

# Estimation conjointe modulations de fréquence/spectre, application à l'analyse de signaux audio

Harold OMER, Bruno TORRÉSANI

Aix-Marseille Université, CNRS, Centrale Marseille, LATP, UMR 7353, 13453 Marseille Cedex 13, France

Harold.Omer@centrale-marseille.fr, Bruno.Torresani@univ-amu.fr

**Résumé** – Nous considérons le problème d'estimation conjointe de la modulation et de la densité spectrale de puissance d'un signal stationnaire (large bande) transformé par une modulation de fréquence inconnue. Ce travail est motivé par des applications en analyse de sons, dans lesquelles une fluctuation de fréquence instantanée peut traduire une modification de vitesse d'un système (mécanique par exemple) à l'origine du signal.

Les techniques d'estimation de fréquence instantanée ont été largement étudiées dans le domaine des télécommunications, mais se focalisent souvent sur des porteuses à bande étroite. L'approche proposée ici se base sur une modélisation explicite sous la forme *timbre*×*dynamique*, faisant intervenir une porteuse Gaussienne stationnaire. L'estimation est basée sur une analyse temps-fréquence et des approximations valables pour une modulation lentement variable.

L'algorithme proposé se base sur une modélisation d'un signal stationnaire sous-jacent pas un signal stationnaire Gaussien de spectre inconnu, modifié par une modulation de fréquence inconnue. Nous obtenons une forme approchée pour la covariance de la transformée de Gabor de tels signaux, exploitée ensuite dans des estimateurs du maximum de vraisemblance.

**Abstract** – The problem of joint estimation of the modulation and power spectrum of a (wideband) signal obtained by modulating a weakly stationary random process. This work is motivated by audio applications, in which time fluctuations of the spectrum can originate from speed fluctuations of some underlying system (for instance a mechanical system).

Instantaneous frequency estimation methods have received very significant attention, for example in telecommunication sciences; however there were mainly limited to the case of narrow band carrier signals. The proposed approach is based upon an explicit modeling of the form *timbre*×*dynamics*, involving a Gaussian stationary random carrier. The estimation relies on time-frequency analysis, together with approximations valid for slowly varying frequency modulations.

The proposed algorithm models the underlying stationary signal as a Gaussian random signal with unknown spectrum, modified by some unknown frequency modulation. We derive an approximate expression for the covariance of fixed time Gabor transform of the modified signal, that is further exploited within maximum likelihood estimators.

## 1 Introduction, position du problème

Les modèles temps-fréquence usuels pour les signaux audio sont généralement basés sur des décompositions de ces signaux sur des dictionnaires temps-fréquence. Ces approches sont pertinentes pour certaines classes de signaux (parole, musique,...) pour lesquels certaines propriétés (localisation temps-fréquence, parcimonie) peuvent être exploitées. Elles le sont moins pour des sons moins structurés tels que les sons large bande (par exemple sons d'environnement, bruits de moteurs,...). Nous nous focalisons sur une approche alternative, destinée à modéliser des sons non-stationnaires, dont la non-stationnarité peut porter une information pertinente (comme par exemple une information de vitesse dans le cas de bruits de moteur en phase d'accélération). Le modèle est de la forme

*timbre*×*dynamique*

où un signal stationnaire (de spectre de puissance inconnu) est modulé par une transformation dynamique. Le problème est l'estimation conjointe de la modulation et du spectre à partir d'une observation.

Dans le cas où le spectre du signal stationnaire sous-jacent

est un spectre à bande étroite, voire discret, c'est à dire lorsque le signal peut être modélisé comme une somme de composantes sinusoïdales modulées en fréquence, ce problème se ramène au problème classique d'estimation de fréquence instantanée, pour lequel existe une abondante littérature (voir par exemple [1, 6, 5] et les références incluses). Le cas de spectres large bande a été moins étudié à notre connaissance. Nous nous limitons ici au cas particulier de signaux Gaussiens, modulés en fréquence par une fonction de modulation régulière. L'approche consiste à caractériser la distribution de la transformée de Gabor de tels signaux, pour formuler l'estimation par maximum de vraisemblance. Elle repose sur une approximation de la covariance de signaux modulés en fréquence valide dans les cas où la modulation est régulière, laquelle approximation conduit à une formulation approchée du maximum de vraisemblance, qui peut être résolue numériquement. L'algorithme fournit en sortie une estimation de la modulation, ainsi qu'une estimation du spectre du signal stationnaire sous-jacent.

