SHORT TIME FOURIER TRANSFORM PROBABILITY DISTRIBUTION FOR TIME-FREQUENCY SEGMENTATION

Fabien Millioz, Julien Huillery, Nadine Martin

SHORT TIME FOURIER TRANSFORM PROBABILITY DISTRIBUTION FOR TIME-FREQUENCY SEGMENTATION

Fabien Millioz, Julien Huillery, and Nadine Martin

Laboratory of Image and Signal (LIS)
961 rue de la Houille Blanche, BP 46
38402 Saint Martin d'Hères Cedex, France
email: 1stname.2ndname@lis.inpg.fr

ABSTRACT

Taking as signal model the sum of a non-stationnary deterministic part embedded in a white Gaussian noise, this paper presents the distribution of the coefficients of the Short Time Fourier Transform (STFT), which is used to determine the maximum likelihood estimator of the noise level. We then propose an automatic segmentation algorithm of the real and imaginary parts of the STFT based on statistical features, which is an alternative to the spectrogram segmentations considered as image segmentations. Examples of segmented time-frequency space are presented on a simulated signal and on a dolphin whistle.

1. INTRODUCTION

Time-Frequency Representations (TFR) are useful tools for nonstationnary signal analysis by determining the time-frequency patterns, which are time-varying areas containing energetic signal. A segmentation task is a helpful step in such a signal characterization by highlighting these patterns. With a general model of signal considered as a nonstationnary deterministic signal $d[m]$ embedded in a white Gaussian noise $n[m]$ of variance σ^2

$$x[m] \quad = \quad d[m] + n[m], \tag{1}$$

we proposed ([1], [2]) a spectrogram segmentation based on statistical features. Given that spectrogram coefficients of signal described in equation (1) have a non-central χ^2 distribution while white Gaussian noise have a central χ^2 distribution, segmentation task consists in discriminating non-central from central χ^2 distribution.

In this paper, we propose a new way of segmentation, by considering real and imaginary part of Short Time Fourier Transform (STFT). Instead of having a χ^2 distribution, TFR coefficients have a Gaussian distribution, which allows a simpler segmentation method. In section 2 we determine the real and imaginary part of STFT distribution and show that their respective variance are not always equal, in order to determine an efficient noise level estimator in section 3.

In a second part, we use these results to propose a new segmentation algorithm based on local statistical features of the STFT, and its application on a simulated signal and a dolphin whistle.

2. DIFFERENCE OF VARIANCE BETWEEN PROBABILITY DISTRIBUTION OF STFT REAL AND IMAGINARY PARTS

The STFT of a discrete signal $x[m]$ is determined by computing the discrete Fourier transform on N overlapping segments centered on n, which describes the spectral contents of x around the instant n. The STFT is defined by

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

where k is the frequency index, ϕ is the M_ϕ-length window function and Z the zero padding. We will consider an energy-normalized window, so

$$\sum \phi[m]^2 = 1. \tag{3}$$

For the signal defined in (1), real and imaginary parts of the STFT, $X^r_\phi[n,k]$ and $X^i_\phi[n,k]$, are sums of M_ϕ independant Gaussian variables. $X^r_\phi[n,k]$ and $X^i_\phi[n,k]$ thus are Gaussian variables, with mean respectively given by

$$\mathrm{E}(X^r_\phi[n,k]) = \sum_{m=n-\frac{M_\phi-1}{2}}^{n+\frac{M_\phi-1}{2}} d[m]\phi[m-n]\cos(-2\pi k\frac{m}{M_\phi+Z}) \tag{4}$$

$$\mathrm{E}(X^i_\phi[n,k]) = \sum_{m=n-\frac{M_\phi-1}{2}}^{n+\frac{M_\phi-1}{2}} d[m]\phi[m-n]\sin(-2\pi k\frac{m}{M_\phi+Z}), \tag{5}$$

and variance by

$$\mathrm{Var}(X^r_\phi[n,k]) = \sigma^2\sum_{m=n-\frac{M_\phi-1}{2}}^{n+\frac{M_\phi-1}{2}} \phi[m-n]^2\cos(-2\pi k\frac{m}{M_\phi+Z})^2 \tag{6}$$

$$\mathrm{Var}(X^i_\phi[n,k]) = \sigma^2\sum_{m=n-\frac{M_\phi-1}{2}}^{n+\frac{M_\phi-1}{2}} \phi[m-n]^2\sin(-2\pi k\frac{m}{M_\phi+Z})^2 \tag{7}$$

