Circularity of the STFT and spectral kurtosis for time-frequency segmentation in white Gaussian environment

Fabien Millioz, Nadine Martin

Circularity of the STFT and spectral kurtosis for time-frequency segmentation in Gaussian environment

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

Abstract—**This paper investigates the circularity of Short Time Fourier Transform (STFT) coefficients noise only, and proposes a modified STFT such that all coefficients coming from white Gaussian noise are circular. In order to use the spectral kurtosis (SK) as a Gaussianity test to check if signal points are present in a set of STFT points, we consider the SK of complex circular random variables, and its link with the kurtosis of the real and imaginary parts. We show that the variance of the SK is smaller than the variance of the kurtosis estimated from both real and imaginary parts. The effect of the non-circularity of Gaussian variables upon the spectral kurtosis of STFT coefficients is studied, as well as the effect of signal presence. Finally, a time-frequency segmentation algorithm based on successive iterations of noise variance estimation and time-frequency coefficients detection is proposed. The iterations are stopped when the spectral kurtosis on non-detected points reaches zero. Examples of segmented time-frequency space are presented on a dolphin whistle and on a simulated signal in non-white and non-stationary Gaussian noise.**

Index Terms—**Short time Fourier transform, circularity, spectral kurtosis, time-frequency segmentation, statistical segmentation, time-frequency analysis**

EDICS: SSP-NSSP

I. INTRODUCTION

Time-Frequency representations are useful tools for non-stationary signal analysis as they describe the spectral energy along time. A segmentation task is helpful in such a signal characterization by highlighting time-frequency patterns containing a signal of interest. Some enhancements to the readability of time-frequency representations already exist. An example is Coates *et al.* [1], which describes the time-frequency plane by a set of basis functions. Chassande-Mottin *et al.* propose a time-frequency segmentation based on the reassignment method [2], which splitting the time-frequency plane into different basins of attraction, where the attractors are the time-frequency locations that are invariant to reassignment.

Considering a deterministic signal embedded in a white stationary Gaussian noise, we have already proposed a spectrogram segmentation ([3],[4]) based on a deterministic model,

Fabien Millioz was is with the GIPSA-Lab, Signal and Image Department (DIS), INPG-CNRS, Grenoble, France. He is currently with the University of Edinburgh, UK.

Nadine Martin is with the GIPSA-Lab, Signal and Image Department (DIS), INPG-CNRS, Grenoble, France.

which has been successfully applied to automatic speech recognition ([5], [6]). More recently, we proposed a second algorithm based directly on the Short Time Fourier Transform (STFT) ([7],[8]).

In this paper, we focus on the STFT segmentation and extend the signal model. We consider any kind of signal $s[m]$ embedded in a centered white Gaussian noise $w[m]$ of variance σ_w^2.

$$x[m] = s[m] + w[m] \tag{1}$$

$s[m]$ is the signal to be segmented, which contains all the non-stationary parts of $x[m]$, which may be either deterministic or stochastic. The only condition on $s[m]$ is that it is not a stationary white Gaussian signal, which is the definition of the noise. Our goal is to determine what time-frequency points contain part of the $s[m]$ energy.

To that end, we need to know the distribution of time-frequency coefficients containing noise only. Section II investigates the non-circularity of some coefficients. A circular complex random variable has a probability density independent of its argument, permitting an easier way to handle it. We proposes a modified STFT where all noise coefficients have the same circular distribution.

A criterion is needed to characterize the noise. Section III recalls the spectral kurtosis of a circular random variable, and studies the link between the spectral kurtosis of a complex variable and the kurtosis of the real or imaginary parts of this variable. The influence of non-circularity for Gaussian coefficient on the spectral kurtosis is detailed, as well as the influence of the presence of signal in the STFT.

Section IV describes the proposed algorithm. Examples of time-frequency segmentations are given on a real-life signal, a dolphin whistle, and on a synthetic signal embedded in non-white and non-stationary Gaussian noise.

II. CIRCULARITY OF THE TIME FOURIER TRANSFORM

The STFT of a discrete signal $x[m]$, denoted as $X_\phi[n,k]$, is a collection of N Fourier Transforms, computed on windowed segments of $x[m]$ centered at time n. $X_\phi[n,k]$ describes the spectral content's change over time. The STFT is a complex-valued transformed defined as

$$X_\phi[n,k] = \sum_{m=n-\frac{M_\phi-1}{2}}^{n+\frac{M_\phi-1}{2}} x[m]\phi[m-n]\exp\left(-2i\pi k\frac{m}{M_\phi+Z}\right) \tag{2}$$

$$= X_\phi^r[n,k] + iX_\phi^i[n,k] \tag{3}$$

n and k are time and frequency indexes respectively, and $\phi[m]$ an M_ϕ-length window. Z is the zero padding, which gives $K = M_\phi + Z$ frequency bins. $X_\phi^r[n,k]$ and $X_\phi^i[n,k]$ are the real and imaginary parts of $X_\phi[n,k]$ respectively. We consider an energy-normalized window, so that

$$\sum_{m=-\frac{M_\phi-1}{2}}^{+\frac{M_\phi-1}{2}} \phi[m]^2 = 1. \tag{4}$$

To simplify notations, the range of summation is omitted in the following.

In a detection context, we need to determine the probability density of time-frequency coefficients containing noise only.

Considering $x[m]$ as a centered white Gaussian noise $w[m]$ of variance σ_w^2, the STFT coefficients of $w[m]$, denoted $W_\phi[n,k]$, are Gaussian complex variables. Their variance is given by

$$\mathrm{Var}(W_\phi[n,k]) = \mathrm{E}\left(W_\phi[n,k]W_\phi[n,k]^*\right) = \mathrm{E}\left(|W_\phi[n,k]|^2\right) \tag{5}$$

$$= \mathrm{E}\left(W_\phi^r[n,k]^2\right) + \mathrm{E}\left(W_\phi^i[n,k]^2\right) \tag{6}$$

According to equation (2), $W_\phi^r[n,k]$ and $W_\phi^i[n,k]$, the real and imaginary parts of $W_\phi[n,k]$ respectively, are the sum of M_ϕ independent centered Gaussian variables. Therefore, $W_\phi^r[n,k]$ and $W_\phi^i[n,k]$ are centered Gaussian variables with second-order moment given by

$$\mathrm{E}(W_\phi^r[n,k]^2) = \sigma_w^2 \sum_m \phi[m-n]^2 \cos(-2\pi k \frac{m}{M_\phi + Z})^2, \tag{7}$$

$$\mathrm{E}(W_\phi^i[n,k]^2) = \sigma_w^2 \sum_m \phi[m-n]^2 \sin(-2\pi k \frac{m}{M_\phi + Z})^2. \tag{8}$$

Substituting (7) and (8) into (6) yields

$$\mathrm{Var}(W_\phi[n,k]) = \mathrm{E}(W_\phi^r[n,k]^2) + \mathrm{E}(W_\phi^i[n,k]^2) = \sigma_w^2. \tag{9}$$

assuming the normalized window (4).

As a result, we get the well-known result that a STFT coefficient of Gaussian white noise of variance σ_w^2 is a Gaussian complex variable with the same variance.

