

ECOLE POLYTECHNIQUE

PROMOTION X2004

BOUDARD Léopold

RAPPORT DE STAGE D'OPTION SCIENTIFIQUE

*Adaptation des tailles de fenêtre pour les  
représentations temps fréquence*

Option : MATHEMATIQUES APPLIQUEES

Champ de l'option : Traitement du signal

Directeur de l'option : Emmanuel Bacry

Directeur de stage : Xavier Rodet – Axel Roebel

Dates du stage : 10/04/2007-20/07/2007

Adresse de l'organisme :

IRCAM – Institut de Recherche et de Coordination Acoustique Musique

1 place Igor Stravinsky

75004 PARIS

## RESUME

L'analyse spectrale des signaux non-stationnaires nécessite de mettre en oeuvre des représentations temps-fréquence spécifiques, s'adaptant aux caractéristiques locales des signaux. Il s'agit ici de déterminer quels critères sont les plus pertinents afin d'évaluer la localisation offerte par une distribution temps fréquence; avec en particulier une application sur un spectrogramme à fenêtres multiples.

## ABSTRACT

Spectral analysis of non-stationary signals calls for specific time-frequency representations, which adapt to time evolution of frequency characteristics. We determine which criteria is the more relevant to evaluate localisation given by a time frequency distribution; we propose an application with an adapted window spectrogram.

## Table des matières

Introduction.....	4
Représentations temps-fréquence.....	6
Transformée de Fourier à Court terme.....	6
Transformée en ondelettes.....	7
Sélection adaptative de la base.....	7
Distributions quadratiques.....	7
Distribution Wigner Ville.....	7
Fonction d'ambiguïté.....	8
Critères d'adaptation locale de la taille de fenêtre pour le spectrogramme.....	10
Critères entropiques, entropies de Rényi et Kurtosis.....	10
« Energy Smearing » - étalement d'énergie.....	11
Partition d'analyse et espace de fenêtres.....	11
Analyse numérique.....	12
Comparaison des critères sur des signaux.....	13
Transitoires de Dirac dans un signal harmonique.....	14
Pseudo-entropie de Rényi.....	14
Étalement d'énergie.....	15
Transitions d'harmoniques.....	16
Pseudo-entropie de Rényi.....	16
Étalement d'énergie.....	18
Modulation linéaire.....	19
Pseudo-entropie de Rényi.....	19
Vibrato.....	21
Pseudo-entropie de Rényi.....	21
Signaux bruités.....	22
Signaux réels.....	25
Adaptation par optimisation du noyau sur une distribution de Wigner Ville.....	27
Conclusion.....	28
Références.....	29

## Introduction

La localisation dans le plan temps fréquence est un des problèmes majeurs du traitement du signal, et plus particulièrement dans le cadre de l'analyse du signal musical pour des applications aussi variées que la détection des fondamentales, la distinction des timbres ou encore le codage. La transformée de Fourier à court terme, dont la résolution est définie par la fenêtre choisie, a été le premier outil d'analyse à introduire la notion de fréquence locale, en accord avec l'intuition physique. J. Ville a par la suite cherché une définition mathématique non ambiguë d'une fonction fréquentielle du temps; enfin Cohen, à partir des travaux de Ville, a introduit une nouvelle classe de distributions, généralisant celle du spectrogramme, qui ont été largement étudiées dans les années 80, notamment pour des signaux stationnaires.

Cependant, la recherche d'une fréquence instantanée a en soi une contradiction qui limite intrinsèquement la résolution d'une représentation; une fréquence n'étant parfaitement définie que pour une sinusoïde pure définie pour tout temps. Pour la transformée de Fourier à court terme, cela se manifeste par l'étalement des atomes de Fourier selon la fenêtre choisie, soumis à l'inégalité d'Heisenberg.

Jusqu'à présent, les techniques d'analyse basées sur ces représentations dont les atomes de base sont prédéfinis, si elles parviennent à la décomposition une mélodie homophone, échouent à l'analyse de signaux plus riches, tels que des accords de plus de trois sons mis en musique selon différents timbres; tandis qu'une oreille humaine éduquée est capable de bien meilleures performances (Mozart, dès l'âge de 11 ans, reconstituait à la première écoute un motet à 9 voix d'Allegri, ceci à la précision d'une partition). Bien sûr, ce que l'on entend par « oreille éduquée » dans le contexte occidental est chargé de sous-entendus: éducation à la tonalité et à l'harmonie, permettant souvent de prévoir les accords ou de les imaginer à partir de la mélodie, éducation aux rythmes, aux signifiés des phrases et des cellules musicales, et également éducation à la distinction des timbres et de leurs fonctionnalités. Mais le travail de formation auditive n'apporte pas seulement celui d'une meilleure compréhension de la musique il contribue également à améliorer l'écoute. Plusieurs interprétations sont possibles: adaptation locale aux caractéristiques du signal, notamment de la sélectivité fréquentielle, sensibilité accrue aux variations d'amplitudes utilisée pour les battements... De récents travaux [1], proposent même l'hypothèse d'une réallocation temps/fréquence, méthode introduite par Flandrin visant à rectifier l'étalement des composantes dans les distributions de la classe Cohen, effectué par le cerveau pour expliquer une telle acuité.

Sur ce modèle, il paraît nécessaire, dans le cadre de l'analyse numérique, de choisir les atomes de Fourier qui corroborent localement le signal le « mieux » possible, sachant qu'il faudra dans certains cas faire des compromis.

Dans cette perspective, plusieurs approches sont envisageables:

- les algorithmes de poursuite, qui cherchent une base d'atomes à l'intérieur d'un dictionnaire de bases.
- tirer profit des propriétés de localisation de la distribution de Wigner-Ville. Celle-ci jouit en effet d'une localisation exacte jusqu'aux fréquences quadratiques dans le plan temps fréquence pour des signaux monocomposants; en contrepartie, sa bilinéarité introduit des interférences entre les composantes qui nécessitent moyennage et ainsi perte de résolution. Le spectrogramme réassigné, dont le principe est de rectifier l'étalement créé par le moyennage avec le noyau en réassignant les « composantes » à leurs centres de gravité, a été utilisé avec succès pour l'analyse du signal musical. Cependant, une telle approche connaît certaines limites, le résultat dépendant

fortement du choix du noyau et l'algorithme étant sensible au bruit.

Une autre approche, introduite par Jones et Baraniuk[11], consiste à chercher localement le noyau de lissage optimal à appliquer à la fonction d'ambiguïté du signal (celle-ci étant également la TF2D de la distribution de Wigner-Ville).

- adapter les fenêtres du spectrogramme *a posteriori*, en s'intéressant davantage à la résolution propre des distributions temps-fréquence. Dans le cas du spectrogramme, il s'agit alors d'adapter localement les tailles de fenêtre correspondant à la distribution la moins étalée. Plusieurs critères, dérivés de concepts en probabilités, ont été envisagés: l'entropie de Rényi, le kurtosis et « l'Energy Smearing » définie telle que dans [2]. Cependant, si cette approche paraît pertinente et rapide pour l'analyse du signal musical, notamment la distinction des transitoires, elle pose également plusieurs problèmes. Premièrement, une multi-représentation, mettant en jeu plusieurs tailles de fenêtre, modifie la redondance d'une transformée à court terme: en particulier, il n'y a pas de conservation d'énergie et il est nécessaire de prendre des précautions avant de reconstruire le signal. Le second problème est celui de la pertinence de la comparaison entre deux spectrogrammes et du choix de localisation effectué lors de cette comparaison. En effet, si cette méthode semble légitime pour les signaux monocomposants, localement stationnaires ou dont les propriétés sont invariantes dans le temps, la situation est nettement plus délicate pour des signaux multicomposants dont les non-stationnarités ont des caractéristiques parfois très différentes. Une fois ces difficultés admises, on peut néanmoins garder l'approche par adaptation des tailles de fenêtres en ayant en tête les principaux enjeux pour les signaux musicaux:

- localisation des transitoires,
- meilleure résolution possible, ie plus large fenêtre possible, pour les sons harmoniques, d'autant plus que les fondamentales sont basses fréquences,
- compromis pour les situations de modulations (linéaires ou non), qui imposent dans les situations de jeu (vibratos, trajectoires de formants...) des variations de fréquence et d'amplitude très rapides,
- faible dépendance aux niveaux du signal et au rapport signal/bruit,
- compatibilité avec les transformations usuelles en traitement du signal (translation, filtrage, modulation...),
- coût de calcul réduit, sachant que l'on doit déjà effectuer plusieurs transformées de Fourier à court terme.

Dans le cadre de ce stage, nous nous sommes concentrés sur les deux dernières approches. Nous commençons par un rappel du cadre théorique des distributions temps fréquence, avant d'étudier les critères évalués dans le cadre d'un spectrogramme adaptatif. Nous terminons avec l'approche par adaptation de noyau sur les distributions de classe Cohen.

## Représentations temps-fréquence

Cette section introduit un panorama non exhaustif des différents types de représentations temps fréquence que l'on peut trouver dans la littérature. Plus de détails et démonstrations sont proposées dans des ouvrages de référence tel que celui de L.Cohen, *Time Frequency Analysis*, celui de S. Mallat, *Une exploration des signaux en ondelettes* ou celui de P. Flandrin, *Temps Fréquence*.

Les articles de Patrick Flandrin, disponibles sur « <http://perso.ens-lyon.fr/patrick.flandrin/publis.html> », constituent également une excellente bibliographie, notamment sur les distributions Wigner-Ville et la réallocation temps-fréquence.

## Transformée de Fourier à Court terme

La plus simple représentation introduisant la dépendance temporelle de la représentation de Fourier est la transformée de Fourier à court terme, appliquant une transformée de Fourier localement sur un segment du signal fenêtré, si  $x$  désigne le signal et  $h$  la fenêtre choisie:

$$F_x^{(h)}(t, \omega) = \int_{-\infty}^{\infty} x(s) \overline{h(s-t)} e^{-i\omega(s-t/2)} ds, \quad (1)$$

On peut interpréter cette représentation comme une décomposition dans le plan temps fréquence sur des atomes de Fourier  $h_{t,\omega}(s) = h(s-t) e^{i\omega(s-t/2)}$ , comme le suggère la formule de reconstruction:

$$\text{Si } f \in L^2(\mathbb{R}), f(t) = \frac{1}{2*\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_x^{(h)}(u, \omega) h(t-u) e^{i\omega t} d\omega du = \frac{1}{2*\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \langle f | h_{u,\omega} \rangle h_{u,\omega}(t) d\omega du$$

Cette formule ressemble à celle d'une décomposition dans une base orthogonale, bien que la famille  $h_{u,\omega}$  soit en fait largement redondante dans  $L^2(\mathbb{R})$ .

