

# Remarques sur l’adaptativité des représentations temps-fréquence

Florent JAILLET<sup>1,2</sup>, Bruno TORRÉSANI<sup>1</sup>

<sup>1</sup>Laboratoire d’Analyse, Topologie et Probabilités  
CMI, 39 rue Joliot-Curie, 13453 Marseille Cedex 13

<sup>2</sup>GENESIS S.A.

Batiment Beltram, Domaine du Petit Arbois - BP 69, F-13545 Aix en Provence cedex 4  
florent.jaillet@genesis.fr, Bruno.Torresani@cmi.univ-mrs.fr

**Résumé** – Nous nous intéressons au problème de l’adaptation au signal analysé de la fenêtre d’une représentation de Fourier à court terme dans deux cas de figure: adaptations *globale* et *locale*. Nous montrons que la représentation “optimale” dépend fortement du critère choisi dans le cas de l’optimisation globale, et qu’une optimisation locale peut permettre de pallier ce problème, au moins dans certaines situations. L’intérêt de telles techniques est illustré par un exemple de séparation de composantes.

**Abstract** – We consider the problem of window optimization for short time Fourier transform in two different frameworks: *global* and *local* optimization. We show that in the global situation, where the window is adapted to the complete signal, the “optimal” representation depends strongly on the chosen criterion. This suggests to turn to procedures which adapt the window to specific “components” of the signal, and we show that such approaches overcome at least partially the above mentioned shortcoming. The interest of such techniques is illustrated by an example of separation of components.

## 1 Introduction

Il est bien connu que les représentations temps-fréquence ne permettent de donner qu’une image relativement “subjective” d’un signal, dans la mesure où celle-ci dépend de choix préalables (type de représentation, fenêtres, lissages,...). Se pose donc le problème de la détermination de la “meilleure” représentation pour un signal donné ou une classe donnée de signaux, si tant est que cette notion ait un sens. Dans un certain nombre de situations simples, comme par exemple la représentation de Wigner-Ville pour les chirps linéaires, la meilleure représentation est connue (voir par exemple [1, 4]). Il ne s’agit cependant que de cas très particuliers, difficilement généralisables à des situations plus complexes.

Nous nous intéressons ici à une approche adaptative, en nous limitant au cas de représentations “atomiques” de type Gabor ou ondelettes, que nous cherchons donc à adapter à un signal donné, en fonction d’un critère à optimiser. Plus précisément, nous considérons deux approches pour la sélection de la fenêtre optimale dans le cas d’une représentation de Fourier à court terme : l’une exploitant des critères d’adéquation de la fenêtre au signal étudié (au sens développé notamment dans [6]), l’autre des critères de parcimonie de la représentation (dans l’esprit des algorithmes de “best basis” [7]). Dans un cadre comme dans l’autre, le domaine d’applicabilité de telles approches s’avère confiné à des classes de signaux extrêmement simples. Ceci apparaît clairement sur quelques exemples de signaux “académiques”, pour lesquels un calcul exact montre une dépendance nette de la fenêtre dans le critère

choisi, ainsi que sur des signaux réels complexes. Ceci nous conduit à proposer des variantes “localisées”, qui adaptent la représentation à des composantes isolées du signal analysé. L’intérêt de telles variantes est illustré par des exemples numériques sur des signaux réels.

Dans un cadre d’analyse et de synthèse de signaux audiophoniques, ce programme est également motivé par la perspective de définir (et estimer) des “objets sonores”, dans l’esprit des représentations hybrides proposées dans [2].

## 2 Critères d’optimalité

Pour simplifier, nous nous placerons dans cet article dans le cadre de signaux fonctions d’une variable continue, et nous nous focaliserons sur la représentation de Fourier à court terme  $x \in L^2(\mathbb{R}) \rightarrow X_g \in L^2(\mathbb{R}^2)$ , définie par

$$X_g(b, \omega) = \int_{-\infty}^{\infty} x(t) \bar{g}(t-b) e^{-i\omega(t-b)} dt$$

( $g \in L^2(\mathbb{R})$  étant la fenêtre d’analyse). L’adaptation de la représentation de Fourier à court terme est effectuée en sélectionnant parmi une famille de fenêtres données celle qui optimise un critère donné d’optimalité. On a souvent recours aux deux familles suivantes de critères, qui exploitent deux idées différentes : l’adéquation (“matching”) de la fenêtre au signal analysé, ou la parcimonie (“sparsity”) de la représentation obtenue.

