MIXTURE DENSITIES FORMULATION OF A SPECTROGRAM SEGMENTATION TASK

Cyril Hory, Nadine Martin
Laboratoire des Images et des Signaux (LIS)
BP 46, 38402 Saint-Martin d'Hères Cedex France
Tel: 33 (0)4 76 82 64 23
e-mail: 1stname.2ndname@lis.inpg.fr

ABSTRACT

This communication addresses the general problem of non-stationary signal interpretation. We show that the Time-Frequency Representation segmentation task can be assimilated to a mixture densities parameter estimation when performed in a statistical Features Space. Two models of mixture are discussed, setting the number of features to be extracted. We propose to resolve moments equations to characterize the energy evolution of the signal components in the Time-Frequency plane. This provides quantitative measures associated to the components which can be used in a decision procedure.

1 Introduction

Time-Frequency Representations (TFR) are natural tools for the interpretation of a non-stationary signal because they highlight the temporal evolution of the signal spectral content. Many methods can be found in the literature. For instance, Altes extends in [1] the signal detection theory to the Time-Frequency domain. Prior to the decision task, Baraniuk and Jones propose to adapt the TFR according to the signal [2]. Davy and Doncarli realizes the classification by means of a fitted TFR kernel in [3].

All of these methods require a priori knowledge on the signal. Our purpose is to develop a procedure of interpretation taking into account the statistical properties of the TFR only, according to a general model of signal. It consists in detecting TFR structures containing non-stationary components energy that we call spectral patterns via a segmentation procedure. We adopt a local approach, perceptible to local non-stationarities, by extracting features from sets of coefficients defined by a gliding cell running over the whole TFR. The TFR coefficients are then represented in terms of their features in the so-called Features Space (FS). A segmentation procedure lead by the statistical properties of the TFR takes place in this new space. Describing the FS structure by extracting pertinent parameters can provide the necessary information for the interpretation step.

In this paper we focus on a particular TFR: the spectrogram. We show in section 2 that spectrogram distribution law can be seen as a mixture density. This leads to construct the segmentation procedure in section 3. In section 4 and 5, we describe how the segmentation algorithm behaves when respectively two and three statistical features chosen according to the mixture model are locally extracted. In final section we propose a first method to characterize the spectral patterns content in terms of the chosen features.

2 Mixture densities formulation

Statistical properties of the Spectrogram are derived from the huge knowledge addressing to periodogram. A complete study of the periodogram, defined as the square modulus of the Fourier transform of a signal, can be found in [9] when the analyzed signal is a white Gaussian process. The case of a stationary process is also derived in [8] but with a matricial approach. This theory broaden the case of a non-stationary time-dependent discrete signal $s[n]$ defined as:

$$s[n] = d[n] + b[n], \tag{1}$$

where $b[n]$ is a white Gaussian process of zero-mean and unknown variance σ^2. Signal $d[n]$ is a deterministic signal which components, represented by spectral patterns in the TF plane, are to be characterized. The spectrogram $S[n,k]$, defined as the square modulus of the short time Fourier transform of the signal computed over a N_w-points window $w[n]$:

$$S[n,k] = \frac{1}{N_w}\left|\sum_{i=1}^{N_w} w[i-n]s[i]e^{-i2\pi\frac{ik}{N_w}}\right|^2, \tag{2}$$

where n and k are the time and frequency indices, has non-central χ^2 distribution with two degrees of freedom, proportionality parameter $\frac{\sigma^2}{2}$ and non-centrality parameter $ncp = D[n,k]$:

$$f_{ncp}(x) = f_{\chi^2(2,ncp,\sigma^2/2)}(x). \tag{3}$$

The non-centrality parameter is the spectrogram coefficient of the deterministic signal alone. Spectrogram coefficients which do not contain deterministic signal energy have a null ncp and can be distinguished as having

a central χ^2 distribution.
When the analyzed signal is a deterministic signal embedded in a white Gaussian noise, the spectrogram distribution is a mixture of central and non-central χ^2 distributions. The segmentation problem can then be reoriented towards the estimation of the mixture parameters σ^2 and $D[n,k]$.

