

# Modèle de Représentations Temps-Fréquence pour une interprétation statistique

Cyril HORY<sup>1</sup>, Nadine MARTIN<sup>2</sup>

<sup>1</sup>Department of Electrical Engineering and Computer Science,  
1301 Beal Avenue Ann Arbor, Mi. 48109 USA.

<sup>2</sup>Laboratoire des Images et des Signaux (LIS)  
BP 46 38402 Saint-Martin d' Hères Cedex France  
hory@eecs.umich.edu, nadine.martin@lis.inpg.fr

**Résumé** – Cette communication présente une méthode de segmentation et de caractérisation d'une représentation temps-fréquence permettant l'interprétation du contenu temps-fréquence du signal analysé. La méthode exploite l'estimation fournie par une version de l'algorithme EM adapté à un mélange de lois du  $\chi^2$  centré et décentré qui permet la caractérisation. La segmentation est réalisée grâce à un test du maximum *a posteriori* puis affinée par une technique de contours actifs statistiques.

**Abstract** – This communication presents a method of segmentation and characterization of a Time-Frequency Representation for the interpretation of the time-frequency content of the analyzed signal. The method exploits the estimation provided by a version of the EM algorithm adapted to a mixture of central and non-central  $\chi^2$  distributions. The segmentation is provided by a maximum a posteriori test extended by an active contours technique.

## 1 Introduction

L'analyse temps-fréquence fournit une information sur l'évolution temporelle du contenu fréquentiel d'un signal. Cette étape conduit à la construction de la Représentation Temps-Fréquence (RTF) qu'il s'agit alors d'interpréter par son contenu.

Une approche descriptive considère la RTF comme composée d'un ensemble de structures énergétiques qu'on appellera motifs spectraux. Interpréter la RTF signifie extraire ces motifs et à les caractériser. Michel *et al.* ont par exemple recours à des graphes minimaux pour extraire le squelette des motifs [10]. La caractérisation consiste dans ce cas à estimer une loi de variation du contenu fréquentiel au cours du temps. Coates et Fitzgerald proposent de caractériser le contenu énergétique du motif extrait en le modélisant par une somme pondérée de fonctions gaussiennes [2]. Cette approche s'appuie sur une hypothèse très forte quant au contenu énergétique des composantes.

Nous suivons ici le même objectif, en partant d'un modèle de loi de distribution pour les coefficients temps-fréquence déduit du processus de construction de la RTF. Considérons le signal temporel discret  $x[m]$ , somme d'une composante déterministe  $d[m]$  et d'un bruit blanc  $b[m]$  de moyenne nulle et

de variance  $\sigma^2$

$$x[m] = d[m] + b[m]. \quad (1)$$

On peut montrer en étendant les résultats concernant le périodogramme, que les coefficients  $S_x[n, k]$  de la RTF définie comme le module carré de la projection du signal sur une base de fonctions analysantes  $\Phi_{n,k}$ :

$$S_x[n, k] = |\langle x, \Phi_{n,k} \rangle|^2, \quad (2)$$

suivent une loi du  $\chi^2$  centré s'ils ne portent que l'énergie de la composante aléatoire  $b[m]$  et une loi du  $\chi^2$  décentré dans le cas contraire. L'estimation des paramètres définissant ces lois permet la caractérisation du contenu énergétique temps-fréquence du signal analysé. L'expression (2) définit le spectrogramme du signal  $x[m]$ , module carré de la Transformée de Fourier à Court Terme. Cette définition s'applique aussi au scalogramme, module carré de la Transformée en Ondelettes Continue de  $x[m]$ . On parle plutôt de Représentation Temps-Echelle dans ce cas puisque la grandeur d'analyse n'est plus la fréquence mais l'échelle de la fonction analysante  $\Phi_{n,k}$  relative à la largeur de son enveloppe [11]. Dans la section 2 nous présentons le modèle de distribution en mélange de lois du  $\chi^2$  centré et décentré fournissant une définition statistique d'un motif spectral. Ce modèle permet de placer le problème

