



HAL
open science

Décomposition parcimonieuse structurée des signaux audio de musique guidée par l'information experte

Hélène Papadopoulos, Matthieu Kowalski

► **To cite this version:**

Hélène Papadopoulos, Matthieu Kowalski. Décomposition parcimonieuse structurée des signaux audio de musique guidée par l'information experte. GRETSI, Sep 2013, Brest, France. hal-00829133

HAL Id: hal-00829133

<https://hal.science/hal-00829133>

Submitted on 2 Jun 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Décomposition parcimonieuse structurée des signaux audio de musique guidée par l'information experte

Hélène PAPADOPOULOS, Matthieu KOWALSKI

Laboratoire des Signaux et Systèmes
3, rue Joliot-Curie, 91192 Gif-sur-Yvette, France

helene.papadopoulos@lss.supelec.fr, matthieu.kowalski@lss.supelec.fr

Thème – 6.2 - Applications: Son, parole, musique

1.3 - Représentations et modèles: Parcimonie, modèles et méthodes

Problème traité – La construction, dans un cadre Bayésien, d'*a priori* musicaux pour les décompositions parcimonieuses des signaux de musiques est étudiée. Ces *a priori* reposent sur la connaissance musicale obtenue sur le signal, résumé dans des vecteurs de chroma contenant les 12 notes de la gamme chromatique.

Originalité – L'originalité principale de ce travail est l'intégration de la connaissance musicale « experte » pour les décomposition parcimonieuse, plutôt que d'*a priori* « physique », comme les persistences en temps ou en fréquence, classiquement utilisés.

Résultats – Les décompositions obtenus sont comparées à l'état de l'art utilisant des approches « physique » sur un problème de débruitage d'un bruit blanc gaussien. Les résultats obtenus en terme de rapport signal sur bruit sont équivalents. Les cartes de significances des coefficients temps-fréquence font apparaître les structures harmoniques attendues.

1 Introduction

Une particularité des signaux audio de musique est, souvent, de faire apparaître différentes composantes de manière superposées, telle qu'une couche transitoire (les attaques des notes, événements bien localisés en temps) et une couche tonale (les partiels des notes, bien modélisé par des sommes de sinusoides variant lentement en amplitude et en fréquence). Les modèles Hybrides [1] permettent une représentation simultanée des différentes composantes aux comportements significativement différents. Suivant cette approche, on considère un modèle en trois couches de la forme $signal = tonals + transients + residual$ afin de décomposer un signal de musique $x \in \mathbb{R}^N$. On utilise aussi l'hypothèse de parcimonie dans un dictionnaire hybride formé par l'union de deux bases MDCT (Modified Discrete Cosine Transform) avec des résolutions temps-fréquences différentes.

Plus spécifiquement, les signaux de musiques sont très structurés et, idéalement, ces structures devraient se refléter dans la décomposition, de sorte que les coefficients aient une interprétabilité physique ou musicale dans la perspective de son analyse. On cherche alors une approximation qui, en plus d'être parcimonieuse, soit aussi structurée en prenant en compte les dépendances entre les coefficients. Ces structures peuvent être modélisées directement sur les coefficients, cependant, il peut être pratique de les modéliser directement sur les indices temps-fréquences directement plutôt que sur les coefficients eux-mêmes.

On se limitera ici à une tâche de débruitage sur des signaux de musiques, cette application servant de « preuve de concept » pour illustrer les nouveaux *a priori* musicaux présentés. Le modèle proposé s'inspire d'un modèle Bayésien précédemment présenté par Févotte *et al.* dans [4], la différence essentielle étant dans la manière de modéliser les dépendances entre coefficients. En effet, dans cette contribution, le contenu musical sert directement à construire les *a priori* au lieu d'utiliser des intuitions physiques.

Comme proposé dans [5], la progression des accords ([6] [7]) fournit une information qui peut-être directement utilisé pour construire un *a priori* musical pour la couche tonale. On donne ici quelques améliorations du modèle, par exemple l'utilisation des chroma [8] comme alternative à la progression des accords pour modéliser la couche tonale. Un modèle pour la couche transitoire utilisant l'information de position des « beats » [9] est aussi proposé.

