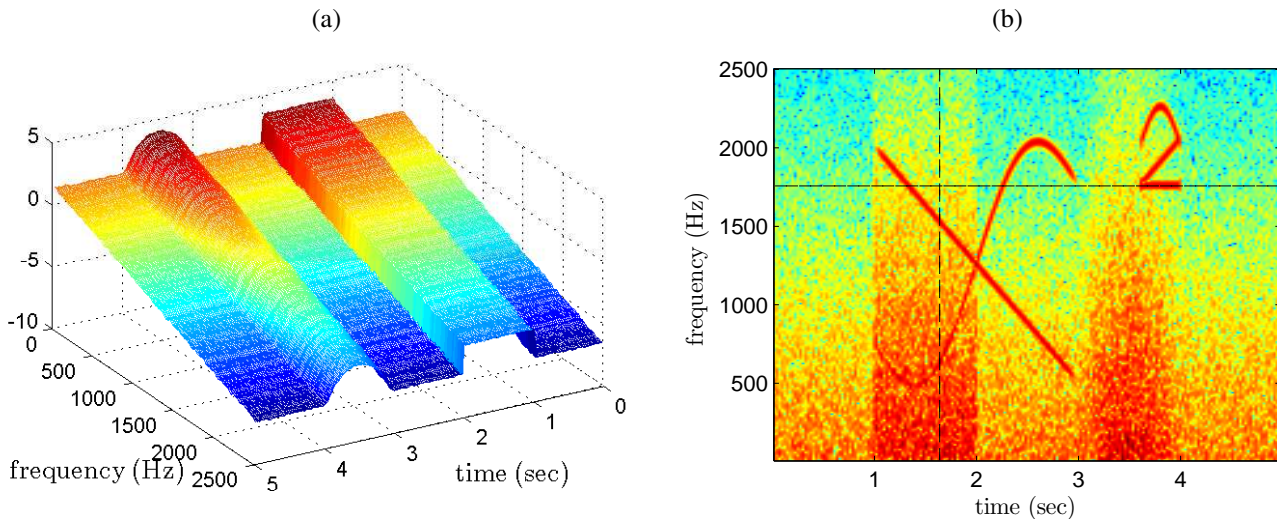
Figure 10. (a) Spectrogram of the synthetic non-white and non-stationary additive noise considered in section VI-B. Note that for a convenience of visualization, the time and frequency axes are reversed. (b) Spectrogram of the noisy signal.

coefficient observed in a recursively averaged STFT. For comparison with the proposed estimator, we use the same 15-bin time window for the minimum searching.

Figure 11 displays the estimation results in the entire time-frequency plane. The time and frequency axes direction are the same as in figure 10-a). Taking benefit of its estimation neighborhood which spreads in both time and frequency, the MiniSMaL estimate is not impacted by the presence of signal. Furthermore, it is observed that the MiniSMaL estimator provides a fair trade-off between the time-frequency smoothness of the estimate and the possibility to track the abrupt changes of the noise power spectrum. For a closer inspection, figure 12-a) displays the noise estimates as a function of time at 1750 Hz (see the horizontal black line in figure 10-b)). The effect of the spread in time of the estimation neighborhood is clearly visible around the abrupt changes of the noise where the MiniSMaL and the *MiniStatRecAv* estimators have the same performance as they use the same time window for the minimum searching. Note however that the proposed estimator is less sensitive to the presence of signal. Figure 12-b) displays the noise estimates along the frequency axis at time 1.6 sec (see the dashed vertical black line in figure 10-b)). All the four estimators provide qualitatively satisfactory estimates of the exponentially decreasing noise power spectrum. The *IterDesc* estimator has the smallest variance but is very sensitive to the presence of signal. On the contrary, the *MiniStatRecAv* estimator is the more robust to the presence of signal but has a higher variance. The MiniSMaL strategy lies in between.

