

# Fenêtre et grille optimales pour la transformée de Gabor

## Exemples d'application à l'analyse audio

H. LACHAMBRE<sup>1</sup>, B. RICAUD<sup>2</sup>, G. STEMPFEL<sup>3</sup>, B. TORRESANI<sup>3</sup>, C. WIESMEYR<sup>4</sup>, D. M. ONCHIS<sup>5</sup>

<sup>1</sup>Genesis, Domaine du Petit Arbois, 13545 Aix-en-Provence, France

<sup>2</sup>Signal Processing Laboratory 2, Ecole Polytechnique Fédérale de Lausanne (EPFL), Station 11, 1015 Lausanne, Switzerland

<sup>3</sup>Aix-Marseille Université, CNRS, Centrale Marseille, LATP, UMR 7353, 13453 Marseille, France

<sup>4</sup>Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Wien, Austria

<sup>5</sup>West University of Timisoara, Blvd. V. Parvan 4, Timisoara, 300223, Romania

helene.lachambre@genesis.fr, benjamin.ricaud@epfl.ch, guillaume.stempfel@genesis.fr  
bruno.torresani@univ-amu.fr, christoph.wiesmeyr@univie.ac.at, donchis@hpc.uvt.ro

**Résumé** – Cet article présente l'utilisation d'une grille optimale et d'une fenêtre optimale pour le calcul de la transformée de Gabor discrète. Dans le cas d'une Gaussienne généralisée, nous étendons des travaux précédents et proposons une fenêtre localement optimale pour des signaux non-stationnaires. Nous présentons des résultats sur trois problèmes d'analyse temps-fréquence, sur des signaux réels et synthétiques : la distinction de composantes temps-fréquence proches, l'estimation de fréquence instantané et l'estimation du Rapport Signal à Bruit.

**Abstract** – This article deals with the use of optimal lattice and optimal window in Discrete Gabor Transform computation. In the case of a generalized Gaussian window, extending earlier contributions, we introduce an additional local window adaptation technique for non-stationary signals. We illustrate our approach and the earlier one by addressing three time-frequency problems: close frequencies distinction, frequency estimation and Signal to Noise Ratio estimation. The results are presented, when possible, with real world audio signals.

## 1 Introduction

Notre travail se place dans le contexte général de l'analyse Temps-Fréquence (T-F) du signal [3]. La représentation T-F est d'une extrême utilité pour l'analyse de signaux non stationnaires, que ce soit pour la recherche scientifique ou dans l'industrie. Cependant, cette représentation et sa précision, à travers la transformation de Gabor discrète (DGT), dépend du choix d'une fenêtre [2] et d'une grille temps fréquence. Chaque choix permet de visualiser avec précision un seul type de composant T-F et n'est donc pas adapté aux différents types pouvant survenir successivement dans des signaux non-stationnaires.

Un travail précédent [5] présentait la recherche de la fenêtre Gaussienne réelle optimale. Le critère d'optimisation était la minimisation de l'entropie de Renyi [6] ou de Shannon.

Dans une étude récente [7], nous avons défini et calculé la fenêtre Gaussienne généralisée optimale pour un signal donné. Le critère d'optimisation est la maximisation de la concentration de l'énergie dans le domaine T-F, ou la parcimonie de la représentation T-F. Nous proposons aussi une optimisation des paramètres de la grille. Cette étude, axée sur la théorie, ne permet de voir que partiellement l'utilité réelle de l'approche.

Par cette contribution, nous souhaitons mettre en valeur les résultats de [7] en la complétant. Premièrement, nous proposons un adaptation *locale* de la fenêtre, évoluant avec les dif-

férents composants fréquentiels rencontrés dans un signal non-stationnaire. Ceci permet la création d'un nouveau type de spectrogramme s'adaptant au signal au cours du temps. Deuxièmement, nous démontrons d'un point de vue pratique l'utilité de ces fenêtres et grilles (globalement ou localement) optimales. Nous nous concentrons sur la représentation T-F et sa précision à des fins d'analyse (et non de synthèse). Nous présentons dans la mesure du possible les résultats sur des signaux réels. La lecture de ce nouveau spectrogramme adapté est plus claire quand les composants fréquentiels du signal évoluent avec le temps.

Dans la suite, nous rappelons brièvement dans la partie 2 le principe de notre approche et présentons l'adaptation locale de la fenêtre. Puis dans la partie 3 nous présentons trois applications sur des signaux réels et synthétiques.