A. Why circularity?

A complex variable z is circular if z and $e^{ja}z$ have the same probability distribution for any a [9]. In other words, the probability distribution of z is independent of its argument. A Gaussian complex variable is circular if its real and imaginary parts have the same probability distribution and are independent. In this case, the spectrogram, defined as

$$\mathrm{Spec}_w[n,k] = W_\phi^r[n,k]^2 + W_\phi^i[n,k]^2, \tag{10}$$

is the sum of two independent squared Gaussian of the same variance, and consequently have a χ^2 distribution.

The continuous Fourier transform leads to circular variables [9], but not the discrete Fourier transform. Koopmans [10] and Johnson and Long [11] noticed that spectrogram coefficients

do not have a χ^2 distribution at $k = 0$ and $k = \frac{M_\phi}{2}$ for a boxcar window without zero padding, and at $k = 0$, $k = 1$, $k = \frac{M_\phi}{2} - 1$ and $k = \frac{M_\phi}{2}$ for a Hanning window without zero padding.

B. Non-circularity of the Short-Time Fourier Transform

To study the circularity of the discrete STFT, we propose to compare the second order moments of $W_\phi^r[n,k]$ and $W_\phi^i[n,k]^2$ and their correlation. $W_\phi[n,k]$ is circular if its real and imaginary parts are uncorrelated and have the same variance. We define $\alpha[n,k]$ as the second-order moment $\mathrm{E}(W_\phi^r[n,k]^2)$ normalized by σ_w^2 so that

$$\alpha[n,k] = \frac{\mathrm{E}(W_\phi^r[n,k]^2)}{\sigma_w^2} \tag{11}$$

$$= \sum_m \phi[m-n]^2 \cos(-2\pi k \frac{m}{M_\phi + Z})^2. \tag{12}$$

Using trigonometric identities, equation (12) becomes

$$\alpha[n,k] = \frac{1}{2} + \frac{1}{2} \sum_m \phi[m]^2 \cos(4\pi k \frac{m+n}{M_\phi + Z}). \tag{13}$$

Given equations (9) and (12), we obtain a new expression of equations (7) and (8)

$$\mathrm{E}(W_\phi^r[n,k]^2) = \alpha[n,k]\sigma_w^2 \tag{14}$$

$$\mathrm{E}(W_\phi^i[n,k]^2) = \left(1 - \alpha[n,k]\right)\sigma_w^2. \tag{15}$$

Parameter $\alpha[n,k]$ characterizes the second-order moments of the Gaussian distributions of $W_\phi^r[n,k]$ and $W_\phi^i[n,k]$.

Figure 1 shows the variation of this parameter for a STFT, computed with a Blackman window of 31 points, an overlap of 30 points between two consecutive windows and 128 frequency bins. The $\alpha[n,k]$ leading to different real and imaginary second-order moments are located around the low and high frequencies.

For discrete frequencies k far enough away from zero and the index of the Shannon frequency $\frac{M_\phi + Z}{2}$, the cosine function in equation (13) has a high enough frequency compared to the variations of $\phi[m]^2$ and correctly nullifies the second sum in equation (13). We thus have $\alpha[n,k] = \frac{1}{2} = 1 - \alpha[n,k]$. In other words, both $W_\phi^r[n,k]$ and $W_\phi^i[n,k]$ have the same Gaussian distribution of variance $\frac{\sigma_w^2}{2}$. In the other case, for k close to zero and $\frac{M_\phi + Z}{2}$, the edges of the window will have an effect on the value of $\alpha[n,k]$, which becomes different from $\frac{1}{2}$, leading to a different Gaussian distribution for $W_\phi^r[n,k]$ and $W_\phi^i[n,k]$.

To complete the characterization of the distribution of $W_\phi[n,k]$, we consider the correlation coefficient between the
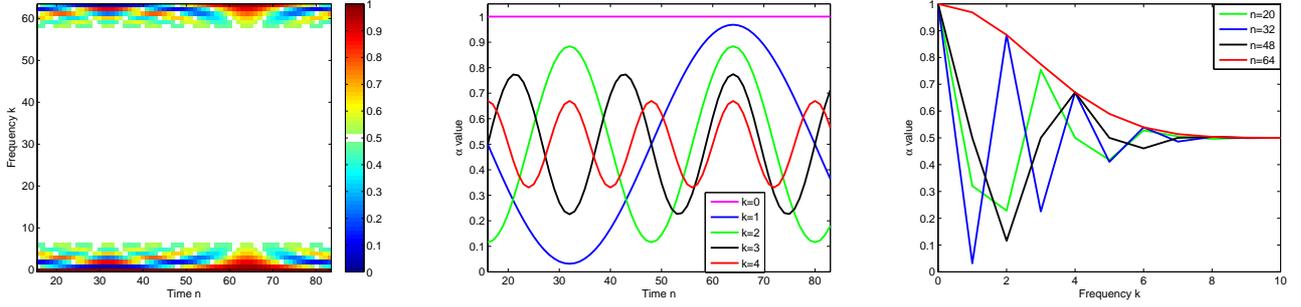
Fig. 1. (left), $\alpha[n,k]$ variations on a STFT with a 31-point Blackman window, a 30-point overlap and 128 frequency bins. White area concerns α values between 0.485 and 0.515. (middle), cross-section of $\alpha[n,k]$ variations at frequencies $k = 0$ to 4. (right), cross-section of $\alpha[n,k]$ values at time $n = 20, 32, 48$ and 64, over $k = 0$ to 10 only.
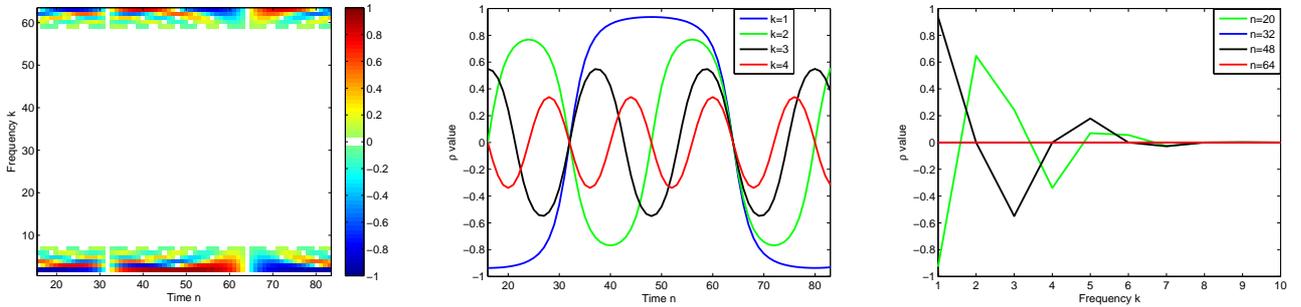

Fig. 2. (left), $\rho[n,k]$ variations on a STFT with a 31-point Blackman window, a 30-point overlap and 128 frequency bins. Time-frequency coefficients whose $\alpha[n,k]$ coefficient is $\sqrt{1}2$ (Figure 1) have a correlation coefficient equal to zero. (middle), cross-section of $\rho[n,k]$ variations at frequencies $k = 1$ to 4. (right), cross-section of $\rho[n,k]$ values at time $n = 20, 32, 48$ and 64, over $k = 0$ to 10 only.

real and imaginary parts of $W_\phi[n,k]$, denoted $\rho[n,k]$

$$\rho[n,k] = \frac{\mathrm{E}(W_\phi^r[n,k]W_\phi^i[n,k])}{\sqrt{\mathrm{E}(W_\phi^r[n,k]^2)\mathrm{E}(W_\phi^i[n,k]^2)}} \qquad (16)$$

$$= \frac{\sigma_w^2 \sum_m \phi[m]^2 \cos(-2\pi k\frac{m+n}{M_\phi+Z})\sin(-2\pi k\frac{m+n}{M_\phi+Z})}{\sigma_w^2\sqrt{\alpha[n,k](1-\alpha[n,k])}} \qquad (17)$$

$$= \frac{1}{2\sqrt{\alpha[n,k](1-\alpha[n,k])}} \sum_m \phi[m]^2 \sin(-4\pi k\frac{m+n}{M_\phi+Z}) \qquad (18)$$

This definition assumes that $\alpha[n,k] \neq 0$ and $\alpha[n,k] \neq 1$. In those cases, either the real or imaginary part of $W_\phi[n,k]$ is zero, therefore the correlation between the two parts is zero while the correlation coefficient is not defined.