2 Modèle

On considère un dictionnaire hybride construit comme l'union de deux bases orthogonales $V = \{v_n, n = 1, \dots, N\}$ et $U = \{u_m, m = 1, \dots, N\}$, avec différentes résolutions temps-fréquences afin de représenter les couches tonales et transitoires du signal. Le modèle est alors de la forme

$$x = \sum_{\lambda \in \Lambda} \alpha_\lambda v_\lambda + \sum_{\delta \in \Delta} \beta_\delta u_\delta + r, \quad (1)$$

où les cartes de significances Λ et Δ sont des petits sous-ensembles de $I = \{1, \dots, N\}$, et r un résidu n'acceptant pas de décomposition parcimonieuse dans le dictionnaire.

Le modèle hybride est défini par un modèle de probabilité discrète sur les cartes de significances et un modèle de probabilité sur les coefficients, conditionnellement aux cartes de significances. On re-écrit (1) comme :

$$x = \sum_{n \in I} \gamma_{ton,n} \alpha_n v_n + \sum_{m \in I} \gamma_{tran,m} \beta_m u_m + r. \quad (2)$$

où les $\gamma_{ton,n}$ et $\gamma_{tran,m}$ sont des variables indicatrices. On construit l'*a priori* proposé sur un modèle Bernoulli-Gaussien :

$$p(\alpha_n | \gamma_{ton,n}, \sigma_{ton,n}) = (1 - \gamma_{ton,n}) \delta_0(\alpha_n) + \gamma_{ton,n} \mathcal{N}(\alpha_n | 0, \sigma_{ton,n}^2). \quad (3)$$

Avec un modèle similaire pour les coefficients β et γ_{trans} .

Afin de sélectionner les atomes de la base MDCT correspondant au contenu harmonique du signal, on commence par construire une relation entre les positions des coefficients MDCT et les 12 demi tons de la gamme chromatique. Étant donné un indice q , on pose $\{p_k^{chroma}\}_{k=1..12}$ les classes des 12 demi-tons de la gamme chromatique. On note alors $\{p_\nu^{MDCT}\}_{\nu=1, \dots, \ell_{ton}}$ les classe de demi tons pour chaque indice fréquentiel de la base MDCT. En supposant un accordage parfait à 440 Hz, un coefficient MDCT à la fréquence ν est converti en chroma p_ν^{MDCT} par :

$$p_\nu^{MDCT} = (12 \log_2 \frac{\nu}{440} + 69) \pmod{12}. \quad (4)$$

Remarquons que la classe d'un demi-ton donné correspond à plusieurs indices fréquentiels consécutif. En effet, l'échelle logarithmique utilisé dans la musique tonale occidentale résulte en un plus grand nombre de coefficients MDCT correspondant à une classe donnée.

Le modèle chroma utilisé pour les indicatrices de la couche tonale est alors

$$P_\Lambda \{\gamma_{ton,(q,\nu)} = 1\} = \begin{cases} p_{ton} & \text{if } \exists k \in [1, N_c] \mid p_\nu^{MDCT} = p_k^{chroma} \\ 1 - p_{ton} & \text{otherwise,} \end{cases}$$

where $0 \leq p_{ton} \leq 1$ et $\{p_k^{chroma}\}_{k=1..12}$ le vecteur de chroma normalisé.

On peut remplacer $\{p_k^{chroma}\}_{k=1..12}$ par un vecteur d'accord $\{p_k^{chroma}\}_{k=1..N_c}$ estimé sur le morceau de musique [7].

Pour la couche transitoire, un modèle similaire peut-être construit en utilisant l'information de « beats » pour leur position temporelle.

Le modèle est ensuite estimé par une méthode MCMC classique utilisant un échantillonneur de Gibbs.

3 Résultats

On présente sur la figure 1 les cartes de signifiante obtenus par le modèle proposé par Févotte *et al.* et la méthode chroma, pour un extrait de Mozart, auquel un bruit blanc gaussien a été ajouté (SNR de 10 dB). D'un point de vue débruitage, les trois méthodes obtiennent un résultat similaire (environ 16 dB).

