



Transformée à Q constant

Jacques Prado

2010D004

février 2010

Département Traitement du Signal et des Images
Groupe AAO : Audio, Acoustique et Ondes

Transformée à Q constant

J.Prado, département TSI

Institut Télécom ; Télécom ParisTech ; CNRS/LTCI
46 rue Barrault, 75683, Paris CEDEX 13

courriel : jacques.prado@telecom-paristech.fr

4 février 2010

Table des matières

1	Principe de la CQT	1
2	Algorithme rapide	1
3	Modifications apportées	2
3.1	Utilisation d'une fenêtre	2
3.2	Synchronisation temporelle	2
4	Résultats de simulation	3
4.1	Simulation 1	3
4.2	Simulation 2	3
5	Conclusion	4

Résumé

Le but de la transformée à Q constant (CQT), est d'obtenir la transformée de Fourier d'un signal de musique pour certaines fréquences espacées logarithmiquement de façon à s'adapter à l'échelle musicale. De plus cette transformée doit respecter une résolution dépendant de la fréquence de manière à conserver un rapport constant entre la résolution et la fréquence pour laquelle s'effectue le calcul. Le principe de calcul a été introduit dans [2], un algorithme rapide a été proposé dans [3] et une version optimisée en code Matlab est développée dans [1]. Le but de ce rapport est d'apporter quelques corrections au code Matlab et d'introduire une liberté de choix sur la position de l'instant auquel l'analyse du signal est effectué permettant ainsi d'améliorer la détection de l'instant d'apparition d'une fréquence.

1 Principe de la CQT

Dans l'article original [2], la CQT est définie comme un équivalent d'un banc de filtre en $\frac{1}{24}$ ème d'octave, mais peut facilement être étendu à d'autres cas. Partant de cette remarque, cela revient à dire que si l'on veut faire l'analyse d'un signal musical avec un espacement du quart de ton sur l'échelle tempérée, alors la m ème composante spectrale est définie par : $f_m = (2^{1/24})^m f_{\min}$, où f_{\min} est choisie légèrement inférieure à la fréquence à laquelle on veut commencer l'analyse.

Lorsque l'on effectue une analyse par transformée de Fourier discrète (TFD), la résolution (ou largeur de bande) δf que l'on peut obtenir est inversement proportionnelle au nombre de points utilisés et dépend du type de fenêtrage choisi. Afin que le rapport, noté Q , entre la fréquence et la résolution soit constant (i.e. $Q = \frac{f}{\delta f} \simeq 34$ dans le cas du quart de ton), il faut que la longueur de la fenêtre d'analyse soit dépendante de la fréquence. En notant f_e la fréquence d'échantillonnage on obtient :

$$N(m) = \frac{f_e}{\delta f_m} = \frac{f_e}{f_m} Q. \quad (1)$$

Sous ces conditions, la m ème composante spectrale calculée à Q constant est exprimée dans [2] sous la forme :

$$X_m^{cq} = \frac{1}{N(m)} \sum_{n=0}^{N(m)-1} p(m, n) x(n) e^{-j2\pi \frac{Qn}{N(m)}} \quad (2)$$

expression dans laquelle $p(m, n)$ est une fenêtre de pondération de longueur $N(m)$. Nous reviendrons un peu plus loin sur cette expression.

Le calcul direct à l'aide de la TFD serait beaucoup trop lourd dans la mesure où il faudrait en calculer autant qu'il y a de $N(m)$ différents :

$$N(m) = \frac{N_{\max}}{(2^{1/24})^m} \quad (3)$$

Où N_{\max} est égal à Q fois la période de la fréquence la plus basse à analyser (i.e. f_{\min}).

Une propriété élémentaire de la TFD permet de contourner la difficulté.

2 Algorithme rapide

En remarquant que :

$$\sum_{n < N} x_n t_n^* = \frac{1}{N} \sum_{m < N} X_m T_m^* \quad (4)$$

où X et T sont les TFD respectives de x et t .

Il devient alors plus simple de calculer les différents X_m^{cq} donné par l'expression (2). Il suffit pour cela de fixer $N = N_{\max}$ et de précalculer une seule fois les transformées T associées aux longueurs de fenêtres d'analyse et différentes selon la résolution.

Un algorithme efficace a été proposé dans [1], laissant en paramètre la finesse d'analyse, autrement dit permettant de changer le $\frac{1}{24}$ ème d'octave en $\frac{1}{b}$ ème d'octave.

Ayant défini les fréquences f_{\min} et f_{\max} de l'analyse, on en déduit le nombre K de composantes, à calculer :

$$K = \left\lceil b \log_2 \left(\frac{f_{\max}}{f_{\min}} \right) \right\rceil \quad (5)$$

où $\lceil y \rceil$ représente le plus petit entier supérieur ou égal à y .