Figure 2 shows the variation of $\rho[n,k]$ for a STFT computed with a Blackman window of 31 points, an overlap of 30 points between two consecutive windows and 128 frequency bins. Like $\alpha[n,k]$, values of $\rho[n,k]$ leading to a non-circularity of $W_\phi[n,k]$, that is non-zero values, are located around the low and high frequencies.

To study $\alpha[n,k]$ and $\rho[n,k]$, we rewrite equations (12) and (18)

$$\alpha[n,k] = \frac{1}{2} + \frac{1}{2}\Re(\Phi[n,k]), \qquad (19)$$

$$\rho[n,k] = \frac{1}{2\sqrt{\alpha[n,k](1-\alpha[n,k])}}\Im(\Phi[n,k]) \qquad (20)$$

with a function $\Phi[n,k]$ such that

$$\Phi[n,k] = e^{(-2j\pi 2k\frac{n}{M_\phi+Z})} \sum_m \phi[m]^2 e^{(-2j\pi 2k\frac{m}{M_\phi+Z})} \qquad (21)$$

The variations of $\alpha[n,k]$ and $\rho[n,k]$ are both described by $\Phi[n,k]$. $\Phi[n,k]$ (21) is made of the product of two terms: a phase term $e^{(-2j\pi 2k\frac{n}{M_\phi+Z})}$ depending on the time and frequency indexes $[n,k]$, and the Discrete Fourier Transform (DFT) of the squared window $\phi[m]^2$.

When $\Phi[n,k]$ is zero, $\alpha[n,k]$ is equal to $\frac{1}{2}$ and $\rho[n,k]$ is zero, that is $W_\phi[n,k]$ is circular. On the other hand, $W_\phi[n,k]$ is not circular on two frequency bands depending on the DFT of a window function, where $\Phi[n,k]$ is not zero. The STFT is not circular for normalized frequencies around zero and $\frac{1}{2}$, depending on the window's length M_ϕ, the zero padding Z and the window's type.

Figure 3 illustrates this phenomenon by showing $\Phi[n,k]$ absolute values for various STFT parameters.

C. Rectified Short Time Fourier Transform

The STFT definition (2) is not unique. The time translation may switch from the window $\phi[m]$ to the signal $x[m]$, leading
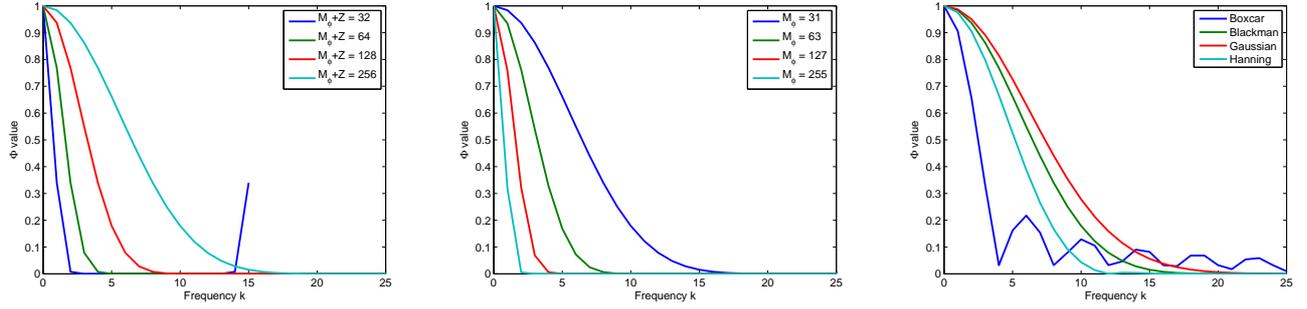
Fig. 3. Variations of $\phi[n,k]$ (21) for various STFT parameters, for the 25 first frequency bins. (left), impact of the number of computed frequencies $M_\phi + Z$, for a Blackman window of 31 points. The last blue point correspond to the $k = \frac{M_{phi}}{2} - 1$. (middle), impact of the length M_ϕ of a Blackman window, for 256 frequency bins. (right), impact of the window's type, for $M_\phi = 31$ and $M_\phi + Z = 256$. When $\Phi[n,k]$ is zero, $\alpha[n,k]$ is equal to $\frac{1}{2}$ and $\rho[n,k]$ is equal zero, so $W_\phi[n,k]$ is circular.

to[1].

$$X_\phi[n,k] = \sum_{m=\frac{M_\phi-1}{2}}^{\frac{M_\phi-1}{2}} x[n+m]\,\phi[m]e^{\left(-2j\pi k \frac{m}{M_\phi+Z}\right)} \quad (22)$$

Using (22) instead of (2) makes the second order moment of the real and imaginary parts of $W_\phi[n,k]$ independent from the time index n. The function $\Phi[n,k]$ of (21) becomes

$$\Phi[n,k] = \sum_m \phi[m]^2 e^{(-2j\pi 2k\frac{m}{M_\phi+Z})} = \Phi[0,k] \quad (23)$$

$\Phi[n,k]$ does not depend on n, so do $\alpha[n,k]$ and $\rho[n,k]$. Additionally, assuming that window functions $\phi[m]$ are even and using Euler's formula, (23) is

$$\Phi[n,k] = \phi[0]^2 + \sum_{m=1}^{+\frac{M_\phi-1}{2}} \phi[m]^2 \Big(\exp(-2j\pi 2k\frac{m}{M_\phi+Z})... \quad (24)$$

$$+ \exp(2j\pi 2k\frac{m}{M_\phi+Z}) \Big) \quad (25)$$

$$= \phi[0]^2 + 2 \sum_{m=1}^{+\frac{M_\phi-1}{2}} \phi[m]^2 \cos(4\pi k\frac{m}{M_\phi+Z}) \quad (26)$$

Therefore, $\Phi[n,k]$ is real. According to equation (20), coefficient correlation $\phi[n,k]$ between the real and imaginary parts is zero for all time-frequency points.

Using STFT defined by (22) and an even window function make the correlation between the real and imaginary parts disappear. However, the variances of the real and imaginary parts are not equal everywhere on the time-frequency plane, depending on $\alpha[n,k]$.