La méthode est validée sur des simulations numériques, avec des exemples de spectres large bande et des spectres de raies.

## 2 Modèle et estimation

### 2.1 Modèle

On considère des signaux aléatoires Gaussiens complexes  $\mathbf{X}$  de longueur  $L$ , centrés, périodiques (i.e.  $\mathbf{X}_{t+L} = \mathbf{X}_t$ ). Nous noterons  $C_{\mathbf{X}}$  la matrice de covariance et  $R_{\mathbf{X}}$  la matrice de relation de  $\mathbf{X}$  (voir [4] par exemple), définies par

$$C_{\mathbf{X}}(t, s) = \mathbb{E} \{ X_t \bar{X}_s \}, \quad \text{et} \quad R_{\mathbf{X}}(t, s) = \mathbb{E} \{ X_t X_s \}, \quad (1)$$

et nous notons  $\mathbf{X} \sim \mathcal{CN}(0, C_{\mathbf{X}}, R_{\mathbf{X}})$ .  $\mathbf{X}$  est *circulaire* si  $R_{\mathbf{X}} = 0$ . Nous noterons enfin  $\mathbf{Z}$  le signal analytique associé à  $\mathbf{X}$ .

L'observation est la partie réelle d'une copie modulée  $\mathbf{Y}$  d'un signal stationnaire de référence  $\mathbf{X}$ , de la forme

$$Y_t = Z_t e^{2i\pi\gamma(t)} + N_t, \quad (2)$$

où  $\gamma \in \mathbb{C}^2$  est une *modulation* inconnue, régulière et lentement variable, et  $\mathbf{N} = \{N_t, t = 0, \dots, L-1\}$  est un bruit blanc complexe Gaussien circulaire de variance  $\sigma_0^2$ . Clairement, quand  $\gamma$  n'est pas constante,  $\mathbf{Y}$  n'est pas stationnaire. Le problème posé est d'estimer la modulation inconnue  $\gamma$  ainsi que le spectre de puissance  $\mathcal{S}_{\mathbf{X}}$  à partir d'une réalisation de  $\mathbf{Y}$ .

Nous nous basons pour cela sur une transformation de Gabor (finie), pour laquelle nous utilisons les définitions suivantes. Partant d'une fenêtre  $\mathbf{g}$ , la TFCT correspondante d'un signal  $\mathbf{x} \in \mathbb{C}^L$  est définie par

$$\mathcal{V}_{\mathbf{g}} \mathbf{x}(m, n) = \sum_{t=0}^{L-1} x[t] \bar{g}[t-n] e^{-2i\pi m(t-n)/L}. \quad (3)$$

Si  $a$  et  $b$  sont deux diviseurs de la longueur  $L$ , la transformée de Gabor correspondante est

$$\mathcal{G}_{\mathbf{x}}[m, n] = \mathcal{V}_{\mathbf{g}} \mathbf{x}(mb, na), \quad m \in \mathbb{Z}_M, n \in \mathbb{Z}_N, \quad (4)$$

où  $M = L/b$  et  $N = L/a$ .  $\mathcal{G}_{\mathbf{X}}$  est une matrice  $M \times N$ . Pour des choix convenables de  $\mathbf{g}$ , et pour  $a$  et  $b$  assez petits, la transformation de Gabor est inversible (voir [5, 7]), et des algorithmes efficaces sont disponibles en dimension finie (voir [9]).

### 2.2 Estimation

Notre procédure d'estimation est basée sur la propriété classique de covariance de la TFCT. La TFCT d'une copie modulée (avec fréquence de modulation constante) d'un signal est égale à une copie translatée en fréquence de la TFCT du signal original. Lorsque la fréquence modulante varie (lentement) au cours du temps, cette propriété reste vraie de façon approchée, ce qu'exprime le résultat suivant [8]

