Une inversion simple de la transformée à Q constant

Jacques Prado

2011D006

mai 2011

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

Une inversion simple de la 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

16 mai 2011

Table des matières

Résumé

Dans ce rapport nous donnons une interprétation de la transformée à Q constant qui permet d'introduire simplement une approche possible de l'inversion de cette transformation. Des exemples sont donnés qui permettent de valider l'approche considérée.

1 Principe de la CQT

Nous rappelons ici que le principe de base de la CQT, appliquée au signal musical, est de pouvoir effectuer une analyse de type temps fréquence sous contrainte d'une résolution en fréquence suffisante pour distinguer les différentes notes d'une octave. Dans l'article original [1], la CQT est définie comme un équivalent d'un banc de filtre en $\frac{1}{24}^{ème}$ d'octave, mais peut facilement être étendue 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 : $f_m = \left(2^{1/24}\right)^m f_{\min}$, où $f_{\min}$ est choisie égale 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 $f_e = 1$ la fréquence d'échantillonnage et $f_m = \frac{F_m}{f_e}$ la fréquence normalisée que l'on veut analyser, on obtient :

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

Sous ces conditions, la $m^{ème}$ composante spectrale calculée à Q constant est exprimée sous la forme :

$$X_m^{cq} = \sum_{n=0}^{N(m)-1} h(m,n) x(n) e^{-j2\pi n f_m} \tag{2}$$

expression dans laquelle $h(m,n)$ est une fenêtre de pondération normalisée par la somme de ses coefficients, de longueur $N(m)$.

Si la fenêtre est une fenêtre de Hamming $ham(m,n)$ alors :

$$h(m,n) = \frac{ham(m,n)}{\sum_{n=0}^{N(m)-1} ham(m,n)},\ 0 \leq n < N(m) \tag{3}$$

L'expression (2) est calculée sous sa forme équivalente par TFD (cf [2]) :

$$X_m^{cq} = \frac{1}{N} \sum_{k=0}^{N-1} X(k) H_m^*(k),\ m = 0, \cdots, K-1 \tag{4}$$

où $K = \left\lceil b \log_2 \left(\frac{f_{\max}}{f \min} \right) \right\rceil$ est le nombre de fréquences par octave, $f_{\min}$ la fréquence de début de la dernière octave et $f_{\max}$ la fréquence finale de cette octave.

avec :

$$X(k) = \sum_{n=0}^{N-1} x_n e^{-j2\pi \frac{nk}{N}} \tag{5}$$

$$H_m(k) = \sum_{n=0}^{N-1} r_{m,n} e^{-j2\pi \left(\frac{nk}{N}\right)} \tag{6}$$

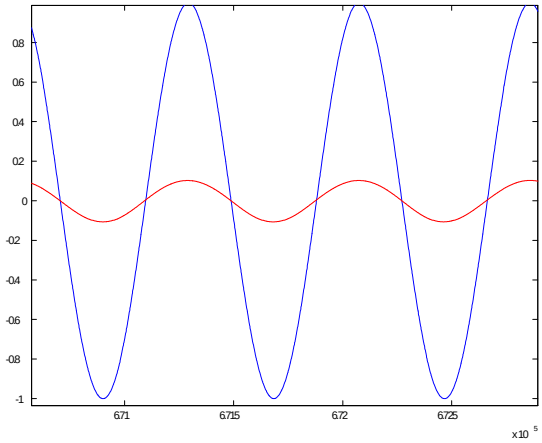
et :

$$r_{m,n} = \begin{cases} 0,\ l < n_1 \\ h_{m,(l-n1)} e^{j2\pi \frac{Q(l-n_1)}{N(m)}},\ l = n_1, \cdots, n_2 \\ 0,\ n_2 < l < \end{cases} \tag{7}$$