In order to get a circular STFT, we propose a *rectified STFT*

[1] This definition is the one used in the *Time-Frequency Toolbox*. This Toolbox is a collection of about 100 free scripts for GNU Octave and Matlab developed for the analysis of non-stationary signals using time-frequency distributions. http://tftb.nongunu.org
Definition (2) is the definition used in Matlab.

denoted $X_\phi[n,k]' = X_\phi^r[n,k]' + iX_\phi^i[n,k]'$, such as

$$X_\phi^r[n,k]' = X_\phi^r[n,k],$$
$$X_\phi^i[n,k]' = X_\phi^i[n,k], \quad \text{if } \alpha[n,k] = 0 \text{ or } \alpha[n,k] = 1 \quad (27)$$
$$X_\phi^r[n,k]' = X_\phi^r[n,k]/\sqrt{2\alpha[n,k]},$$
$$X_\phi^i[n,k]' = X_\phi^i[n,k]/\sqrt{2(1-\alpha[n,k])} \quad \text{otherwise.} \quad (28)$$

Like classic STFT, all rectified STFT coefficients of a white Gaussian noise $W_\phi[n,k]'$ have a variance equal to σ_w^2, but now have the variances of their real $W_\phi^r[n,k]'$ and imaginary $W_\phi^i[n,k]'$ both equal to $\frac{\sigma_w^2}{2}$.

The rectified STFT with an even window function is circular for all time-frequency coefficients. Consequently, *rectified spectrogram* coefficients of noise only

$$\text{Spec}_w[n,k]' = W_\phi^r[n,k]'^2 + W_\phi^i[n,k]'^2, \quad (29)$$

have a χ^2 distribution, with one degree a freedom if $\alpha[n,k] = 0$ or 1, or with two degree of freedom otherwise.

Huillery *et al* [12] shows that a χ^2 distribution with two degree of freedom is a good approximation of a single spectrogram coefficient, even with correlated noise, conditionally that the window function has null boundaries. Here, we aim to use a single χ^2 distribution with two degree of freedom to approximate the distribution of all spectrogram coefficients, and to compare it with the distribution of all rectified spectrogram coefficients.

We use the Kullback-Leibler divergence $D_{KL}(p,q)$ to measure the distance between the χ^2 distribution and the distribution of the classic or rectified spectrogram. It is calculated as [13]

$$D_{KL}(p,q) = \int p(x) \log \left(\frac{p(x)}{q(x)} \right) dx \quad (30)$$

As the natural logarithm is used in this expression, the Kullback-Leibler divergence unit is the 'nat'. This divergence is null if the two distributions p and q are the same.

Figure 4 show the Kullback-Leibler divergence between respectively the spectrogram (10) and the rectified spectrogram
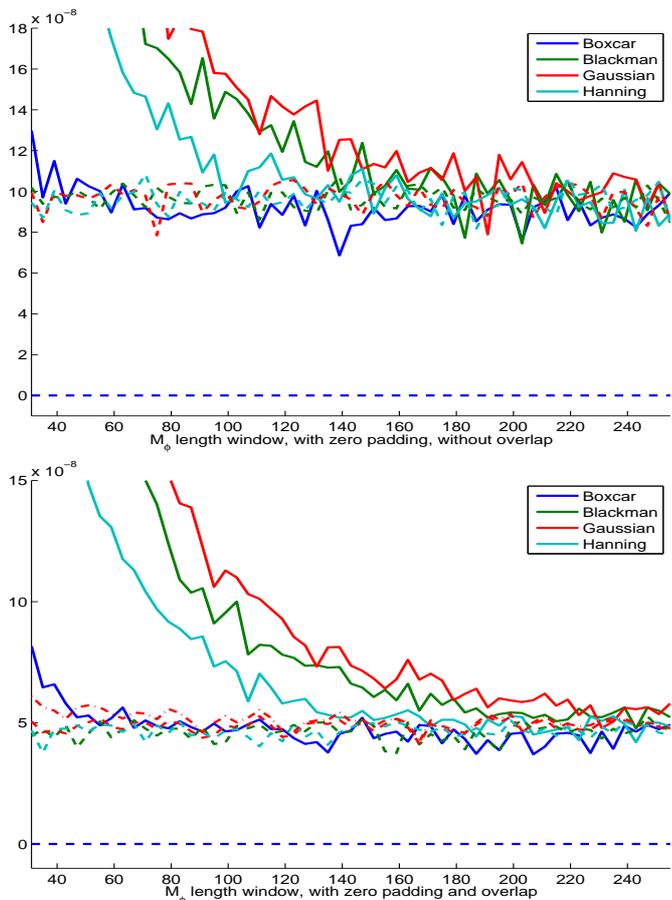
Fig. 4. Kullback-Leibler divergence between a χ^2 distribution with two degree of freedom and the distribution of a spectrogram of a 2000 points white Gaussian noise (solid line) and the distribution of a rectified spectrogram (dashed line), for several windows. On left, STFT are computed without overlap, and a zero padding such as $M_\phi + Z = 4M_\phi$. On right, STFT are computed with an overlap of $0.75M_\phi$ points, and a zero padding such as $M_\phi + Z = 4M_\phi$. Rectified spectrogram leads to a better approximation by a χ^2 distribution.

(29), and the χ^2 distribution with two degree of freedom which is the reference. The divergences are computed from 1000 runs of the spectrogram of a white Gaussian noise $b[m]$ of 2000 points.

These results allows us to conclude that the rectified spectrogram distribution is better approximated by a χ^2 distribution than the usual spectrogram distribution, especially for low window length M_ϕ.

III. SPECTRAL KURTOSIS FOR NOISE CHARACTERIZATION

This section investigates the use of the spectral kurtosis of the STFT in the framework of noise characterisation.

The kurtosis $\mathcal{K}(x)$ of a random variable x is the fourth normalized cumulant. Considering a real centered random variable, the kurtosis is

$$\mathcal{K}(x) = \frac{\mathrm{E}(x^4)}{\mathrm{E}(x^2)^2} - 3 \qquad (31)$$

A kurtosis for a complex random variable z has been defined

in [14], and is called spectral kurtosis $\mathcal{KS}(x)$

$$\mathcal{KS}(z) = \frac{\mathrm{E}(z^2 z^{*2})}{\mathrm{E}(zz^*)^2} - 2 \qquad (32)$$

A real Gaussian random variable has a kurtosis equal to zero, whereas a complex Gaussian variable has a spectral kurtosis which is equal to zero. We propose to use kurtosis for characterization if a set of STFT points is issued from a Gaussian complex random variable, namely noise. First, the link between kurtosis and spectral kurtosis is made, and their relative variance is described. Second, the influence of the non-circularity of noise coefficients on the spectral kurtosis is studied, which shows the interest of the rectified STFT presented in section II. Finally, we show the influence of signal in a set of noise coefficients on the spectral kurtosis in order to determine a way of using it in a time-frequency segmentation algorithm.

A. Link between kurtosis and spectral kurtosis of circular variable

STFT coefficients of noise may only be approximated to be circular. In this section, we consider a random circular complex variable $z = a + jb$. Spectral kurtosis of z can be given in regard to the kurtosis of its real or imaginary parts.