3 Spectrogram Segmentation

We propose to identify the deterministic components of the signal in the Time-Frequency space by discriminating coefficients of different non-centrality parameters. For this purpose features are extracted from local sets of spectrogram coefficients, able to describe the random behavior of their samples in the Features Space (FS).

3.1 Local Mixture Densities

Define a gliding cell C of N spectrogram coefficients running over the whole time-frequency plane. The distribution $f(x)$ of the parent random variable associated to C is a mixture of N χ^2 densities $f_{D[n,k]}(x)$ defined by (3):

$$f(x) = \frac{1}{N} \sum_{(n,k)\in C} f_{D[n,k]}(x), \tag{4}$$

with N unknown parameters $D[n,k]$. The method consists in associating characteristic features to the central point of each cell C. Assuming ergodicity, we define the features as statistics of the spectrogram coefficients inside C:

$$m_q = \frac{1}{N} \sum_{(n,k)\in C} S[n,k]^q, \tag{5}$$

in order to estimate the qth moments to zero μ_q of the parent variable of C. The first and second order moments of m_q are expressed in terms of the μ_q's:

$$E\{m_q\} = \mu_q, \tag{6a}$$
$$Var\{m_q\} = \frac{1}{N}(\mu_{2q} - \mu_q^2). \tag{6b}$$

These expressions provide a description of the FS in terms of the point location and dispersion.
For instance, we call noise cell a cell of the spectrogram containing only noise energy. Its parent variable is a $\chi^2(2, 0, \frac{\sigma^2}{2})$ variable. Its qth moment to zero, written μ_q^b, is:

$$\mu_q^b = q!(\sigma^2)^q. \tag{7}$$

Equations (6) and (7) show that noise cells aggregate in the FS around point $(\mu_1^b, \mu_2^b, ..., \mu_q^b)$ with a dispersion along the qth axis given by $Var\{m_q^b\} = ((2q!) - (q!)^2)(\sigma^2)^{2q}/N$.
The behavior of cells which contain deterministic components coefficients[1] also depends on the non-null non-centrality parameters. We assume hypothesis on their variations in order to reduce the number of unknown parameters in equation (4). These hypothesis define the mixture model of the cell which leads to the segmentation procedure. We derive two cases in sections 4 and 5.

[1] We call deterministic components coefficients the spectrogram coefficients which contain the deterministic component and the noise contributions in opposition to noise coefficient which contain the only noise contributions.

3.2 Region Growing Algorithm

The segmentation procedure we define takes into account the structure of the FS described by equations (6). A complete description of the algorithm is proposed in [6] and [5]. It is a region growing algorithm which consists in determinating in the FS spectrogram coefficients called seeds which are characteristic of a spectral pattern. A label is associated to the seeds. These seeds contaminate their neighbors by associating the same label. Our approach ensures that coefficients with same label present similar properties regarding to the extracted features.
Coefficients with the highest distance to the noise cluster are determined as seeds. In the FS, the distance $d_{i,j}$ between coefficients $S(n_i, k_i)$ and $S(n_j, k_j)$ is the Mahalanobis distance defined by [4]:

$$d_{i,j}^2 = (m_1^i, m_2^i, ..., m_q^i)\Sigma^{-1}(m_1^j, m_2^j, ..., m_q^j)^T, \tag{8}$$

where m_q^i is the qth features extracted from the cell centered on (n_i, k_i), $.^T$ stands for the matrix transposition and Σ is the diagonal matrix of the features variances. This distance takes into account the different variabilities of the FS axis. Note also that parameter σ^2 is not considered as a mixture parameter but is iteratively estimated by maximum likelihood estimator of a central χ^2 distribution, assuming that each unlabeled coefficient is a noise coefficient. Under this assumption the white Gaussian noise variance σ^2 is efficiently estimated by the empirical mean of m_1 [5]. The segmentation stops when the likelihood of the estimated central χ^2 converges. The two next sections illustrate the algorithm with mixture models of respectively two and three unknown characteristic parameters.