4 Conclusion

La principale contribution de ce travail est de montrer que les *a priori* basé exclusivement sur la connaissance musicale sont une alternative viable aux modèles plus classiquement utilisés tels que les HMM, tout en apparaissant « naturels ». D'un point de vue performance en débruitage, les résultats obtenus sont tout à fait comparable. Cependant, les cartes de significances obtenues font plus clairement apparaître les structures attendues, ces structures étant directement prise en compte dans le modèle. Les futurs travaux consisteront à tester ce type d'approche sur des problèmes plus difficiles comme la séparation de sources.

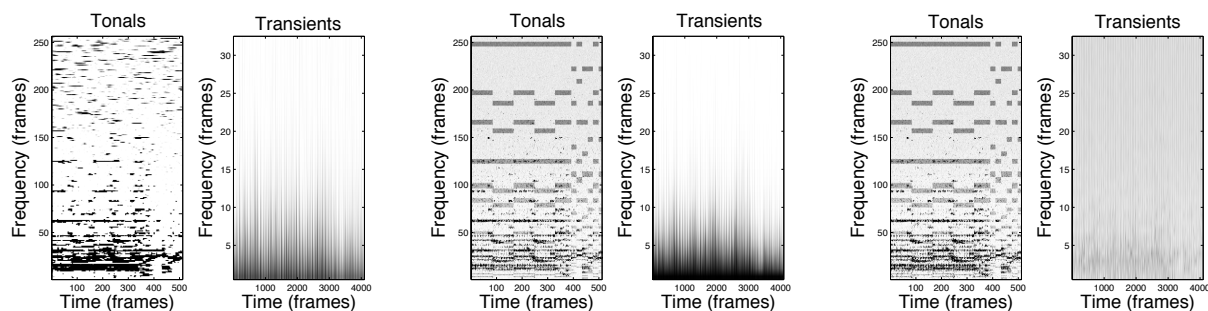


FIGURE 1 – Carte de signifiante pour chaque base (estimation au sens MMSE) pour un extrait de Mozart, et un rapport signal sur bruit de 10dB en entrée. Gauche : approche [4] ; Milieu : approche avec a priori musical (méthode chroma) pour la couche tonale seule. Droite : approche avec a priori musicale pour la couche tonale (méthode chroma) et la couche transitoire.

Références

- [1] L. Daudet and B. Torrèsani. *Hybrid representations for audiophonic signal encoding*, Sig. Proc. J., 2002.
- [2] M. D. Plumbley, T. Blumensath, L. Daudet, R. Gribonval M. E. Davies, *Sparse Representations in Audio and Music : from Coding to Source Separation*, Proc. IEEE vol. 98, no 6, pp 995-12005, 2010
- [3] P. Wolfe, S. Godsill, and W.-J. Ng . *Bayesian variable selection and regularization for time-frequency surface estimation*, J. of the Royal Stati. Soc. Serie B, 2004.
- [4] C. Févotte, B. Torrèsani, L. Daudet, and S. Godsill. *Sparse Linear Regression With Structured Priors and Application to Denoising of Musical Audio*, IEEE Trans. on Au. Sp., and Lang. Proc., 2008.
- [5] H. Papadopoulos and M. Kowalski. *Sparse Signal Decomposition on Hybrid Dictionaries Using Musical Priors* ISMIR 2011.
- [6] A. Sheh and D. Ellis. *Chord segmentation and recognition using EM-trained HMM*, ISMIR 2003.
- [7] H. Papadopoulos, and G. Peeters. *Joint estimation of chords and downbeats*, IEEE Trans. on Au. Sp., and Lang. Proc., 2011.
- [8] T. Fujishima. *Real-time chord recognition of musical sound : a system using common lisp music*, ICMC 1999.
- [9] A. Klapuri, A. Eronen, and J. Astola. *Analysis of the meter of acoustic musical signals*, IEEE Trans. on Au. Sp., and Lang. Proc. 2006.