The variance of z is

$$\mathrm{E}(zz^*) = \mathrm{E}(a^2 + b^2) = \mathrm{E}(a^2) + \mathrm{E}(b^2) = 2\mathrm{E}(a^2) \qquad (33)$$

The fourth moment is

$$\mathrm{E}(z^2 z^{*2}) = \mathrm{E}(a^4 + b^4 + 2a^2 b^2) = 2\mathrm{E}(a^4) + 2\mathrm{E}(a^2 b^2) \qquad (34)$$

Considering that $a = \frac{z+z^*}{2}$ and $b = \frac{z-z^*}{2i}$, (34) becomes

$$\mathrm{E}(z^2 z^{*2}) = 2\mathrm{E}(a^4) - \frac{1}{8}\mathrm{E}(z^4 + z^{*4} - 2z^2 z^{*2}) \qquad (35)$$

From the circularity of z we deduce that both $\mathrm{E}(z^4)$ and $\mathrm{E}(z^{*4})$ are both zero. Consequently, the fourth moment of z is

$$\mathrm{E}(z^2 z^{*2}) = \frac{3}{8}\mathrm{E}(a^4) \qquad (36)$$

Substituting (36) and (33) into (31) and (32), the ratio of the spectral kurtosis relative to the kurtosis of the real part is

$$\frac{\mathcal{KS}(z)}{\mathcal{K}(a)} = \frac{\mathrm{E}(z^2 z^{*2}) - 2\mathrm{E}(zz^*)^2}{\mathrm{E}(zz^*)^2} \times \frac{\mathrm{E}(a^2)^2}{\mathrm{E}(a^4) - 3\mathrm{E}(a^2)^2} \qquad (37)$$

$$= \frac{\frac{8}{3}\mathrm{E}(a^4) - 8\mathrm{E}(a^2)^2}{4(\mathrm{E}(a^4) - 3\mathrm{E}(a^2)^2)} = \frac{2}{3} \qquad (38)$$

The spectral kurtosis of a complex circular random variable is equal to $\frac{2}{3}$ of the kurtosis of its real or imaginary parts. Consequently, spectral kurtosis or usual kurtosis have the same behaviour.

The variance of a usual kurtosis estimator is approximated to $\frac{24}{N}$ [15], while the variance of spectral kurtosis is approximated by $\frac{4}{N}$ [15], where N is the number of independent realizations of the random variable. Two phenomena explain this difference of variance: first, a complex coefficient present a real and a imaginary part, so there are twice the real

independent realizations as the complex realizations. Second, the ratio between the spectral kurtosis and the kurtosis of its real or imaginary part is $\frac{2}{3}$ (38). Taking into account these two phenomena, we define an equivalent variance of usual kurtosis, such that

$$\text{Var}_{eq}(\mathcal{K}(x)) = \frac{1}{2}\left(\frac{2}{3}\right)^2 \frac{24}{N} = \frac{4}{3}\frac{4}{N} = \frac{4}{3}\text{Var}(\mathcal{KS}(z)) \quad (39)$$

Considering the kurtosis estimated on real and imaginary parts of a complex variable leads to a greater variance than considering the spectral kurtosis. As a result, only spectral kurtosis will be used.

B. Influence of non-circularity of the STFT

We now consider a centered complex Gaussian variable $z = a + jb$ of variance σ^2, such that its real and imaginary parts are independent, but with different variances. From (14) and (15) of II-B, $\alpha\sigma^2$ is the variance of the real part a, and $(1-\alpha)\sigma^2$ is the variance of the imaginary part b.

The fourth moment of z (36) becomes

$$\text{E}(z^2 z^{*2}) = \text{E}(a^4 + b^4 + 2a^2 b^2) \quad (40)$$
$$= \text{E}(a^4) + \text{E}(b^4) + 2\text{E}(a^2 b^2) \quad (41)$$

Given that a and b are centered independent real Gaussian variables, their kurtosis is zero. Then applying equation (31) to variables a and b yields

$$\text{E}(z^2 z^{*2}) = 3\text{E}(a^2)^2 + 3\text{E}(b^2)^2 + 2\text{E}(a^2)\text{E}(b^2) \quad (42)$$

Substituting the variances of a (14) and b (15), $\text{E}(z^2 z^{*2})$ is

$$\text{E}(z^2 z^{*2}) = 3\alpha^2\sigma^4 + (1-\alpha)^2\sigma^4 + 2\alpha(1-\alpha)\sigma^4 \quad (43)$$
$$= \sigma^4(3 - 4\alpha + 4\alpha^2) \quad (44)$$

The spectral kurtosis becomes

$$\mathcal{KS}(z) = \frac{\text{E}(z^2 z^{*2})}{\text{E}(zz^*)^2} - 2 = (1 - 2\alpha)^2 \quad (45)$$

Spectral kurtosis of a Gaussian complex variable is zero only for circular variables. The greater the difference in variance between the real and imaginary parts, the more the spectral kurtosis differs from zero. Simulations confirm this result. Spectral kurtosis estimated from a usual STFT of white Gaussian points of 1000 points, computed with a Hanning window of 63 points, an overlap of 31 points, and 128 frequency bins leads to 0.055. The spectral kurtosis estimated from the rectified STFT of the same parameters leads to -0.006. The absolute value of the bias is divided by ten when using rectified STFT instead of the usual one.

C. Kurtosis in a STFT with an added signal of interest

The previous section describes the spectral kurtosis for the Gaussian distribution, corresponding to a STFT containing noise only. In this paper, the use of the spectral kurtosis aims to determinate if a set of time-frequency coefficients contains noise only or a part of the signal. Consequently, the behaviour of the spectral kurtosis in a STFT with a signal of interest must be known.
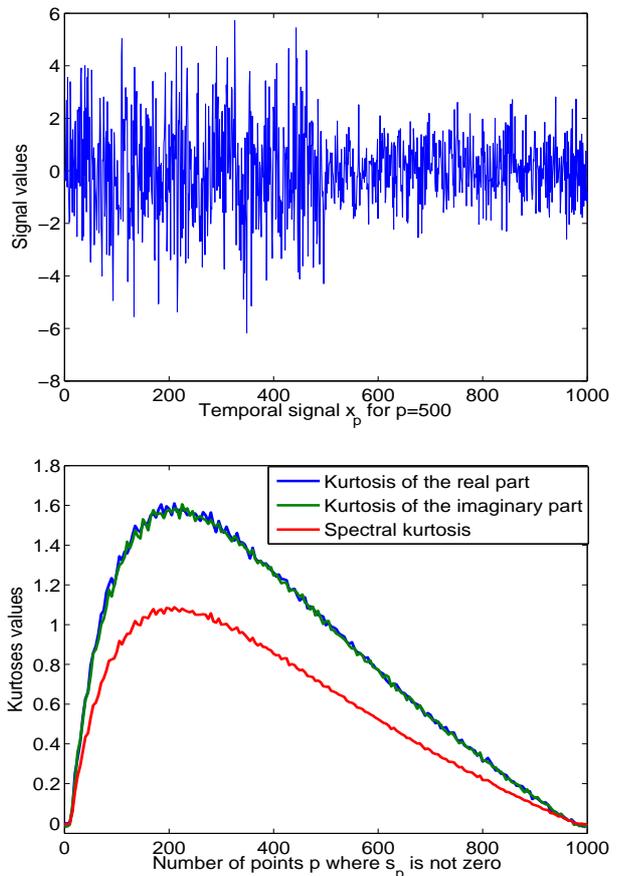

Fig. 5. On the left, signal $x_p[m]$ (46) for $p = 500$. On the right, evolution of the kurtoses of the STFT of $x_p[m$, in function of p. When $x_p[m]$ is a white Gaussian noise, that is for $p = 0$ or $p = 1000$, the kurtoses are null. In other cases, in presence of non-stationary signal, the kurtoses are positive. Spectral kurtosis is equal to $\frac{2}{3}$ of the kurtosis of real or imaginary part for any value of p.