4 Bi-dimensional Features Space

The segmentation purpose is to distinguish between coefficients having central distribution, that is noise coefficients, and non-central coefficients, that is deterministic components coefficients. The dimension of the FS on which is processed the segmentation is the number of unknown parameters of the mixture model. In this section we describe the bi-dimensional FS (m_1, m_2).
Let consider that C is filled of P deterministic component coefficients and $(N-P)$ noise coefficients. The local approach allows one to assume that the P coefficients have the same *ncp* we define as equal to the mean M_s of the P non-null *ncp*'s in the cell:

$$M_s = \frac{1}{P} \sum_{(n,k)\in C} D[n,k]. \tag{9}$$

The distribution $f(x)$ of equation (4) is then considered as the mixture of a central distribution $f_0(x)$ and a non-central distribution $f_{M_s}(x)$:

$$f(x) = (1-p)f_0(x) + pf_{M_s}(x), \qquad (10)$$

with $p = \frac{P}{N}$. Feature m_q is an unbiased estimator of μ_q of expression [6], [7]:

$$\mu_q = q!(\sigma^2)^q \left[1 + p \sum_{i=1}^{q} C_q^i \frac{1}{i!} r^q \right]. \qquad (11)$$

where $r = M_s/\sigma^2$ can be interpreted as a local SNR. The algorithm is applied to segment the spectrogram of figure 1(a). The analyzed signal is a sum of three nearby linear chirps which spectral pattern presents sharp edges, and of eleven truncated sines presenting smoother energy variations. The additive white Gaussian noise is zero-mean and of variance $\sigma^2 = 10.42$. Figure 1(b) shows that the algorithm detects two classes corresponding to spectral patterns of different energy behavior and a noise class. One can see with figure 1(c) that the clusters corresponding to the two classes overlap in the FS. The region growing approach allows to separate the clusters by introducing a neighboring constraint in the TFR. The shapes of the clusters depend on the local evolution of the mixture parameters. The noise cluster is centered near the origin. Fitting cluster i to an order 2 polynomial reg_i permits to quantify this evolution and to characterize the spectral patterns by assuming: $\mu_2 = reg_i(\mu_1)$. Figure 1(d) shows the total least-square regressions [4], reg_1 and reg_2 of clusters 1 and 2 . The polynomial regression of cluster 2 presents a smallest derivative near the noise mode (μ_1^b, μ_2^b) which is characteristic of a widely spread and smoothed pattern. Note to conclude that the iterative procedure of segmentation provides a final estimation of the noise power $\hat{\sigma^2} = 10.53$.

5 Three-dimensional Features Space

The mixture model presented in the previous section does not take into account the dispersion of the non-centrality parameters inside the cell. In this section, we propose to introduce a third mixture parameter to make the model more accurate.

Let approximate now the $P/2$ lowest non-centrality parameters of the cell by their mean M_{s-} and the $P/2$ other by their mean M_{s+}. The mixture model is then a mixture of the central distribution $f_0(x)$ and of two non-central distributions $f_{M_{s-}}(x)$ and $f_{M_{s+}}(x)$:

$$\begin{aligned} f(x) &= (1-p)f_0(x) \\ &+ \frac{p}{2}(f_{M_{s-}}(x) + f_{M_{s+}}(x)). \end{aligned} \qquad (12)$$

Following this model, moment μ_q of the parent variable depends on three mixture parameters p, $r_- = M_{s-}/\sigma^2$