La constante Q est définie par :

$$Q = \left(2^{\frac{1}{b}} - 1 \right)^{-1} \quad (6)$$

Les longueurs nécessaires aux différentes résolutions sont données par :

$$N(m) = \left\lceil \frac{f_e}{f_m} Q \right\rceil \quad (7)$$

avec :

$$f_m = f_{\min} 2^{\frac{m}{b}}, m = 0, \dots, K-1 \quad (8)$$

Avec :

$$\begin{aligned} t_{nm} &= \frac{1}{N(m)} p(m, n) e^{-j2\pi \frac{Q_n}{N(m)}}, 0 \leq n < N(m) \\ T_{km} &= TFD(t_{nm}), \text{ de longueur } N_{\max} \\ X_k &= TFD(x_n), \text{ de longueur } N_{\max} \end{aligned} \quad (9)$$

Le calcul des coefficients X_m^{cq} se ramène à :

$$X_m^{cq} = \frac{1}{N_{\max}} \sum_{k=0}^{N_{\max}-1} X_k T_{km}^*, m = 0, \dots, K-1 \quad (10)$$

3 Modifications apportées

3.1 Utilisation d'une fenêtre

Si l'on reprend l'expression (2), $p(m, n)$ représente une fenêtre d'observation permettant de limiter la longueur du signal utile à la résolution souhaitée. Afin que l'influence de cette fenêtre soit indépendante du type de fenêtre et de sa longueur $N(m)$, la normalisation doit être faite par la somme des coefficients de la fenêtre. Ainsi l'expression (9) doit être modifiée en :

$$t_{m,n} = \frac{1}{\sum_{n=0}^{N(m)-1} p(m, n)} p(m, n) e^{-j2\pi \frac{Q_n}{N(m)}}, 0 \leq n < N(m) \quad (11)$$

3.2 Synchronisation temporelle

Selon le principe développé plus haut, le signal x_n utilisé pour le calcul de X_m^{cq} est de longueur N_{\max} , par contre les fenêtres appliquées sont de longueur variable $N(m)$ et sont appliquées à partir de l'instant 0 de la fenêtre de signal. Il s'ensuit que l'instant de détection considéré comme étant le point milieu de la fenêtre de signal ne correspond pas à l'instant milieu de la fenêtre de pondération appliquée, instant qui lui est variable en fonction de la longueur de cette fenêtre.

Il est possible assez simplement de modifier l'algorithme précédent pour définir dans la fenêtre de signal à partir de quel instant on applique la fenêtre de pondération et faire en sorte que de cette manière les K fenêtres utilisées soient centrées au même instant.

Reprenons l'expression du calcul d'un coefficient X_m^{cq} :

$$X_m^{cq} = \sum_{n=0}^{N(m)-1} w_{m,n} x_n e^{-j2\pi \frac{Q_n}{N(m)}} \quad (12)$$

où $w_{m,n}$ est donné par $w_{m,n} = \frac{1}{\sum_{n=0}^{N(m)-1} p(m, n)} p(m, n)$.

Si l'on veut que la fenêtre de pondération soit appliqué sur le signal à l'instant n_1 et non à l'instant zéro, l'expression (12) devient :

$$X_m^{cq} = \sum_{n=0}^{N(m)-1} w_{m,n} x_{n+n_1} e^{-j2\pi \frac{Q_n}{N(m)}} \quad (13)$$

par changement d'indice $l = n + n_1$, il vient :

$$X_m^{cq} = \sum_{l=n_1}^{n_2} w_{m,(l-n_1)} x_l e^{-j2\pi \frac{Q(l-n_1)}{N(m)}}, \text{ où } n_2 = n_1 + N(m) - 1 \quad (14)$$

En notant $r_{m,n}$ une fenêtre définie par :

$$r_{m,n} = \begin{cases} 0, & l < n_1 \\ w_{m,(l-n_1)} e^{j2\pi \frac{Q(l-n_1)}{N(m)}}, & l = n_1, \dots, n_2 \\ 0, & n_2 < l < N_{\max} - 1 \end{cases} \quad (15)$$