We emphasize that the above comparison has been made as fair as possible. Hours of experiment have been spent to adjust the parameters of the four estimators so as to obtain their best results. Table II summarizes those parameters. This points out that the MiniSMaL estimator is clearly the easiest to use. It relies on only two parameters: the initial estimation neighborhood and the spectral kurtosis κ. We suggest the choice $\kappa = -0.4$ which provides good results for a wide class of signals.
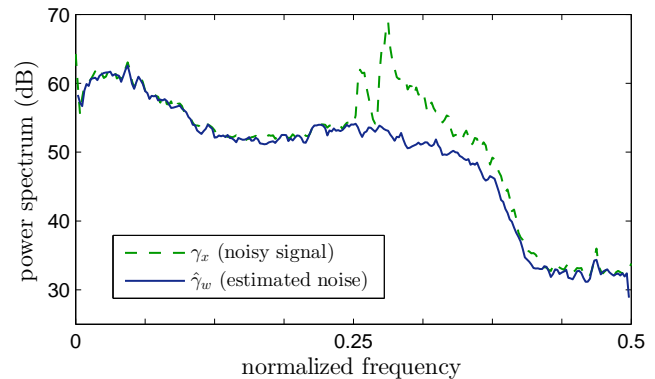

Figure 14. Power spectrum γ_x of the noisy signal (dashed green) and estimate of the power spectrum $\hat{\gamma}_w$ of the noise alone (plain blue). The noise is estimated with the MiniSMaL algorithm. The initial time-frequency estimation neighborhood is 9 frequency bin-wide and contains all time bins (the noise is stationary). The minimal subset is selected according to a spectral kurtosis $\kappa = -0.4$.

C. Time-frequency detection of real-world signals

The estimation of the noise time-varying power spectrum allows one to pursue a detection task in the time-frequency domain [28]. As an example, a Neyman-Pearson detection strategy [29] can be used to determine a time and frequency-dependent detection threshold $S^{th}[n, k]$ satisfying

$$\mathbb{P}\left\{S_w[n, k] > S^{th}[n, k]\right\} = P_{fa}, \qquad (26)$$

where $S_w[n, k]$ is a noise-only spectrogram coefficient and P_{fa} a chosen Probability of False Alarm (PFA). Assuming the signal of interest is embedded in a Gaussian noise, the probability distribution of $S_w[n, k]$ is that of a χ^2 variable with two degrees of freedom (see section III-C) and the time-frequency detection threshold $S^{th}[n, k]$ which satisfies

