HAL
open science

Outil d'analyse temps-fréquence multi-résolution appliqué aux signaux audio

Thomas Fillon, Jacques Prado, Roland Badeau

To cite this version:

Thomas Fillon, Jacques Prado, Roland Badeau. Outil d'analyse temps-fréquence multi-résolution appliqué aux signaux audio. Colloque GRETSI 2013, 2013, Brest, France. hal-00945239

HAL Id: hal-00945239

https://inria.hal.science/hal-00945239v1

Submitted on 25 Mar 2014

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.

Outil d'analyse temps-fréquence multi-résolution appliqué aux signaux audio

Thomas FILLON[1,3], Jacques PRADO, Roland BADEAU[2*]

[1]Parisson, France

[2]Institut Mines-Telecom, Telecom ParisTech, CNRS LTCI,
37/39 rue Dareau, 75014 Paris, France

[3]LAM, Institut Jean Le Rond d'Alembert, UPMC Univ. Paris 06, UMR CNRS 7190,
11 rue de Lourmel, 75015 Paris, France

thomas.fillon@parisson.com, roland.badeau@telecom-paristech.fr

Résumé – Cet article présente un outil d'analyse temps-fréquence multi-résolution généralisant la transformée à Q constant. Grâce à cette approche, il est possible de spécifier de manière ad-hoc les paramètres de fréquences centrales et de résolutions fréquentielles.

Abstract – In this article, a multi-resolution time-frequency analysis framework generalizing the Constant Q Transform is proposed. The purpose of this framework is to provide a flexible tool that enable to specify center frequencies and frequency resolution according to the application requirements.

1 Introduction

La transformée à Q constant (TQC) est un outil d'analyse temps-fréquence largement utilisé pour les signaux musicaux. Proposée par Brown [1], la TQC offre l'avantage de réaliser l'analyse fréquentielle des signaux selon une échelle géométrique alignée avec les notes de la gamme tempérée avec une résolution d'1/12 ou d'1/24 d'octave.

Dans cet article, dans la section 2, certaines propriétés et limites de la TQC seront détaillées et des améliorations seront proposées en conséquence. Dans la section 3, à partir de la définition de la TQC, nous développerons un formalisme plus général intégrant par ailleurs les améliorations proposées. Ce formalisme permet de définir un nouvel outil d'analyse temps-fréquence pour lequel le choix des fréquences centrales et des résolutions fréquentielles est libre et paramétrable. A partir d'une interprétation en banc de filtres, nous proposerons une méthode pour inverser cette nouvelle transformée.

Finalement, dans la section 4, nous exposerons différents exemples de paramétrisation de ce nouvel outil afin d'en illustrer la pertinence et les avantages pour des applications d'analyse du signal audio.

*Ce travail a été réalisé dans le cadre du programme de recherche Quaero financé par OSEO.

2 Transformée à Q constant

2.1 Mise en œuvre de la TQC

La résolution fréquentielle de la Transformée de Fourier Discrète est inversement proportionnelle à la longueur de la fenêtre d'analyse. La TQC, proposée par Brown [1, 2] autorise une variation de la longueur de la fenêtre pour chaque composante fréquentielle. Cela permet alors d'ajuster la résolution de façon à obtenir un facteur de qualité Q constant pour chaque composante spectrale. Le facteur de qualité est définit par $Q = \frac{f}{\Delta f}$ où f est la fréquence centrale du filtre et Δf sa largeur de bande à $-3dB$. L'aspect multi-résolution de la TQC présente quelques similitudes avec une analyse par ondelettes [3].

La TQC est définie pour chaque composante fréquentielle k selon la formule suivante :

$$x^{tqc}[k] = \frac{1}{N_k} \sum_{n=0}^{N-1} w_k[n] x[n] e^{-j2\pi n \frac{f_k}{f_s}} \tag{1}$$

où

- f_s est la fréquence d'échantillonnage
- $\{w_k[n]\}_{n\in[0,N_k-1]}$ est une fenêtre d'analyse de longueur N_k avec $N_k \in \mathbb{N}$,
- N est la longueur d'une trame de signal telle que $\forall k \in [0; K-1], N \geq N_k$,
- $f_k = (2^{1/B})^k . f_{min}$ sont les fréquences centrales géométriquement espacées et alignées avec les notes de la gamme tempérée,
- f_{min} est la fréquence analysée minimale pour $k = 0$,

– B est le nombre de composantes par octave de la transformée (typiquement $B = 12$ ou 24).

