Estimation locale des modulations AM/FM: applications à la modélisation sinusoïdale audio et à la séparation de sources aveugle

Dominique FOURER[1], François AUGER[2], Geoffroy PEETERS[1]

[1]UMR STMS (IRCAM - CNRS - UPMC)
1, place Igor Stravinsky, 75004 Paris, France

[2]IREENA - CRTT
37, boulevard de l'université, 44600 Saint-Nazaire, France
dominique@fourer.fr, francois.auger@univ-nantes.fr

Résumé – Dans cet article, plusieurs nouveaux estimateurs de modulation d'amplitude (AM) et de fréquence (FM), dédiés aux signaux linéairement modulés sont étudiés. Premièrement, ces estimateurs sont comparés entre eux, puis avec d'autres méthodes issues de l'état de l'art en termes d'efficacité statistique. Deuxièmement, nous proposons d'appliquer ces nouveaux estimateurs à la modélisation sinusoïdale des signaux audio. Enfin, une nouvelle technique de séparation de sources s'appuyant sur les modulations cohérentes d'amplitude et de fréquence est proposée et évaluée sur des signaux musicaux.

Abstract – In this paper, several new amplitude (AM) and frequency modulation (FM) estimators designed for linear modulated signals (linear chirps) are investigated. First they are compared together in terms of statistical efficiency with other state-of-the-art methods. Second, they are used to improve sinusoidal modeling of audio signals. Finally, a new source separation technique based on coherent amplitude and frequency modulation is proposed and evaluated on musical signals.

1 Introduction

L'analyse et la transformation des signaux non-stationnaires et multicomposantes, est omniprésente en traitement audio, notamment pour l'extraction d'informations musicales ou pour la séparation de sources [1]. Pour être traités, ces problèmes nécessitent le plus souvent l'utilisation d'outils adaptés reposant sur l'analyse temps-fréquence ou temps-échelle [2]. La Transformée de Fourier à Court Terme (TFCT) compte parmi les outils les plus couramment utilisés, bien qu'elle ne possède pas toujours une concentration suffisante dans le plan temps-fréquence. Ainsi, la réallocation [3] offre une solution efficace qui améliore la localisation d'une Représentation Temps-Fréquence (RTFR) bien qu'elle ne permette pas une reconstruction de celle-ci. C'est pour cela que le synchrosqueezing [4,5] a été proposé pour améliorer la localisation d'une représentation temps-fréquence tout en admettant une reconstruction du signal analysé. Une autre approche, la modélisation sinusoïdale [6, 7] se concentre sur l'estimation locale des paramètres du signal et permet sa reconstruction. De nombreux travaux [8, 9] ont permis d'améliorer la précision de ces estimateurs. Dans cet article nous proposons plusieurs estimateurs originaux du taux de variation de la fréquence et de l'amplitude. Dans la section 2, nous présentons le modèle de signal choisi pour lequel nous introduisons plusieurs nouveaux estimateurs de ses paramètres, en s'appuyant sur nos précédents travaux [10]. Nous proposons d'appliquer ces nouveaux estimateurs à la modélisation sinusoïdale audio, puis de les évaluer comparativement à l'existant en termes de performances dans la section 3. Dans la section 4, nous proposons d'appliquer ces nouveaux outils à la séparation aveugle de sources audio. Quelques conclusions sont finalement exposées en section 5 ainsi que de nouvelles perspectives de recherche.

Cette recherche a été financée par le projet H2020 ABC-DJ (688122) ainsi que le projet ANR ASTRES (ANR-13-BS03-0002-01).

2 Estimation locale des modulations

2.1 Modèle de signal et propriétés

Nous proposons ici d'estimer en tout point d'une représentation temps-fréquence les paramètres d'un signal modulé en amplitude et en fréquence. Nous considérons donc le modèle de signal suivant :

$$x(t) = \mathbf{e}^{\lambda_x(t)+j\phi_x(t)} \tag{1}$$

$$\text{avec} \quad \lambda_x(t) = l_x + \mu_x t + \nu_x \frac{t^2}{2} \tag{2}$$

$$\text{et} \quad \phi_x(t) = \varphi_x + \omega_x t + \alpha_x \frac{t^2}{2} \tag{3}$$