La résolution de la stft (short time Fourier transform) dépend de l'étalement de la fenêtre en temps et en fréquence.

Dans les applications numériques, il est nécessaire de choisir une fenêtre à support compact, dont la transformée de Fourier est à support infini; cela est à l'origine des oscillations de Gibbs observées au niveau du spectre (sidelobes), d'autant plus grandes et étroites que la taille de la fenêtre est grande. Les propriétés des différentes fenêtres couramment utilisées (Hamming, Blackman...) sont discutées dans le livre de S. Mallat.

La discrétisation du calcul donne, pour des signaux discrets de période  $N$ , :

$$F_x^{(h)}[m, l] = \sum_{n=0}^{N-1} f[n] h[n-m] e^{\frac{-i2\pi ln}{N}} \quad (2)$$

En pratique, on ne calcule la stft sur un nombre de points inférieur à  $N$  (en temps). Si le pas d'avancement (incrément) entre chaque trame  $m$  est tel que

$$\forall n, \sum_m h(n-m) \approx cst \quad (3)$$

alors, on peut tirer profit de la linéarité de la transformée de Fourier pour obtenir une reconstruction quasi-parfaite. Dans le cas où

$$\forall n, \sum_m h(n-m) \neq cst$$

on peut même pondérer la reconstruction sur chaque trame afin d'obtenir une reconstruction quasi-parfaite (ceci pourra en particulier être exploité pour des transformées à court terme combinant plusieurs tailles de fenêtre).

## **Transformée en ondelettes**

Les ondelettes ont été introduites dans le cadre du traitement d'images, avec notamment des applications très intéressantes du point de vue du codage (à l'origine du format JPEG2000). L'idée principale est de construire une famille d'atomes de Fourier dont les étalements sont étroits en temps dans les hautes fréquences, correspondants aux zones de transitions ou de rupture des images, et au contraire larges dans les basses fréquences (fenêtres dilatées). Sans détailler l'aspect mathématique, que l'on peut trouver dans le livre de S.Mallat, cette approche paraît plus difficilement applicable dans le cadre du signal sonore, notamment du fait de la richesse de l'information contenue dans les hautes fréquences. Cependant, les ondelettes peuvent tout de même trouver leur applications pour un signal audio, notamment pour la détection de transitoires.

## **Sélection adaptative de la base**

Les premières adaptations de bases au signal ont été effectuées par programmation dynamique avec des paquets d'ondelettes et de cosinus locaux, particulièrement pertinentes pour des images [3]. Mais pour les signaux audios, comprenant au même instants des notes de durées différentes, il est nécessaire d'introduire un cadre d'optimisation plus général.

Étant donné un dictionnaire redondant  $g_p$  de P vecteurs (invariants par translation en temps et en fréquence), comprenant au moins N vecteurs indépendants, il s'agit de trouver la meilleure approximation du signal de taille N:

$$f_M = \sum_{m=0}^M a[p_m] g_{p_m} \quad (4)$$

Le problème d'optimisation général, minimisant  $\|f - f_m\|$ , constitue un problème NP difficile [4]. Il existe cependant de bonnes approximations par poursuite de vecteurs; l'idée principale de la poursuite de vecteurs est de chercher itérativement un vecteur de projection qui maximise le produit scalaire avec le signal et ses résidus successifs :

$$f = \langle f | g_{y_0} \rangle g_{y_0} + Rf, |\langle f | g_{y_0} \rangle| \geq \alpha \sup_y |\langle f | g_y \rangle|;$$
$$\dots |\langle R^m f | g_{y_m} \rangle| \geq \alpha \sup_y |\langle R^m f | g_y \rangle|$$

A l'heure actuelle, les algorithmes et les dictionnaires de matching pursuit n'ont pu obtenir des représentations réellement avantageuses et restent relativement coûteuses par rapport à des représentations classiques, d'autant plus que le dictionnaire est large. Cependant, la voie est encore largement inexplorée et on peut penser qu'avec les performances croissantes des ordinateurs, notamment en matière de parallélisation, cette approche puisse concurrencer les autres types de représentations; on peut par exemple imaginer la construction de dictionnaires spécialement adaptés aux structures musicales.

## **Distributions quadratiques**

### **Distribution Wigner Ville**

Toutes les représentations étudiées plus haut dépendent intrinsèquement de la famille d'atomes choisie pour la décomposition. S'inspirant du filtrage adapté, la distribution de Wigner-Ville prend pour fenêtre d'analyse l'image retournée dans le temps du signal analysé afin d'utiliser des atomes ayant une structure de phase adaptée

localement au signal :

$$\boxed{W_x(t, w) = \int_{-\infty}^{\infty} x(t + \tau/2) x(t - \tau/2) e^{-i w \tau} d \tau} \quad (5)$$

Plusieurs propriétés importantes découlent cette définition:

-si  $f$  est analytique de la forme  $f(t) = a(t) e^{i \phi(t)}$ , alors

$$\phi'(u) = \frac{\int_{-\infty}^{\infty} \xi P_V f(u, \xi) du}{\int_{-\infty}^{\infty} P_V f(u, \xi) du}$$

-localisation des phases quadratiques

$$x(t) = e^{i(w_0 t + \alpha t^2)} \Rightarrow W_x(t, w) = \delta(w - w_0 + \alpha t),$$

-pseudo-densité d'énergie

$$|\hat{f}(\xi)|^2 = \int_{-\infty}^{\infty} P_V f(u, \xi) du$$

$$|f(u)|^2 = \frac{1}{2 * \pi} \int_{-\infty}^{\infty} P_V f(u, \xi) d \xi$$

En revanche, la distribution de Wigner-Ville étant une forme quadratique, on a également des termes d'interférence entre les composantes qui annihilent du point de vue de la représentation tout le bénéfice apporté par la localisation des composantes.

Ces termes d'interférences étant nécessairement oscillants (comme le prouvent les propriétés de pseudo-densité d'énergie), on peut les atténuer par moyennage local. Ce lissage de la distribution est effectué avec un noyau intégral de la forme :

$$\theta(u, u', \xi, \xi') = \theta(u - u', \xi - \xi')$$

conservant les propriétés d'invariance par translation en temps ou en fréquence. On obtient ainsi les distributions de la classe Cohen:

$$\boxed{P_\theta f(u, \xi) = \frac{1}{2\pi} P_V f * \theta(u, \xi) = \iint P_V f(u', \xi') \theta(u - u', \xi - \xi') du' d \xi'} \quad (6)$$

Le spectrogramme, amplitude de la transformée de Fourier à court terme, est un cas particulier de distribution de la classe Cohen, pour laquelle le noyau de lissage correspond à la distribution Wigner de la fenêtre.

## Fonction d'ambiguïté

La fonction d'ambiguïté a été introduite dans le cadre de la théorie de la détection radar. En effet, c'est une mesure de corrélation du signal retardé et modulé (comme par exemple dans le cas de l'effet Doppler) :

$$\boxed{A_x(v, \tau) = \int_{-\infty}^{\infty} x(t + \tau/2) x^*(t - \tau/2) e^{i 2\pi v t} dt} \quad (7)$$

Elle correspond à la transformée de Fourier bi-dimensionnelle de la distribution Wigner-Ville:

$$\iint_{-\infty}^{\infty} A_x(v, \tau) e^{-i 2\pi(vt + \tau f)} d v d \tau = W_x(t, f)$$

On peut donc également construire une distribution de la classe Cohen à partir de la fonction d'ambiguïté,

$$P(t, f) = \frac{1}{4\pi^2} \iint A_x(v, \tau) \Phi(v, \tau) e^{-j v t - j \tau f} d \theta d \tau$$

où  $\Phi$  est la fonction d'ambiguïté de la fenêtre dans le cas du spectrogramme.

La fonction d'ambiguïté hérite de propriétés similaires à la distribution Wigner, notamment l'invariance de volume:

$$\iint_{-\infty}^{\infty} A_x(v, \tau) A_y^*(v, \tau) d v d \tau = \left| \int_{-\infty}^{\infty} x(t) y^*(t) dt \right|^2$$

La décroissance de la fonction d'ambiguïté mesure l'étalement d'une composante en



temps et en fréquence. Une discussion des différentes propriétés et des applications aux radars est proposée dans [5].

Les représentations citées ci-dessus sont les plus usitées en analyse spectrale, et offrent des algorithmes de calculs relativement rapide; dus notamment à la fft (la distribution Wigner-Ville, la stft ont des algorithmes rapides en  $N^2 \log N$ , la transformée en ondelettes en  $N(\log N)^2$ ).

D'autres types de distributions, telles que la généralisation des distributions quadratiques de Bertrand ou des distributions allant au delà du quadratique, sont données dans [6]. A noter également qu'une discussion des représentations temps-fréquence pour des signaux non-stationnaires est proposée dans [7].

Critères d'adaptation locale de la taille de fenêtre pour le spectrogramme

### Critères entropiques, entropies de Rényi et Kurtosis

Pour déterminer une représentation qui soit la plus localisée possible, c'est à dire la moins étalée, une première idée légitime est de comparer les intégrales de distributions élevées à une puissance  $>1$ , afin de privilégier les distributions maximisant, à énergie totale donnée, les grandes amplitudes. Cette première approche, étayée par [8], conduit au critère de maximisation d'un pseudo Kurtosis (coefficient d'aplatissement); si  $h$  désigne la fenêtre,  $x$  le signal, et  $F$  la transformée de Fourier à court terme :

$$K_x^{(h)} = \frac{\iint |F_x^{(h)}(t, f)|^\alpha dt df}{\|F_x^{(h)}\|^{\alpha/2}} \rightarrow \max_h (K_x^{(h)}), \alpha > 1 \quad (7)$$

Une borne inférieure est donnée par le théorème de Lieb :

$$\text{Pour } x \in L^2(\mathbb{R}), \frac{\iint |F_x^{(h)}(t, f)|^\alpha dt df}{\|F_x^{(h)}\|^{\alpha/2}} \leq (2/\alpha) * \|h\|^{\alpha/2}$$

avec égalité si et seulement si  $x$  et  $h$  sont deux gaussiennes identiques.