Les fenêtres d'analyse $\{w_k\}_{n\in[0,N_k-1]}$ sont construites à partir d'une même fonction de fenêtrage (par exemple rectangulaire, Hamming ou Blackman). Le facteur Q devant être constant pour chaque composante de la TQC, N_k varie selon la valeur de la fréquence centrale f_k. Dans [1], N_k est défini par $N_k = \widehat{Q} \cdot \frac{f_s}{f_k}$ où $\widehat{Q} = 1/(2^{1/B} - 1)$ dépend du nombre de composantes par octave B.

2.2 Discussion

La définition de la TQC présente quelques imprécisions. En particulier, elle ne prend pas en compte le type de fenêtre d'analyse qui influence pourtant sensiblement la résolution fréquentielle [4, 5]. Ainsi le paramètre $\widehat{Q}$ n'est pas le vrai facteur de qualité Q. De même, ne pas prendre en compte le type de fenêtre peut poser un problème de normalisation.

Dans la section suivante, nous proposons une généralisation de la TQC qui intègre par ailleurs certaines modifications visant à prendre en compte les problèmes relevés ci-dessus et permettre ainsi de respecter rigoureusement les spécifications en terme de résolution fréquentielle.

3 Outil paramétrable d'analyse multi-résolution

3.1 Définition

La définition de la TQC dans (1) peut être facilement étendue à un choix arbitraire de fréquences centrales et de résolutions fréquentielles $\{f_k; \delta f_k\}$ sans se limiter à un cadre musical. Selon le contexte applicatif, il peut être intéressant d'avoir cette flexibilité et de pouvoir spécifier une analyse temps-fréquence *ad-hoc*.

A cet effet, nous proposons d'établir un cadre formel qui généralise le principe de la TQC et définit un outil d'analyse temps-fréquence que nous désignerons sous le terme *Transformée de Fourier Multi-Résolution à court terme* (TF-MR). Dans ce contexte, la TF-MR est déterminée par :

– un ensemble arbitraire de K fréquences $\{f_k\}_{k\in[0;K-1]}$, avec $K \in \mathbb{N}$.
– un ensemble arbitraire $\{\delta f_k\}_{k\in[0;K-1]}$ de résolutions fréquentielles
– une fonction de fenêtrage $\mathcal{W}$ (par exemple rectangulaire, Hamming, Blackmann, ...)

En s'inspirant du formalisme de la TQC, la TF-MR est définie par :

$$x^{tf-mr}[k] = \frac{1}{\gamma_k} \sum_{n=0}^{N-1} a_k[n] x[n] e^{-j2\pi n \frac{f_k}{f_s}} \tag{2}$$

où

– N est la longueur d'une trame de signal (pour une implémentation par fenêtres recouvrantes, on définit également le pas d'avancement, R)
– $\{a_k[n]\}_{n\in[0,N]}$ est une fenêtre d'analyse de longueur N, contenant une fenêtre d'analyse éventuellement plus courte $\{w_k[n]\}_{n\in[0,N_k-1]}$ de longueur N_k et qui intègre un facteur de décalage afin de régler l'alignement des fenêtres $\{w_k\}$
– γ_k est un facteur de normalisation : $\gamma_k = \sum_{n=0}^{N_k-1} w_k[n]$

Pour le calcul de la TF-MR, le signal est décomposé en trames recouvrantes de longueur N.

Résolution fréquentielle Dans l'équation (2), il convient de souligner que la résolution fréquentielle de la TF-MR est définie par le type de fenêtre d'analyse et par leur longueur N_k. Afin d'assurer que la résolution fréquentielle δf_k spécifiée pour chaque composante est respectée, il convient de définir N_k de façon rigoureuse. Ceci est réalisé en se basant sur les méthodes classiques de synthèse de filtres numériques par fenêtrage [4, 5]. Pour chaque composante k, la longueur des fenêtres N_k est déterminée en fonction de la résolution fréquentielle δf_k spécifiée et de la fonction de fenêtrage choisie. Afin d'offrir encore plus de flexibilité dans cette spécification, la résolution spectrale peut être indiquée au choix en fonction de la largeur du lobe principale, de la largeur à $-3dB$ ou de la largeur de bande de bruit équivalente.