## 2.1 Critères d'adéquation

La fenêtre  $g$  est sélectionnée de manière à ce que son "étalement" dans le plan temps-fréquence soit le plus proche possible de celui du signal. Comme l'a montré W. Kozek [6], cette démarche aboutit à une relation liant la fonction d'ambiguïté de la fenêtre à celle du signal analysé. Dans les formulations les plus simples, la fenêtre optimale est solution d'un problème d'optimisation du type (voir par exemple [6])

$$g = \arg \max_{\gamma} \langle K * |A_x|^2, |A_{\gamma}|^2 \rangle ,$$

où  $A_x$  est la fonction d'ambiguïté du signal  $x$

$$A_x(b, \omega) = \int_{-\infty}^{\infty} x(t) \bar{x}(t-b) e^{-i\omega(t-b)} dt ,$$

$A_{\gamma}$  celle de la fenêtre  $\gamma$  et  $K$  un noyau de lissage bidimensionnel. Cette optimisation est alors effectuée numériquement. L'extension au cas des signaux  $x$  aléatoires s'effectue en substituant à  $A_x$  la "fonction d'ambiguïté moyenne"  $\mathbb{E}\{A_x\}$ . Cette dernière doit généralement être estimée à partir d'une réalisation du signal.

## 2.2 Critères de parcimonie

On recherche ici la fenêtre qui rend la représentation la plus parcimonieuse, c'est à dire telle que l'énergie du signal soit concentrée sur un petit nombre de coefficients significatifs. Cette démarche est proche de celle utilisée dans des algorithmes de choix de meilleure base [7]. Les mêmes types de critères peuvent être utilisés, par exemple des critères d'entropie. Pour cela, on définit un spectrogramme normalisé  $\rho_{\gamma}$  du signal  $x$  par

$$\rho_{\gamma}(\tau, \omega) = \frac{|X_{\gamma}(\tau, \omega)|^2}{2\pi \|x\|^2 \|\gamma\|^2} ,$$

avec  $X_{\gamma}$  la transformée de Fourier à court terme du signal  $x$  calculée avec la fenêtre  $\gamma$ .  $\rho_{\gamma}$  est alors une densité de probabilités, et on introduit l'entropie de Shannon

$$S(\gamma) = - \iint \rho_{\gamma}(\tau, \omega) \log(\rho_{\gamma}(\tau, \omega)) d\tau d\omega ,$$

et la famille des entropies de Renyi par

$$R_{\alpha}(\gamma) = \frac{1}{1-\alpha} \log \left( \iint \rho_{\gamma}(\tau, \omega)^{\alpha} d\tau d\omega \right) , \quad \alpha \in (0, 1) .$$

La recherche de la fenêtre optimale  $g = \arg \min S(\gamma)$  ou  $g = \arg \min R_{\alpha}(\gamma)$  (à l'intérieur d'une famille paramétrée de fenêtres) est effectuée numériquement.

## 3 Optimisation globale

Il s'avère que cette approche est rapidement mise en difficulté lorsque la composition temps-fréquence du signal se complexifie, comme le montrent quelques calculs explicites dans les cas les plus simples. Par exemple, dans le cas d'un logon Gaussien  $x(t) = e^{i\lambda t} e^{-\frac{t^2}{2\sigma^2}}$  analysé avec une famille de fenêtres Gaussiennes d'écart-type  $s$  (donc  $\gamma(t) = \exp\{-t^2/2s^2\}$ ), le calcul explicite fournit

$$\rho(b, \omega) = \frac{1}{\pi} \frac{\sigma s}{\sigma^2 + s^2} \exp \left( -\frac{b^2 + \sigma^2 s^2 (\omega - \lambda)^2}{\sigma^2 + s^2} \right) ,$$