## 2 Aspects théoriques

La théorie complète pour l'optimisation de fenêtre et de grille a été présentée dans un article distinct [7]. Nous n'en rappelons que les principaux points. Nous travaillons dans l'espace des signaux de dimension finie  $\mathbb{C}^N$ . Soit  $g \in \mathbb{C}^N$  une fenêtre d'analyse et  $\Lambda \subset \mathbb{Z}^N \times \mathbb{Z}^N$  une grille T-F. La DGT du signal  $f \in \mathbb{C}^N$  est définie comme  $\mathcal{V}_g f(x, \xi) = \sum_{n=0}^N f(n)g(n-x)e^{-j2\pi\xi n/N}$ , avec  $(x, \xi) \in \Lambda$  un point de la grille.

## 2.1 Optimisation de la fenêtre

Dans un premier temps, notre but est de trouver la fenêtre optimale en termes de concentration de l'énergie de la DGT [5]. Pour mesurer cette concentration, nous considérons la norme  $l_p$  de la DGT. La norme  $l_p$  est une mesure d'étalement pour  $p < 2$ , et de parcimonie pour  $p > 2$ . Ici, nous choisissons de maximiser la norme  $l_p$  avec  $p > 2$  ( $p = 2.5$ ), ce qui mène à un critère différentiable [7].

Le problème que nous devons résoudre s'écrit :

$$\operatorname{Argmax}_{g, \|g\|_2=1} \|\mathcal{V}_g f\|_p^p = \operatorname{Argmax}_{g, \|g\|_2=1} \sum_{\xi, x \in \Lambda} |\mathcal{V}_g f(x, \xi)|^p \quad (1)$$

De manière à obtenir une fenêtre interprétable, nous restreignons notre problème à la famille des Gaussiennes «chirpées» :

$$g_{\sigma, s}(t) = \sqrt[4]{\frac{2}{N\sigma}} e^{-\pi \frac{t^2}{N\sigma} + j\pi s t^2 \frac{N+1}{N}} \quad (2)$$

avec  $\sigma > 0$  l'étalement dans le temps,  $s \in \mathbb{R}$  le paramètre de «chirp» et  $N$  la longueur du signal. Le problème à optimiser devient alors :

$$\operatorname{Argmax}_{\sigma, s} \|\mathcal{V}_{g_{\sigma, s}} f\|_p^p. \quad (3)$$

L'optimisation est réalisée en utilisant la méthode BFGS [4], qui est une approche Quasi-Newton.

## 2.2 Optimisation de la grille

Le second problème est de déterminer, pour une redondance  $R = |\Lambda|/N$  donnée, la meilleure grille, c'est à dire le placement optimal des atomes temps-fréquence.

Une grille correctement choisie est définie comme une grille qui mène à une DGT inverse stable numériquement. La grille optimale est donc celle qui donne la DGT inverse la plus stable.

En partant de l'hypothèse que la grille hexagonale est optimale pour une Gaussienne standard [8], alors la grille optimale pour la fenêtre  $g_{\sigma, s}$  optimale est [7] :

$$\Lambda = \sqrt{\frac{N}{R}} \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix} \begin{pmatrix} \sqrt{\sigma} & 0 \\ 0 & 1/\sqrt{\sigma} \end{pmatrix} \begin{pmatrix} \frac{\sqrt[4]{3}}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt[4]{3}\sqrt{2}} & \frac{\sqrt{2}}{\sqrt[4]{3}} \end{pmatrix} \mathbb{Z}_N^2 \quad (4)$$

où nous avons, de gauche à droite, les matrices de «shearing» (ou cisaillement), de dilatation et d'échantillonnage hexagonal. La constante à gauche donne la bonne redondance.

## 2.3 Optimisation locale de la fenêtre

Une fenêtre d'analyse unique n'est pas forcément adaptée sur la totalité du signal (voir les figures 1.a et 1.b), mais uniquement sur une région T-F limitée. Dans ce travail, nous proposons d'adapter localement la fenêtre d'analyse, en utilisant une grille constante pour la totalité du signal. En partant de l'hypothèse que dans les signaux réels l'évolution temporelle des amplitudes et des fréquences sont régulières, nous cherchons une fenêtre qui évolue régulièrement au cours du temps. Nous estimons les paramètres  $\sigma_t$  et  $s_t$  de la façon suivante :

- partitionner le signal en  $M$  segments,
- sur chaque segment, trouver la fenêtre optimale comme proposé dans la partie 2.1,
- considérer que cette fenêtre est optimale pour la trame centrale de chaque segment, et interpoler les paramètres  $\sigma$  et  $s$  sur toutes les trames,
- calculer la DGT, en considérant une fenêtre différente à chaque trame.