Selon ce principe, N_k est alors défini par :

$$N_k = \begin{cases} \delta\omega \cdot f_s/\delta f_k & \text{, largeur du lobe principal} \\ \Delta\omega \cdot f_s/\delta f_k & \text{, largeur à } -3dB \\ \Pi\omega \cdot f_s/\delta f_k & \text{, largeur de bande de bruit équivalente} \end{cases} \tag{3}$$

où $\delta\omega$, $\Delta\omega$ et $\Pi\omega$ sont des coefficients réels dépendants du type de fenêtre d'analyse. Des valeurs typiques pour ces coefficients sont données à titre d'exemple dans le tableau 1.

Fenêtre	$\Delta\omega$	$\delta\omega$	$\Pi\omega$
Rectangulaire	0.89	2	1.00
Hann	1.44	4	1.50
Hamming	1.30	4	1.36
Blackman	1.68	6	1.73

TABLE 1 – Caractéristiques de quelques fonctions de fenêtrage classiques [4, 5]. $\Delta\omega$, $\delta\omega$ et $\Pi\omega$ correspondent respectivement à la largeur à $-3dB$, à la largeur du lobe principal et à la largeur de bande de bruit équivalente du filtre exprimées en échantillons.

Alignement temporel des noyaux Le support temporel et l'alignement entre eux des noyaux a_k est défini par :

$$a_k[n] = \begin{cases} 0 & \text{if } n < \tau_k \\ w_k[n - \tau_k]\, e^{j2\pi\tau_k \frac{f_k}{f_s}} & \text{if } \tau_k \le n \le \tau_k + N_k - 1 \\ 0 & \text{if } \tau_k + N_k \le n \end{cases} \tag{4}$$

où $\tau_k \in \mathbb{N}$ est un terme de décalage qui permet de contrôler l'alignement temporel des K atomes au sein de la trame d'-

analyse globale de taille N tel que :

$$\tau_k = \begin{cases} 0, & \text{les atomes sont alignés au } \textit{gauche} \\ \frac{N-N_k}{2}, & \text{les atomes sont } \textit{centrés} \\ N - N_k, & \text{les atomes sont alignés à } \textit{droite} \end{cases} \quad (5)$$

3.2 Interprétation banc de filtres et inversion

Plusieurs méthodes ont été proposées pour effectuer une inversion partielle de la TQC [6]. Notamment, [7, 8] présentent une méthodologie très intéressante basée sur les trames de Gabor.

La liberté de spécification introduite pour la TF-MR empêche de recourir aux mêmes hypothèses simplificatrices et doit s'inscrire également dans un cadre plus général.

La difficulté majeure tient au fait que le recouvrement fréquentiel entre bandes et temporel entre trames est éventuellement non négligeable et doit être compensé. En considérant une représentation par banc de filtres [9] avec une décomposition polyphase d'ordre R où R est le pas d'avancement contrôlant le recouvrement entre les trames, il est possible de déterminer les conditions de reconstruction parfaite. A partir de ces conditions, on obtient un système linéaire faisant intervenir une matrice Toeplitz par blocs de grande dimension et qu'il est possible de résoudre en ayant recours à des méthodes itératives [10, 11].

La partie du signal dont le contenu spectral n'est pas couvert par les spécifications de la TF-MR ne peut pas être reconstruit. Si l'on souhaite malgré tout restaurer cette partie du signal, il convient de le spécifier et de compléter la matrice de la transformée pour la rendre inversible.

4 Exemples de spécifications particulières

Une boîte à outils Matlab sous licence GPL et contenant l'implémentation de la TF-MR et quelques exemples d'utilisation est disponible [1] librement.

4.1 Liens avec d'autres transformées

La définition très générale et flexible de la TF-MR la relie avec d'autres transformées usuelles comme la STFT, la TQC, la FFT multi-résolution [12]. Des exemples de paramétrisation sont donnés à titre indicatifs dans le tableau 2

4.2 TQC avec une résolution hybride

Un avantage de la TQC est sa forte résolution fréquentielle en basse fréquence. En contrepartie la résolution temporelle pour ces fréquences devient très faible et la TQC a donc recours à de très longues fenêtres d'analyse (de l'ordre d'une seconde) ce qui peut être inapproprié pour l'analyse de signaux musicaux si la durée d'une note est dépassée ou si on souhaite détecter l'instant d'attaque. L'illustration de ce problème de résolution temporelle de la TQC en basse fréquence es illustrée sur la figure 1 qui représente la TQC d'un extrait de guitare basse de $2,5s$.