The spectral kurtosis has already been used as a signal characterization in the spectrum [14], its value permits a conclusion to whether a given frequency is stationary or not.

In our case, non-stationary signals in a time-frequency approach, we only want to differentiate a STFT containing noise only from a STFT containing noise and signal, regardless of the analyzed signal.

Given the model of the signal (1), any kind of signal should be studied. We only consider as a generic example a non-stationary signal $x_p[m]$ of 1000 points, such that

$$x_p[m] = s_p[m] + w[m], \quad (46)$$

$$\text{with } s_p[m] = \begin{cases} g[m] \text{ for } m \leq p \\ 0 \text{ otherwise,} \end{cases} \quad (47)$$

where $w[m]$ is a white Gaussian Noise of variance $\sigma_w^2 = 1$ and $g[m]$ a white Gaussian Noise of variance $\sigma_x^2 = 3$. Signal $s_p[m]$ is non-stationary, described by its parameter p, indicating the number of points where $x[m]$ exists. Given that $s[m]$ is white, a proportion of $p/1000$ points in the time domain where $s_p[m]$

is non-zero leads to a proportion of $p/1000$ points in the time-frequency domain where the STFT of $s_p[m]$ is non-zero.

Figure 5 shows on the left $x_p[m]$ for $p = 500$ and on the right the evolution of the kurtoses of all STFT coefficients with respect to p. For $p = 0$, $x_0[m] = w[m]$ is a white Gaussian Noise of variance 1, the kurtoses are null. On the other hand, for $p = 1000$, $x_p[m] = g[m] + w[m]$ is a white Gaussian noise of variance $3 + 1 = 4$, its kurtosis is null. Between these two extreme cases, the signal $x_p[m]$ is a mixture of two Gaussian variables, that is to say it is not Gaussian anymore, and the kurtoses are positive. As stated in (38), this figure shows that the spectral kurtosis is $\frac{2}{3}$ of the kurtosis of the real or imaginary parts of the STFT.

Time-frequency representations are meant to concentrate energy on some time-frequency areas; a few points, containing signal energy, will have great values while other will contain small values, only noise energy. In other words, a set of STFT coefficients tends to be super-Gaussian, and to have a positive kurtosis, as illustrated on figure 5.

Consequently, we assume that a positive spectral kurtosis of a set of STFT points means that some of these points contain signal.

Next section proposes the use of the spectral kurtosis of the STFT coefficients as a Gaussianity test for a time-frequency segmentation algorithm.

IV. STATISTICAL SEGMENTATION OF THE RECTIFIED STFT

In this section, we propose an update of the time-frequency segmentation algorithm proposed in ([3],[7]). Considering the model of signal $x[m] = s[m] + w[m]$ (1), where $s[m]$ is an unknown signal to segment, and $w[m]$ a white Gaussian noise of unknown variance σ_w^2, the goal of the segmentation is to determine automatically what are the time-frequency locations where there is a part of the signal s energy. In other words, we aim to determine if a time-frequency coefficient contains noise only or not.

A. Principle

a) Signal detection: Section II shows that a coefficient of the rectified STFT containing noise only has a complex circular Gaussian distribution. In the presence of a signal $s[m]$, due to the linearity of the STFT, consequently of the rectified STFT, the time-frequency coefficients become

$$X_\phi[n,k]' = S_\phi[n,k]' + W_\phi[n,k]' \qquad (48)$$

The problem is to discriminate coefficients for which $S_\phi[n,k]'$ is non-zero. Given that $s[m]$ is unknown, $S_\phi[n,k]'$ is also unknown, and may have a zero-mean. Consequently, we consider the second-order moment of (48)

$$E(X_\phi[n,k]'^2) = \gamma_s[n,k]'^2 + \sigma_w^2 \qquad (49)$$

where $\gamma_s[n,k]'^2$ is the second order moment of $S_\phi[n,k]'$ only. A time-frequency coefficient containing a signal energy has a higher second-order moment than coefficients containing noise only.

Given that the probability distribution of the signal $s[m]$ is unknown, the probability distribution of the second-order moment of rectified STFT coefficients containing signal is also unknown. However, the distribution of the second-order moment of the coefficients containing noise only is known (29) to be a χ^2 distribution with two degrees of freedom[2].

Consequently, a detection based on a Neyman-Pearson approach is achievable to determine if a STFT coefficient (48) contains only noise, $S_\phi[n,k]' = 0$, called hypothesis $\mathcal{H}_0$, of if it contains non-zero signal, called hypothesis $\mathcal{H}_1$. A threshold $t_{\sigma_w^2}$ given a chosen false alarm probability p_{fa} is defined by

$$t_{\sigma_w^2} \quad / \quad Prob\left(W_\phi[n,k]'^2 > t_{\sigma_w^2}\right) = p_{fa} \qquad (50)$$

$$t_{\sigma_w^2} = \sigma_w^2 \ln(p_{fa}^{-1}) \qquad (51)$$

where $\ln(x)$ is the natural logarithm. All time-frequency coefficients whose second-order moment is higher than this threshold will be detected as a signal, and called $\mathcal{H}_1$ set. The coefficients lower than the threshold are called $\mathcal{H}_0$ set.

The threshold $t_{\sigma_w^2}$ (51) depends on the chosen false alarm probability, and the noise level, which has to be estimated.

b) Noise variance estimation: To estimate a noise variance, a Maximum Likelihood (ML) estimator is unbiased and optimal in the case of noise only. The estimation $\hat{\sigma}_w^2$ is

$$\widehat{\sigma_w^2} = \frac{1}{NK} \sum_{n,k} \left(X_\phi^r[n,k]'^2 + X_\phi^i[n,k]'^2\right) \qquad (52)$$

where K and K are the number of time and frequency indexes.

When the signal $s[m]$ is present, the estimator becomes biased

$$E\left(\widehat{\sigma_w^2}\right) = \sigma_w^2 + \frac{1}{NK} \sum_{n,k} |\gamma_s[n,k]'|^2 \qquad (53)$$

The presence of the signal $S_\phi[n,k]$ to segment results in an overestimation of the noise level. Unfortunately, given that this signal is unknown, its time-frequency location is also unknown.

The idea is consequently to use an iterative algorithm: a first noise variance is overestimated, but allows a first threshold (51) to be defined and then leads to a first separation of the STFT into two sets, the $\mathcal{H}_1$ set corresponding to the time-frequency coefficients with a second-order moment than the threshold, and the $\mathcal{H}_0$ set. At this iteration, the $\mathcal{H}_0$ set contains some STFT coefficients containing signal. The next iterations re-estimate the noise variance from the $\mathcal{H}_0$ set only. As iterations go on, the noise variance is less and less overestimated, leading to a lower detection threshold, and thus time-frequency coefficients containing signal are detected in the $\mathcal{H}_1$ set.

c) Stop criterion: Section III-B shows that the spectral kurtosis estimated from a rectified STFT is either positive when containing signal, or zero when the coefficients contain noise only. Therefore we propose to use it as a stop criterion

---

[2]The cases $\alpha[n,k] = 0$ or 1 is ignored here.

for the iterative algorithm. The algorithm will stop when the spectral kurtosis of the $\mathcal{H}_0$ set becomes smaller than a threshold t_{KS}, and considers that the remaining points have a complex Gaussian distribution corresponding to noise only. To define this threshold, we will consider the variance of the spectral kurtosis estimator.