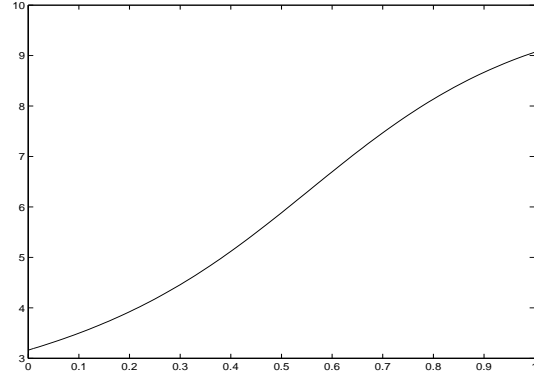


FIG. 1 – Influence du critère sur le choix de l'échelle optimale dans le cas mixte : valeur de l'échelle optimale pour l'entropie de Renyi en fonction de  $\alpha$ , pour deux logons d'échelles  $\sigma = 1$  et  $\sigma' = 10$

d'où on déduit aisément

$$R_{\alpha}(\gamma) = -\frac{\log(\alpha)}{1-\alpha} + \log(\pi) + \log \left( \frac{\sigma}{s} + \frac{s}{\sigma} \right)$$

et

$$S(\gamma) = 1 + \log(\pi) + \log \left( \frac{\sigma}{s} + \frac{s}{\sigma} \right) .$$

On voit apparaître dans tous les cas la fonction  $u \rightarrow u + 1/u$ , qui est minimale en  $u = 1$ , ce qui montre que dans ce cas d'école, la fenêtre optimale est effectivement celle dont l'échelle  $s$  coïncide avec l'échelle  $\sigma$  de  $x$ , et ce, quel que soit le critère utilisé.

En revanche, dans un cas à peine plus complexe où le signal est la somme de deux logons d'échelles différentes (localisés dans des régions éloignées du plan temps-fréquence),  $x(t) = e^{i\lambda t} e^{-\frac{(t-b)^2}{2\sigma^2}} + e^{i\lambda' t} e^{-\frac{(t-b')^2}{2\sigma'^2}}$ , les différentes entropies de Renyi et de Shannon conduisent à des fenêtres "optimales" très significativement différentes. L'hypothèse de séparation des localisations temps-fréquence des deux logons permet d'approximer avec une bonne précision (à un terme d'erreur près, décroissant exponentiellement en fonction de  $b-b'$  et  $\lambda-\lambda'$ ) le spectrogramme de la somme par la somme des spectrogrammes, ce qui simplifie le calcul explicite des entropies. On obtient alors

$$R_{\alpha}(\gamma) = \log(\pi) - \frac{\log(\alpha)}{1-\alpha} - \frac{\alpha}{1-\alpha} \log(\sigma + \sigma') + \frac{1}{1-\alpha} \log \left( \sigma^{\alpha} \left( \frac{\sigma}{s} + \frac{s}{\sigma} \right)^{\alpha-1} + \sigma'^{\alpha} \left( \frac{\sigma'}{s} + \frac{s}{\sigma'} \right)^{\alpha-1} \right)$$

pour les entropies de Renyi, ainsi que

$$S = 1 + \log(\pi) + \frac{\sigma}{\sigma + \sigma'} \log \left[ \left( \frac{\sigma}{s} + \frac{s}{\sigma} \right) \left( 1 + \frac{\sigma'}{\sigma} \right) \right] + \frac{\sigma'}{\sigma + \sigma'} \log \left[ \left( \frac{\sigma'}{s} + \frac{s}{\sigma'} \right) \left( 1 + \frac{\sigma}{\sigma'} \right) \right] .$$

Ces expressions ne se prêtent pas facilement à une analyse théorique, mais sont aisées à étudier numériquement. Comme on peut le voir sur les résultats présentés en FIG. 1, la valeur de l'échelle optimale dépend maintenant du critère utilisé.

Cette situation se retrouve également dans des situations plus complexes où de plus grands nombres de logons "séparés" sont considérés. Le cas plus réaliste de logons "non-séparés" n'est accessible que numériquement, et conduit aux mêmes conclusions.