Notons que la DGT ainsi calculée est en réalité une transformée de Gabor non-stationnaire [1], et que son inversibilité n'est pas acquise *a priori*. Notons également que la première étape est actuellement réalisée manuellement.

## 3 Applications

Nous évaluons les avantages de l'approche proposée sur trois applications : la distinction de composantes temps-fréquences (T-F) proches, l'estimation de la fréquence instantanée, et du rapport signal à bruit.

Pour chaque application, nous proposons une méthode naïve basée sur l'analyse de la DGT. Nous comparons alors les performances obtenues par cette méthode, avec des DGTs calculées avec différentes fenêtres et sur différentes grilles. La grille optimale sera comparée à d'autres grilles de même redondance. La fenêtre Gaussienne chirpée optimale sera comparée à la fenêtre Gaussienne réelle optimale. Cette dernière est estimée de manière similaire à l'équation 3, avec un seul paramètre (la dilatation) pour l'optimisation :

$$\operatorname{Argmax}_{\sigma} \|\mathcal{V}_{g_{\sigma}}(f)\|_p^p. \quad (5)$$

### 3.1 Distinction de composantes T-F proches

Dans cette première application, la validation de l'approche est faite visuellement. L'exemple présenté sur les figures 1 et 2 est l'enregistrement d'un passage d'avion. Ce passage produit deux fréquences très proches qui varient au cours du temps (effet Doppler). Sur la figure 1.b, le signal est analysé avec la fenêtre Gaussienne réelle, optimisée pour le début du signal (changement de fréquence lent). Sur la figure 1.a, la fenêtre est la Gaussienne chirpée optimale, optimale pour la fin du signal (changement de fréquence rapide). Enfin, sur la figure 1.c, la fenêtre est adaptée localement (avec un découpage du signal en segments de une seconde) : les deux composantes sont visibles tout du long. Ces DGTs ont été calculées avec la même grille.

La figure 2 montre l'intérêt d'utiliser la grille optimale. Le même signal d'avion est analysé avec la fenêtre optimale et trois grilles de même redondance : la grille optimale (2.a), une grille avec plus de points en temps et moins en fréquence (2.b), une grille avec moins de points en temps et plus en fréquence (2.c). Dans les cas (b) et (c), le manque de résolution fréquentielle ou temporelle fait que les deux composantes proches ne sont pas visibles.

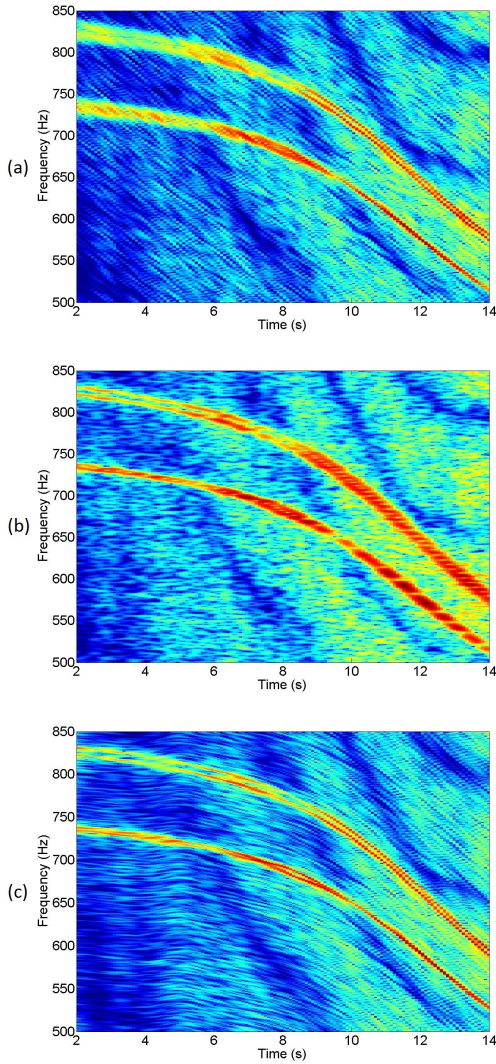


FIGURE 1 – Passage d’avion, présentant deux fréquences proches. (a) Gaussienne chirpée optimale, (b) Gaussienne réelle optimale, (c) Gaussienne chirpée localement optimale.