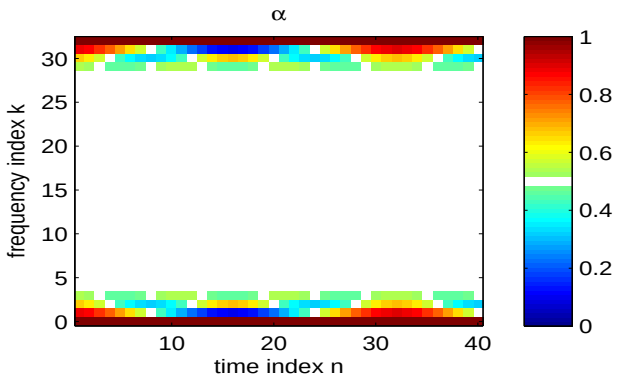


Fig. 1. Variations of $\alpha[n,k]$, for a STFT with a Blackman window of 31 points, with a zero padding of 33. White points are values of α between 0.485 and 0.515.

We notice that means of $X_\phi^r[n,k]$ and $X_\phi^i[n,k]$ are respectively real and imaginary part of the STFT of $d[m]$.

To compare the variances, we then define $\alpha[n,k]$ as the ratio of the variance of the real part of the STFT to the sum of the two variances

$$\alpha[n,k] = \frac{\text{Var}(X_\phi^r[n,k])}{\text{Var}(X_\phi^r[n,k]) + \text{Var}(X_\phi^i[n,k])} \quad (8)$$

$$= \sum_{m=n-\frac{M_\phi-1}{2}}^{n+\frac{M_\phi-1}{2}} \phi[m-n]^2 \cos(2\pi k \frac{m}{M_\phi+Z})^2. \quad (9)$$

Relation (3) induces

$$\text{Var}(X_\phi^r[n,k]) + \text{Var}(X_\phi^i[n,k]) = \sigma^2. \quad (10)$$

Using trigonometric identities, equation (9) writes

$$\alpha[n,k] = \frac{1}{2} + \frac{1}{2} \sum_{p=-\frac{M_\phi-1}{2}}^{\frac{M_\phi-1}{2}} \phi[p]^2 \cos(4\pi k \frac{p+n}{M_\phi+Z}). \quad (11)$$

When the frequency value is far enough from 0 and $\frac{M_\phi+Z}{2}$, the frequency of the cosine function will be high enough compared to the window variations to cancel the second term, so the value of $\alpha[n,k]$ will be $\frac{1}{2}$. Otherwise, $\alpha[n,k]$ will discriminate variances of $X_\phi^r[n,k]$ and $X_\phi^i[n,k]$. Consequently, a bias in the variance estimation appears and the spectrogram coefficients do not have a χ^2 distribution anymore.

Fig. 1 shows the variations of α along the time and frequency indexes for a STFT computed with a Blackman window of length $M_\phi = 31$ and with a zero padding of $Z = 33$.

This parameter extends the works of L.H. Koopmans [3], N.L. Johnson and D.G. Long [4], who only determined that the frequency bins where the spectrogram distributions computed without zero padding do not match a χ^2 distribution are $k = 0$ and $k = \frac{M_\phi}{2}$ for a rectangular window and $k = 0$, $k = 1$, $k = \frac{M_\phi}{2}$ and $k = \frac{M_\phi}{2} - 1$ for a Hanning window.

3. NOISE ESTIMATION FROM REAL PART OF STFT

In order to segment the time-frequency representation, we need to know the noise level of $x[m]$. In this section we determine a variance estimator using real part of STFT coefficients considering first that the deterministic part is null. Then we study the effect of the unknown deterministic part on the estimator.

3.1. Centered white Gaussian noise

We consider a signal $x[m]$ (1) of length N_0, where the deterministic part $d[m]$ is null.