où $\lambda_x(t)$ et $\phi_x(t)$ sont respectivement la log-amplitude et la phase qui varient au court du temps, et $j^2{=}{-}1$. Ce signal vérifie

$$\frac{dx}{dt}(t) = \left(\frac{d\lambda_x}{dt}(t) + j\frac{d\phi_x}{dt}(t)\right) x(t) = (q_x t + p_x)x(t) \tag{4}$$

avec $q_x = \nu_x + j\alpha_x$ et $p_x = \mu_x + j\omega_x$. Nous définissons $F_x^h(t,\omega)$, la TFCT du signal x, utilisant la fenêtre d'analyse h

dérivable plusieurs fois, exprimée par

$$F_x^h(t,\omega) = \int_{\mathbb{R}} x(u)h(t-u)^* \, \mathbf{e}^{-j\omega u} \, \mathrm{d}u \tag{5}$$

$$= \mathbf{e}^{-j\omega t} \int_{\mathbb{R}} x(t-u)h(u)^* \, \mathbf{e}^{j\omega u} \mathrm{d}u. \tag{6}$$

En dérivant l'Eq. (6) par rapport à t, nous obtenons

$$\frac{\partial F_x^h}{\partial t}(t,\omega) = \int_{\mathbb{R}} x(u) \frac{\mathrm{d}h}{\mathrm{d}t}(t-u)^* \, \mathbf{e}^{-j\omega u} \, \mathrm{d}u \tag{7}$$

$$= -j\omega F_x^h(t,\omega) + \mathbf{e}^{-j\omega t} \int_{\mathbb{R}} \frac{\mathrm{d}x}{\mathrm{d}t}(t-u)h(u)^* \, \mathbf{e}^{j\omega u} \mathrm{d}u. \tag{8}$$

En remplaçant $\frac{\mathrm{d}x}{\mathrm{d}t}(t-u)$ par $\left(q_{x\,(t-u)} + p_x\right)\, x(t-u)$, on obtient

$$F_x^{\mathcal{D}h}(t,\omega) = -q_x F_x^{\mathcal{T}h}(t,\omega) + (q_x t + p_x - j\omega) F_x^h(t,\omega), \tag{9}$$

où $F_x^{\mathcal{D}h}(t,\omega)$ et $F_x^{\mathcal{T}h}(t,\omega)$ sont deux TFCT utilisant comme fenêtre d'analyse $\mathcal{D}h(t) = \frac{\mathrm{d}h}{\mathrm{d}t}(t)$ et $\mathcal{T}h(t) = t\,h(t)$.

En dérivant une seconde fois par rapport à t, nous obtenons

$$F_x^{\mathcal{D}^2 h}(t,\omega) = -q_x F_x^{\mathcal{T}\mathcal{D}h}(t,\omega) + (q_x t + p_x - j\omega) F_x^{\mathcal{D}h}(t,\omega) \tag{10}$$

et plus généralement, pour $n \geq 1$ [10] :

$$F_x^{\mathcal{D}^n h}{}_{(t,\omega)} = -q_x F_x^{\mathcal{T}\mathcal{D}^{n-1}h}{}_{(t,\omega)} + (q_x t + p_x - j\omega) F_x^{\mathcal{D}^{n-1}h}{}_{(t,\omega)}. \tag{11}$$

De même, si on dérive l'Eq. (6) par rapport à ω, nous obtenons $\frac{\partial F_x^h}{\partial \omega}(t,\omega) = j(F_x^{\mathcal{T}h}(t,\omega) - t\, F_x^h(t,\omega))$ qui nous permet d'écrire l'expression suivante, pour $n \geq 1$ [10] :

$$F_x^{\mathcal{T}^{n-1}\mathcal{D}h}(t,\omega) + (n-1)\, F_x^{\mathcal{T}^{n-2}h}(t,\omega) = -q_x\, F_x^{\mathcal{T}^n h}(t,\omega) + (q_x t + p_x - j\omega)\, F_x^{\mathcal{T}^{n-1}h}(t,\omega). \tag{12}$$