1. https://github.com/thomasfillon/MRT-toolbox

δf_k	f_k	
	Uniforme	1/B-octave
Constante	STFT	log-STFT
Constante par bande	MR-FFT	-
Q constant	-	TQC

TABLE 2 – Exemples de correspondance entre différentes spécifications de la TF-MR en terme de fréquences $\{f_k\}_{k\in[0;K-1]}$ et de résolutions $\{\delta f_k\}_{k\in[0;K-1]}$ avec quelques outils classiques d'analyse temps-fréquence audio

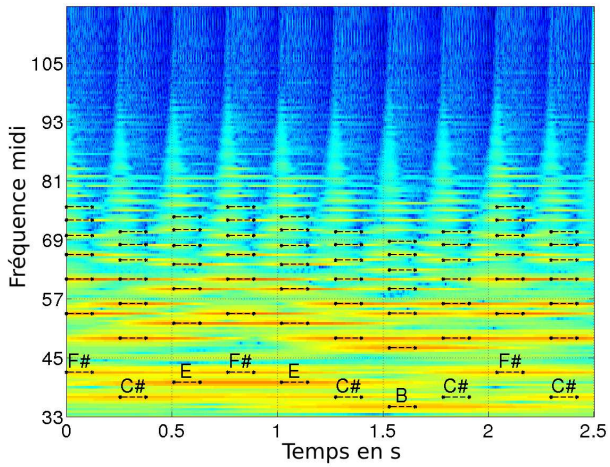

FIGURE 1 – Analyse par TF-MR d'un extrait musical (guitare basse) avec une résolution à Q constant. Les notes, les temps d'attaque et les durées sont indiquées pour les harmoniques correspondantes.

A partir de la TF-MR, il est possible de modifier, par exemple, de relâcher la contrainte sur le facteur de qualité de la TQC pour les basses fréquences afin d'améliorer la résolution temporelle. A titre d'exemple, une implémentation de la TF-MR avec un facteur de qualité constant au dessus de $500Hz$ et une résolution fréquentielle constante au dessous de $500Hz$ a été appliqué au signal de guitare basse de l'exemple précédent. Le résultat est illustrée sur la figure 2.

4.3 Analyse par filtres auditifs

Pour des applications psychoacoustiques, il est possible de spécifier une répartition des fréquences centrales et une résolution fréquentielle conformes aux filtres auditifs [13] en alignant les fréquences centrales et les résolutions fréquentielle sur l'échelle en bande rectangulaire équivalente. Un exemple d'une telle implémentation de la TF-MR est illustré par la figure 3 sur un signal de parole.

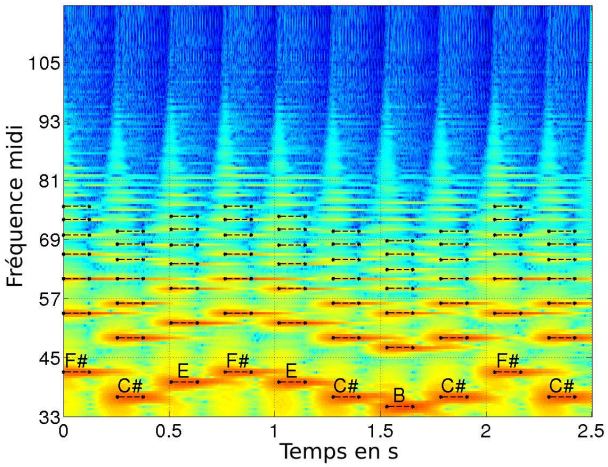

FIGURE 2 – Analyse par TF-MR d'un extrait musical (guitare basse) avec une résolution fréquentielle hybride. Les notes, les temps d'attaque et les durées sont indiquées pour les harmoniques correspondantes.