The Maximum Likelihood (ML) estimator of the variance σ^2 is unbiased and optimal, and writes

$$\widehat{\sigma^2} = \frac{1}{N_0} \sum_{m=1}^{N_0} x[m]^2. \quad (12)$$

We want to estimate σ^2 with the real part of the STFT coefficients. In section 2 we saw that $X_\phi^r[n,k]$ has a non constant variance equals to $\alpha[n,k]\sigma^2$, where $\alpha[n,k]$ is deterministic. We thus define a new random variable of constant variance

$$X_\phi^r[n,k]' = \frac{X_\phi^r[n,k]}{\sqrt{\alpha[n,k]}}. \quad (13)$$

$X_\phi^r[n,k]'$ is a centered white Gaussian noise of variance σ^2. The variance can now be estimated as equation 12. The so-build estimator from the real part of STFT coefficients is a ML estimator, which is optimal and remains unbiased.

3.2. Deterministic signal embedded in a white Gaussian noise

We now consider a general case where $x[m]$ is a white Gaussian noise of unknown mean $d[m]$ (1).

$X_\phi^r[n,k]$ is a Gaussian variable of mean $D_\phi^r[n,k]$, the real part of the STFT of $d[m]$. The mean of the rectified random variable (13) is consequently

$$E\left(X_\phi^r[n,k]'\right) = \frac{D_\phi^r[n,k]}{\sqrt{\alpha[n,k]}} \quad (14)$$

which is unknown. In the context of a TFR segmentation of an unknown deterministic signal, we cannot specify which points are centered and which are not. When we estimate the noise level, we will take non-centered points. ML estimator thus overestimate σ^2 as

$$\text{E}\left(\widehat{\sigma^2}\right) = \frac{1}{NK} \sum_{n,k} \text{E}\left(\left(X_\phi^r[n,k]'\right)^2\right) \quad (15)$$

$$= \sigma^2 + \frac{1}{NK} \sum_{n,k} \frac{D_\phi^r[n,k]^2}{\alpha[n,k]}, \quad (16)$$

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

4. TFR SEGMENTATION

For the model of signal (1), we showed in section 2 that real and imaginary parts of STFT coefficients have a Gaussian distribution, where the mean depends on the deterministic part. The segmentation task consists in identifying coefficients with non-zero mean, which are points containing deterministic signal, in order to reconstruct time-frequency regions called spectral patterns. As seen in equation (16), non-zero means overestimate the value of the noise variance. The idea is to estimate local variances of (n,k) sites, and select with a threshold depending on the estimated noise level the points of highest variance.

4.1. Local variance distribution and threshold

As in [1], we consider a small cell of P points $\mathcal{C}_{n,k}$, centered on the (n,k) site of the real part of STFT. Local variance estimator (12) of the rectified random variable (13) writes

$$\widehat{\sigma^2}[n,k] \quad = \quad \frac{1}{P}\sum_{\mathcal{C}_{n,k}} \left(X^r_\phi[n,k]'\right)^2 . \qquad (17)$$

The knowledge of the local variance distribution of points without deterministic part allows us to propose a suitable threshold to discriminate (n,k) sites without deterministic part from others with a given false alarm probability p_{fa}.

For (n,k) sites without deterministic part, (17) is a sum of P squared centered Gaussian variables. If they are independant, $\widehat{\sigma^2}[n,k]$ have a central $\frac{\sigma^2}{P}\chi^2_P$ distribution. Due to STFT construction, $\{X^r_\phi[n,k]'\}$ are correlated, we thus have

$$\widehat{\sigma^2}[n,k] \sim \frac{\sigma^2}{\delta}\chi^2_\delta \qquad (18)$$

where δ is an unknown degree of freedom, verifying $\delta \leq P$.

This distribution has two unknown parameters, σ^2 which depends on the analyzed signal, and δ which depends on the STFT construction. By computing the STFT of a centered white Gaussian noise of known variance, the only unknown parameter of the χ^2 distribution is δ, which can be estimated with a maximum likelihood approach [2].