d'interprétation dans un contexte de classification de données issues d'un mélange de lois [1]. Nous présentons dans la section 3 l'algorithme d'interprétation. Il s'appuie sur une version de l'algorithme d'Expectation-Maximization (EM) spécifique au mélange de lois du  $\chi^2$  [6], [7]. Nous proposons, dans la section 4, une procédure d'affinement de la segmentation obtenue par un test du maximum *a posteriori*. Le taux d'erreur de segmentation est réduit par application d'une technique de contours actifs statistiques dont le terme d'attache aux données n'est défini que par la vraisemblance du bruit. Enfin nous concluons cette communication en présentant et en commentant un exemple d'interprétation de spectrogramme.

## 2 Modèle statistique

Considérons la RTF du signal (1) composée de  $I$  motifs spectraux non-recouvrants. Le processus de segmentation associé à chacun des motifs spectraux une classe  $\mathcal{C}_i$ ,  $i \in [0, I]$  de coefficients temps-fréquence identifiée par une loi de distribution paramétrée. La classe  $\mathcal{C}_0$  est réservée aux coefficients ne portant que l'énergie du bruit. Le nombre  $I$  est appelé ordre de la RTF. La caractérisation consiste à estimer les paramètres des lois associées.

Une RTF d'ordre  $I$  non-nul est composée de  $P$  coefficients suivant une loi du  $\chi^2$  décentré et de  $N - P$  coefficients suivant une loi du  $\chi^2$  centré à deux degrés de liberté et de même coefficient de proportionnalité  $\sigma^2/2$  [8]. Les  $P$  lois du  $\chi^2$  décentré diffèrent par leur paramètre de décentrage  $S_d[n, k]$  non-nul. C'est le coefficient de la RTF du signal déterministe non-bruité. Il caractérise le contenu énergétique du motif spectral dans le plan temps-fréquence. Afin de réduire le nombre de paramètres du modèle, nous faisons l'hypothèse que l'énergie des motifs spectraux varie lentement suivant les axes temporels et fréquentiels. Nous supposons donc que l'ensemble des  $P_i$  coefficients temps-fréquence formant le  $i$ -ième motif spectral suivent une loi du  $\chi^2$  décentré de même paramètre de décentrage  $S_i$ , moyenne des  $S_d[n, k]$  sur le motif. Tout coefficient  $S_x[n, k]$  est alors considéré comme issu d'une loi de la forme :

$$f_{S_x[n, k]}(s) = (1 - p)f_{\chi^2(\sigma^2/2, 2)}(s) + \sum_{i=1}^I p_i f_{\chi^2(\sigma^2/2, 2, S_i)}(s), \quad (3)$$

avec  $p = \frac{P}{N}$ ,  $p_i = \frac{P_i}{N}$  et  $\sum_i p_i = p$ .

Le théorème de Paley-Wiener montre qu'il n'existe pas de motif spectral à support compact dans le plan temps-fréquence. La notion de segmentation en termes de traitement bas-niveau de la pseudo-

image temps-fréquence perd alors son sens. En considérant tout coefficient de décentrage  $S_i$  inférieur à la puissance du bruit comme nul, le modèle (3) définit chaque motif spectral sur un support fini.

L'objectif de la caractérisation est l'estimation de la puissance du bruit  $\sigma^2$ , des  $I - 1$  paramètres de décentrage  $S_i$  et des  $I - 1$  proportions  $p_i$ . L'estimation de ces  $2I - 1$  paramètres fournit une description de la RTF basée sur le modèle (3) dont la pertinence est discutée dans cette communication.