The variance of a spectral kurtosis built from R independent Gaussian realizations is approximated by $\frac{4}{R}$ [15]. A rectified STFT is a set of NK non-independent time-frequency points. In order to approximate a number of independent realizations from non-dependant ones, we define a *redundancy factor* f_r such that

$$f_r = \frac{NK}{M} \qquad (54)$$

where M is the point number of the signal $s[m]$ in time.

At each iteration i of the algorithm, the $\mathcal{H}_0$ set has $N_i K_i$ non-independent realizations. The number of independent realizations is approximated by $\frac{N_i K_i}{f_r}$.

Finally, considering that the spectral kurtosis estimator of a Gaussian complex variable has a Gaussian distribution, 97.5% of these realizations are within twice its standard-deviation. We thus choose twice the standard-deviation of the spectral kurtosis estimator as a threshold t_{SK}

$$t_{SK} = 2\sqrt{\frac{4f_r}{N_i K_i}} \qquad (55)$$

d) Region growing: To create, we add at each iteration a region growing algorithm [16], as in [3]. The main advantage of this region growing is the possibility to limit the number of detected points becoming segmented at each iteration. By segmentation only a proportion of p_{cand} points, one can avoid classifying as signal small isolated high realization of Gaussian noise.

The region growth is executed by choosing the coefficient of the $\mathcal{H}_1$ set with the highest second-order moment as a "seed" for the region growing algorithm, labelled l. A search for other detected points is carried out in its 8-connectivity neighborhood in the time-frequency representation, assigning to these points the same label. The spectral pattern labeled l is thus created. When all neighbouring points are classified, a new seed is chosen among the remaining detected points, and a new pattern is created. Consequently, a segmented spectral pattern is a set of coefficients connected in the time-frequency plane.

An analysis of the influence of the algorithm parameters p_{fa}, p_{cand} and t_{SK} may be found in [17] and [18].

In summary, the time-frequency segmentation proposed in this paper is an iterative process: at each iteration, the noise variance is overestimated (53) due to the presence of the signal; then this noise level permits the detection of some signal points (51). A region growing algorithm is applied to create spectral patterns in the $\mathcal{H}_1$ set, and put isolate coefficients back into the $\mathcal{H}_0$ set. The iterations are stopped when the non-detected points have a Gaussian complex distribution, in other words when the spectral kurtosis of these points becomes smaller than the threshold t_{SK}.
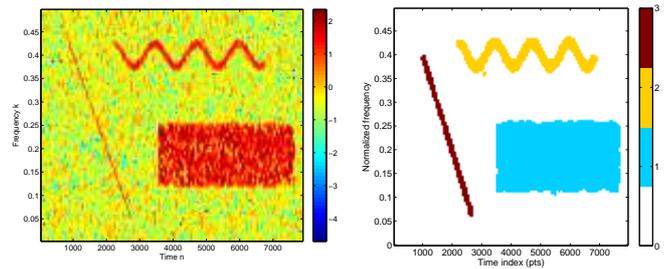

Fig. 6. Spectrogram of a synthetic signal at left. Segmentation results with $\sigma_w^2 = \frac{1}{2}$ at right. All time-frequency patterns are correctly detected.
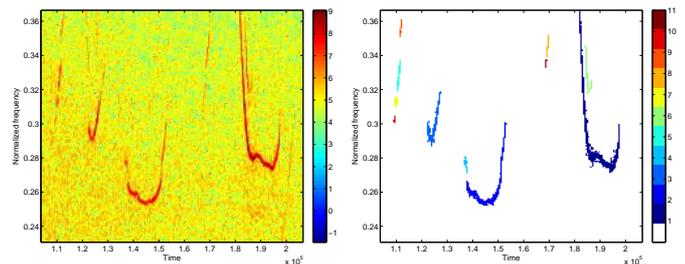

Fig. 7. Spectrogram of a dolphin whistle, bounded to a frequency band where the noise is approximately white. The segmentation gives a pattern for each frequency modulation.

B. Illustrations

Figure 6 shows the spectrogram of a synthetic signal of 8000 points made of three non-stationary signals: a linear chirp of amplitude 0.5 and a sinusoidal frequency modulation of amplitude 1, both windowed in time; the additive Gaussian noise has a variance $\sigma_s^2=32$ windowed in time and frequency. The additive noise $w[m]$ has a variance $\sigma_w^2 = \frac{1}{2}$. All time-frequency patterns are well segmented, while the estimated noise variance is $\hat{\sigma}_w^2 = 0.494$. The spectral kurtosis of the $\mathcal{H}_0$ set is 0.032, and the final threshold 0.049.

The hypothesis of white Gaussian noise may be limited to the hypothesis of white Gaussian noise on a given frequency band only. Figure 7 illustrates this with a dolphin whistle in non-white underwater noise. The spectrogram here is limited to a frequency band where noise is roughly white, permitting the segmentation of the dolphin whistle into different time-frequency patterns.

Another way deal with non-white and/or non-stationary Gaussian noise is to apply the algorithm locally in the time-frequency domain. In this case, the noise variance is estimated separately for each time-frequency coefficient, on a given local neighbourhood. A unique probability of false alarm allows the algorithm to detect signal coefficients in relation to the local noise level (51).

However, the $\mathcal{H}_0$ set has not a single Gaussian distribution anymore, so the spectral kurtosis criterion is not directly usable. By using the coefficients of the $\mathcal{H}_0$ set normalized by the local estimation of the variance, we get a set having a single Gaussian distribution, on which the spectral kurtosis criterion is applied.
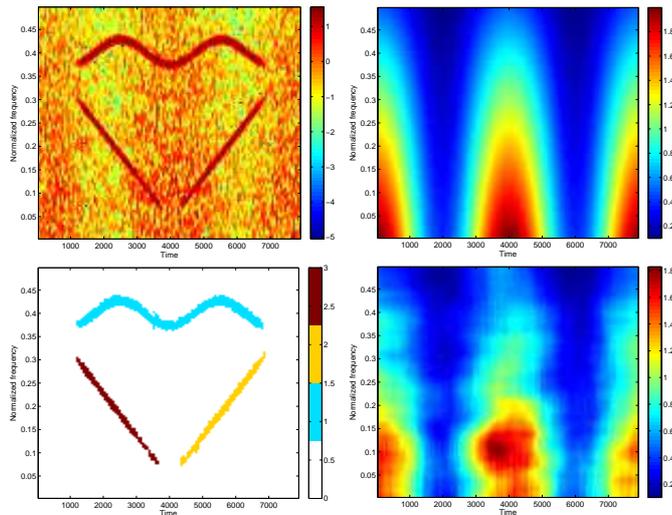
Fig. 8. Segmentation a synthetic signal and a non-white non-stationary Gaussian noise. Top left: spectrogram. Top right: theoretical variance of the added Gaussian noise. Bottom left: segmentation result. Bottom right: estimated variance.