The second unknown parameter σ^2 is estimated by ML with equation (16). When the distribution (18) is fully estimated, we define a threshold t_{σ^2}

$$t_{\sigma^2} \quad / \quad \mathcal{P}rob\{\widehat{\sigma^2}[n,k] > t_{\sigma^2}\} = p_{fa}, \qquad (19)$$

where p_{fa} is a given false alarm probability. The use of this threshold in a segmentation algorithm is described in the next subsection.

4.2. Segmentation algorithm

The proposed algorithm is a region growing algorithm, applied to the TFR.

We first overestimate the noise variance over all the STFT real part coefficients, which give the last unknown parameter of the local variance distribution (18), and enable us to compute the threshold t_{σ^2} (19).

We then select (n,k) sites whose local variance is higher than the threshold t_{σ^2} to be candidates to the segmentation. These sites are supposed to contain deterministic mean due to equations (16) and (18).

Then, a "seed" with the highest local variance is choosen among the candidates, associated with a given label l. If some of its neighbours in the TFR are candidates, they become new seeds of same label, which contaminate then their own neighbours. Iteratively, we create so a spectral pattern of label l.

Once most of the candidates have been segmented, we estimate noise variance again with only the unlabelized coefficients. We thus obtain a less overestimated value. A new threshold is then computed on the new estimated $\widehat{\sigma^2}[n,k]$ distribution, which gives new candidates to the segmentation. Consequently, at each iteration the estimated noise level comes closer to σ^2, which allow to segment more points containing deterministic part.

4.3. Segmentation control

We use two criterions in this algorithm in order to supervise its performance. The first one is the Kolmogorov distance d_k [5] defined as

$$d_k \quad = \quad \sup\left|F^*_n(x) - F(x)\right|, \qquad (20)$$

where $F(x)$ represents the theoretical cumulative distribution function and $F^*_n(x)$ the empirical cumulative distribution function. The Kolmogorov distances on the unlabelized points before and after contamination are compared to validate a seed contamination. If the algorithm has effectively segmented (n,k) sites containing deterministic signal, the unlabelized points will converge to a Gaussian distribution and d_k will decrease.

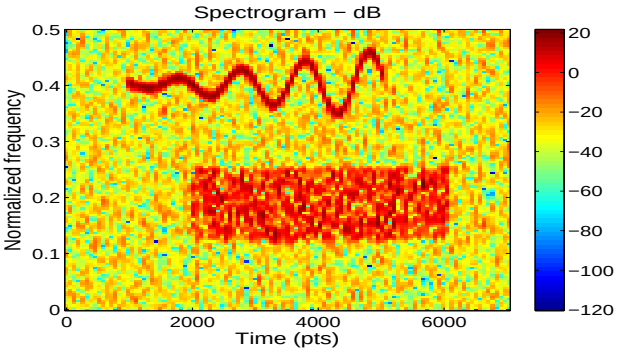
Secondly, the kurtosis [6] defined as

$$K \quad = \quad \frac{\mu_4}{\left(\sigma^2\right)^2} - 3, \qquad (21)$$

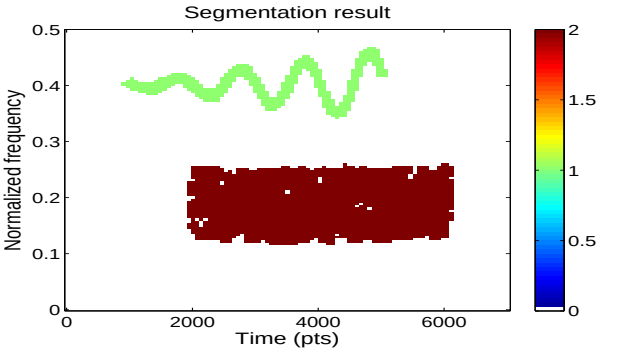
where μ_4 is the fourth centered moment, is estimated on the unlabelized points at each iteration, in order to stop the segmentation when it reaches 0. Indeed, when the algorithm does not have anymore deterministic signal to segment, the unlabelized points have a zero mean Gaussian distribution, with a null kurtosis. Moreover, it provides an indicator of execution of the algorithm. If the algorithm ends before the kurtosis reaches zero, we know that all spectral patterns do not have been segmented.