(a) RecAv

(b) IterDesc
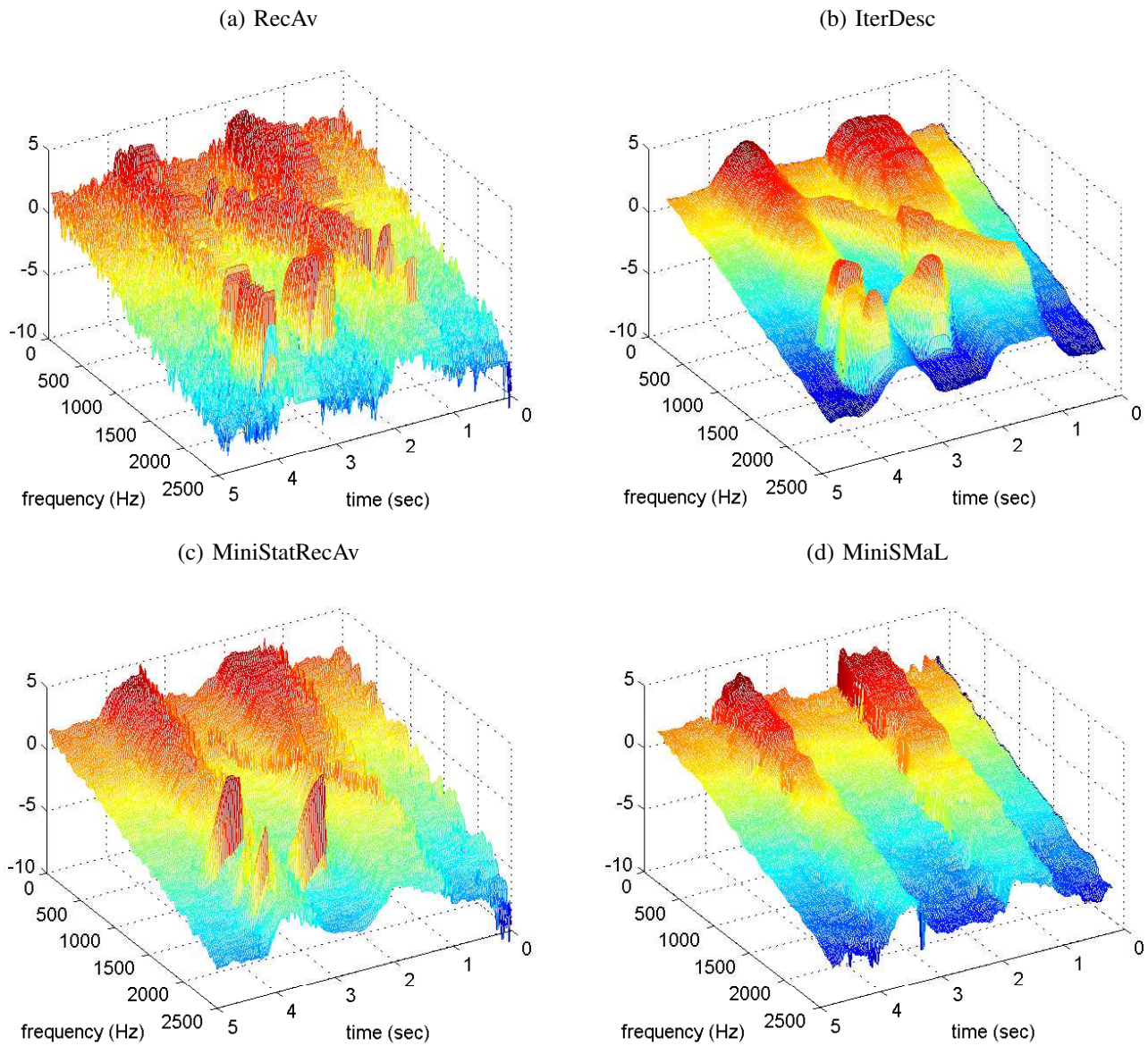

(c) MiniStatRecAv

(d) MiniSMaL

Figure 11. Non-white and non-stationary noise estimates obtained with the four estimators considered in the section VI-B. Results should be compared with the figure 10-a). For convenience of visualization, the time and frequency axes are also reversed.

equation (26) is

$$S^{th}[n,k] = \gamma_w[n,k] \ln(P_{fa}^{-1}), \qquad (27)$$

with $\ln(.)$ the natural logarithm. In the case the noise is unknown, we use the MiniSMaL strategy so as to estimate its time-varying power spectrum $\gamma_w[n,k]$.

Figure 13 illustrates the result of such a detection procedure applied to an underwater acoustic signal made of dolphin whistles. Figure 13-a) displays the spectrogram of the 225000-sample signal computed with a 1023-sample Hanning window, an overlap of 511 samples and 2048 frequency bins. The spectrogram has 1024 (positive) frequency bins and 438 time bins. The MiniSMaL estimator is used to estimate the power spectrum of the non-white but stationary noise. The initial estimation neighborhood spans over 9 frequency bins and all the time bins, which represents a total of $N = 3942$ time-

frequency coefficients. The minimal subset threshold u was computed according to a spectral kurtosis fixed at $\kappa = -0.4$. In this configuration, the minimal subset has an average size of $Z \sim 2200$ coefficients. Figure 14 shows the estimated noise power spectrum $\hat{\gamma}_w[k]$ superimposed on the noisy signal power spectrum $\gamma_x[k]$ estimated with the Welch method [30]. Incorporating this estimate into the detection threshold defined in (27) with a probability of false alarm $P_{fa} = 10^{-5}$ yields to the detection result shown in figure 13-b). The frequency modulations forming the dolphin whistles are correctly detected.