Figure 8 illustrates the segmentation of a synthetic signal embedded in a non-white and non-stationary Gaussian noise, for a neighbourhood of 21 time points and 21 frequency points, and a probability of false alarm of 0.1. The segmentation result is correct (top left), while the noise variance is well estimated (bottom right).

Finally, this algorithm relies only on the linearity of the STFT, and can be extended to other time-frequency transforms [18] (in french). An application of this segmentation to the Capongram of Barkhausen noise is detailed by Padovese *et al* [19].

V. CONCLUSION

This papers firstly deals with the circularity of the STFT coefficients of white Gaussian noise. These coefficients are not circular for normalized frequencies close to 0 and 0.5, and we have described their non-circularity depending on their position in the time-frequency plane. We propose a modified STFT, where all coefficients are circular. We check that the spectrogram of white Gaussian noise coming from this new STFT has a distribution closer to a χ^2 distribution with two degrees of freedom than usual spectrogram.

We have then considered spectral kurtosis: first, we have shown that the spectral kurtosis of a circular coefficient is equal to $\frac{2}{3}$ of the kurtosis of its real or imaginary part, and that the variance of the spectral kurtosis is smaller than the variance of the kurtosis estimated from both the real and imaginary parts of the variable. We have also proved that a non-circular Gaussian coefficient has a positive spectral kurtosis, defending the interest of the rectified STFT. The influence of the presence of signal in the STFT on the spectral kurtosis has been illustrated, and we have shown that the signal leads to positive spectral kurtosis.

These results are used to update a time-frequency algorithm, based on successive iterations of noise variance

over-estimation and signal coefficient detection. Instead of using a criterion to stop the algorithm based on the convergence of the statistics of non-segmented points, we use the spectral kurtosis as a Gaussianity test of non-segmented coefficients. This new criterion is not open to local convergences. When this spectral kurtosis becomes zero, all non-segmented points have a single Gaussian distribution, meaning that they comes from noise only.

The algorithm has been extended to non-white and non-stationary Gaussian noise by applying the algorithm locally in the time-frequency domain, and estimating the spectral kurtosis on non-detected points normalized by the local noise variance estimation. Results of the algorithm have been illustrated on a real-case signals for a stationary white Gaussian noise, and on a synthetic signal embedded in a non-white and non-stationary Gaussian noise.

REFERENCES

[1] M. Coates and W. Fitzgerald, "Regionally optimised time-frequency distributions using finite mixture models," *Signal Processing*, vol. 77, pp. 247–260, 1999.

[2] E. Chassande-Mottin, "Méthodes de réallocation dans le plan temps-fréquence pour l'analyse et le traitement de signaux non stationnaires," Thèse de doctorat, Université de Cergy-Pontoise, Sep. 1998.

[3] C. Hory, N. Martin, and A. Chehikian, "Spectrogram Segmentation by Means of Statistical Features for Non-stationary Signal Interpretation," *IEEE Transactions on Signal Processing*, vol. 50, no. 12, pp. 2915–2925, Dec. 2002.

[4] N. Martin, C. Hory, and M. Huchard, "Chi-square Law Mixture for an Unsupervised Time-Frequency Pattern Extraction," in *Proceedings of International Congress on Sound and Vibration 10*, Stockholm, Sweden, Jul. 2003.

[5] F. R. Rodriguez, W. Liu, N. Evans, and J. Mason, "Morphological Filtering of Speech Spectrograms in the Context of Additive Noise," in *Procedings of Eurospeech*, Geneva, Switzerland, Sep. 2003.

[6] W. Liu, V. R. Bastante, F. R. Rodriguez, N. Evans, and J. Mason, "Morphological Filtering of Spectrograms for Automatic Speech Recognition," in *Proceedings of Visualization, Imaging, and Image Processing*, Marbella, Spain, Sep. 2004.

[7] F. Millioz, J. Huillery, and N. Martin, "Short Time Fourier Transform Probability Distribution for Time-Frequency Segmentation," in *Proceedings of IEEE Conference on Acoustics, Speech and Signal Processing*, Toulouse, France, May 2006.

[8] F. Millioz and N. Martin, "Time-Frequency Segmentation for Engine Speed Monitoring," in *Proceedings of International Congress on Sound and Vibration 13*, Vienna, Austria, Jul. 2006.

[9] B. Picinbono, "On Circularity," *Proceedings of the IEEE*, vol. 42, no. 12, pp. 3473–3482, Dec. 1994.

[10] L. Koopmans, *The Spectral Analysis of Time Series*. Academic Press, 1974.

[11] P. Johnson and D. Long, "The Probability Density of Spectral Estimates Based on Modified Periododram Avergares," *IEEE Transactions on Signal Processing*, vol. 47, no. 5, pp. 1255–1261, May 1999.

[12] 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. 6, pp. 2249–2258, Jun. 2008.

[13] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York, NY, USA: Wiley and Sons, 1991.

[14] 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.

[15] M. Kendall and A. Stuart, *The advanced theory of statistics*. Griffin, 1994.

[16] R. Adams and L. Biscof, "Seeded Region Growing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 6, pp. 641–647, 1994.

[17] F. Millioz and N. Martin, "Reassignment Vector Field for Time-Frequency Segmentation," in *Proceedings of International Congress on Sound and Vibration 14*, Cairns, Australia, Jul. 2007.

[18] F. Millioz, "Deux approches de segmentation temps-fréquence : détection par modèle statistique et extraction de contour par le champ de vecteurs de réallocation," Thèse de doctorat, INP Grenoble, Sep. 2009.

[19] L. Padovese, N. Martin, and F. Millioz, "Time-Frequency and Time-Scale analysis of Barkhausen noise signals," *Proc. IMechE, Part G: J. Aerospace Engineering*, no. 223(G5), pp. 577–588, 2009.

Fabien Millioz Fabien Millioz graduated from the cole Normale Suprieure 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 is currently a postdoctoral research associate at the Institute for Digital Communications at the University of Edinburgh, UK. His research interests are time-frequency analysis, statistical signal processing, and signal segmentation.

Nadine Martin 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 Centre of Scientific Research (France), and in charge of the research team SA-IGA, Signal and Automatic for surveillance, diagnostic and biomechanics, a team within GIPSA-lab, Grenoble Image Speech Signal & Automatic, Grenoble, France. In the signal-processing domain, her research interests are the analysis and the interpretation of non-stationary signals. She is now working on signal detection, models of polynomial phase signals, and on time-frequency decision based on time-frequency random models. She is leading the AStrion project on an automatic spectral analyzer, a research project involved in an industry transfer. In addition of acoustic signals, vibratory signals are more particularly studied in relation with physical models. In 2004-2005, she was a scientific advisor for an automotive industry. Dr. N. Martin is the author of more than 100 papers, and of a book on Time-frequency decision. She was member of the National Council of Universities (CNU) in 2002. She was co-organizer of the Fourth European Signal Processing Conference (EUSIPCO88), and of the Sixth French Symposium on Signal and Image Processing (GRETSI97). She regularly organizes special sessions on advanced signal processing for diagnosis (IEEE-SDEMPED, ACD-Advanced Control and Diagnosis, BINDT-Condition Monitoring and Machinery Failure Prevention). Dr. Martin is a member of IEEE society.