**Théorème 1** 1. Pour chaque valeur fixée de l'indice temporel  $n$ , la transformée de Gabor peut être approximée comme

$$\mathcal{G}_{\mathbf{Y}}[m, n] = \mathbf{G}^{(n; \gamma'(na)/b)}[m] + R[m], \quad \text{où} \quad (5)$$

$$\mathbf{G}^{(n; \delta)}[m] = \sum_{t=0}^{L-1} Z_t \bar{g}[t-na] e^{-2i\pi[m-\delta][t-an]/M} + \mathcal{G}_{\mathbf{N}}[m, n], \quad (6)$$

et le reste  $R[m]$  est borné par

$$\mathbb{E} \{ |R[m]|^2 \} \leq \sigma_Z^2 \left( \frac{\pi e}{L} \|\gamma''\|_{\infty} \mu_2 + 2\mu_1 \right)^2, \quad (7)$$

avec

$$\mu_1 = \sum_{t \in I_T^c} |g(t)|, \quad \mu_2 = \sum_{t \in I_T} t^2 |g(t)|, \quad T = \sqrt{\frac{L}{\pi \|\gamma''\|_{\infty}}}, \quad (8)$$

et où  $I_T = [-T, T]$  et où son complémentaire est noté  $I_T^c$ .

2. Pour  $\delta$  donné, et pour  $n$  fixé,  $\mathbf{G}^{(n; \delta)}$  est distribué suivant une loi Gaussienne complexe circulaire de matrice de covariance

$$C_{\mathbf{G}^{(n; \delta)}}[m, m'] = C_{\mathcal{G}_{\mathbf{Z}}}[m-\delta, m'-\delta] + C_{\mathcal{G}_{\mathbf{N}}}[m-\delta, m'-\delta]. \quad (9)$$

Nous utilisons un estimateur du maximum de vraisemblance approché basé sur cette approximation. L'estimation nécessite de pouvoir inverser la matrice de covariance des vecteurs  $\mathbf{G}^{(n; \delta)}$ , cette dernière est semi-définie positive, mais souvent singulière. Le résultat suivant fournit une condition suffisante sur le bruit et sur  $\mathbf{g}$  pour l'inversibilité de cette matrice.

**Proposition 1** On suppose que la fenêtre  $\mathbf{g}$  est telle que

$$K_{\mathbf{g}} := \min_{t=0 \dots L-1} \left( \sum_{k=0}^{b-1} |g[t+kM]|^2 \right) > 0. \quad (10)$$

Alors la matrice de covariance est inversible à inverse borné : pour tout  $\mathbf{x} \in \mathbb{C}^M$ ,

$$\mathbf{x}^* C_{\mathbf{G}} \mathbf{x} \geq \sigma_0^2 K_{\mathbf{g}}. \quad (11)$$

Notons que cette condition, qui pousse à considérer de faibles valeurs de  $M$ , donc à limiter la redondance fréquentielle, assure uniquement la non-singularité de la matrice de covariance, ce qui simplifie l'estimation par maximum de vraisemblance. Nous verrons cependant plus tard que la précision de l'estimation est potentiellement d'autant meilleure que le nombre  $M$  de canaux fréquentiels est élevé. Le problème est contourné en considérant une nouvelle définition de la transformée de Gabor

$$\mathcal{G}_{\mathbf{x}}^c[m, n] = \mathcal{V}_{\mathbf{g}} \mathbf{x}(mb + c, na), \quad m \in \mathbb{Z}_M, n \in \mathbb{Z}_N, \quad (12)$$

où  $c \in [0, b-1]$ . Ceci fournit un ensemble de  $b$  transformées de Gabor, correspondant chacune à un différent échantillonnage de la TFCT. Les résultats précédents restent alors valides avec cette nouvelle définition, les équations (5) et (6) devenant

$$\mathcal{G}_{\mathbf{Y}}^c[m, n] = \mathbf{G}^{(n; \gamma'(na)/b - c/b)}[m] + R, \quad (13)$$

où

$$\mathbf{G}^{(n; \delta^c)}[m] = \sum_{t=0}^{L-1} Z_t \bar{g}[t-na] e^{-2i\pi[m-\delta^c][t-an]/M} + \mathcal{G}_{\mathbf{N}}^c[m, n]. \quad (14)$$

L'équation (9) devient alors :

$$C_{\mathbf{G}^{(n; \delta^c)}}[m, m'] = C_{\mathcal{G}_{\mathbf{Z}}}[m-\delta^c, m'-\delta^c] + C_{\mathcal{G}_{\mathbf{N}}}[m-\delta^c, m'-\delta^c]. \quad (15)$$