Alors X_m^{cq} s'écrit :

$$X_m^{cq} = \sum_{n=0}^{N_{\max}-1} x_n r_{m,n}^* \quad (16)$$

Et l'expression (16) peut se calculer à partir de TFD de longueur N_{\max} :

$$X_m^{cq} = \frac{1}{N_{\max}} \sum_{k=0}^{N_{\max}-1} X(k) R_m^*(k), \quad m = 0, \dots, K-1 \quad (17)$$

où :

$$X(k) = \sum_{n=0}^{N_{\max}-1} x_n e^{-j2\pi \frac{nk}{N_{\max}}} \quad (18)$$

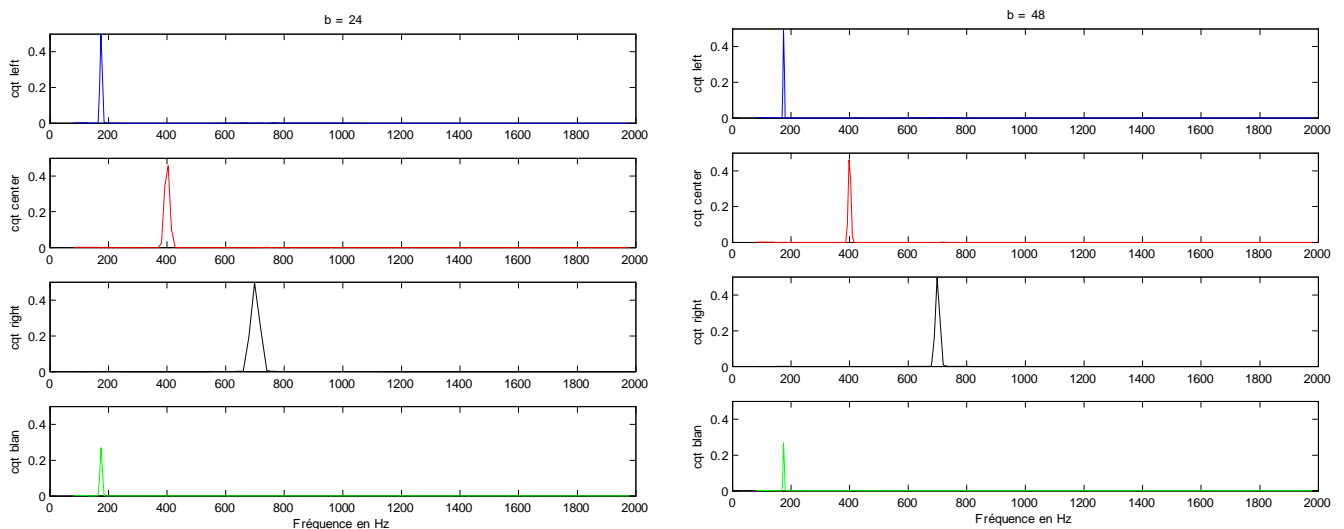
et :

$$R_m(k) = \sum_{n=0}^{N_{\max}-1} r_{m,n} e^{-j2\pi \left(\frac{nk}{N_{\max}}\right)} \quad (19)$$

4 Résultats de simulation

4.1 Simulation 1

Pour cette première simulation, le signal x_n est constitué de 3 sinusoïdes successives représentant chacune 1/3 de la fenêtre d'analyse et situées aux fréquences 175, 400 et 700 hertz pour une fréquence d'échantillonnage de 12000Hz. L'analyse est effectuée avec 2 valeurs de b , $b = 24$ ce qui correspond à un N_{\max} de 8192 et $b = 48$ qui correspond à un N_{\max} de 16384. Dans les deux cas les fenêtres sont positionnées au début, ce qui correspond à l'analyse proposée par l'algorithme cité dans [1], au centre de la fenêtre de signal et à gauche de la fenêtre de signal. Les résultats sont comparés à celui donné par l'algorithme de Blankertz et l'on peut vérifier que pour une fenêtre de signal donnée l'analyse de Blankertz (notée *cqt_blan* sur les figures suivantes) montre la présence de la 1^{ère} sinusoïde et que, selon que la fenêtre est située à gauche (*cqt_left*), au centre (*cqt_center*) ou à droite (*cqt_right*) on distingue la présence de la 1^{ère}, la 2^{ème} ou la 3^{ème} sinusoïde.



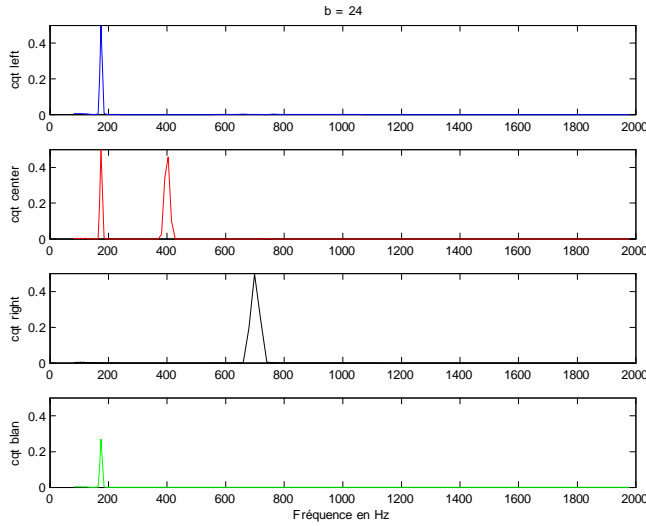
3 sinusoïdes successives.