Looking at equation (27), the detection threshold S^{th} is a linear function of the noise power γ_w. When the estimated version $\hat{\gamma}_w$ is used to calculate the detection threshold, an error in the estimation directly impacts the result of the detection procedure. In the following we analyze the relation between an estimation error and the PFA of the detection test. Lets
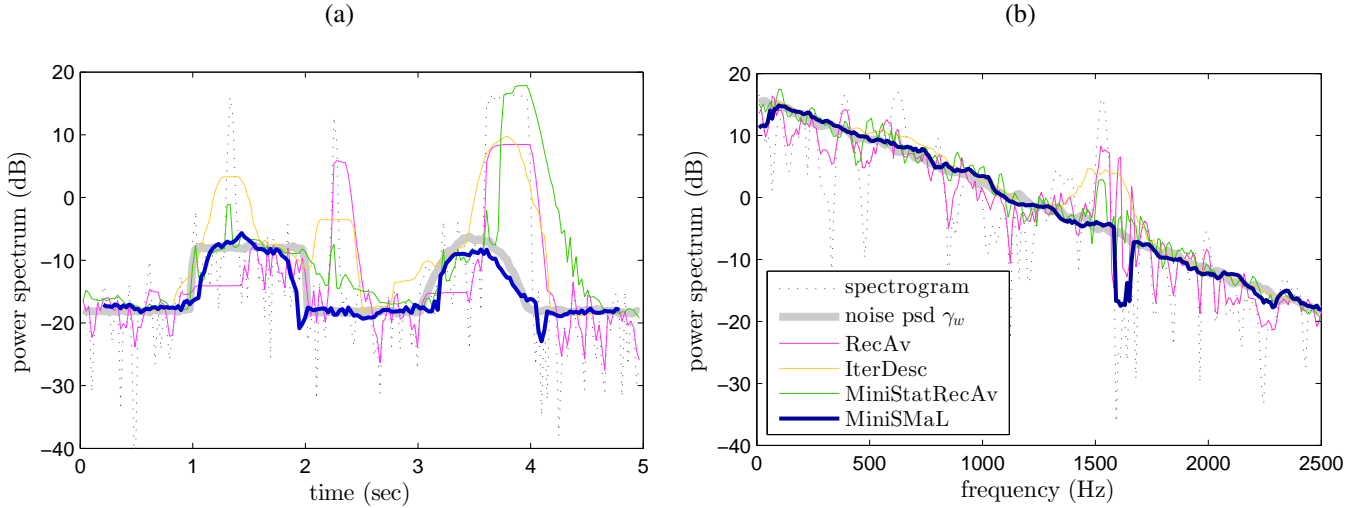
(a)

(b)

Figure 12. Noise estimates obtained with the four methods along (a) the time line and (b) the frequency line indicated in figure 10-b). The legend displayed in the subfigure (b) also holds for the subfigure (a).

estimator	RecAv [5]	IterDesc [6]	MiniStatRecAv [9]	MiniSMaL
parameters	$\eta = 0.7$; $\gamma = 0.95$; $\beta = 0.5$; $\alpha_p = 0.5$; $\alpha_d = 0.3$; $\delta = 1.5$;	neighborhood $= 15 * 15$; $pfa = 10^{-3}$; $Nb_{\min} = 4$;	$D = 15$; $M = 0.668$; $H = 1.55$; $\alpha_{max} = 0.9$; $\alpha_{min} = 0.1$;	neighborhood $= 15 * 15$; $\kappa = -0.4$;