**Estimation de la modulation.** Avec les notations ci-dessus, pour chaque indice temporel  $n$ , on notera  $\mathcal{G} = \mathcal{G}^{(n)}$ . La log-vraisemblance est donnée par

$$\mathcal{L}_\delta(\mathcal{G}) = \mathcal{G}^* (C_{\mathbf{G}(n;\delta^c)})^{-1} \mathcal{G} + \ln(\pi^M \det(C_{\mathbf{G}(n;\delta^c)})) . \quad (16)$$

et l'estimateur du maximum de vraisemblance est obtenu en minimisant cette expression. Comme,  $\det(C_{\mathbf{G}(n;\delta^c)})$  ne dépend pas du paramètre de modulation  $\delta^c$ , l'estimateur a pour expression

$$\hat{\delta}^c = \arg \min_{\delta^c} \left[ \mathcal{G}^* (C_{\mathbf{G}(n;\delta^c)})^{-1} \mathcal{G} \right] , \quad (17)$$

problème que nous résolvons par recherche exhaustive. Comme le montrent les équations (13) et (14),  $\delta^c(n) \approx (\gamma'(an) - c) / b$ , et l'estimation de  $\delta^c$  fournit ainsi une estimation quantifiée de  $\gamma'$ , à savoir  $\hat{\gamma}'(an) \in [c, b + c, 2b + c, \dots, (M - 1)b + c]$ . Notons que la condition nécessaire de la Proposition 1 implique que le paramètre  $b$  soit grand, et l'estimée de  $\gamma'$  est par conséquent fortement quantifiée. Ce problème est contourné en utilisant la collection de transformées de Gabor définies en (12), en effectuant une nouvelle recherche exhaustive sur les  $\delta^c$  :

$$\hat{\delta} = \arg \min_c \left[ \mathcal{G}^* (C_{\mathbf{G}(n;\delta^c)})^{-1} \mathcal{G} \right] . \quad (18)$$

Ainsi,  $\hat{\gamma}'(an) \in [0, L - 1]$  et l'effet de quantification sur l'estimée est fortement atténué. Notons que l'estimation de la modulation requiert la connaissance préalable de la matrice de covariance  $C_{\mathbf{G}(n;0)}$  correspondant à la transformée de Gabor du signal stationnaire bruité. Cette dernière n'est généralement pas disponible et il sera nécessaire de l'estimer.

**Estimation de la matrice de covariance.** Ayant une estimation de  $\gamma'(an)/b$  pour tout  $n$ , on peut en déduire, par interpolation, une estimation de  $\gamma$ . Le signal peut alors être démodulé via cette estimation

$$U = Y e^{-2i\pi\hat{\gamma}/L} . \quad (19)$$

Cela nous permet, après transformation de Gabor, d'en déduire une estimation de la matrice de covariance  $C_{\mathbf{G}(n;0)}$ . La méthode est décrite en détail dans [8], notons néanmoins que l'estimation de  $C_{\mathbf{G}(n;0)}$  requiert la connaissance préalable de la fonction de modulation  $\gamma$ .

**Résumé de la procédure d'estimation.** La procédure d'estimation repose sur l'alternance des deux procédures ci-dessus. Etant donné une première estimation de la modulation, nous pouvons effectuer une première estimation de la matrice de covariance, cette dernière nous permettant alors d'effectuer une nouvelle estimation de la modulation, ainsi de suite. L'itération est stoppée lorsque la norme de la différence entre deux estimées consécutives de la modulation est inférieure à un seuil fixé

$$\frac{\|\hat{\delta}^{(k)} - \hat{\delta}^{(k+1)}\|_2}{\|\hat{\delta}^{(k+1)}\|_2} < \epsilon . \quad (20)$$

L'algorithme est initialisé en faisant une première estimation de la modulation par le calcul du barycentre spectral de la transformée de Gabor du signal modulé

$$\hat{\delta}^{(0)}(n) = \frac{\sum_{m=0}^{M-1} m |\mathbf{G}^{(n;\delta)}|^2[m]}{\sum_{m=0}^{M-1} |\mathbf{G}^{(n;\delta)}|^2[m]} . \quad (21)$$