Une idée similaire, inspirée de l'article [9] discutant l'information contenue dans les représentations temps-fréquence, est d'utiliser un critère entropique. Dans le cadre probabiliste, on introduit l'entropie généralisée de Rényi, définie par:

$$H_\alpha^R(P, \Omega) = \frac{1}{1-\alpha} \log_2 \frac{\int_\Omega P^\alpha(X) dX}{\int_\Omega P(X) dX}$$

Où  $X$  est une variable aléatoire,  $P$  est une mesure de probabilité et  $\Omega$  un sous-ensemble des résultats possibles. On retrouve dans la limite où  $\alpha \rightarrow 1, P(\Omega) \rightarrow 1$  l'entropie de Shannon; on a également l'importante propriété de moyenne:

$$H(P, \Omega_1 \cup \Omega_2) = \frac{P(\Omega_1)H(P, \Omega_1) + P(\Omega_2)H(P, \Omega_2)}{P(\Omega_1) + P(\Omega_2)}$$

En choisissant des distributions définies positives (une discussion sur les représentations non positives est proposée dans [9]), telles que le spectrogramme, on peut interpréter ces distributions comme mesure de probabilité.

Cela nous amène à définir une pseudo-entropie de Rényi pour le spectrogramme:

$$H_\alpha(S_x^{(h)}, \Omega) = \frac{1}{1-\alpha} \log_2 \left( \iint_\Omega \left( \frac{S_x^{(h)}(t, f)}{\iint_\Omega S_x^{(h)}(t', f') dt' df'} \right)^\alpha dt df \right) \quad (8)$$

Cette définition varie de celle donnée dans le cas probabiliste afin de confondre les cas de composantes de même nature mais d'énergie différente.

De la propriété de la moyenne découle une conséquence importante pour une représentation temps fréquence: dans le cas où on a  $n$  structures de même nature (ie translattées en temps ou en fréquence) et représentées dans des régions différentes

$\Omega_1 \dots \Omega_n$  avec  $\Omega_i \cap \Omega_j = \emptyset$  et  $\iint_{\Omega_i} dt df = \iint_{\Omega_j} dt df$ , la pseudo entropie de Rényi est augmentée de  $n$ ,

$$H_{\alpha}(S_x^{(h)}, \Omega_1 \cup \dots \cup \Omega_n) = \frac{1}{1-\alpha} \log_2 \left( \frac{1}{\sum_{\Omega_i} \left( \iint_{\Omega_i} S_x^{(h)}(t', f') dt' df' \right)^{1-\alpha}} * \sum_{\Omega_i} \iint_{\Omega_i} \frac{(S_x^{(h)}(t, f))^{\alpha}}{\iint_{\Omega_i} S_x^{(h)}(t', f') dt' df'} \right)$$

$$= n + \frac{H_{\alpha}^R(S_x^{(h)}, \Omega_1) * \iint_{\Omega_1} dt df + \dots + H_{\alpha}^R(S_x^{(h)}, \Omega_n) * \iint_{\Omega_n} dt df}{\iint_{\Omega_1} dt df + \dots + \iint_{\Omega_n} dt df}$$

$$= n + H_{\alpha}(S_x^{(h)}, \Omega_1)$$

Revenant à notre problème initial de comparaison de représentations, on en vient ainsi au critère de minimisation de la pseudo entropie de Rényi, privilégiant un maximum d'amplitude dans une région temps fréquence minimale (découle de la propriété de moyenne) :

$$H_{\alpha}(S_x^{(h)}, \Omega) = \frac{1}{1-\alpha} \log_2 \left( \iint_{\Omega} \left( \frac{S_x^{(h)}(t, f)}{\iint_{\Omega} S_x^{(h)}(t', f') dt' df'} \right)^{\alpha} dt df \right), \quad \boxed{\min_h H_{\alpha}(S_x^{(h)}, \Omega)}$$

Du fait du facteur  $\frac{1}{1-\alpha}$ , on peut choisir  $\alpha > 0, \alpha \neq 1$  selon la sensibilité que l'on veut donner au critère; pour  $\alpha < 1$ , on cherche à minimiser les coefficients de faible amplitude, tandis que pour  $\alpha > 1$  on cherche davantage à maximiser les coefficients de grande amplitude.

### « Energy Smearing » - étalement d'énergie

Cette deuxième approche, issue de [2] où elle est exploitée dans le cadre d'une représentation additionnant plusieurs résolutions de spectrogrammes, est en fait assez similaire à celle du critère entropique. L'idée principale est de sélectionner les spectrogrammes dont la moyenne de la distribution d'amplitude est maximale. Etant donné un spectrogramme  $S_x^{(h)}$ , une région temps fréquence  $\Omega$  et une distribution d'amplitudes locale  $a_i(S_x^{(h)}, \Omega)$  triée par ordre décroissant, le critère proposé est:

$$ES(S_x^{(h)}, \Omega) = \frac{\sum_i i * a_i(S_x^{(h)}, \Omega)}{\sqrt{\sum_i a_i(S_x^{(h)}, \Omega)}} \quad \boxed{\min_h (ES(S_x^{(h)}, \Omega))} \quad (9)$$

La pondération permet ainsi de minimiser les faibles et moyennes amplitudes. D'autres critères sont également envisageables sur le même principe; on peut par exemple s'intéresser à la dispersion de la distribution d'amplitude autour de sa moyenne.

### Partition d'analyse et espace de fenêtres

1. Dans ce qui précède, nous avons omis de discuter trois questions cruciales : la partition d'analyse  $\Omega_i$ , qui définit le caractère local de l'approche
2. la puissance  $\alpha$  pour la pseudo entropie de Rényi ou Kurtosis
3. l'espace des fenêtres -ou éventuellement bases d'atomes- sur lequel évaluer ces critères

Dans le cas d'un signal constitué d'une composante dont la distribution d'amplitude est invariante selon la partition pour chaque représentation (signal harmonique, sinusoïde modulée linéairement...), la situation est relativement simple puisqu'on peut choisir

tout le plan temps fréquence pour région d'analyse.

En revanche dès que le signal contient plusieurs composantes ou une composante dont la distribution d'amplitude varie en temps et/ou en fréquence sur une représentation (modulation non linéaire...), la situation devient inextricable. Ou bien on parvient a posteriori à déterminer une partition d'analyse telle que chaque  $\Omega_i$  distingue des composantes à distribution d'amplitude invariante, ou bien on est contraint de réaliser des compromis localement entre différentes composantes.

Par ailleurs, si une région d'analyse contient plusieurs composantes, l'optimum choisi par un critère entropique dans un espace de fenêtre dépend également de la puissance  $\alpha$ .

En ce qui concerne l'espace des fenêtres, il semble délicat de mener une approche par optimisation sur le modèle des algorithmes de matching pursuit, vu la non linéarité des critères, et la dépendance en la partition d'analyse. Dans le cadre de ce stage, nous nous sommes donc restreints à un panel de tailles de fenêtres réduit.

### Analyse numérique

On choisit tout d'abord un panel de tailles de fenêtres à partir desquelles on va effectuer une comparaison des spectrogrammes. Pour une comparaison pertinente, les STFT sont calculées avec la fonction spectrogram de MATLAB sur une même grille temps-fréquence, à savoir qu'on choisit un même pas d'avancement et une même taille de DFT pour les différentes fenêtres. Les trames obtenues avec la fonction spectrogram correspondent temporellement aux centres des fenêtres.

Pour les tests effectués, les fenêtres choisies vont de 128 samples (~3ms pour un échantillonnage à 44100Hz, relativement adaptées pour les transitoires d'attaques), jusqu'à 4096 samples (~95 ms à 44100Hz).

Sur chaque spectrogramme, on calcule ensuite les différents critères sur la partition d'analyse, choisie a priori comme un quadrillage uniforme du plan temps fréquence. Comme discuté précédemment, la partition d'analyse est un élément crucial de l'étude menée; dans un premier temps, nous nous sommes principalement concentrés sur la dimension temporelle en comparant les entropies sur toute la distribution d'amplitude du spectre. Afin d'appréhender l'influence des paramètres de manières plus fine, nous avons opté pour un « overlapping » des analyses, à savoir qu'on réalise une analyse locale sur chaque trame.

L'algorithme utilisé est assez coûteux, avec un nid de boucles inextricable :

```
for windowsize = [panel]
    spec<-spectrogram(signal, windowsize);
    for trame=1:length(spec)
        ->critere  $\Omega(trame)$ 
```

Sachant que le calcul du spectrogramme est en  $O(N^2 \log(N))$  et que les critères réclament une élévation à une puissance de la distribution dans le cas des entropies ou un tri local ( $\log|\Omega|$ ) dans le cas de l'étalement en énergie. Cependant, en dehors d'une parallélisation, qui pourrait être envisagée pour des applications, il n'y a a priori pas de solution triviale pour améliorer la complexité.

## Comparaison des critères sur des signaux

Dans le cadre de ce stage, un ensemble de signaux de synthèse correspondant à différentes situations que l'on peut trouver dans un signal musical ont été mis à l'épreuve des critères de sélection de fenêtre.

Parmi les types de signaux étudiés, on s'est particulièrement intéressé aux transitoires et aux modulations fréquentielles. En effet, ces types de signaux revêtent une importance particulière pour le signal musical; on aimerait d'une part avoir la meilleure localisation temporelle possible pour les attaques de notes, ceci notamment afin d'éviter les effets de pré-échos qui interviennent dans la reconstruction, et d'autre part garantir une bonne visibilité de la représentation pour un signal modulant (vibrato, trajectoires de formants...), généralement de façon non-linéaire.

Cette section détaille les résultats obtenus pour l'entropie de Rényi et « l'étalement d'énergie ». Leur analyse nous permettra de discuter la validité de notre approche ainsi que les paramètres et notamment la partition d'analyse à adopter.

Afin d'appréhender l'efficacité de l'approche, on représente dans ce qui suit les valeurs de l'entropie pour le panel de fenêtres, le spectrogramme adapté (on choisit la représentation correspondant à la meilleure fenêtre sur la région d'analyse).

Pour l'optimisation, on aura choisi des fenêtres de type Blackman qui ont de faibles coefficients sidelobes par rapport aux autres fenêtres usuelles, qui jouent fortement sur les critères adoptés.

## Transitoires de Dirac dans un signal harmonique.

Il s'agit d'un signal particulièrement éclairant pour comprendre le comportement de l'algorithme dans une situation d'attaque de note, que l'on souhaite la plus localisée possible en temps, en détériorant un minimum l'information pour les harmoniques qui précèdent ou qui suivent.