## 3 Interprétation

L'estimation des paramètres du mélange est réalisée par maximisation de la vraisemblance des données. L'algorithme EM fournit une procédure analytique de résolution de ce problème lorsque les lois impliquées dans le mélange sont de forme exponentielle [3], [5]. C'est le cas de la loi du  $\chi^2$  centré. En écrivant la loi du  $\chi^2$  décentré comme une somme infinie de lois du  $\chi^2$  centré [9], le mélange (3) prend la forme, pour  $I = 1$  :

$$f_{S_x[n, k]}(s) = \sum_{j=0}^{\infty} \beta_j f_{\chi^2(\sigma^2/2, 2+2j)}(s), \quad (4)$$

où  $\beta_j = \sum_{i=0}^I p_i \gamma_{ij}$ ,  $p_0 = 1 - p$  et  $\gamma_{ij}$  est la probabilité pour qu'une variable aléatoire suivant une loi de Poisson de paramètre  $S_i/\sigma^2$  soit égale à  $j$ . L'algorithme EM appliqué à ce mélange de lois exponentielles réalise l'estimation itérative des paramètres jusqu'à convergence vers un maximum local de la vraisemblance [6], [7]. Un motif est extrait à partir d'un test du Maximum *a posteriori* et l'algorithme est de nouveau appliqué sur les données non-extraites. L'algorithme EM est ensuite appliqué à chaque ensemble de données  $\mathcal{C}_0 \cup \mathcal{C}_i$  précédemment identifié afin d'améliorer l'estimation des paramètres du mélange (3), c'est à dire la caractérisation énergétique de chaque motif.

La statistique  $\nu_2 = \mu_2 - 2\mu_1^2$  du mélange (4), où  $\mu_i$  est le moment non-centré d'ordre  $i$ , permet de contrôler l'application de l'algorithme d'interprétation. En effet, cette statistique est nulle pour un mélange d'ordre 0. Elle est utilisée pour définir un critère d'arrêt de la segmentation. D'autre part,  $\nu_2$  est strictement positive si  $p_1 < 1/2$  et strictement négative si  $p_1 > 1/2$ . Le contrôle du signe de  $\nu_2$  permet de réduire l'espace des paramètres à l'un des deux sous-espaces séparés par le sous-espace  $p_1 = 1/2$  à chaque extraction de motif. Le risque de convergence vers un maximum local de la vraisemblance est ainsi réduit par une initialisation de l'algorithme proche de la solution.

## 4 Affinement de la segmentation

L'extraction des motifs spectraux est réalisée grâce à un test du Maximum *a posteriori*. Comme tout test bayésien, la probabilité de non-détection n'est pas nulle. Il en résulte une segmentation morcellée (voir figure 1). Nous proposons de remédier à ce phénomène en propageant les motifs par contours actifs statistiques. La méthode proposée par Germain [4], associe une loi de distribution exponentielle à chacune des classes existantes et fait évoluer la population de chaque classe par maximisation de la vraisemblance. Dans le cas de la segmentation d'une RTF, les équations de la vraisemblance impliquant les lois du  $\chi^2$  décentré ne peuvent être résolues que numériquement. Nous choisissons d'imposer une contrainte aux contours basée uniquement sur la maximisation de la vraisemblance normalisée des coefficients de la classe  $\mathcal{C}_0$ . A cette énergie externe uni-latérale est associée une énergie interne assurant la régularité des contours. Il s'agit d'un critère empirique d'homogénéité locale défini sur le support du noyau  $\mathcal{N}_{n,k}[n', k']$  de la RTF en chaque point  $(n, k)$ . Ce noyau est défini comme la RTF de la fonction analysante elle-même. C'est une mesure de l'échange d'énergie entre coefficients temps-fréquence [11]. La largeur du support du noyau varie de manière inverse en temps et en fréquence dans le cas du scalogramme alors qu'il est constant pour le spectrogramme, du fait de la résolution temps-fréquence. Un coefficient  $S_x[n, k]$  n'est reconnu susceptible de changer de classe qu'à la condition que le changement de classe de l'ensemble  $\mathcal{V}_{n,k} = \mathcal{C}_0 \cup \{(n', k'), \mathcal{N}_{n,k}[n', k'] > 0\}$  fasse augmenter la vraisemblance normalisée de la classe  $\mathcal{C}_0$ . L'introduction de ce critère empirique dans l'expression de la fonction d'énergie à maximiser conduit au critère de changement de classe :

$$S_x[n, k] \in \mathcal{C}_i \quad \text{ssi} \quad \overline{\mathcal{C}_0} < \overline{\mathcal{C}_0 \setminus \mathcal{V}_{n,k}}, \quad (5)$$

où la barre supérieure indique la moyenne des coefficients sur l'ensemble, statistique exhaustive de la loi du  $\chi^2$  centré. Notons que la définition d'une énergie externe unilatérale a tendance à privilégier la propagation des motifs. L'évolution des contours est arrêtée lorsque le rapport entre nombre de points formant le motif et nombre total de points de la RTF est égal à la proportion estimée pour la classe associée.