2.2 Estimation de tous les paramètres du signal

En combinant les Eqs. (9) et (10) (*i.e.* Eq. (11) pour $n = 1$ et $n = 2$), nous obtenons un système d'équations avec q_x et p_x comme inconnues (où (t,ω) est absent par souci de place) :

$$\begin{pmatrix} tF_x^h - F_x^{\mathcal{T}h} & F_x^h \\ tF_x^{\mathcal{D}h} - F_x^{\mathcal{T}\mathcal{D}h} & F_x^{\mathcal{D}h} \end{pmatrix} \begin{pmatrix} q_x \\ p_x \end{pmatrix} = \begin{pmatrix} F_x^{\mathcal{D}h} + j\omega F_x^h \\ F_x^{\mathcal{D}^2 h} + j\omega F_x^{\mathcal{D}h} \end{pmatrix}. \tag{13}$$

En supposant l'Eq. (13) inversible, on obtient l'égalité suivante :

$$\begin{pmatrix} q_x \\ p_x \end{pmatrix} = \begin{pmatrix} tF_x^h - F_x^{\mathcal{T}h} & F_x^h \\ tF_x^{\mathcal{D}h} - F_x^{\mathcal{T}\mathcal{D}h} & F_x^{\mathcal{D}h} \end{pmatrix}^{-1} \begin{pmatrix} F_x^{\mathcal{D}h} + j\omega F_x^h \\ F_x^{\mathcal{D}^2 h} + j\omega F_x^{\mathcal{D}h} \end{pmatrix}$$

qui nous permet d'obtenir les estimateurs suivants, appelés $(t2)$ car impliquant les dérivées secondes par rapport à t de $F_x^h(t,\omega)$:

$$\hat{q}_x^{(t2)}(t,\omega) = \frac{F_x^{\mathcal{D}^2 h}{}_{(t,\omega)} F_x^h{}_{(t,\omega)} - F_x^{\mathcal{D}h}{}_{(t,\omega)}{}^2}{F_x^{\mathcal{T}h}{}_{(t,\omega)} F_x^{\mathcal{D}h}{}_{(t,\omega)} - F_x^{\mathcal{T}\mathcal{D}h}{}_{(t,\omega)} F_x^h{}_{(t,\omega)}} \tag{14}$$

$$\begin{aligned} \hat{p}_x^{(t2)}(t,\omega) =& j\omega - t \frac{F_x^{\mathcal{D}^2 h}{}_{(t,\omega)} F_x^h{}_{(t,\omega)} - F_x^{\mathcal{D}h}{}_{(t,\omega)}{}^2}{F_x^{\mathcal{T}h}{}_{(t,\omega)} F_x^{\mathcal{D}h}{}_{(t,\omega)} - F_x^{\mathcal{T}\mathcal{D}h}{}_{(t,\omega)} F_x^h{}_{(t,\omega)}} \\ &+ \frac{F_x^{\mathcal{T}\mathcal{D}h}{}_{(t,\omega)} F_x^{\mathcal{D}h}{}_{(t,\omega)} - F_x^{\mathcal{T}h} F_x^{\mathcal{D}^2 h}{}_{(t,\omega)}}{F_x^{\mathcal{T}\mathcal{D}h}{}_{(t,\omega)} F_x^h{}_{(t,\omega)} - F_x^{\mathcal{T}h}{}_{(t,\omega)} F_x^{\mathcal{D}h}{}_{(t,\omega)}} \\ =& \tilde{\omega}(t,\omega) - \hat{q}_x^{(t2)}(t,\omega)\tilde{t}(t,\omega) \end{aligned} \tag{15}$$

où $\hat{\omega}$ et $\hat{t}$ sont les opérateurs de réallocation [3, 5] tels que

$$\hat{t}_{(t,\omega)} = \mathrm{Re}\left(\tilde{t}_{(t,\omega)}\right), \text{avec} \quad \tilde{t}_{(t,\omega)} = t - \frac{F_x^{\mathcal{T}h}{}_{(t,\omega)}}{F_x^h{}_{(t,\omega)}}, \tag{16}$$

$$\hat{\omega}_{(t,\omega)} = \mathrm{Im}\left(\tilde{\omega}_{(t,\omega)}\right), \text{avec} \quad \tilde{\omega}_{(t,\omega)} = j\omega + \frac{F_x^{\mathcal{D}h}{}_{(t,\omega)}}{F_x^h{}_{(t,\omega)}}. \tag{17}$$