### 3.2 Estimation de fréquence instantanée

Dans ce deuxième exemple, la tâche que nous nous fixons est de suivre une unique fréquence. Ainsi, nous proposons d’estimer, sur chaque trame, la fréquence cherchée comme celle correspondant au maximum d’amplitude. Ici, le signal est un enregistrement de nombreuses bulles qui explosent dans de l’eau. Le son de l’explosion de chaque bulle produit un motif très reconnaissable dans le plan T-F (voir la figure 3). Sur la figure 3.a, la DGT est calculée avec la fenêtre Gaussienne réelle optimale. Sur la figure 3.b, elle est calculée avec la fenêtre «chirpée» optimale. Sur chaque figure, la fréquence estimée est mise en évidence à l’aide de points noirs. Sur la figure 3.a, contrairement à ce attendu, les motifs ne sont ni similaires ni réguliers. Au contraire, sur la figure 3.b, les motifs sont similaires et réguliers. Comme ce motif dans le plan T-F ne change pas au cours du temps, une adaptation locale de la fenêtre n’apporte-

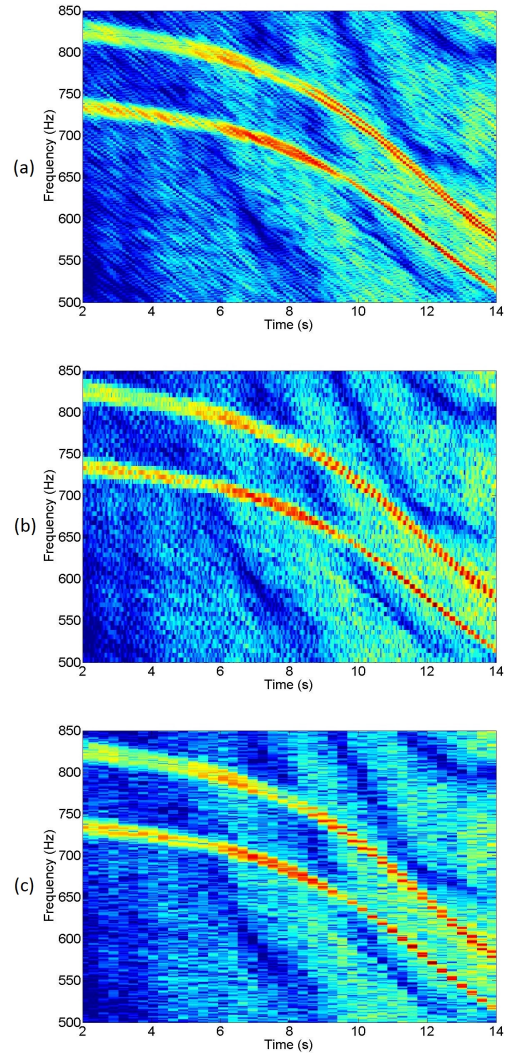


FIGURE 2 – DGTs d’un passage d’avion, calculées avec les mêmes fenêtre et redondance, mais des grilles différentes : (a) Grille optimale, (b) Plus de points en temps, moins en fréquence, (c) Moins de points en temps, plus en fréquence.

rait pas d’amélioration significative de la représentation. La fenêtre optimale permet donc de révéler la signature acoustique des bulles, ainsi que le suivi de leur fréquence instantanée.

### 3.3 Estimation du rapport signal à bruit

Pour cette dernière application, nous estimons le rapport signal à bruit (RSB) d’un signal bruité. Dans notre exemple, le signal est un chirp quadratique dont l’amplitude décroît de manière log-linéaire. Le bruit est un bruit blanc Gaussien constant. Nous avons choisi un signal artificiel dans ce cas, pour disposer des vraies valeurs et pouvoir mesurer les performances.

Comme l’estimation de la puissance du bruit ne dépend pas de la fenêtre (tant que sa norme est unitaire, ce qui est vrai ici), nous l’estimons donc directement comme la valeur moyenne de la partie non harmonique du spectre.

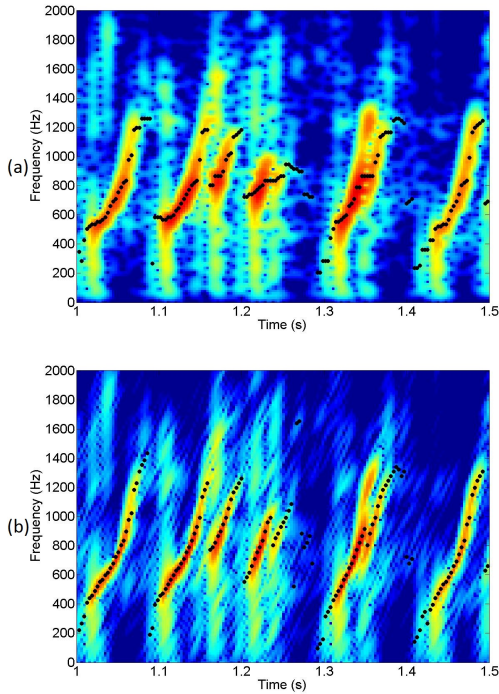


FIGURE 3 – Bulles dans de l’eau. La fréquence estimée est mise en évidence à l’aide des points noirs. (a) Gaussienne réelle optimale, (b) Gaussienne chirpée optimale.