## 5 Exemples et conclusion

La figure 1 illustre les étapes successives de la méthode proposée. La RTF présentée en figure 1(a), est le spectrogramme d'une somme de trois com-

posantes aux propriétés énergétiques différentes. Le résultat de la segmentation après application du test du maximum *a posteriori* pour chaque loi du  $\chi^2$  décentré estimée est donné en figure 1(c). Cette segmentation a associé la classe 1 à la loi de variation de fréquence instantanée sinusoïdale dont le motif présente de grandes variations énergétiques. Le motif large-bande (classe 2) est plus homogène mais il présente une enveloppe gaussienne. Enfin le motif 3 est une modulation de fréquence linéaire d'amplitude constante. La propagation par contours actifs des motifs extraits conduit au résultat final présenté en figure 1(b). La figure 1(d) permet de comparer ce résultat à la segmentation de référence obtenue en seuillant la RTF de la composante déterministe non bruitée. Le seuil est la puissance du bruit additif. Les résultats de la caractérisation sont présentés dans le tableau 1. Ils sont comparés à la segmentation de référence.

Il apparaît que la méthode donne des résultats très satisfaisants tant au niveau de la segmentation que de l'estimation des paramètres lorsque l'énergie du motif extrait est homogène et ce même pour de faibles rapports signal-sur-bruit (motif 3). On entend ici par rapport signal-sur-bruit le rapport entre paramètre de décentrage à l'instant  $n$  et à la fréquence  $k$  et la puissance du bruit  $rsb[n, k] = S_d[n, k]/\sigma^2$ . En revanche, lorsque  $rsb[n, k]$  varie à l'intérieur du motif, ce qui est le cas pour la loi de variation de fréquence instantanée sinusoïdale ou la composante large-bande d'enveloppe gaussienne, l'algorithme sur-estime les paramètres de décentrage  $S_i$  moyens et sous-estime les proportions  $p_i$ . Il s'en suit une mauvaise détection des régions de faibles  $rsb[n, k]$  alors que l'application des contours actifs statistiques a permis la propagation des motifs. Ce comportement est à mettre en relation avec la sur-estimation de la puissance du bruit. Il semble que l'approximation de la loi de l'ensemble des coefficients d'un motif à une seule loi du  $\chi^2$  décentré trouve là ses limites. Des travaux sont en cours pour intégrer un algorithme EM appliqué à un mélange d'ordre  $I$  supérieur à 1 afin de prendre en compte la diversité énergétique de chaque motif.

## Références

- [1] G. Celeux. Classification et modèles. *R. S. A.*, **36**(4):43–58, 1988.
- [2] M. J. Coates and W. J. Fitzgerald. Regionally optimised time-frequency distributions using finite mixture models. *Signal Processing*, **77**(3):247–260, 1999.