Table II
PARAMETERS OF THE FOUR NOISE ESTIMATORS WITH THEIR VALUES USED IN SECTION VI-B.
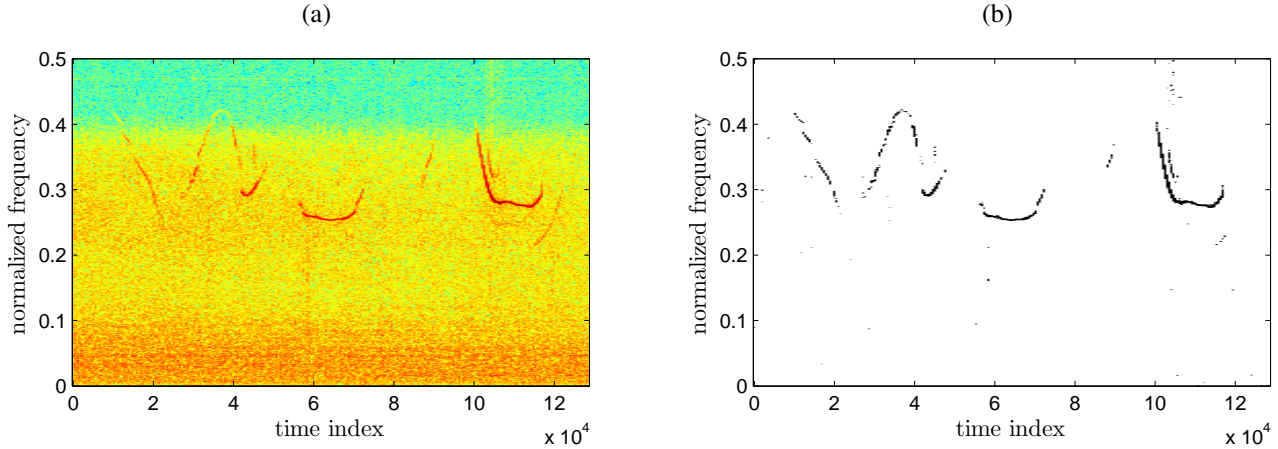
(a)

(b)

Figure 13. (a) Spectrogram of a dolphin whistle in underwater noise. (b) Time-frequency detection result for $P_{fa} = 10^{-5}$.

consider the detection threshold S^{th} obtained according to equation (27) with a nominal PFA P_{fa0} and a noise estimate $\hat{\gamma}_w$. The same threshold is obtained with the true noise power γ_w and an other 'equivalent' PFA denoted P_{faeq}. According to equation (27) we can write

$$S^{th} = \hat{\gamma}_w \ln(P_{fa0}^{-1}) = \gamma_w \ln(P_{faeq}^{-1}). \qquad (28)$$

It follows that the relation between the estimation error and the equivalent PFA is

$$P_{faeq} = P_{fa0}^{\left(\frac{\hat{\gamma}_w}{\gamma_w}\right)}. \qquad (29)$$

Denoting $\Delta\gamma_w = \hat{\gamma}_w - \gamma_w$ the difference between the estimated and the true value of the noise, equation (29) can be rewritten as

$$P_{faeq} = P_{fa0}^{\left(1 + \frac{\Delta\gamma_w}{\gamma_w}\right)}, \qquad (30)$$

which highlights the impact of an uncertainty on the noise power. We illustrate this relation on the following example: we consider the MiniSMaL estimation of the noise given a $15 * 15$ initial time-frequency neighborhood ($N = 225$) and a spectral kurtosis criterion $\kappa = -0.4$. For the ideal case of a noise-only neighborhood, those parameters lead to a minimal

subset normalized threshold $u_n = 2.12$ (see table I) and the size of the minimal subset is $Z = N(1 - exp(-u_n)) = 198$. With those values we calculated the normalized variance of the MiniSMaL estimator[4] which we found equal to 0.017. The associated normalized standard deviation $stdn\{\hat{\gamma}_w\} = 0.13$ determines the range of errors typically encountered at the output of the estimation process. Taking this value for $\Delta\gamma_w/\gamma_w$ in equation (30), the equivalent PFA of a detection test with a nominal PFA of 10^{-3} ranges between $10^{-3.39}$ and $10^{-2.61}$. In regard to the detection of the dolphin signal presented above, the normalized standard deviation of the MiniSMaL estimator is equal to 0.0396 and the equivalent PFA ranges between $10^{-5.2}$ and $10^{-4.8}$ (the nominal PFA was $P_{fa0} = 10^{-5}$).

VII. CONCLUSION

The minimal statistics approach represents an interesting alternative for the estimation problem of a non-stationary noise power spectral density. However, two elements have restricted the practical use of the method. First, it relies on a selection of the smallest coefficients which have generally a high variance. As a result, the noise estimators built upon these coefficients have a high variance. Second, it has appeared difficult to formulate some automatic procedures in regard to the determination of its constitutive elements, *i.e.* the choice of the initial set of time-frequency coefficients (the estimation neighborhood) and the determination of the threshold to select the minimal values subset.