La puissance du chirp quadratique est proportionnelle à l’amplitude maximum du spectre. Le facteur de normalisation à appliquer dépend de la fenêtre, il est estimé avec un signal connu (ici un sinus pur d’amplitude connue).

Avec ces deux valeurs (puissance instantanée du chirp et puissance du bruit), nous alors pouvons calculer le RSB instantané. Les courbes de RSB calculées avec différentes fenêtres, de même que le RSB de référence, sont présentées sur la figure 4. La figure montre clairement que le meilleur estimateur est obtenu avec la fenêtre optimale adaptée localement.

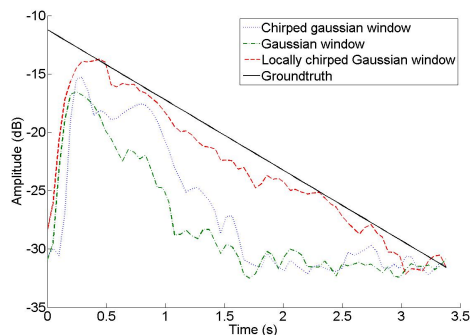


FIGURE 4 – Chirp quadratique d’amplitude décroissant log-linéairement avec un bruit blanc Gaussien constant. Le RSB est estimé sur des DGTs calculées avec diverses fenêtres.

## 4 Conclusion et perspectives

Nous avons présenté ici l’application d’une procédure d’optimisation de la fenêtre et de la grille de calcul de la DGT dans trois cas réels : la distinction de fréquences proches, l’estimation de fréquence, l’estimation du RSB. Pour les deux premiers cas, nous présentons des résultats qualitatifs sur des signaux réels, dans le dernier cas des résultats quantitatifs sur un signal synthétique. L’utilisation de la fenêtre et la grille optimales mène à des DGTs plus interprétables pour ces applications.

Le spectrogramme adapté présenté ici apporte une vraie plus value par rapport au spectrogramme standard. Notre approche est à la fois adaptée automatiquement, rapide à implémenter et précise dans la discrimination fréquentielle de signaux non-stationnaires.

## 5 Remerciements

Ces travaux ont été réalisés dans le cadre du projet EU FET Open grant UNLocX (255931), et du projet EU grant no. FP7-REGPOT-CT-2011-284595 (HOST) pour le dernier auteur.

## Références

- [1] P. Balazs, M. Dorfler, F. Jaillet, N. Holighaus, and G. A. Velasco. Theory, Implementation and Applications of Nonstationary Gabor Frames. *J. of Computational and Applied Mathematics*, 236(6) :1481–1496, 2011.
- [2] B. Boashash. *Time-Frequency Signal Analysis and Processing : A Comprehensive Reference*. Elsevier, 2003.
- [3] P. Flandrin. *Time-Frequency/Time-Scale Analysis*. Academic Press, San Diego, 1999.
- [4] D. Goldfarb. A Family of Variable-Metric Methods Derived by Variational Means. *Mathematics of Computation*, 24(109) :23–26, 1970.
- [5] F. Jaillet and B. Torrèsani. Time-Frequency Jigsaw Puzzle : Adaptive multiwindow and multilayered Gabor expansions. *International Journal of Wavelets Multiresolution and Information Processing*, 5(2) :293–315, 2007.
- [6] A. Rényi. On Measures of Entropy and Information. In *4th Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 547–561, 1960.
- [7] B. Ricaud, G. Stempf, B. Torrèsani, C. Wiesmeyr, H. Lachambre, and D. Onchis. An optimally concentrated Gabor transform for localized time-frequency components. *Advances in Computational Mathematics*, 40(3) :683–702, 2013.
- [8] C. Wiesmeyr, N. Holighaus, and P. Sondergaard. Efficient Algorithms for Discrete Gabor Transform on a Non-separable Lattice. *IEEE Trans. on Signal Processing*, 61(20) :5131–5142, 2013.