Le pseudo-code de l'algorithme est donné ci-dessous

---

### Algorithme 1 Estimation conjointe covariance/modulation

---

**Initialiser** suivant (21)

**tant que** le critère (20) est faux **faire**

- Calculer  $\hat{\gamma}^{(k)}$  par interpolation de  $\delta^{(k)}$ .
- Démoduler le signal suivant (19)
- Calculer la transformée de Gabor du signal démodulé  $\hat{\mathbf{G}}^{(k;n)}[m] = \mathcal{G}_{U^{(k)}}[m, n]$
- Estimer  $\hat{\delta}^{(k+1)}$  en utilisant la matrice de covariance de  $\hat{\mathbf{G}}^{(k;n)}$  suivant (17) et (18)
- $k := k + 1$

**fin tant que**

---

## 3 Tests et résultats

L'algorithme a été implémenté en MATLAB/OCTAVE, et s'appuie sur les transformations temps-fréquence mises en oeuvre dans la boîte à outils LTFAT [10]. Nous décrivons ici deux exemples d'application sur signaux synthétiques.

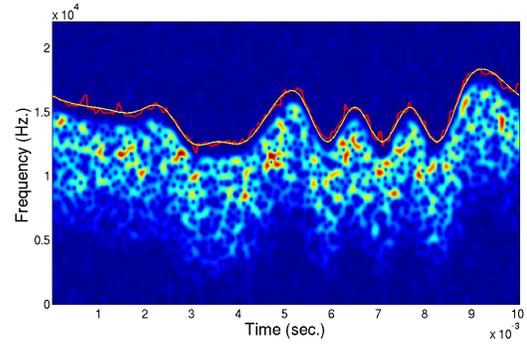


FIG. 1 – Transformée de Gabor d'un signal synthétique de type bande de bruit, superposé avec les courbes de modulation de fréquence vraie (pointillés, jaune) et estimée (traits pleins, rouge).

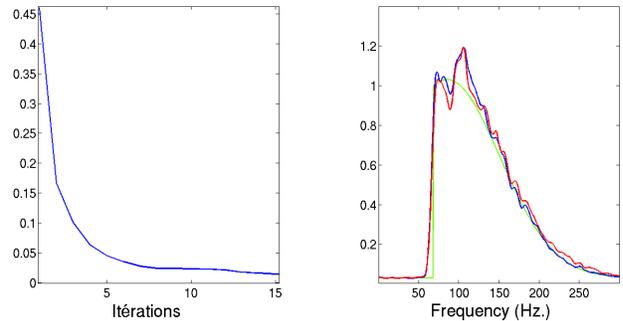


FIG. 2 – Gauche : Evolution du critère d'arrêt en fonction des itérations. Doite : Spectre du signal stationnaire (en vert) et ses estimations, à partir du signal démodulé (en rouge) et du signal stationnaire initial (en bleu).

Un premier exemple d'estimation se trouve en FIG. 1, où est représentée la transformée de Gabor d'un signal aléatoire stationnaire (large bande), modulé en fréquence, superposée avec la courbe de modulation de fréquence estimée, et la modulation réelle. La convergence de l'algorithme est illustrée par la courbe de gauche en FIG. 2, où est exhibée l'évolution du critère d'arrêt en fonction des itérations (moyenné sur 20 réalisations du signal). La courbe de droite représente la densité spectrale du signal stationnaire sous-jacent, superposée avec la densité spectrale estimée à partir de la réalisation considérée, et la densité estimée à partir du signal stationnarisé. Comme on peut le voir, ces deux dernières sont proches.

Un second exemple d'estimation est donné en FIG. 3. Le signal stationnaire sous-jacent est cette fois un signal à spectre de raies, et la modulation est lisse par morceaux, avec quelques discontinuités. La modulation de fréquence estimée suit encore fidèlement la vérité terrain, y compris en présence de discontinuités, comme on peut notamment le voir sur la figure du bas qui représente un "zoom" sur l'une de ces discontinuités.