In this paper, the Minimal Statistics Maximum Likelihood (MiniSMaL) estimator of the noise has been presented and evaluated. According to the signal to noise ratio and to the presence of signal within the minimal subset, this estimator may be strongly biased or may possess a high variance. We observed that the proportion of signal contained in the initial neighborhood impacts more deeply the estimator performance than the overall SNR. However, it has also been observed that the bias of the MiniSMaL estimator always admits a minimum for some specific value of the minimal subset threshold. This specific threshold is generally close to the value of the noise power itself.

So as to pick out this threshold, the use of the spectral kurtosis is proposed. It is shown that the spectral kurtosis of a minimal subset only depends on the ratio between the threshold and the noise power. This property implies that in an ideal context where the minimal subset contains noise-only coefficients, the spectral kurtosis can be used to determine a threshold leading to a noise estimate close to the minimum bias. If the presence of signal within the minimal subset can not be avoided (at low SNRs or for wide time-frequency supports of the signal), it is shown that the variations induced on the spectral kurtosis have a small impact on the performance of the estimator. As a consequence, the spectral kurtosis may figure as a potential robust tool so as to determine the minimal subset required by the minimal statistics approach.

Two elements will be considered for further research studies: first, the MiniSMaL estimator is rooted in a maximum

---

4As for the results reported in figure 2, the variance was calculated over 10^4 noise realizations.
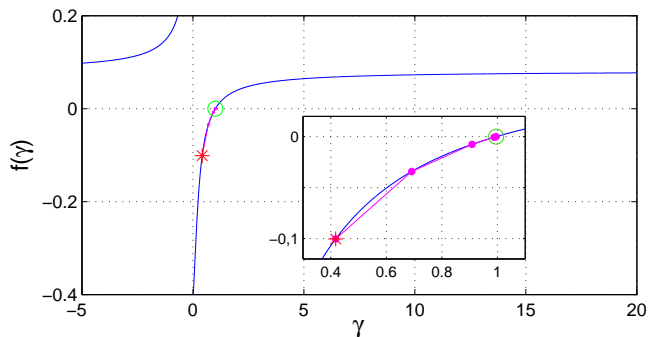

Figure 15. Example of function $f(\gamma)$ (see equation (31)) whose zero is sought for using the Newton-Raphson method. In this example, the solution is $\gamma^* = 1$. The initialisation point (red star) located at the left of the desired root (green circle) ensures the fast convergence of the method. The subfigure emphasises this convergence through the progress of the first three iterations (purple points).

likelihood strategy which is not optimal for the small estimation sample sizes considered here. Other kind of estimators, like the Best Linear Unbiased, may achieve better performances in this context. This estimator requires some additional knowledge (the total number of noise coefficient within the estimation neighborhood, *i.e.* $N(1-P)$) that needs to be fulfilled with further investigations. Second, the choice of the initial set of time-frequency coefficients is still an open issue. This choice must depend on the time and frequency variability of the noise power spectral density on one hand, and on the time-frequency region of support of the signal on the other hand.

APPENDIX A
MAXIMUM LIKELIHOOD SOLUTION WITH THE NEWTON-RAPHSON METHOD

Given a threshold u, a set of Z coefficients $S_{(i)}$ and their arithmetic mean $m_u = \frac{1}{Z}\sum_{i=1}^{Z} S_{(i)}$, the Newton-Raphson method is used to find the solution γ^* to the ML equation

$$f(\gamma) = \gamma - m_u - \frac{u}{\exp\left(\frac{u}{\gamma}\right) - 1} = 0, \text{ for } \gamma > 0. \quad (31)$$