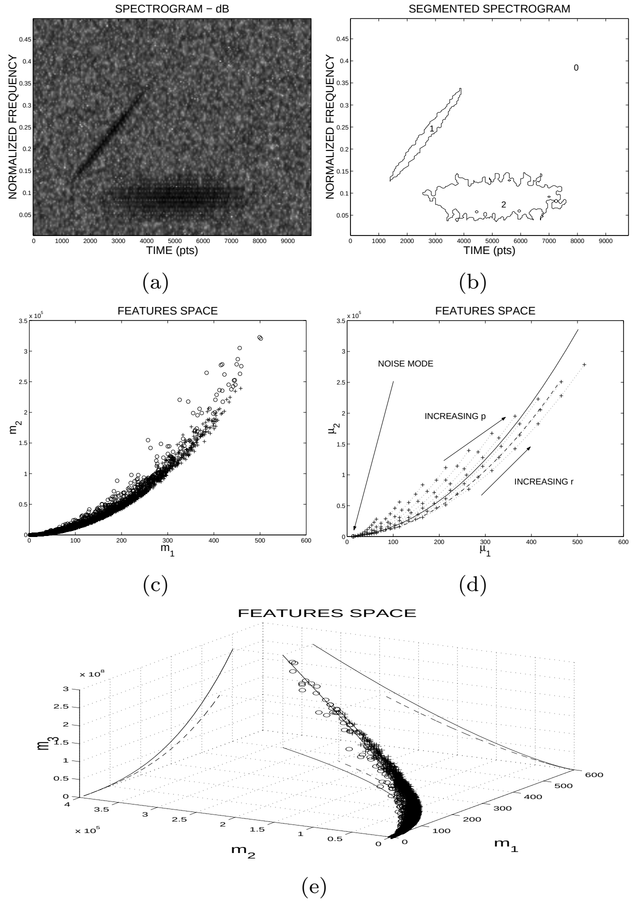

Figure 1: Sum of seven truncated sines and three linear chirps. The spectrogram (a) of 129×197 coefficients is segmented (b) in three regions with labels ranging from 0 to 2. Features are extracted from 5×7 coefficients cells. In figure (c) is presented the observed FS (m_1, m_2) with cluster 1 (o) and 2 (+). Figure (d) is the corresponding theoretical FS (dotted lines) with the polynomial regressions reg_1 (plain line) and reg_2 (dashed line). Figure (e) is the three-dimensional FS with polynomial regressions of the clusters and their projections to the three subspaces of the two polynomial regressions.

and $r_+ = M_{s+}/\sigma^2$:

$$\mu_q = q!(\sigma^2)^q \left[1 + \frac{p}{2} \sum_{i=1}^{q} C_q^i \frac{1}{i!} (r_-^q + r_+^q) \right]. \qquad (13)$$

The segmentation result is similar to the one obtained with the bi-dimensional FS but the spectral patterns shapes are now described by three parameters. Polynomial regressions of order two and three are respectively performed in subspaces (m_1, m_2) and (m_1, m_3) to fit the clusters. In next section we propose a method to estimate the mixture parameters from the polynomial regressions.

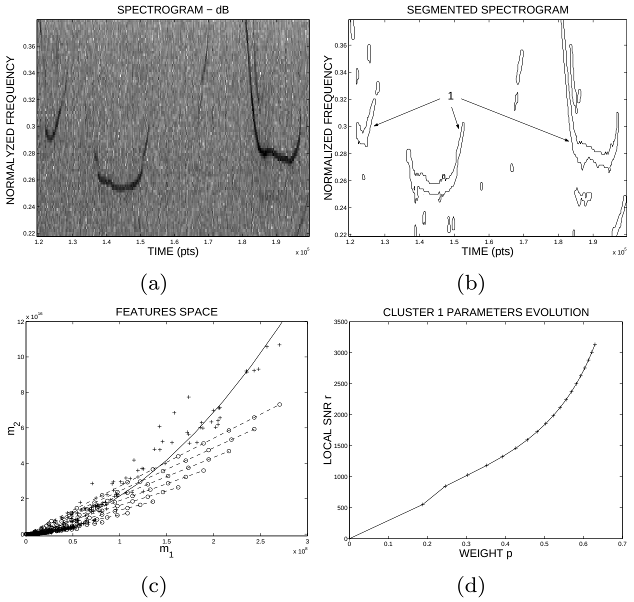

Figure 2: Dolphin whistle. The spectrogram (a) is segmented in two classes (b). First one is composed of three narrow spectral patterns of high energy density. Label 2 is associated to lower energy patterns. The cell contains 3×5 points. In figure (c) a theoretical sample of points $(\mu_1(p,r), \mu_2(p,r))$ with varying p and r (o) is superimposed to the Features extracted from class 1 patterns (+) and to the total least squares polynomial regression of this cluster (plain line). Figure (d) shows the evolution of the estimated parameters p and r.