### Pseudo-entropie de Rényi

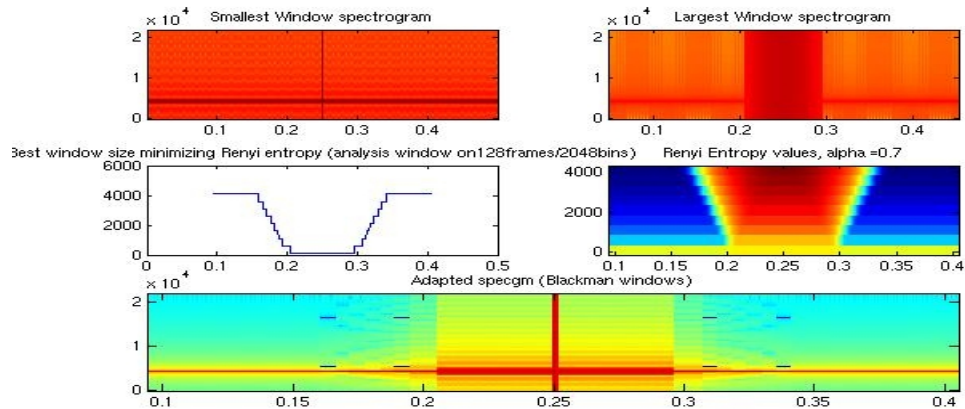


Figure 1 a : adaptation de la taille de fenêtre minimisant localement l'entropie de Rényi pour un dirac avec harmoniques (analyse sur 128 trames)

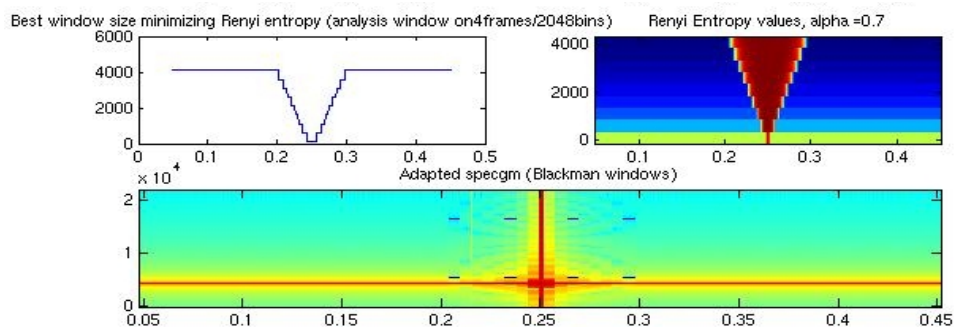


Figure 1 b : adaptation de la taille de fenêtre minimisant localement l'entropie de Rényi pour un dirac avec harmoniques (analyse sur 4 trames)

La partition d'analyse choisie correspond respectivement à la taille de la plus grande fenêtre (4096 samples) et celle de la plus petite fenêtre (128 samples pour les deux derniers spectrogrammes adaptés). On remarque tout d'abord que pour une taille d'analyse plus faible, l'adaptation vers des fenêtres plus petites est davantage localisée au niveau de la transitoire; de même lorsque  $\alpha > 1$  (Fig. 1c), le critère devient tend à maximiser les grandes amplitudes, ce qui explique les débuts de « pré-échos » observés. Pour  $\alpha > 1$ , on a également une augmentation brusque de la taille de fenêtre au niveau de la transitoire; en effet, si le spectre d'un dirac est uniforme et ne devrait par conséquent privilégier aucune fenêtre, ce phénomène vient de l'harmonique davantage perçue par les grandes fenêtres, qui rend le spectre non uniforme et donc d'entropie minimale.

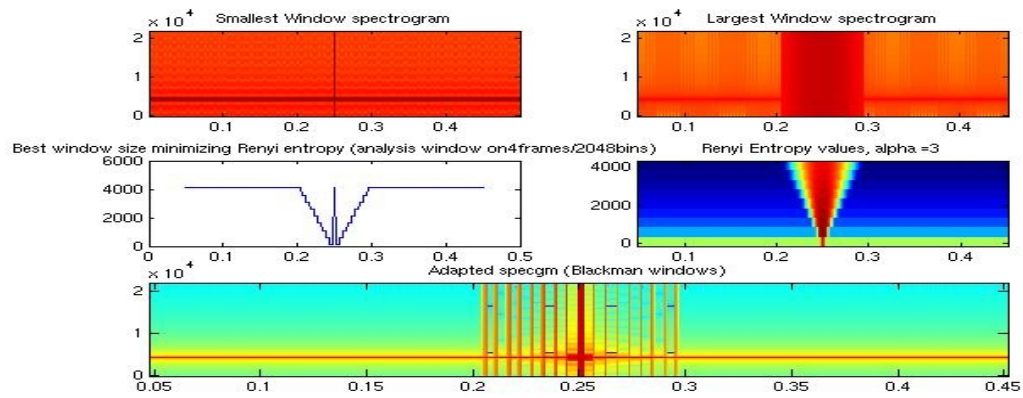


Figure 1c : adaptation de la taille de fenêtre minimisant localement l'entropie de Rényi pour un dirac avec harmoniques (analyse sur 4 trames,  $\alpha=3$ )

## Etalement d'énergie

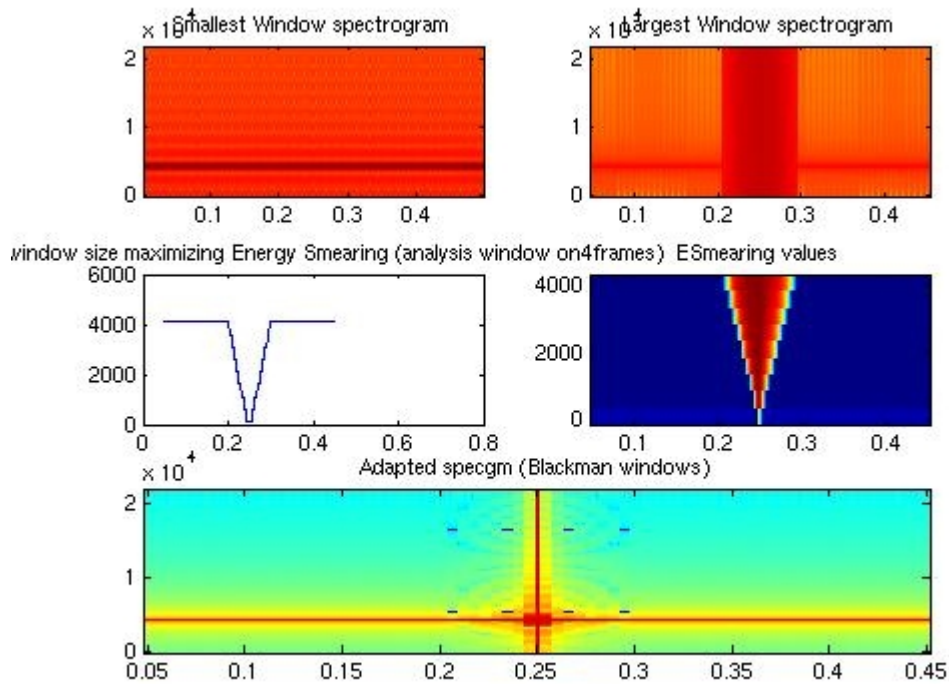


Figure 1d : adaptation minimisant l'étalement d'énergie pour dirac avec harmoniques  
 Le critère d'étalement d'énergie, minimisant les faibles amplitudes, semble localiser au mieux la transitoire et n'a pas le désavantage de la remontée de fenêtre pour une partition d'analyse fine.

## Transitions d'harmoniques

On considère ici une brève harmonique, fenêtrée afin d'éviter une transition de type dirac, encadrée par d'autres harmoniques. On est dans le cas typique où on doit faire un compromis entre temps et fréquence qui dépend fortement du pavage d'analyse choisi.

## Pseudo-entropie de Rényi

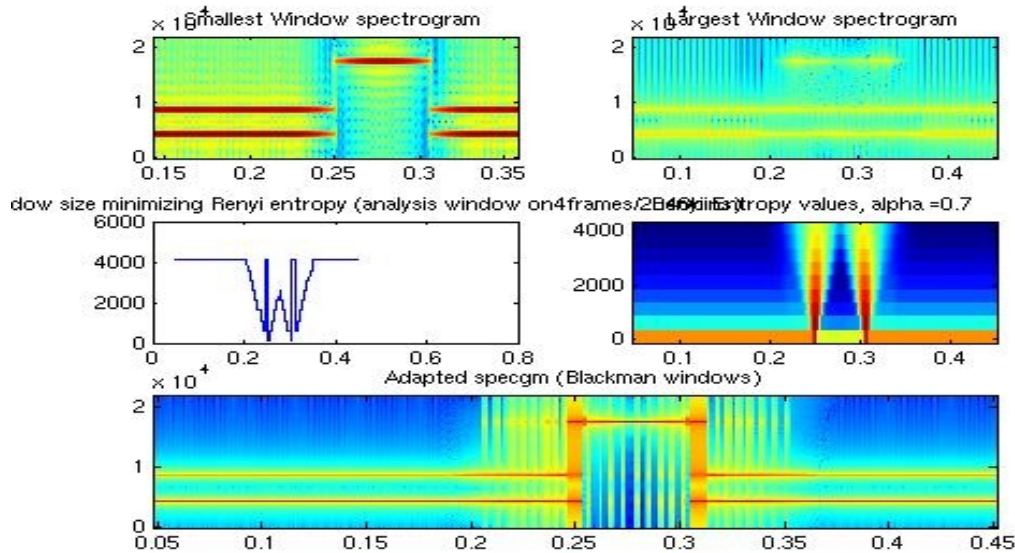


Figure 2a : adaptation des tailles de fenêtre minimisant localement l'entropie Rényi pour une transition d'harmoniques

Ici une analyse fine en temps paraît plus judicieuse afin d'obtenir la meilleure résolution pour la transitoire; on observe bien une amélioration de la résolution fréquentielle au centre de la transitoire, cependant, celle ci reste limitée du fait des échos et des sidelobes produits par les composantes voisines pour les grandes fenêtres. Cela suggère d'introduire une dimension fréquentielle dans l'analyse :