4.4. Segmentation results

Fig. 2 shows the result of a synthetic signal TFR segmentation. It's a sum of a filtered noise of variance 1 and a frequency-

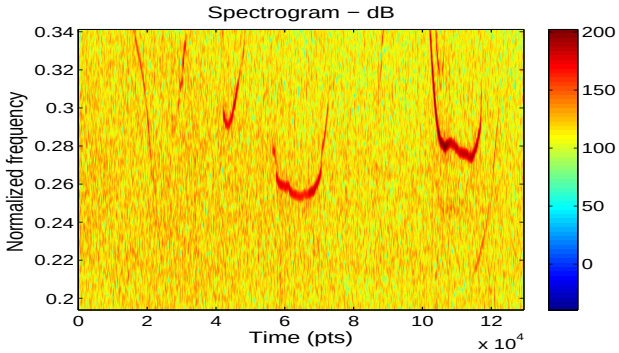


(a) Spectrogram of a simulated signal

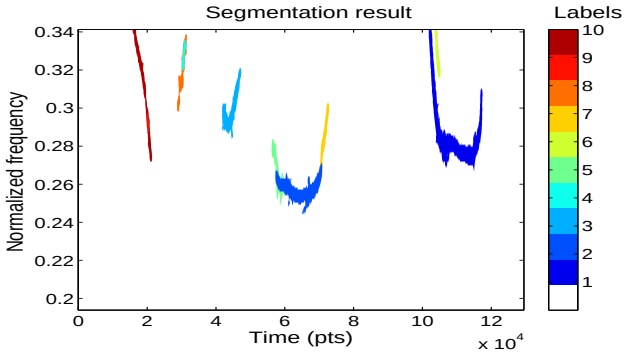


(b) Segmentation result of the simulated signal

Fig. 2. Sum of a frequency-varying signal and a large band signal. The spectrogram (a) is segmented (b) in three regions from 0, set of points without deterministic signal, from 2.



(a) Spectrogram of a dolphin whistle



(b) Segmentation result of the dolphin whistle

Fig. 3. Whistle of a dolphin. The spectrogram (a) is limited to the normalized frequency $[0.2, 0.34]$ to have a white Gaussian noise on the coefficients. Nine patterns are segmented (b).

varying cosine function of amplitude 0.5, embedded in a white Gaussian noise of variance $\sigma^2 = \frac{1}{16}$. The two spectral patterns are correctly segmented, assigning label "1" to the frequency-varying signal and label "2" to the large band signal.

Fig. 3 presents a dolphin whistle segmentation. Given that the recording noise is not white, we limited the TFR to a frequency band of $[0.2, 0.34]$ in order to have approximatively a white noise. Six patterns are segmented, three of them having more than one label.

5. CONCLUSION

We showed that STFT real and imaginary parts of a deterministic signal embedded in a white Gaussian noise have two different Gaussian distributions. The variances depend on the (n, k) point of the STFT. An efficient estimator of the noise variance in the real part was proposed. This estimator is used in a new non-stationnary signal TFR segmentation, based on local statistics of the STFT. Exemples with a simulated signal and a dolphin whistle prove the efficiency of this approach. Current works shows that this new algorithm provides less false alarm patterns than [1].

6. REFERENCES

[1] C. Hory, N. Martin and A. Chehikian, "Spectrogram segmentation by means of statistical features of non-stationnary signal interpretation," *IEEE Transactions on Signal Processing*, vol. 50, no. 12, pp. 2915–2925, december 2002.

[2] C. Hory and N. Martin, "Maximum likelihood noise estimation for spectrogram segmentation control," in *Proceedings of ICASSP02*, Orlando, USA, 2002, pp. 1581–1584.

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

[4] P.E. Johnson and D.G. Long, "The Probability Density of Spectral Estimates Based on Modified Periododram Avergares," *IEEE Transactions on Signal Processing*, vol. 47, may 1999.

[5] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, Wiley, 1973.

[6] J. F. Kenney and E. S. Keeping, *Mathematics of Statistics*, Van Nostrand, 1951.