As plotted in figure 15, the function $f(\gamma)$ is monotonic increasing and admits a single root (except for the pathological case discussed in section IV-A). Consequently, the Newton-Raphson method can be used to find the unique solution. The initialisation of the Newton-Raphson method is a crucial step which may lead to non-convergence or aberrant results. In the present case, the initial value of γ, noted $\gamma^{(0)}$, is set to the arithmetic mean m_u of the Z time-frequency coefficients $S_{(i)}$. As m_u is necessarily smaller than the solution γ^* (the coefficients $S_{(i)}$ are right-truncated exponentially distributed), this initial value ensures the convergence of the method (see figure 15) with less than 10 iterations. The iterative procedure

is summarised as follows:

$$\left[\begin{array}{l} \gamma^{(0)} = m_u; \\ \text{for } i = 1 : 10 \\ \quad f = \gamma^{(i-1)} - m_u - \frac{u}{\exp\left(u/\gamma^{(i-1)}\right)-1}; \\ \quad f' = 1 - \frac{u^2 \exp\left(u/\gamma^{(i-1)}\right)}{\left(\gamma^{(i-1)}\left(\exp\left(u/\gamma^{(i-1)}\right)-1\right)\right)^2}; \\ \quad \gamma^{(i)} = \gamma^{(i-1)} - \frac{f}{f'}; \\ \text{end} \end{array} \right. \qquad (32)$$

REFERENCES

[1] C. Ris and S. Dupont, "Assessing local noise level estimation methods: Application to noise robust asr," *Speech Communication*, vol. 34, pp. 141–158, 2001.

[2] F. Hlawatsch and G. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Processing Magazine*, vol. 9, no. 2, pp. 21–67, April 1992.

[3] I. Cohen and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Signal Processing Letters*, vol. 9, no. 1, pp. 12–15, January 2002.

[4] I. Cohen, "Noise estimation in adverse environments: improved minima controlled recursive averaging," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 5, pp. 466–475, September 2003.

[5] S. Rangachari and P. Loizou, "A noise-estimation algorithm for highly non-stationary environments," *Speech Communication*, vol. 48, pp. 220–231, 2006.

[6] C. Hory and N. Martin, "Maximum likelihood noise estimation for spectrogram segmentation control," *In Proceedings of the ICASSP'02, Orlando, USA*, pp. 1581–1584, 2002.

[7] C. Yeh and A. Robel, "Adaptive noise level estimation," in *Workshop on Computer Music and Audio Technology (WOCMAT'06)*, Taipei, Taiwan, March 2006.

[8] F. Millioz, J. Huillery, and N. Martin, "Short-time fourier transform probability distribution for time-frequency segmentation," *In proceedings of the ICASSP'06, Toulouse, France*, May 2006.

[9] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. on Speech and Audio Processing*, vol. 9, no. 5, pp. 504–512, July 2001.

[10] ——, "Bias compensation methods for minimum statistics noise power spectral density estimation," *Signal Processing*, vol. 86, pp. 1215–1229, 2006.

[11] D. Mauler and R. Martin, "Noise power spectral density estimation on highly correlated data," *In Proc. of the Int. Workshop on Acoustic Echo and Noise Control, Paris*, September 2006.

[12] V. Stahl, A. Fisher, and R. Bippus, "Quantile based noise estimation for spectral subtraction and wiener filtering," *In proc. of the ICASSP*, vol. 3, pp. 1875–1878, 2000.

[13] N. W. Evans and J. S. Mason, "Time-frequency quantile-based noise estimation," in *European Signal and Image Processing Conference*, Toulouse, France, September 2002.

[14] J. Huillery, "Time-frequency regions of support for an unknown signal in additive gaussian noise," PhD thesis (in French), http://tel.archives-ouvertes.fr/tel-00301324/, University of Grenoble, 2008.

[15] F. Millioz and N. Martin, "Estimation of a white Gaussian noise in the Short Time Fourier Transform based on the spectral kurtosis of the minimal statistics: application to underwater noise," in *Proceedings of IEEE Conference on Acoustics, Speech and Signal Processing*, Dallas, Texas, USA, Mar. 2010.

[16] L. K. Koopman, *The spectral analysis of time series*. Academic Press, 1974.

[17] T. S. Durrani and J. Nightingale, "Probability distributions for discrete fourier spectra," *Proc. Inst. Elec. Eng*, vol. 120, no. 2, pp. 299–311, 1973.

[18] F. Millioz and N. Martin, "Circularity of the STFT and Spectral Kurtosis for Time-Frequency Segmentation in Gaussian Environment," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 515 – 524, Feb. 2011.