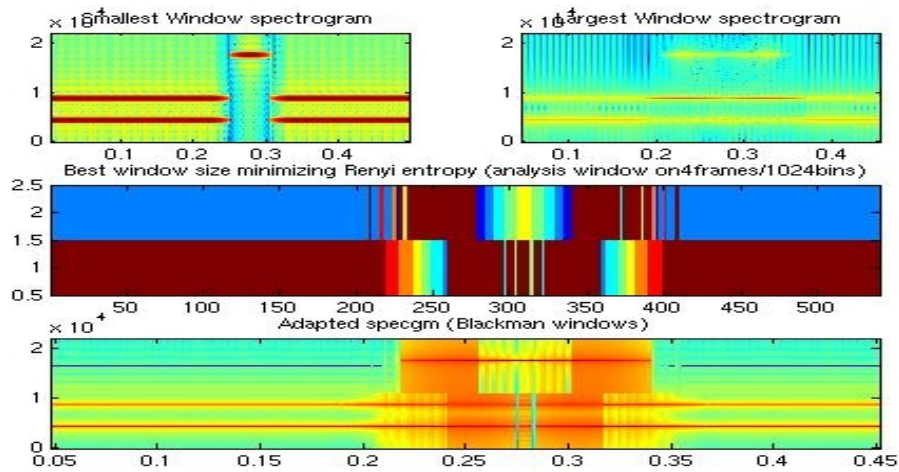


Figure 2b : adaptation locale avec division fréquentielle (échelle de couleur-> en rouge resp. bleu les fenêtres de grandes resp. de petites tailles,  $\alpha=0.7$  )

Pour la composante transitoire, l'analyse conduit à une représentation dégradée par rapport à la partition uniquement temporelle, notamment avec une augmentation plus marquée de la taille de fenêtre juste avant et après la transitoire. Ce phénomène est probablement dû au sidelobes des composantes voisines. On a également un comportement assez imprévisible au niveau des échos (harmoniques basse fréquence) avec des sauts de fenêtre optimale qui montrent l'impossibilité pour le critère de faire un compromis tranché. Pour éviter ce phénomène, on peut choisir une taille d'analyse plus large en temps:

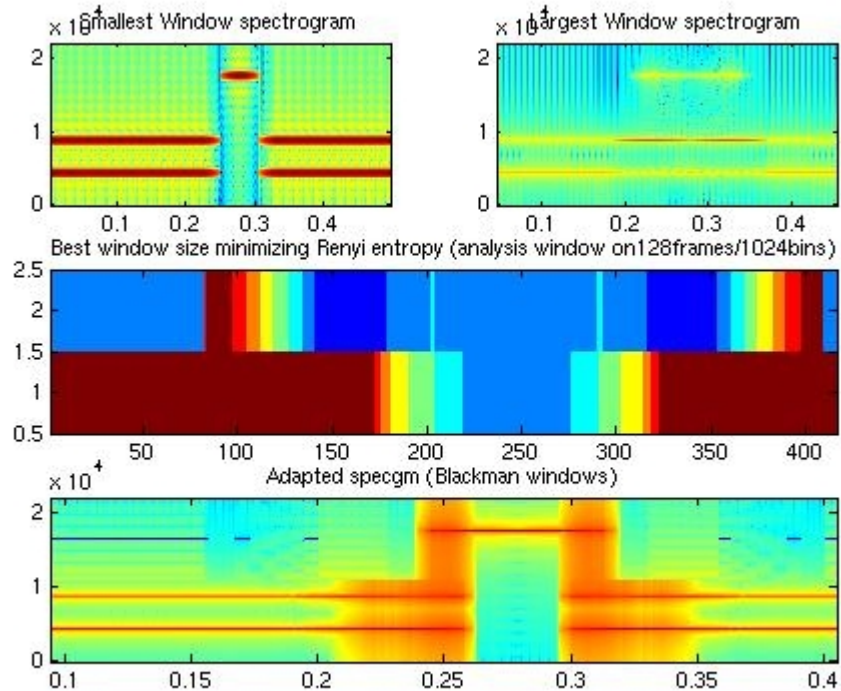


Figure 2c : adaptation locale avec division fréquentielle et analyse sur 128 trames

Si on évite les échos et qu'on localise mieux la transitoire, l'adaptation de la taille de fenêtre n'est clairement pas optimale. On constate également que le choix des tailles de fenêtres est perturbé avant même la transitoire, dans une zone fréquentielle où l'on a que des sidelobes des composantes voisines.

## Etalement d'énergie

Une partition uniquement temporelle donne une représentation très similaire à celle obtenue avec le critère de minimisation d'entropie Rényi:

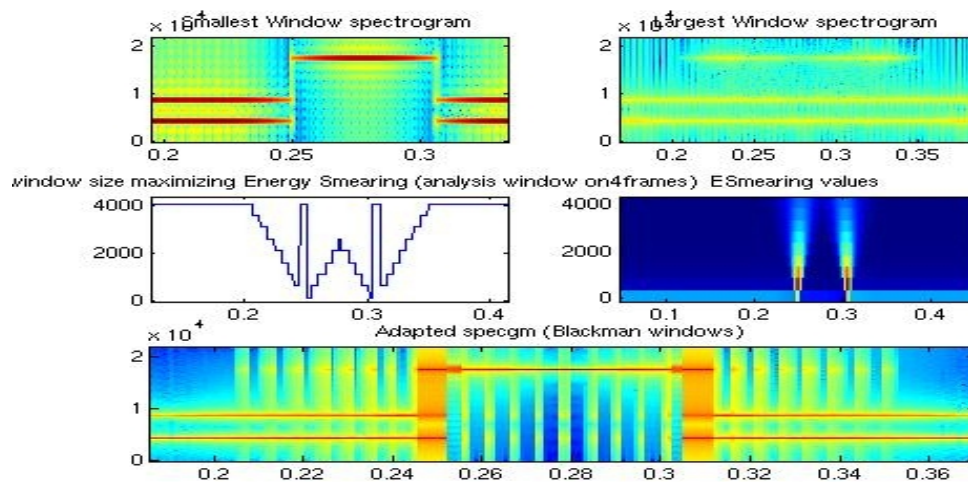


Figure 2d : adaptation des tailles de fenêtre minimisant localement l'étalement d'énergie pour une transition d'harmoniques

Si on divise également le domaine fréquentiel, le critère semble encore plus sensible à l'influence des sidelobes, il faut également prendre une taille d'analyse assez large temporellement; l'optimisation obtenue est à nouveau peu optimale :

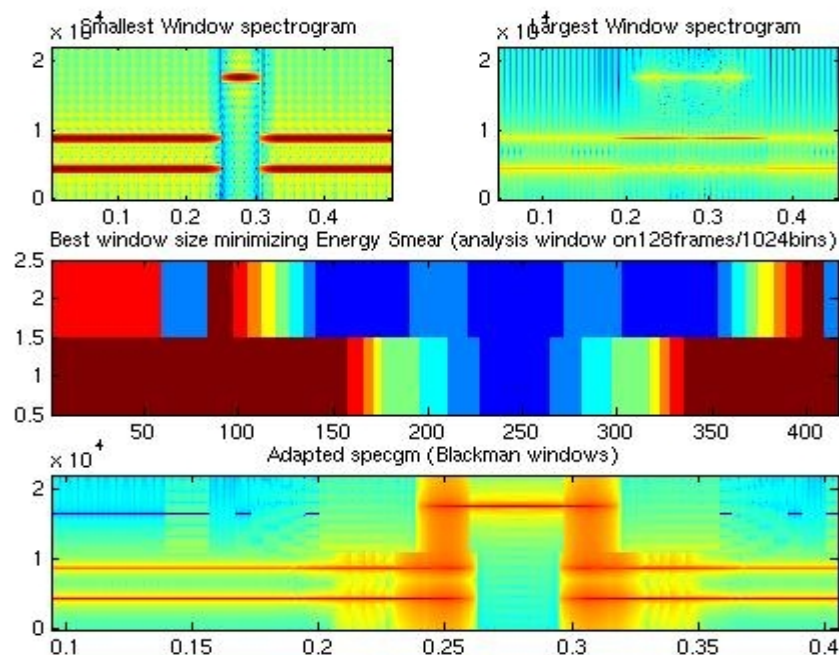


Figure 2e adaptation avec division fréquentielle

Pour éviter le problème des sidelobes, certains critères ont été déterminés [10] afin de discriminer les pics sinusoidaux des sidelobes et du bruit dans un spectre. Les essais menés en tentant d'éliminer les sidelobes avec ces critères n'ont pas donné des résultats réellement pertinents pour la stratégie d'optimisation locale de la taille de fenêtre.

## Modulation linéaire

Ce type de signal, si il est peu réaliste dans le domaine musical, permet en revanche de vérifier la validité du critère de choix, indépendamment de la partition d'analyse.

## Pseudo-entropie de Rényi

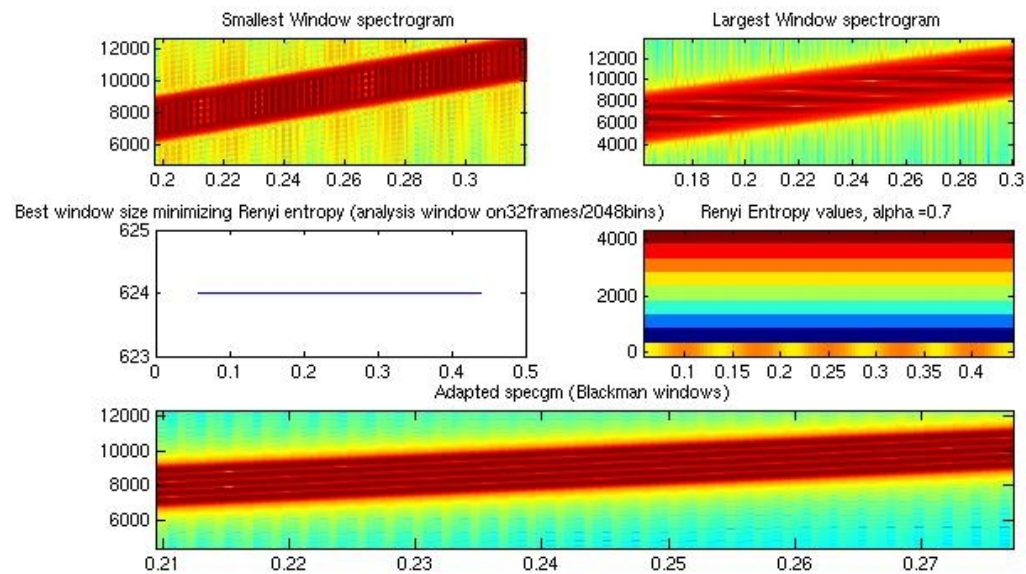


Figure 3a adaptation des tailles de fenêtre minimisant localement l'entropie Rényi pour une modulation linéaire

Pour ce signal, le critère paraît relativement fiable, favorisant des fenêtres intermédiaires en fonction de la pente. Les valeurs d'entropie sont nettement différenciées. Différentes tailles d'analyse conduisent au même choix de fenêtre. Voici la meilleure taille de fenêtre selon ce critère en fonction de la pente de la modulation:

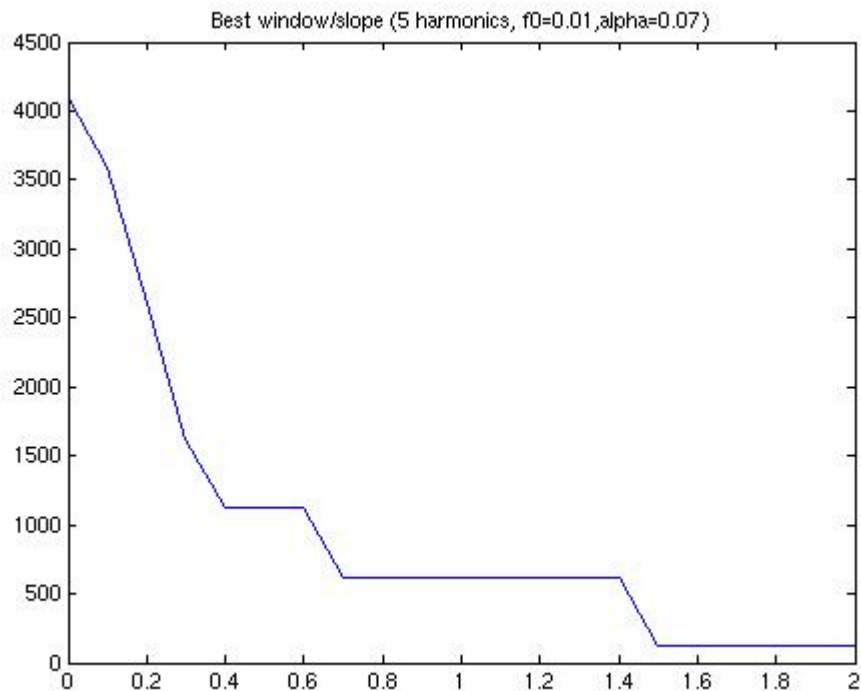


Figure 3b : taille de fenêtre optimale en fonction de la pente de modulation

On peut également s'intéresser au cas plus réaliste d'harmoniques modulant linéairement :

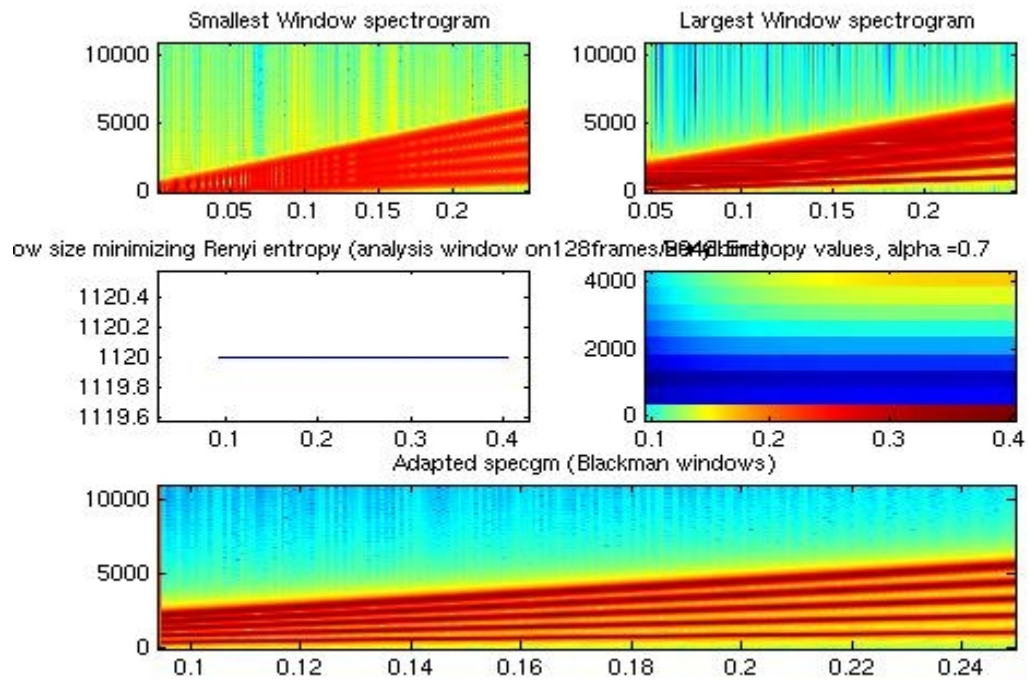


Figure 3c: adaptation locale pour harmoniques avec modulation linéaire

Ici encore, le critère semble choisir un bon compromis qui permet de distinguer les différentes harmoniques.

Les résultats obtenus avec l'étalement d'énergie sont tout à fait similaires.



## Vibrato

Le signal est composé d'harmoniques avec modulation sinusoïdale, proche d'un vibrato naturel (période de  $\sim 0.1$  s).

## Pseudo-entropie de Rényi

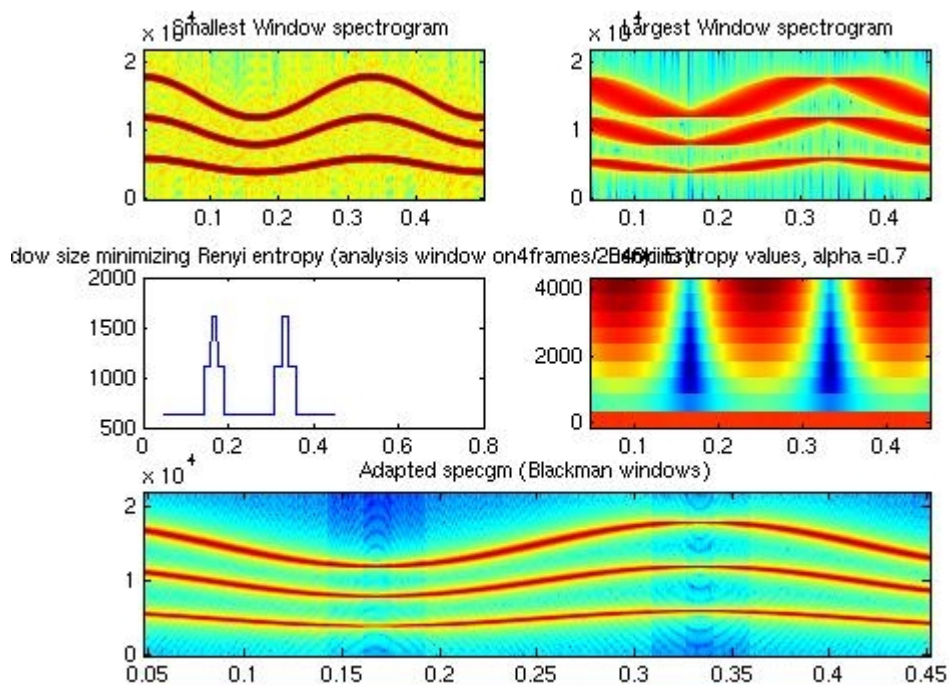


Figure 4a : adaptation des tailles de fenêtre minimisant localement pour vibrato de synthèse

Pour ce signal, le critère paraît tout à fait pertinent, choisissant des grandes tailles de fenêtre là où le signal est peu modulant. On a une plus grande liberté de partition d'analyse que dans le cas des transitoires, l'analyse étant d'autant plus efficace que l'on choisit une partition adaptée au structures du signal (séparant en fréquence le harmoniques et temporellement les temps caractéristiques de modulation).

Le critère d'étalement en énergie donne encore des résultats similaires.

## Signaux bruités

Comme nous l'avons constaté, les critères de choix semblent très sensibles au pics non sinusoïdaux du spectre, et il est devenu nécessaire d'ajuster la partition d'analyse en conséquence.

Dans le cas de signaux bruités, l'étalement en énergie, donnant d'autant plus de poids aux amplitudes faibles et moyennes, devient un mauvais critère, quelle que soit la partition d'analyse:

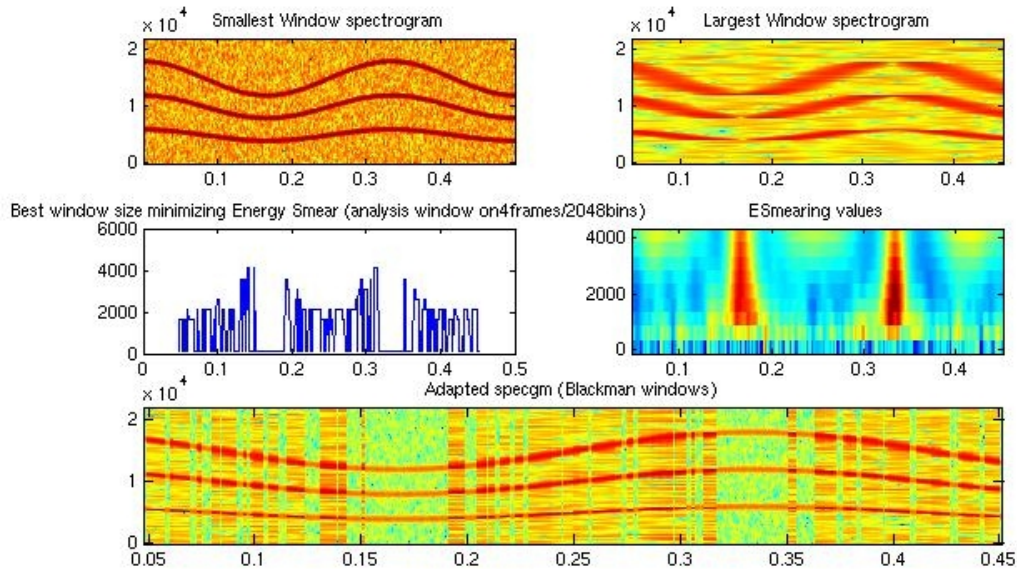


Figure 5a: adaptation minimisant localement l'étalement d'énergie pour un signal de vibrato bruité (bruit blanc gaussien SNR = 10)

De même, le critère de minimisation d'entropie Rényi est fortement affecté par le bruit pour  $\alpha < 1$  (cas où on donne davantage de poids aux faibles amplitudes)