Ainsi, $\dot{\lambda}_{x(t)} = \frac{d\lambda_x}{dt}(t) = \mu_x + \nu_x t$ et $\dot{\phi}_{x(t)} = \frac{d\phi_x}{dt}(t) = \omega_x + \alpha_x t$, sont obtenus en utilisant $\Psi = \dot{\lambda}_x + j\dot{\phi}_x = q_x t + p_x$, pouvant être estimé par l'expression suivante [5, 10] :

$$\hat{\Psi}_x(t,\omega) = \tilde{\omega}(t,\omega) + \hat{q}_x^{(t2)}(t,\omega)(t - \tilde{t}(t,\omega)). \tag{18}$$

Finalement, nous déduisons les estimateurs suivants pour le modèle local de signal donné par l'Eq. (1) :

$$\hat{\nu}_x(t,\omega) = \mathrm{Re}\left(\hat{q}_x(t,\omega)\right), \qquad \hat{\alpha}_x(t,\omega) = \mathrm{Im}\left(\hat{q}_x(t,\omega)\right) \tag{19}$$

$$\hat{\dot{\lambda}}_x(t,\omega) = \mathrm{Re}\left(\hat{\Psi}_x(t,\omega)\right), \quad \hat{\dot{\Phi}}_x(t,\omega) = \mathrm{Im}\left(\hat{\Psi}_x(t,\omega)\right) \tag{20}$$

où l'amplitude et la phase, pour $t = 0$, sont estimés par [8] :

$$\hat{l}_x(t,\omega) = \ln\left(\left|\frac{F_x^h{}_{(t,\omega)}}{F_h(\omega - \hat{\dot{\Phi}}_{x(t,\omega)}, \hat{\dot{\lambda}}_{x(t,\omega)}, \hat{q}_{x(t,\omega)})}\right|\right) \tag{21}$$

$$\hat{\varphi}_x(t,\omega) = \arg\left(\frac{F_x^h{}_{(t,\omega)}}{F_h(\omega - \hat{\dot{\Phi}}_{x(t,\omega)}, \hat{\dot{\lambda}}_{x(t,\omega)}, \hat{q}_{x(t,\omega)})}\right) \tag{22}$$

$$\text{avec} \quad F_{h(\omega,\mu,q)} = \int_{\mathbb{R}} h(t)\, \mathbf{e}^{(\mu + j\omega)t + q\frac{t^2}{2}}\, dt. \tag{23}$$

Une infinité de nouveaux estimateurs $\hat{q}_x(t,\omega)$, appelés (tn), (ωn) et (rn), $\forall n \geq 2$, utilisant les dérivées d'ordre n par rapport à t et ω, permettent aussi de retrouver tous les autres paramètres du signal en utilisant les Eqs. (19)-(22).

Ainsi (tn) est obtenu en combinant les Eqs. (9) et (11), (ωn) utilise l'Eq. (12) pour $n = 1$ et $n \geq 2$:

$$\hat{q}_x^{(\omega n)}{}_{(t,\omega)} = \frac{(F_x^{\mathcal{T}^{n-1}\mathcal{D}h} + (n-1)F_x^{\mathcal{T}^{n-2}h})F_x^h - F_x^{\mathcal{T}^{n-1}h} F_x^{\mathcal{D}h}}{F_x^{\mathcal{T}^{n-1}h} F_x^{\mathcal{T}h} - F_x^{\mathcal{T}^n h} F_x^h}. \tag{24}$$

Enfin, l'estimateur (rn) est obtenu en combinant les Eqs. (11) et (12), $\forall n \geq 2$.

3 Modélisation sinusoïdale audio

La modélisation sinusoïdale [6, 7] fournit une représentation paramétrique d'un signal quelconque permettant ainsi d'appliquer des transformations avant sa resynthèse. Nous donnons ici une brève description de l'algorithme d'analyse et de synthèse, avant de réaliser une évaluation comparative.