- [3] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *J. R. Statist. Soc.*, **B-39**:1–37, 1977.
- [4] O. Germain. *Segmentation d'images radar : Caractérisation des détecteurs de bord et apport des contours actifs statistiques*. PhD thesis, Université Aix-Marseille III, Janvier 2001.
- [5] V. Hasselblad. Estimation of Finite Mixtures of Distributions from the Exponential Family. *J. Amer. Statist. Assoc.*, **64**:1459–1471, 1969.
- [6] C. Hory. *Mélanges de distributions du  $\chi^2$  pour l'interprétation d'une représentation temps-fréquence*. PhD thesis, Institut National Polytechnique de Grenoble, Décembre 2002.
- [7] C. Hory. Formulation of the EM algorithm for a mixture of central and non-central  $\chi^2$  distributions. *Stat. and Prob. Letters*, soumis.
- [8] C. Hory, N. Martin, and A. Chehikian. Spectrogram segmentation by means of statistical features for non-stationary signal interpretation. *IEEE Trans. on signal proc.*, **50**(12):2915–2925, Décembre 2002.
- [9] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions*, volume 2. Wiley and sons, second edition, 1995.
- [10] O. Michel, P. Flandrin, and A. Hero. Détection de structures dans le plan temps-fréquence à l'aide de graphes minimaux. In *17ème colloque GRETSI*, pages 713–716, Vannes, Fr., Septembre 1999.
- [11] B. Torrèsani. *Analyse continue par ondelettes*. InterEditions/CNRS Editions, 1995.

	Référence	Estimé	
Décentrage $S_i$			
Motif	1	75.04	94.37
	2	60.27	74.22
	3	53.09	49.6
Puissance du bruit $\sigma^2$			
	10.41	11.14	
Proportions $p_i$			
Motif	1	0.036	0.022
	2	0.06	0.038
	3	0.018	0.02

TAB. 1: Résultats de la caractérisation

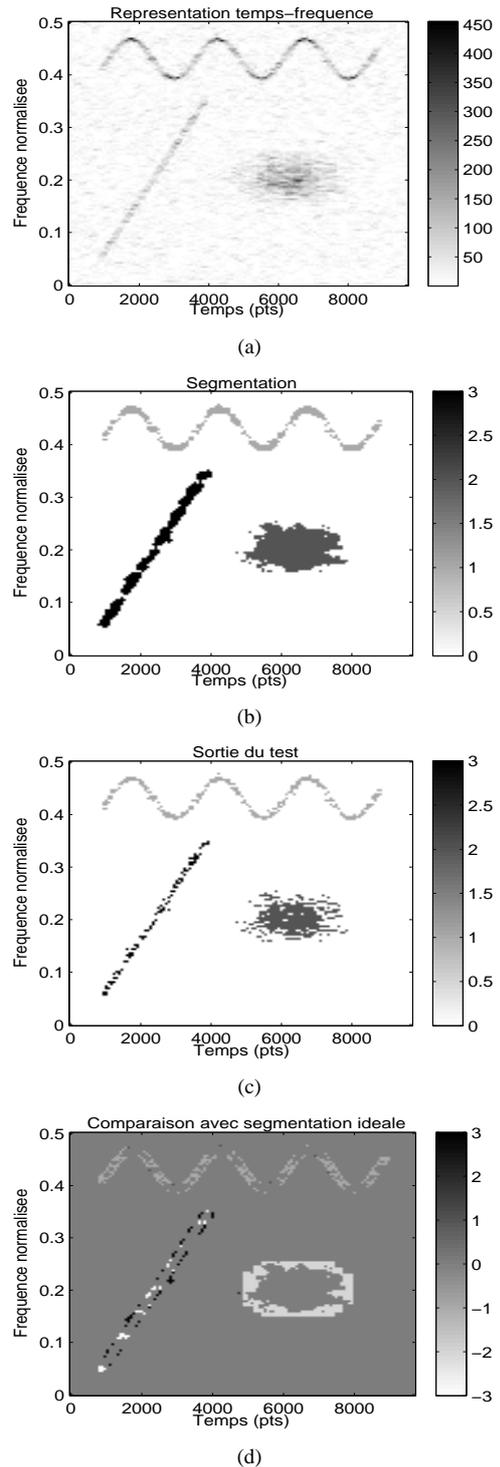


FIG. 1: Le résultat de la segmentation du spectrogramme (a) est présenté en figure (b). L'apport des contours actifs statistiques pour propager les motifs extraits est observé en comparant (b) et la sortie du test (c). En figure (d), le résultat de la segmentation est superposé à la segmentation de référence. Les coefficients négatifs indiquent les points non-segmentés et les coefficients positifs indiquent les fausses alarmes.