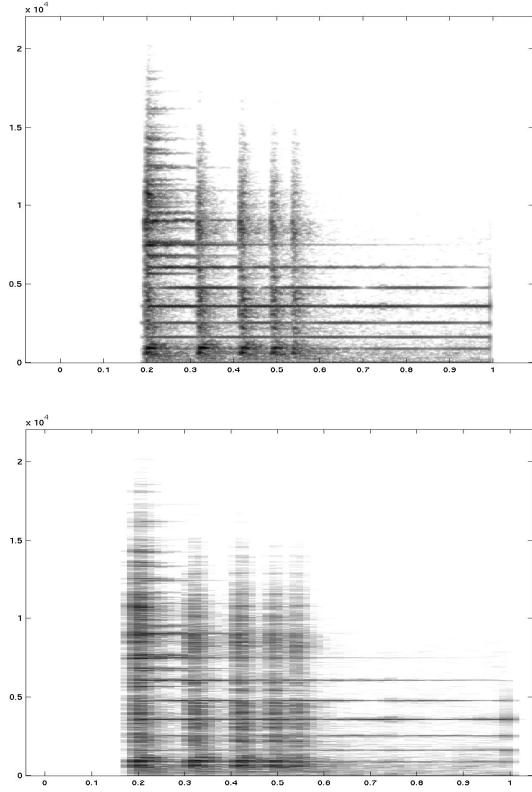


FIG. 2 – Optimisation globale de fenêtre : entropie de Rényi pour deux choix différents de  $\alpha$  ( $\alpha = 0.1$  et  $\alpha = 0.9$ )

Cet effet est confirmé par les tests sur des signaux réels illustrés en FIG. 2 (son de cloche frappée par un marteau qui rebondit) : la taille optimale de fenêtre dépend amplement du critère utilisé (ici l’entropie de Rényi, avec deux valeurs significativement différentes du paramètre  $\alpha$ ), dès que le signal contient des composantes ayant des caractéristiques temps-fréquence différentes. Comme on peut le constater, la fenêtre “optimale” résulte d’un compromis entre les différentes “composantes” présentes dans le signal, et ici de faibles valeurs de  $\alpha$  ont tendance à privilégier la localisation temporelle, alors que les grandes valeurs privilégient la localisation fréquentielle. Ceci limite donc l’intérêt d’une optimisation globale de la fenêtre pour ces signaux.

## 4 Optimisation locale dans le plan temps-fréquence

Lorsque l’optimisation globale n’a pas un sens bien défini, l’adaptation peut être rendue *locale* dans le plan temps-fréquence, limitant ainsi l’influence du critère utilisé.

Après identification (et sélection) d’une région d’intérêt  $\Omega$  du plan temps-fréquence, l’optimisation de fenêtre est effectuée en se limitant aux composantes correspondantes du signal. Plus précisément, étant donnée une fenêtre d’analyse initiale  $h$ , un signal correspondant  $x_{h,\Omega}$  est reconstruit par

$$x_{h,\Omega}(t) = \frac{1}{2\pi\|h\|^2} \int_{\Omega} X_h(b,\omega) e^{i\omega(t-b)} h(t-b) , dbd\omega ,$$