3.1 Algorithme proposé

Nous considérons le cas d'un signal multicomposante :

$$x(t) = \sum_{i \in \mathcal{I}} x_i(t) = \sum_{i \in \mathcal{I}} \mathbf{e}^{\lambda_i(t) + j\phi_i(t)}. \tag{25}$$

Après discrétisation (*i.e.* $F_x^h[k,m] = F_x^h(kT_e, 2\pi\frac{m}{MT_e})$, avec T_e la période d'échantillonnage, $k \in \mathbb{Z}$ et $m \in [0, M]$), on considère une fenêtre d'analyse de longueur L et un pas d'avancement $\Delta k = \lfloor (1-\rho)L \rfloor$ (où $\rho \in [0, 1[$ correspond au taux de recouvrement entre deux fenêtres successives). L'algorithme d'analyse-synthèse est alors décrit comme suit.

Analyse :

1. Pour chaque fenêtre centrée sur l'instant k, les maxima locaux $m \in]0, M/2[$ sont détectés (*i.e.* m vérifie : $|F^h_{x\,[k,m]}| > |F^h_{x\,[k,m+1]}|$ et $|F^h_{x\,[k,m]}| > |F^h_{x\,[k,m-1]}|$).
2. Pour chaque maximum local m, on estime le vecteur $P_m[k] = (l_{x[m]}, \varphi_{x[m]}, \dot{\lambda}_{x[m]}, \dot{\Phi}_{x[m]}, \alpha_{x[m]})^T$ (*i.e.* on fixe arbitrairement $\nu_{x[m]} = 0$, donc on a $\mu_x[m] = \dot{\lambda}_x[m]$) .
3. Par ordre décroissant de $l_{x[m]}$, chaque composante m est synthétisée seule en utilisant P_m et l'Eq. (1). La composante m est ignorée si l'énergie du signal résiduel ne diminue pas après soustraction de la composante. Le cas échéant, on conserve P_m ainsi que le résidu du signal après soustraction de la composante.
4. On incrémente $k \leftarrow k + \Delta k$, puis on itère depuis 1, tant que k est inférieur à la longueur du signal.

Synthèse : On synthétise chaque trame du signal reconstruit $\hat{x}$, associée à une fenêtre d'analyse centrée à un instant k, à partir des $P_m[k]$ et de l'Eq. (25). (*i.e.* chaque composante vérifie $\omega_x[m] = \dot{\Phi}_x[m] - \alpha_x[m]t$).

3.1.1 Simulations numériques

Afin d'évaluer la précision des estimateurs proposés, nous choisissons de les comparer avec deux méthodes issues de l'état de l'art, respectivement appelées (SM08) [8] et (SM12) [9]. Dans cette expérience, nous considérons un signal échantillonné à $F_e = 44.1$ kHz, contenant une seule sinusoïde modulée aléatoirement (loi uniforme), synthétisée à partir de l'Eq. (1), pour $l_x = 0.18$, $\varphi_x \in [-\pi; +\pi]$, $\mu_x \in [0, 100]$, $\omega_x = 2\pi 440$ rad.s^{-1} et $\alpha_x \in [0, 10000]$ rad.s^{-2}. Ce signal est mélangé avec un bruit blanc gaussien, dont le Rapport Signal sur Bruit (RSB) varie entre -10 dB et $+80$ dB. Pour chaque RSB, nous mesurons l'Erreur Quadratique Moyenne (EQM) ainsi que la qualité de reconstruction, donnée par le *Reconstruction Quality Factor (RQF)* : $\text{RQF}(x, \hat{x}) = 10 \log_{10}\left(\frac{||x||^2}{||x-\hat{x}||^2}\right)$. Ainsi, pour chaque RSB, on génère 1000 signaux aléatoires qui sont analysés par une fenêtre de Hann, de taille $L = 1023$. Nos résultats (*cf.* Fig. 1) montrent que l'utilisation des estimateurs d'ordre n, plus élevé, a un effet négligeable sur la précision pour (tn) et (rn), ou alors cela réduit les performances pour (wn). Ainsi, les meilleurs résultats sont obtenus avec (wn) (*i.e.* $n = 2$) suivi par (tn) et (rn) qui obtiennent tous les trois des performances nettement meilleures que (SM08) et (SM12). Cela reste particulièrement vrai pour les RSB élevés.