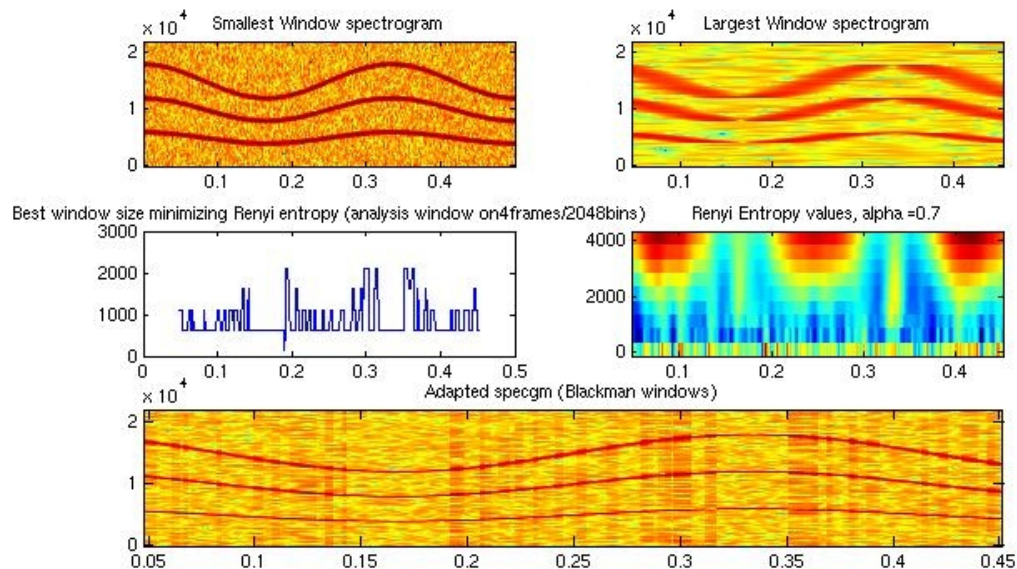


Figure 5b : adaptation minimisant localement l'entropie Rényi pour un signal de vibrato bruité

En revanche, pour  $\alpha > 1$ , on cherche davantage à maximiser les grandes

amplitudes; on a donc une meilleure adaptation en présence de bruit:

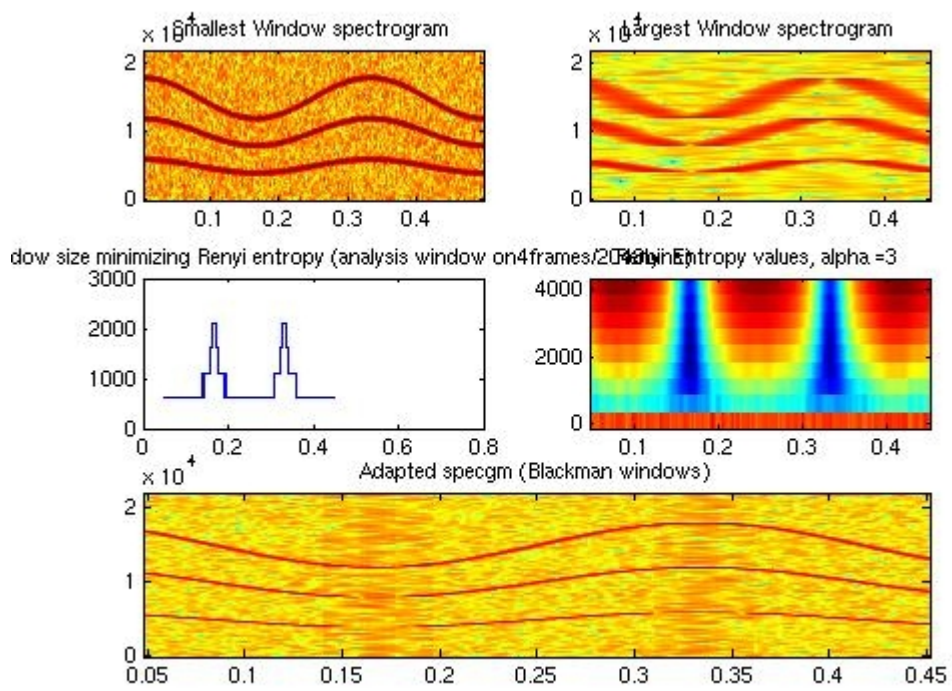


Figure 5c: adaptation minimisant localement l'entropie de Rényi ( $\alpha > 1$ ) pour signal de vibrato bruité

Les critères de choix de fenêtre ont également été appliqués aux spectrogrammes réassignés, sans parvenir à des résultats vraiment concluants. Dans ce cas, l'interprétation de la localisation de la représentation suivant les tailles de fenêtres est délicate; une approche d'adaptation par distribution d'amplitudes n'est pas forcément à privilégier.

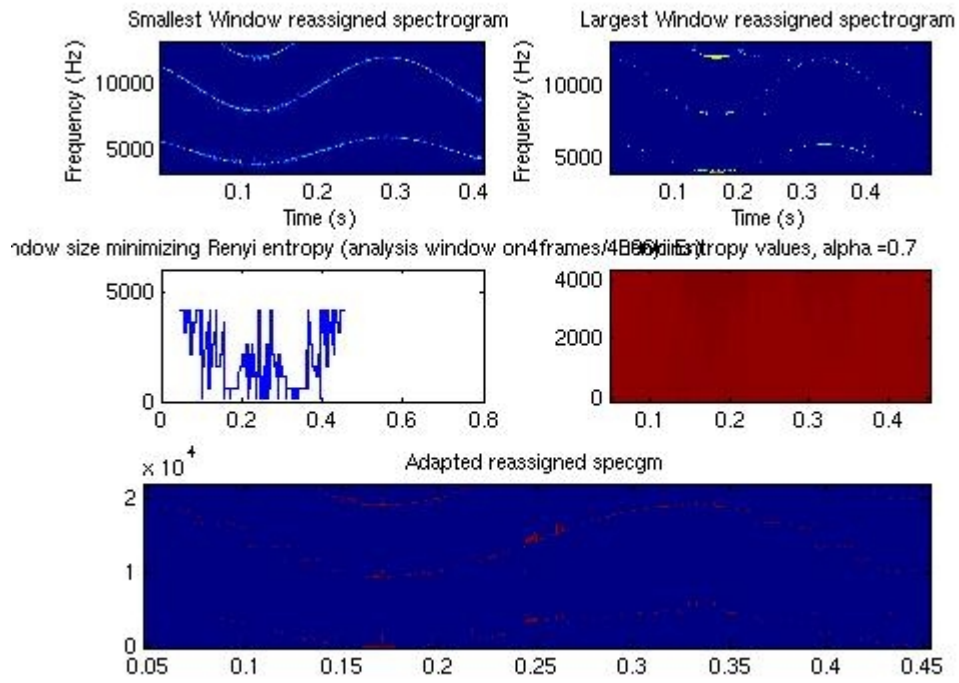


Figure 6 : adaptation locale de la taille de fenêtre minimisant l'entropie de Rényi sur le spectrogramme réassigné (vibrato de synthèse)



## Signaux réels

Les sons de synthèse testés jusqu'à présent ne permettent pas de réellement appréhender le comportement des critères choisis pour des signaux réels. La richesse de ces signaux rend difficile la paramétrisation de l'analyse: partition d'analyse et puissance à adopter. Dans le cadre de ce stage, nous avons simplement considéré une partition en temps fixe, testée sur des signaux de différentes natures.

Pour un signal de glockenspiel monophonique, dont les « composantes » transitoires/harmoniques non modulées sont relativement bien distinctes dans le plan temps fréquence, le choix des tailles de fenêtres selon une partition temporelle fixée a priori semble alors pertinent. On est alors dans une situation tout à fait analogue à celle étudiée pour les pics de diracs:

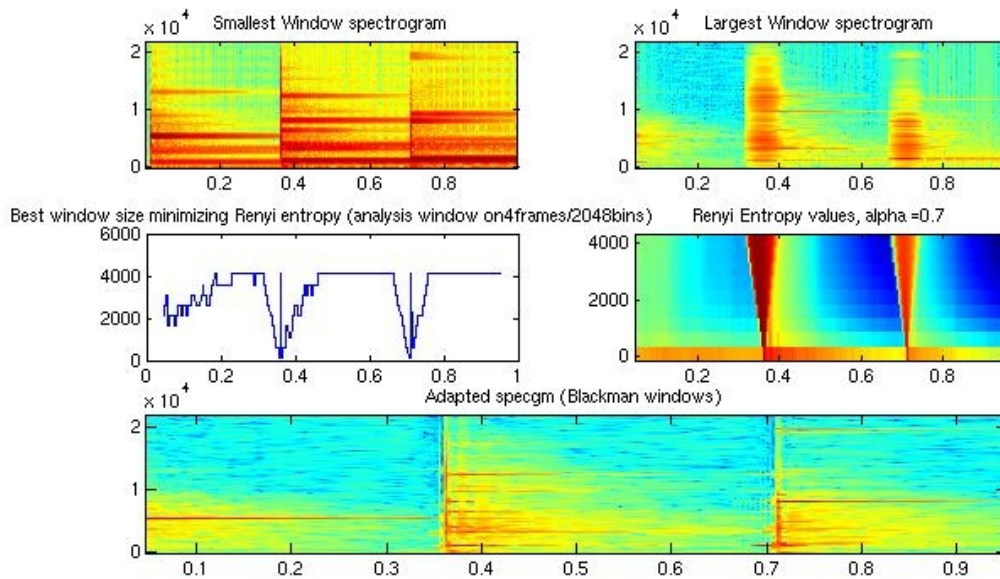


Figure 7a : adaptation de la fenêtre minimisant localement l'entropie de Rényi pour un signal de glockenspiel

Le spectrogramme adapté offre ainsi une très bonne résolution fréquentielle au niveau des harmoniques et une bonne résolution temporelle au niveau de al transitoire.