et la recherche de fenêtre optimale est effectuée comme

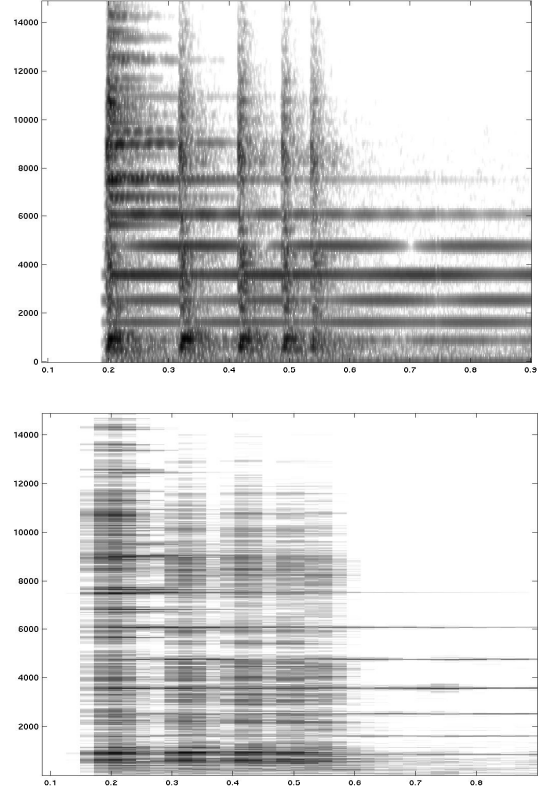


FIG. 3 – Optimisation locale de fenêtre : mise en évidence de structures transitoires et tonales (entropie de Shannon)

précédemment, mais à partir du signal  $x_{h,\Omega}$  cette fois. De nouveau, la fenêtre optimale pour la région  $\Omega$  est celle qui optimise le critère choisi, et fournit une représentation temps-fréquence pour le signal (complet)  $x$ , qui a été optimisée pour une composante donnée de  $x$ . On affine ainsi la représentation non seulement de la composante sélectionnée, mais aussi de toutes les autres composantes présentant des caractéristiques similaires. Dans toutes les situations où sont présentes des composantes “séparées”, des calculs similaires à ceux présentés dans la section précédente peuvent être effectués. Essentiellement, dès qu’une composante peut être isolée, les différents critères fourniront des résultats concordants pour la fenêtre optimale.

Cette technique a été mise en œuvre dans un logiciel, permettant à l’utilisateur de sélectionner graphiquement une région d’intérêt dans le plan temps-fréquence, et de déterminer la fenêtre optimale correspondante. L’exemple de la FIG. 3 (avec le même signal qu’en FIG. 2) présente les représentations ainsi obtenues avec une adaptation de la fenêtre à différentes régions (ici transitoire et tonale) du plan temps-fréquence.

## 5 Perspectives

L’approche locale présentée ici constitue une bonne alternative aux approches globales. Elle permet notamment d’envisager, via des resynthèses partielles de composantes que l’adaptation aura permis de définir précisément, des représentations adaptatives faisant intervenir plusieurs fenêtres simultanément (dans l’esprit des modèles hybrides de [2]). En cela, elle offre une alternative simple et intéres-

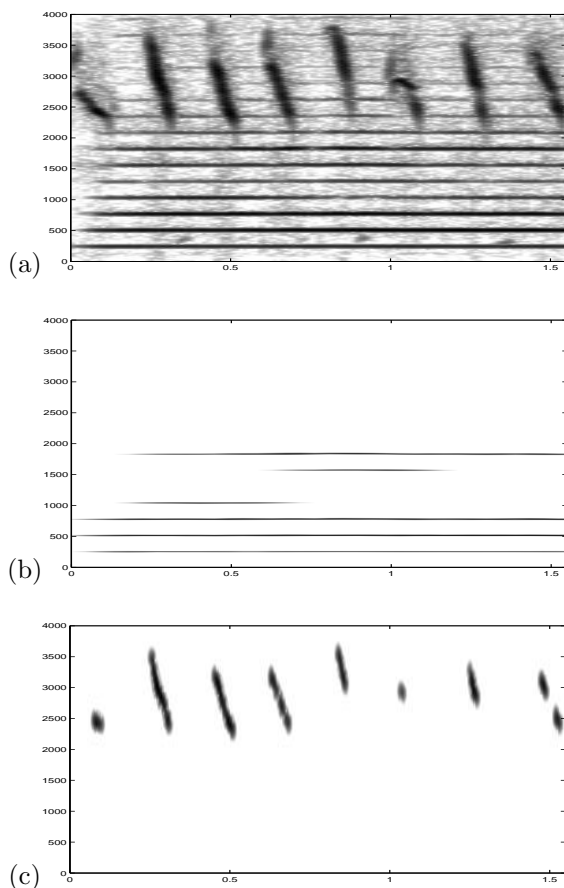


FIG. 4 – Adaptation locale pour la resynthèse partielle : (a) spectrogramme du son original (flûte + chant d’oiseau) ; (b) spectrogramme du son obtenu par resynthèse après adaptation à la partie tonale de la flûte puis seuillage ; (c) spectrogramme du son obtenu par resynthèse après adaptation au chant d’oiseau puis seuillage

sante aux représentations “multifenêtre” proposées notamment dans [3, 8, 9].