4 Application à la séparation de sources

Nous considérons ici le problème de séparation de sources aveugle [1], dans le cas d'un seul mélange (mono-canal) instantané, contenant $C \geq 2$ sources. Ainsi, nous souhaitons retrouver les sources s_c en disposant pour seule et unique observation, le mélange donné par :

$$x(t) = \sum_{c=1}^{C} s_c(t) = \sum_{c=1}^{C} \sum_{i_c \in \mathcal{I}_c} \mathbf{e}^{\lambda_{i_c}(t) + j\phi_{i_c}(t)}. \qquad (26)$$

En supposant que chaque composante sinusoïdale est affectée à une seule source, nous proposons de résoudre ce problème en utilisant comme critère de regroupement, les paramètres de modulation de chaque composante qui sont estimés directement depuis le mélange x.

4.1 Méthode proposée

L'approche *Computational Auditory Scene Analysis (CASA)* [11] suggère que les paramètres de chaque source perçue, évoluent indépendamment des autres sources au cours du temps. Ainsi, nous proposons de regrouper les composantes du signal à chaque instant en utilisant la Modulation de Fréquence Cohérente (MFC) [12] et la Modulation d'Amplitude Cohérente (MAC) définis pour un signal x par :

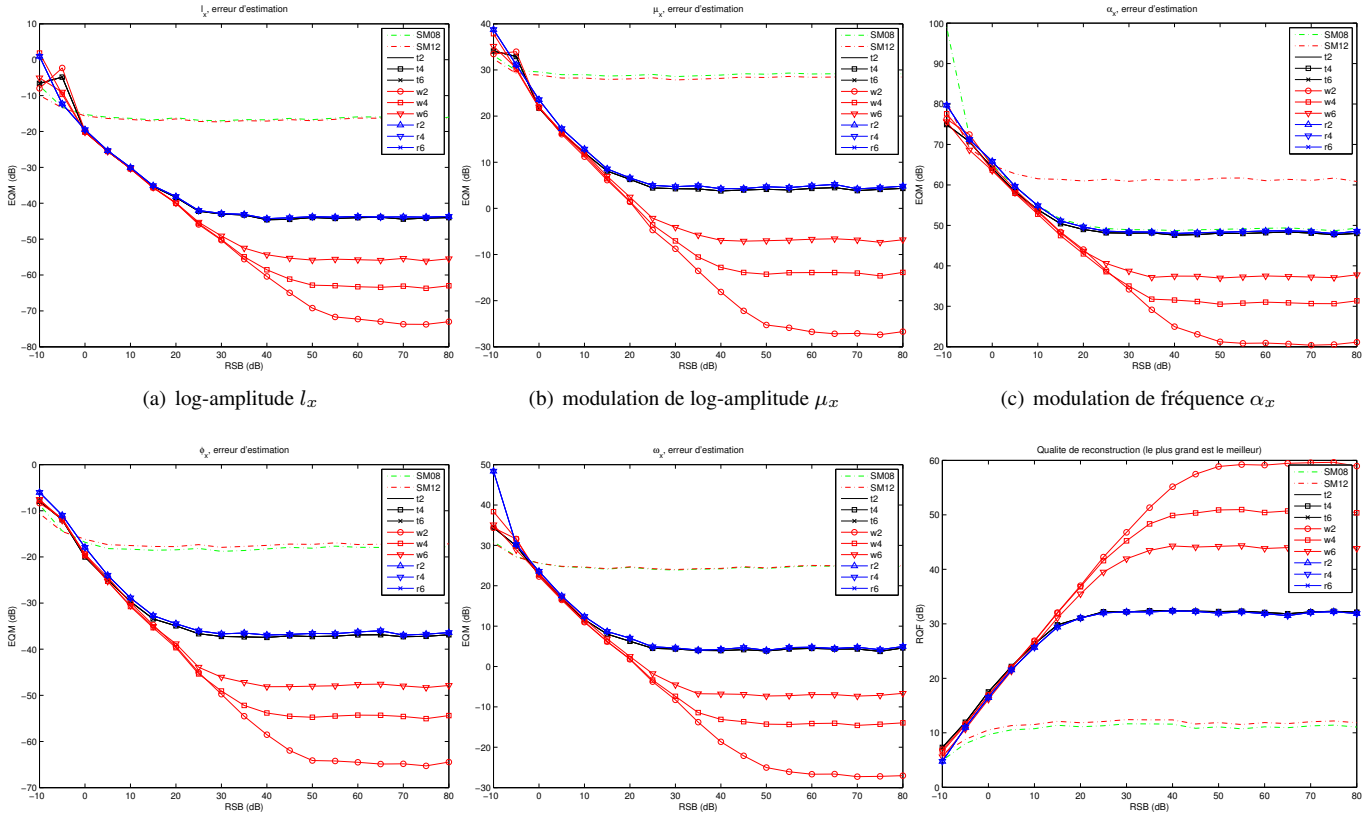
$$\text{MFC}_x(t,\omega) = \frac{\hat{\alpha}_x(t,\omega)}{\hat{\dot{\Phi}}_x(t,\omega)}, \quad \text{MAC}_x(t,\omega) = \frac{\hat{\dot{\lambda}}_x(t,\omega)}{\hat{l}_x(t,\omega)}. \quad (27)$$