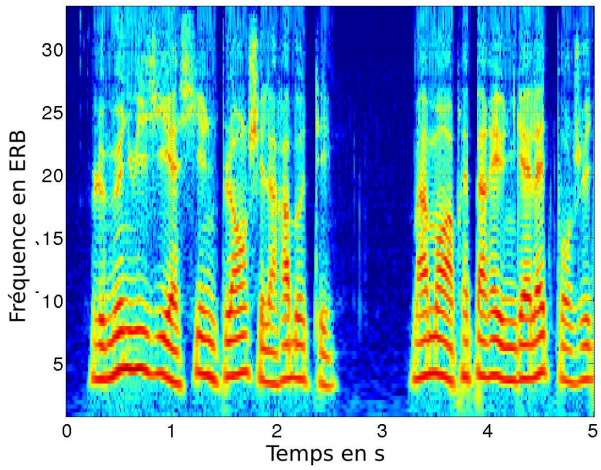

FIGURE 3 – Analyse TF-MR d'un signal de parole selon une résolution ERB. $f_{min} = 25Hz$, $f_{max} = 8kHz$, nombre de canaux $K = 100$

5 Conclusion

A travers ce travail, nous souhaitons introduire un nouvel outil d'analyse temps-fréquence très flexible et qui permet de spécifier librement un ensemble de fréquences à analyser et, pour chaque fréquence, la résolution fréquentielle correspondante. Grâce au formalisme des bancs de filtres, nous pouvons définir les conditions de reconstruction qui permettent une inversion de cette nouvelle transformée à court terme. Plusieurs exemples d'implémentations dans le domaine du traitement du signal audio ont été explicités et permettent d'envisager une utilisation dans d'autres domaines (biomédical, bio-acoustique). Les perspectives d'amélioration de cet outil portent prioritairement sur la mise en œuvre d'une implémentation rapide et le développement d'une méthode d'inversion efficace dans le cas où la taille des matrices est trop importante.

Références

[1] J. C. BROWN, « Calculation of a constant Q spectral transform », *Journal of the Acoustical Society of America*, vol. 89, no. 1, p. 425–434, 1991.

[2] J. C. BROWN, « An efficient algorithm for the calculation of a constant Q transform », *Journal of the Acoustical Society of America*, vol. 92, no. 5, p. 2698, 1992.

[3] S. MALLAT, *A Wavelet Tour of Signal Processing, Second Edition (Wavelet Analysis & Its Applications)*. Academic Press, 2 éd., 1999.

[4] A. OPPENHEIM et R. SCHAFER, *Discrete-time signal processing*. Prentice-Hall signal processing series, Prentice Hall, 2010.

[5] F. J. HARRIS, « On the use of windows for harmonic analysis », *Proceedings of the IEEE*, vol. 6, no. 1, p. 51–83, 1978.

[6] C. SCHÖRKHUBER et A. KLAPURI, « Constant-Q Transform Toolbox For Music Processing », *in 7th Sound and Music Conf*, 2010.

[7] M. DÖRFLER, N. HOLIGHAUS, T. GRILL et G. A. VELASCO, « Constructing an invertible constant-Q transform with nonstationary Gabor frames », *in Proceedings of the 14th International Conference on Digital Audio Effects (DAFx 11), Paris, France*, 2011.

[8] N. HOLIGHAUS, M. DÖRFLER, G. A. VELASCO et T. GRILL, « A framework for invertible, real-time constant-Q transforms », *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, no. 4, p. 775–785, 2013.

[9] P. VAIDYANATHAN, « Multirate digital filters, filter banks, polyphase networks, and applications : a tutorial », *Proceedings of the IEEE*, vol. 78, p. 56 –93, jan 1990.

[10] T. KAILATH et A. H. SAYED, éds, *Fast reliable algorithms for matrices with structure*. Philadelphia, PA, USA : Society for Industrial and Applied Mathematics, 1999.

[11] M. K. NG, *Iterative Methods for Toeplitz Systems (Numerical Mathematics and Scientific Computation)*. New York, NY, USA : Oxford University Press, Inc., 2004.

[12] K. DRESSLER, « Sinusoidal extraction using an efficient implementation of a multi-resolution FFT », *in Proceeding of the 9th Int. Conference on Digital Audio Effects (DAFx-06), Montreal, Canada, September 18-20*, p. 247–252, 2006.

[13] B. R. GLASBERG et B. C. MOORE, « Derivation of auditory filter shapes from notched-noise data », *Hearing Research*, vol. 47, no. 1-2, p. 103 – 138, 1990.