Elle est en particulier bien adaptée aux applications de transformation des sons par modification de leurs représentations temps-fréquences, puisqu’elle permet d’utiliser une représentation optimisée pour chacune des composantes à modifier.

En améliorant ainsi la résolution dans le plan temps-fréquence, on se donne également la possibilité d’améliorer la qualité de resynthèses partielles de diverses composantes du signal. On exploite ainsi la parcimonie de la représentation pour effectuer une séparation de source “unicapteur”.

L’exemple présenté en FIG. 4 illustre les possibilités offertes par l’adaptation locale pour isoler des composantes spécifiques. Les exemples sonores correspondants et les représentations couleur sont disponibles sur le site <http://www.cmi.univ-mrs.fr/~jaillet/GRETSI03>.

Le son original est composé d’une note de flûte et d’un chant d’oiseau. Sur la représentation du son original, les composantes tonale quasi-sinusoidales de la flûte sont bien identifiables sous forme de traits horizontaux, alors que le chant d’oiseau apparaît sous forme de composantes chirpées bien localisées entre 2000 et 4000 Hz.

On réalise une adaptation sur le fondamental du son

de flûte d’une part, et sur un des éléments du chant d’oiseau d’autre part. Par simple seuillage sur le module des coefficients des transformées obtenues, nous pouvons alors extraire les composantes pour lesquelles l’adaptation a été réalisée.

Ainsi sur la deuxième représentation de la FIG. 4, le signal reconstruit après seuillage de la transformée adaptée à la partie tonale de la flûte ne contient que les principaux partiels de la composante tonale du son de flûte et aucun élément du chant d’oiseau, alors que la troisième représentation montre qu’après adaptation sur le chant d’oiseau, on reconstruit uniquement des éléments du chant d’oiseau et aucune composante de la flûte.

Une alternative à ce type d’approche consiste à adapter dynamiquement les fenêtres dans des régions déterminées du plan temps-fréquence, par exemple via un premier pavage de ce dernier en “superpavés”, à l’intérieur desquels la fenêtre optimale peut être recherchée. Des résultats dans ce sens, dans un contexte de repères (or *frames*) de Gabor seront présentés dans [5].

## Références

- [1] L. COHEN. *Time-Frequency Analysis*. Prentice Hall, New York, 1994.
- [2] L. DAUDET et B. TORRÉSANI. “Hybrid Representations for Audiophonic Signal Encoding”. *Signal Processing*, 82(11) :1595–1617, 2002. Special issue on Image and Video Coding Beyond Standards.
- [3] M. DÖRFLER. “*Gabor Analysis for a Class of Signals called Music*”. PhD thesis, NuHAG, University of Vienna, 2002.
- [4] P. FLANDRIN. *Temps-Fréquence*. Traité des Nouvelles Technologies, série Traitement du Signal. Hermès, Paris, 1993.
- [5] F. JAILLET, B. TORRÉSANI et P. WOJDYLLLO. “An adaptive Gabor representation for hybrid signal models”. 2003. in preparation.
- [6] W. KOZEK. “*Matched Weyl–Heisenberg Expansions of Nonstationary Environments*”. PhD thesis, NuHAG, University of Vienna, 1996.
- [7] M. V. WICKERHAUSER. *Adapted Wavelet Analysis from Theory to Software*. AK Peters, Boston, MA, USA, 1994.
- [8] R. WILSON, A. CALWAY et E. PEARSON. “A generalized wavelet transform for Fourier analysis : the multiresolution Fourier transform and its application to image and audio analysis”. *IEEE Transactions on Information Theory*, 38(2) :674–690, March 1992.
- [9] M. ZIBULSKI et Y. ZEEVI. “Analysis of multiwindow gabor-type schemes by frame methods”. *Applied and Computational Harmonic Analysis*, 4(2) :188–212, 1997.