Résolution supérieure.

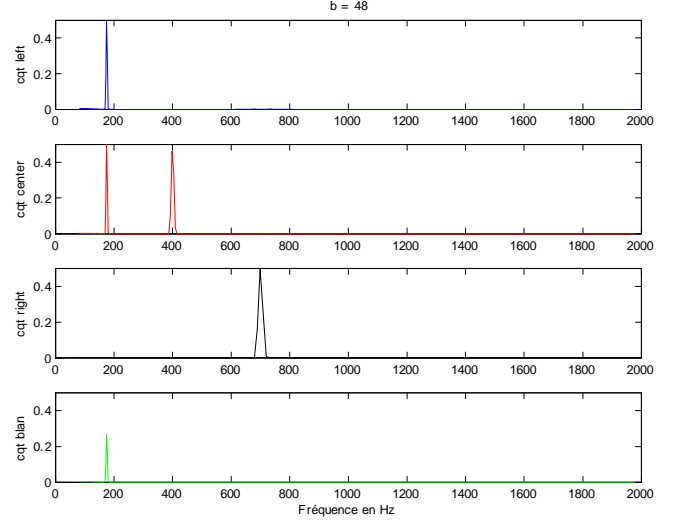
La possibilité de modifier la position de la fenêtre de pondération offre une souplesse supplémentaire pour mieux déterminer l'instant d'apparition d'une fréquence.

4.2 Simulation 2

Dans cette simulation, le signal x_n est comme précédemment constituée de trois sinusoïdes, à cette différence près que la 1^{ère} sinusoïde à une durée égale au 2/3 de la fenêtre d'analyse.



Détection de 2 sinusoïdes au centre.



Résolution supérieure.

Comme il était prévisible, les coefficients `cqt_center` montrent la présence des deux fréquences au centre de la fenêtre d'analyse.

5 Conclusion

La modification possible de la position de la fenêtre de pondération permet d'obtenir un instant de synchronisation identique pour tous les coefficients calculés sur une même fenêtre de signal, ce qui n'est pas le cas lorsque la fenêtre de pondération, de longueur variable, est systématiquement calée au début du signal. De plus le fait de pouvoir se déplacer au sein d'une même fenêtre de signal permet d'éventuellement mieux localiser l'apparition d'une fréquence. En effet l'incertitude sur la position temporelle et d'autant plus grande que la fenêtre d'analyse est longue, ce qui est nécessaire pour la résolution en basse fréquence. Dans ces conditions les coefficients calculés pour les hautes fréquences ne prennent en compte qu'une petite partie des échantillons de la fenêtre de signal il est donc possible, comme c'est le cas dans les simulations précédentes de rater l'apparition de certaines fréquences. Ceci est généralement compensé par le fait que l'analyse s'effectue en faisant se recouvrir les fenêtres successives d'un facteur qui est en général choisi égal à la moitié de la longueur de la fenêtre. Avec les notations utilisées, les fenêtres successives se chevauchent donc de $N_{\max}/2$ échantillons. Si la longueur de la fenêtre de pondération la plus courte $N_{\min}(m)$ est telle que $N_{\min}(m) < N_{\max}/2$ et que l'on a calé les fenêtres de pondération au début de la fenêtre de signal, alors pour chaque fenêtre, les échantillons compris entre $N_{\min}(m)$ et $N_{\max}/2$ ne sont jamais analysés pour les hautes fréquences, ce qui peut se traduire par des manques de détection. La plupart des mises en œuvre de la transformée à Q constant utilisent le recouvrement de $N_{\max}/2$, alors qu'il serait plus approprié d'utiliser un recouvrement de $N_{\min}(m)/2$.

Références

- [1] B. Blankertz, "<http://wwwmath.uni-muenster.de/logik/personen/blankertz/constq/constq.html>."
- [2] J. C. Brown, "Calculation of a constant q spectral transform," *J. Acoust. Soc. Am.*, pp. 425–434, 1991.
- [3] J. C. Brown and M. S. Puckette, "An efficient algorithm for the calculation of a constant q transform," *J. Acoust. Soc. Am.*, pp. 2698–2701, 1992.