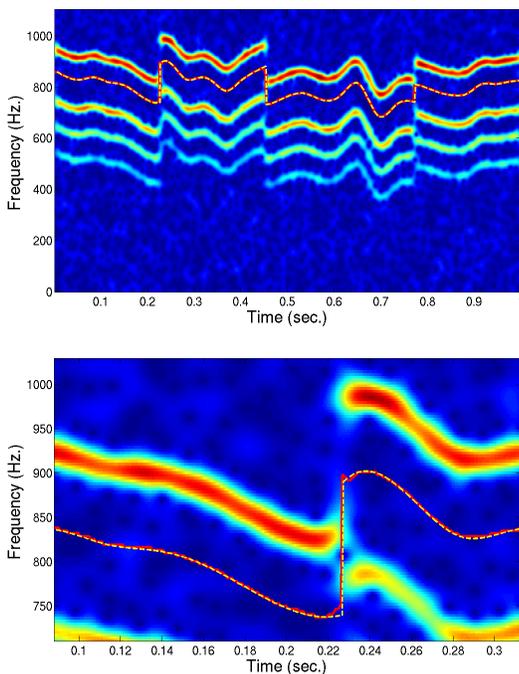


FIG. 3 – Haut : Transformée de Gabor d'un signal synthétique de type spectre de raie, superposé avec les courbes de modulation de fréquence vraie (pointillés) et estimée (trait plein). Bas : Zoom sur un détail.

## 4 Conclusions et perspectives

Dans cet article, nous avons décrit dans ses grandes lignes une nouvelle approche pour l'estimation conjointe de la loi de modulation de fréquence et d'un spectre dans un modèle *timbre*×*dynamique*. Nous nous sommes limités à des modula-

tions de fréquence, la transformation de Gabor nous fournissant dans ce cas un cadre naturel pour l'estimation.

Dans un cadre plus général, il est envisageable de transposer ces résultats à des transformations dynamiques plus générales et/ou complexes que des modulations. Par exemple, des changements d'horloge locaux peuvent être traités de façon similaire en remplaçant la représentation temps-fréquence par une représentation temps échelle, le changement d'horloge se traduisant en première approximation par une translation logarithmique en échelle [3, 2]. Ceci introduit néanmoins de nouvelles difficultés techniques, liées à la nécessité de ré-échantillonner le signal. L'extension à d'autres transformations plus complexes est à l'étude.

Le contexte applicatif de ce programme de recherches est l'application de cette approche et de ses extensions à l'estimation de dynamique à partir de sons réels.

**Remerciements.** Ce travail a bénéficié du soutien de l'ANR, à travers le projet Metason (ANR-10-CORD-010).

## Références

- [1] H. L. Van Trees, *Detection, estimation and modulation theory*. Wiley Interscience, 2003.
- [2] M. Clerc and S. Mallat, "Estimating deformations of stationary processes," *Ann. of Stat.*, vol. 31, no. 6, pp. 1772–1821, 2003.
- [3] M. Clerc and S. Mallat, Shape from Texture and Shading with Wavelets, *Progress in Systems and Control Theory*, vol. 25, pp. 394-417, Birkhauser, 1999.
- [4] B. Picinbono, "Second-order complex random vectors and normal distributions," *IEEE Trans. on Sig. Proc.*, vol. 44, no. 10, pp. 2637–2640, 1996.
- [5] R. Carmona, W. L. Hwang, and B. Torrèsani, *Practical time-frequency analysis : Gabor and Wavelet Transforms With an Implementation in S*, C. K. Chui, Ed. Academic Press, 1998.
- [6] P. Flandrin, *Temps-Fréquence*, Hermes (1998).
- [7] K. Gröchenig, *Foundations of time-frequency analysis*, ser. Applied and Numerical Harmonic Analysis. Boston, MA : Birkhäuser Boston Inc., 2001.
- [8] H. Omer and B. Torrèsani, "Estimation of frequency modulations on wideband signals ; applications to audio signal analysis", Proceedings of SAMPTA 2013, 10th International Conference on Sampling Theory and Applications, Bremen, Germany (2013); <http://hal.archives-ouvertes.fr/hal-00822186>
- [9] P. Soendergaard, "Finite discrete Gabor analysis," Ph.D. thesis, Inst. for Matematik – Denmark Technical University, 2007.
- [10] P. Soendergaard, B. Torrèsani, and P. Balazs, "The linear time frequency analysis toolbox," *International Journal of Wavelets and Multiresolution Information Processing*, vol. 10, no. 4, pp. 1 250 032–1 – 1 250 032–27, 2012.