6 Equating the moments

Let focus on the bi-dimensional FS (m_1, m_2). The polynomial regression can be considered as an estimation of the average evolution of moments μ_1 and μ_2 in terms of p and r. Equation (11) with $q = 1$ and $q = 2$ gives:

$$\mu_1 = \sigma^2(1 + pr), \quad (14a)$$

$$\mu_2 = reg(\mu_1) = 2(\sigma^2)^2(1 + 2pr + \frac{p}{2}r^2). \quad (14b)$$

Equating this system of moment equations provides an estimation of the average evolution of p and r and thus characterizes the energy evolution of the related cluster. Let consider the spectrogram of an acoustic recording of dolphin whistles presented in figure 2(a). The algorithm discriminates two classes of whistles spectral patterns in terms of their energy level mainly, and a noise class, as can be seen on figure 2(b). Let focus on the higher energy class. It is composed of three different thick spectral patterns with sharp edges. Figure 2(c) shows the theoretical grid (μ_1, μ_2) superimposed to the polynomial regression $\mu_2 = reg(\mu_1)$. The variations of parameters p and r presented in figure 2(d) are estimated by equating system of equations (14). It shows that the local SNR r increases rapidly while the proportion p of deterministic components increases. This is characteristic of a sharp edge spectral pattern. Moreover p is inferior to 0.7 which shows that these patterns are thick regarding to the cell size.

7 Conclusion

This paper describes a new method for non-stationary signal interpretation via its spectrogram segmentation. It is based on a mixture density modeling of the spectrogram coefficients distribution laws. This modeling allows to describe theoretically a Features Space on which is projected the spectrogram to perform the segmentation. We show how a moment method leads to estimate the mixture parameters associated to the segmented patterns in a post-segmentation processing. The procedure provides an estimation of the embedding noise power and a characterization of the deterministic components of the signal, whatever the signal is (frequency or amplitude modulated, narrow-band or wide-band, multicomponent signal). An example is given with an acoustic recording of a dolphin whistle, modeling the spectrogram coefficients distribution by a mixture of two distributions. Works are in progress to extend the post-processing method to Features Space of higher dimensions and obtain a more accurate description of the patterns energy evolution.

References

[1] R. A. Altes. Detection Estimation and Classification with Spectrograms. *Journ. Acoust. Soc. Am.*, **67**(4):1232–1246, 1980.

[2] R. G. Baraniuk and D. L. Jones. A Signal Dependent Time-Frequency Representation: Optimal Kernel Design. *IEEE Trans. on Signal Processing*, **41**(4):1589–1601, April 1993.

[3] M. Davy and C. Doncarli. Optimal kernels of t-f representations for signal classification. In *IEEE Int. Symp. on TF and TS analysis*, pages 581–584, Pittsburgh, USA, 1998.

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

[5] C. Hory and N. Martin. Maximum likelihood noise estimation for spectrogram segmentation. In *Proceedings of ICASSP02*, Orlando, USA, June 2002.

[6] C. Hory, N. Martin, and A. Chehikian. Spectrogram segmentation by means of statistical features for non-stationnary signal interpretation. *Submitted to IEEE Trans. on signal proc.*, July 2001.

[7] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions*, volume II. Wiley and sons, second edition, 1995.

[8] P. E. Johnson and D. G. Long. The probability density of spectral estimates based on modified periodogram averages. *IEEE trans. on signal processing*, 47(5):1255–1261, May 1999.

[9] L. H. Koopmans. *The spectral analysis of time series.* Academic press, 1974.