Ces deux descripteurs capturent les facteurs de modulation linéaires en fréquence et en amplitude, supposés presque identiques entre toutes les composantes appartenant à la même source [11]. Cette idée à été mise en oeuvre dans plusieurs approches issues de la littérature telles que [12, 13]. Nous proposons l'algorithme de séparation de sources suivant :

1. Calcul des paramètres $P_i[k]$ à partir du mélange x. Chaque $P_i[k] = (l_i, \varphi_i, \dot{\lambda}_i, \dot{\Phi}_i, \alpha_i)^T$ étant associé à la sinusoïde i, estimée à l'instant k (*cf.* Section 3).
2. Calcul de $\text{MFC}_i[k]$ et de $\text{MAC}_i[k]$ pour chaque composante à partir de l'Eq. (27).
3. À chaque instant k, calcul des ensembles $\mathcal{I}_{c'}$ associés à une source, en appliquant l'algorithme *k-means* [14] sur les composantes i, représentées par $(\text{MFC}_i,\ \text{MAC}_i)^T$, pour un nombre maximum de C groupes.
4. Modélisation de chaque source c à chaque instant par un représentant $v_c[k] = \left(\frac{\sum_{i\in\mathcal{I}_c} l_i^2\ \text{MFC}_i[k]}{\sum_{i\in\mathcal{I}_c} l_i^2}, \frac{\sum_{i\in\mathcal{I}_c} l_i^2\ \omega_i}{\sum_{i\in\mathcal{I}_c} l_i^2}\right)^T$.
5. Si $k > 1$, on affecte chaque source instantanée c' à la source c par $\arg\min_{c'} ||v_{c'}[k] - v_c[k-1]||$.
6. On synthétise chaque source estimée $\hat{s}_c$ à partir de ses $P_i[k]$ pour $i \in \mathcal{I}_c$ en utilisant l'Eq. (25).

4.1.1 Simulation sur des signaux naturels

Dans cette expérience, nous analysons un mélange musical instantané échantillonné à $F_e = 44.1$kHz, composés de 2 sources musicales (guitare/voix) extraites de la base MedleyDb [15]. Pour l'analyse, nous utilisons l'estimateur $(\omega 2)$ avec une fenêtre de Hann de 23 ms et un recouvrement de $\rho = \frac{11}{12}$ (ce qui permet d'obtenir le meilleur RQF pour la modélisation du mélange). La table 1 présente les scores RQF, SIR, SDR et SAR [16] calculés à partir des signaux d'origine et des sources estimées. L'oracle correspond à la meilleure segmentation des composantes, obtenue à partir du support temps-fréquence des sources de référence. Nos résultats montrent que la méthode proposée parvient à séparer de manière non supervisée, les sources qui composent ce mélange grâce au descripteur MFC seul. Cependant, la séparation obtenue en utilisant MAC semble beaucoup moins efficace sur cet exemple.



(a) log-amplitude l_x (b) modulation de log-amplitude μ_x (c) modulation de fréquence α_x

(d) phase ϕ_x (e) fréquence ω_x (f) RQF pour les paramètres estimés

FIGURE 1 – Comparaison des EQMs (a)-(e) et des RQFs entre les méthodes (SM08), (SM12), (tn), (wn) et (rn) avec $n \in \{2, 4, 6\}$, pour l'estimation des paramètres d'une sinusoide modulée en fréquence et en amplitude, mélangée à un bruit blanc.