Dans le cas d'un vibrato instrumental (Fig. 7b), les harmoniques restent bien distinguées par des grandes fenêtres; on a également une variation d'amplitude qui donne un abaissement de la taille fenêtre pour  $\alpha < 1$  (moins d'harmoniques présentes de manière locale) :

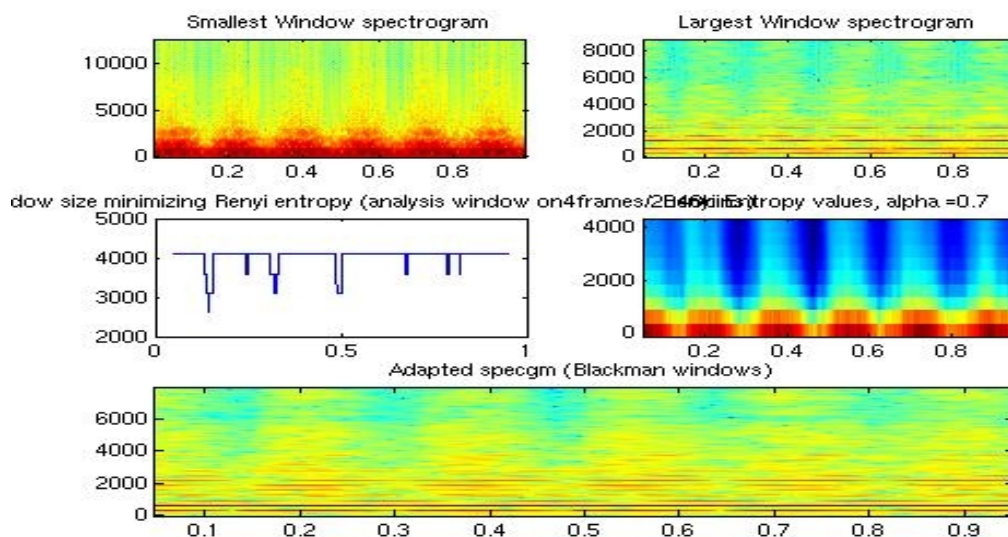


Figure 7a : adaptation de la fenêtre minimisant localement l'entropie de Rényi pour un vibrato de flûte

En revanche, dans le cas d'un vibrato chanté, le contenu spectral est beaucoup plus riche et évolutif, le vibrato souvent plus prononcé. On a donc plus de mal à obtenir une représentation adaptée pertinente, d'autant plus avec une partition d'analyse fixée a priori sans tenir compte des caractéristiques « locales » du signal :

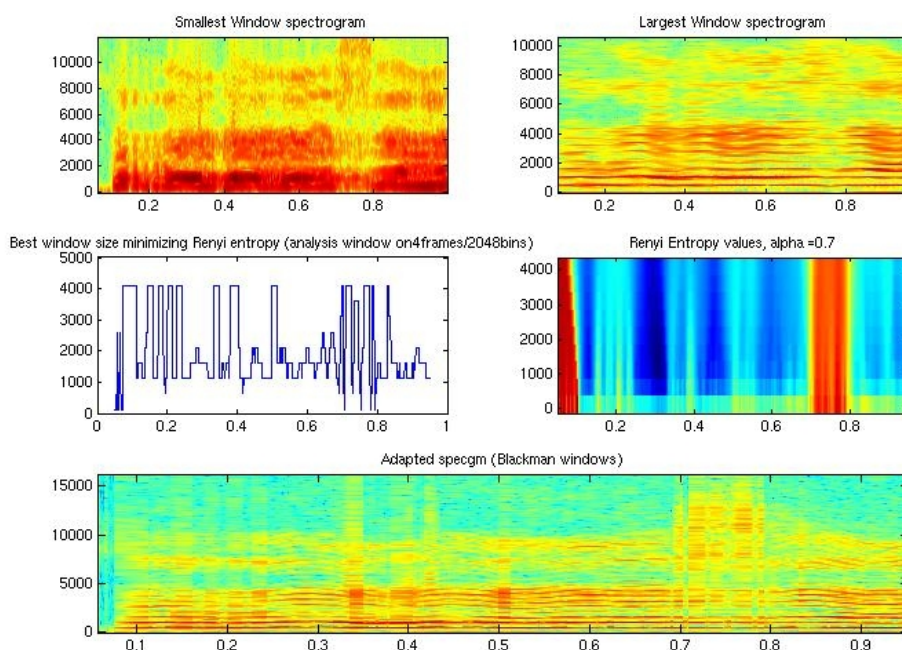


Figure 7a : adaptation de la fenêtre minimisant localement l'entropie de Rényi pour un vibrato chanté

Même si dans le spectrogramme adapté on peut distinguer les harmoniques, on ne tire réellement aucun avantage par rapport à une représentation avec une fenêtre de grande taille. Les valeurs d'entropie sont localement très proches et le choix de fenêtre en est d'autant moins déterminé. Pour ce signal, il paraît même délicat de pouvoir définir une partition d'analyse sans privilégier a priori certaines composantes et certaines fenêtres. De même, les signaux de paroles donnent des choix de fenêtre très imprévisibles.

## Adaptation par optimisation du noyau sur une distribution de Wigner Ville

Cette approche, inspirée de [11], consiste à choisir un noyau de lissage optimal, se concentrant au mieux localement sur une composante, pour une représentation de la classe Cohen. Par passage à la transformée de Fourier 2D, on a vu qu'on pouvait définir une distribution de la classe Cohen comme un lissage par le noyau  $\Phi$  de la fonction d'ambiguïté du signal:

$$P(t, \omega) = \frac{1}{4\pi^2} \iint \left( \int s(u + \frac{\tau}{2}) \bar{s}(u - \frac{\tau}{2}) e^{j\theta u} du \right) \Phi(\theta, \tau) e^{-j\theta t - j\tau \omega} d\theta d\tau \quad (10)$$

Or la fonction d'ambiguïté a la propriété de concentrer toutes les « auto-composants » (composants quadratiques) au centre de la distribution, tandis que les composantes d'interférences, produisant des variations d'amplitude très rapides sur la distribution Wigner-Ville, sont rejetées à vers les fréquences élevées de la fonction d'ambiguïté [12]. L'idée de Jones est de construire un noyau radial maximisant la corrélation avec les composants quadratiques sur la fonction d'ambiguïté en résolvant en coordonnées polaires le problème d'optimisation suivant :

$$\max_{\Phi} \int_0^{2\pi} \int_0^{\infty} |A(r, \psi) \Phi(r; \psi)|^2 dr d\psi \quad (11)$$

Le noyau étant soumis lui même à une contrainte de volume.

Si on interprète le spectrogramme comme une distribution de la classe Cohen, le noyau détermine la fonction d'ambiguïté de la fenêtre choisie. On peut donc, au lieu de résoudre le problème d'optimisation, mener un choix sur un panel de fenêtres, comme dans les approches proposées plus haut.

Pour des fenêtres gaussiennes, le noyau de lissage correspondant est également gaussien:

$$f(t) = \frac{1}{(\sigma^2 \pi)^{1/4}} e^{-\frac{t^2}{2\sigma^2}}$$

$$\rightarrow Af(\tau, \gamma) = \int f(u + \frac{\tau}{2}) f(u - \frac{\tau}{2}) e^{-iyu} du = \frac{e^{-\frac{\tau^2}{4\sigma^2}}}{\sigma \sqrt{\pi}} \int e^{-\frac{u^2}{\sigma^2}} e^{-iyu} du = e^{-\frac{\tau^2}{4\sigma^2}} e^{-\frac{\sigma^2 \gamma^2}{4}}$$

on peut donc réaliser facilement un choix d'étalement de fenêtre, sans passer par les coordonnées polaires.

Cependant, l'approche ne proposée ne revêt pas encore de caractère local pour s'adapter selon les non stationnarités du signal. Pour ce faire Jones et Baraniuk introduisent la fonction d'ambiguïté à court terme:

$$A(t; \theta, \tau) = \int s^*(u - \frac{\tau}{2}) w^*(u - t - \frac{\tau}{2}) s(u + \frac{\tau}{2}) w(u - t + \frac{\tau}{2}) e^{j\theta u} du \quad (12)$$

où  $w$  est une fenêtre rectangulaire donnée à priori, qui joue le même rôle que la partition d'analyse pour les précédents algorithmes.

Dans le cadre de ce stage, seule l'approche globale a été implémentée à partir du code de Jones et Baraniuk (disponible sur [www.dsp.rice.edu/software](http://www.dsp.rice.edu/software)); elle donne comme plus haut une représentation satisfaisante pour des signaux simples.

Cependant, les algorithmes discrets, non discutés ici, sont très coûteux, même pour une approche à court terme. Avant de s'engager plus avant dans cette voie pour des travaux ultérieurs, il sera nécessaire de bien comprendre l'influence de la partition d'analyse.

## Conclusion

Les critères d'adaptation expérimentés au cours de ce stage revêtent des caractéristiques relativement similaires, et peuvent donner de bonnes multi-représentations pour certaines catégories de signaux (successions d'harmoniques et de transitoires d'attaques).

En revanche, pour des signaux plus complexes, tels que des signaux chantés ou parlés, dont les variations fréquentielles sont rapides, le problème crucial du caractère local, à savoir de la partition considérée pour comparer les distributions temps-fréquence, reste à résoudre.

D'autres approches de la localisation dans le plan temps fréquence tels que la réallocation multi-fenêtres [13], le cadre d'optimisation plus général des algorithmes de poursuites ou l'adaptation des atomes suivant la fonction d'auto-correlation [14] sont également à envisager pour des travaux ultérieurs.

## Références

- [1] T. Gardner, M. Magnasco , *Sparse time frequency representations*
- [2] Lukin, Todd, *Adaptative Time-Frequency Resolution for Analysis and Processing of Audio*, AES Convention Paper, May 2006
- [3] R. R. Coifman, M. V. Wickerhauser. *Entropy based algorithms for best basis selection*. IEEE Trans. Info. Theory, mars 1992
- [4] G.M Davis, S. Mallat, M.Avelanedo. *Greedy adaptative approximations*. J. of Constr. Approx., 13 :57-98,1997
- [5] P. Flandrin, *Ambiguity functions* , Time Frequency Signal Analysis and Processing, 2003
- [6] P. Flandrin, *Localisation dans le plan temps fréquence*
- [7] P. Flandrin, *Représentations temps-fréquence des signaux non stationnaires*
- [8] Jones, Baraniuk *A simple scheme for adapting time frequency representations*
- [9] Baraniuk, Flandrin, Janssen, Michel, *Measuring Time-Frequency Information Content Using the Rényi Entropies*
- [10] A. Roebel, M. Zivanovic, X Rodet, *Signal decomposition by means of spectral peaks*
- [11] Jones, Baraniuk, *An adaptative Optimal-Kernel Time-frequency Representation*, IEEE transaction on Signal Processing, vol.43 oct 1995
- [12] Flandrin, *Some features of time-frequency representations of multicomponents signals*, IEEE Int. Conf. Acoust., Speech, Signal Processing, 1984
- [13] J. Xiao et P. Flandrin, "Multitaper time-frequency reassignment," Proc. EUSIPCO-06, Florence (I), September 2006
- [14] Greïtans, *Adaptative STFT-like Time-Frequency analysis from arbitrary distributed signal samples*

## **Ouvrages généraux sur les représentations temps fréquence et le traitement du signal:**

L.Cohen, *Time Frequency Analysis*. Editions Prentice-Hall.

P. Flandrin, *Temps Fréquence*. Editions Hermès.

S. Mallat, *Une exploration des signaux en ondelettes*. Editions de l'Ecole Polytechnique.