[19] J. Huillery, F. Millioz, and N. Martin, "On the description of spectrogram probabilities with a chi-squared law," *IEEE Transactions on Signal Processing*, vol. 56, no. 9, pp. 2249–2258, June 2008.

[20] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous univariate distributions*, 2nd ed. Wiley and sons, 1995, vol. 2.

[21] F. M. Al-Athari, "Estimation of the mean of truncated exponential distribution," *Journal of Mathematics and Statistics*, vol. 4, no. 4, pp. 284–288, 2008.

[22] W. Martin and P. Flandrin, "Wigner-ville spectral analysis of nonstationary processes," *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol. 33, no. 6, pp. 1461–1470, December 1985.

[23] W. D. Mark, "Spectral analysis of the convolution and filtering of non-stationary stochastic processes," *Journal of sound and vibration*, vol. 11, no. 1, pp. 19–63, 1970.

[24] S. Blumenthal and R. C. Dahiya, "Estimating scale and truncation parameters for the truncated exponential distribution with type-1 censored sampling," *Communications in Statistics - Theory and Methods*, vol. 34, pp. 1–21, 2005.

[25] U. J. Dixit and P. N. Nasiri, "Estimation of parameters of a right truncated exponential distribution," *Statistical Papers*, vol. 49, pp. 225–236, 2008.

[26] V. Vrabie, P. Granjon, and C. Servière, "Spectral kurtosis: from definition to application," in *IEEE-EURASIP International Workshop on Nonlinear Signal and Image Processing*, Grado, Italie, jun 2003.

[27] D. C. Hoaglin, F. Mosteller, and J. W. Tuckey, *Understanding Robust and Exploratory Data Analysis*. Wiley-Interscience, 1983.

[28] J. Huillery and N. Martin, "Detection of time-frequency components of signals: Bayes and focus SNR," in *Proceedings of the IMA*, Cirencester, United-Kingdom, Dec. 2006, pp. 132–136.

[29] H. Van Trees, *Detection, Estimation, and Modulation Theory - Part 1*. Wiley, 2004.

[30] P. D. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, June 1967.

Julien Huillery received the M.S. degree in Electrical Engineering from the École Supérieure d'Électricité (France) and the University of Wollongong (Australia) in 2002, the master of research degree in Signal Processing from the university of Rennes (France) in 2004 and the Ph.D. degree in Signal Processing from the university of Grenoble (France) in 2008. After a one year post-doctoral research position in underwater acoustic communications at the university of Algarve (Portugal), he served as an assistant professor in Signal Processing and Control at the university of Grenoble. Since 2010, he holds an associate professor position at the Electrical Engineering department of the École Centrale de Lyon, France.

His research interests concern non-stationary signals and systems analysis, statistical signal processing and information theory. His current projects include signal design for wireless energy transmission, identification and filtering of time-varying systems and stochastic optimization in biological systems.

Fabien Millioz graduated from the École Normale Supérieure de Cachan, France and received the M.Sc. and Ph.D. degree in signal, image, and speech processing from the Institut National Polytechnique of Grenoble, France, in 2005 and 2009, respectively. He was a postdoctoral research associate at the Institute for Digital Communications at the University of Edinburgh, UK, from 2009 to 2011. He is currently lecturer at the university of Lyon 1, associated to the Lyon Neuroscience Research Center.

His research interests are time-frequency analysis, statistical signal processing, and signal segmentation.

Nadine Martin graduated in Physics and Electronics in 1980 from the Institute of Chemistry and Physics of Lyon (France), received the MS degree in 1981 and the Ph.D. degree in 1984 both in Signal Processing and Control from Grenoble Institute of Technology (France). Currently, she is a senior researcher at the CNRS, National Center of Scientific Research (France), and in charge of the research team SAIGA, Signal and Automatic for surveillance, diagnostic and biomechanics, a team within GIPSA-lab, Grenoble Image Speech Signal and Automatic, Grenoble, France.

In the signal-processing domain, her research interests are the analysis, detection and modeling of non-stationary signals. She is project manager of KAStrion, a European KIC InnoEnergy project on current and vibration analysis for a preventive and predictive condition based maintenance in offshore wind farms.