TABLE 1 – Qualité de séparation des sources pour un mélange audio : guitare/voix, en fonction de la méthode proposée.

Méthode	RQF (dB)	SIR (dB)	SDR (dB)	SAR (dB)
Oracle	11.67/10.33	23.84/26.86	11.41/9.91	11.69/10.01
MFC-kmeans	**5.42/7.42**	**9.31/12.65**	**3.97/6.60**	5.96/**8.06**
MAC-kmeans	0.67/2.64	0.60/**12.26**	-0.47/0.23	**8.80**/0.76
MFC/MAC-kmeans	0.78/2.73	0.67/**13.01**	-0.28/0.78	**9.40**/1.26

5 Conclusion et travaux futurs

Dans cet article, nous avons proposé une infinité de nouveaux estimateurs appliqués à la modélisation sinusoïdale audio et à la séparation de sources aveugle. Nous avons montré que ces nouveaux estimateurs obtiennent une meilleure précision que d'autres méthodes issues de l'état de l'art. Nous avons également proposé une application de ces outils sur des signaux naturels de musique. Pour la suite, nous prévoyons une étude plus approfondie de la méthode de séparation de sources. Nous souhaitons également améliorer sa robustesse pour le traitement de signaux du monde réel, fortement bruités.

Références

[1] P. Comon and C. Jutten, *Handbook of Blind Source Separation : Independent component analysis and applications*. Academic press, 2010.

[2] P. Flandrin, *Time-frequency/time-scale analysis*. Academic press, 1998.

[3] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1068–1089, May 1995.

[4] I. Daubechies and S. Maes, "A nonlinear squeezing of the continuous wavelet transform," *Wavelets in Medecine and Bio.*, pp. 527–546, 1996.

[5] R. Behera, S. Meignen, and T. Oberlin, "Theoretical analysis of the second-order synchrosqueezing transform," *Applied and Computational Harmonic Analysis*, Nov. 2016.

[6] R. McAulay and T. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 4, pp. 744–754, Aug. 1986.

[7] J. Smith and X. Serra, "PARSHL an analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation," in *Proc. ICMC*, Urbana, Illinois, USA, Aug. 1987, pp. 290–297.

[8] S. Marchand and P. Depalle, "Generalization of the derivative analysis method to non-stationary sinusoidal modeling," in *Proc. DAFx'08*, Espoo, Finland, Sep. 2008, pp. 281–288.

[9] S. Marchand, "The simplest analysis method for non-stationary sinusoidal modeling," in *Proc. DAFx'12*, York, UK, Sep. 2012, pp. 23–26.

[10] D. Fourer, F. Auger, K. Czarnecki, Meignen, and Flandrin, "Chirp rate and instantaneous frequency estimation : Application to recursive vertical synchrosqueezing," *IEEE Signal Process. Lett.*, Jun. 2017.

[11] A. S. Bregman, *Auditory scene analysis*. MIT Press : Cambridge, MA, 1990.

[12] E. Creager, N. D. Stein, R. Badeau, and P. Depalle, "Nonnegative tensor factorization with frequency modulation cues for blind audio source separation," in *Proc. International Society for Music Information Retrieval (ISMIR) Conference*, New York, USA, Aug. 2016.

[13] F. R. Stöter, A. Liutkus, R. Badeau, B. Edler, and P. Magron, "Common fate model for unison source separation," in *Proc. IEEE ICASSP*, Mar. 2016, pp. 126–130.

[14] G. A. F. Seber, *Multivariate Observations*. Hoboken, NJ : John Wiley & Sons, Inc.

[15] R. Bittner, J. Salamon, M. Tierney, M. Mauch, C. Cannam, and J. P. Bello, "Medleydb : A multitrack dataset for annotation-intensive MIR research," in *Proc. International Society for Music Information Retrieval (ISMIR) Conference*, Taipei, Taiwan, Oct. 2014.

[16] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing (TASLP)*, vol. 14, no. 4, pp. 1462–1469, Jul. 2006.