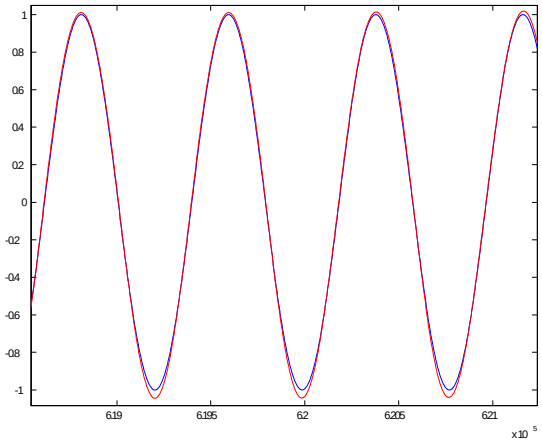
2 Remarque sur la CQT

Si l'on s'en tient à la définition de la CQT, le coefficient X_m^{cq} représente la composante à la fréquence f_m du spectre de x_n pondérée par la fenêtre $h(m,n)$. Si la sélectivité en fréquence est respectée et si la note jouée possède une fréquence entre f_m et f_{m+1}, alors les coefficients X_m^{cq} seront mal estimés.

On donne ci-dessous la reconstruction d'une sinusoïde à $56.3Hz$ calculée sur 1 octave et 12 bins par octave avec une CQT respectant la sélectivité, à comparer à la reconstruction dans les mêmes conditions avec une CQT ne respectant pas la sélectivité :



Fréquences 56.3 Hz CQT exacte.



Fréquences 56.3 Hz CQT approchée.

Pour prendre en compte un possible décalage sur une note jouée, il suffit de ne pas respecter la résolution fréquentielle et d'utiliser un coefficient $Q/2$ au lieu d'un coefficient Q, c'est ce qui est utilisé par la majorité des algorithmes CQT existant.

On peut donc espérer en utilisant une résolution moindre, obtenir des résultats d'analyse utilisables même sur des instruments mal accordés.

3 Interprétation de la CQT

Si l'on reprend l'expression (2) et en notant $z_n = h(m,n)x(n)$, alors on peut interpréter X_m^{cq} comme la composante à la fréquence 0 de la TFD du signal complexe $z_n e^{-j2\pi n f_m}$ ou, ce qui revient au même, la composante spectrale à la fréquence f_m du signal z_n.

$$\begin{aligned} X_m^{cq} &= \sum_{n=0}^{N-1} z_n e^{-j2\pi n f_m} \\ Y(0,m) &= \sum_{n=0}^{N-1} z_n e^{-j2\pi n f_m} e^{-j2\pi \frac{nk}{N}} \Big|_{k=0} \end{aligned} \tag{8}$$

N'ayant qu'une composante spectrale à sa disposition, il est illusoire de vouloir retrouver x_n, sauf si l'on considère que $Y(0)$ est la seule composante non-nulle de la partie utile du spectre du signal x_n. Autrement dit on interprète z_n comme étant une sinusoïde de fréquence f_m pour laquelle on a ramené la fréquence à 0 en modulant à l'aide de l'exponentielle complexe $e^{-j2\pi n f_m}$. C'est l'hypothèse sous-jacente que l'on émet dans le calcul de la CQT pour laquelle on ajuste la fenêtre de pondération afin qu'elle soit équivalente à un filtre suffisamment sélectif pour ne prendre en compte que la composante à la fréquence f_m du signal de départ x_n. Tenter de reconstruire z_n sous la forme d'une sinusoïde à la fréquence f_m devient alors possible simplement en effectuant la démodulation et en prenant la TFD inverse de la séquence $Z(k,m)$ définit par :

$$Z(k,m) = Y(0,m)P(k,m) \tag{9}$$

où :

$$P(k,m) = \sum_{n=0}^{N-1} p_n e^{j2\pi \frac{nk}{N}},\ k = 0, \cdots, N-1 \tag{10}$$

et p_n est une fenêtre de longueur $N(m)$ centrée dans la fenêtre de longueur N et qu'il reste à définir.

Ayant analysé le signal x_n pour différentes valeurs de m, c'est à dire les différentes fréquences f_m censées le composer, on pourra, reconstituer une approximation de x_n, notée $\hat{x}_n$ qui sera une estimation par un polynôme trigonométrique de la forme :

$$\hat{x}_n = \sum_{m=1}^{M} z(n,m) = \sum_{m=1}^{M} \alpha_m \cos(2\pi n f_m + \theta_m) \tag{11}$$

Expression dans laquelle α_m est pris constant sur une durée égale à $N(m)$, largeur de la fenêtre d'analyse, et où θ_m dépend de la phase de $Y(0,m)$.

$Y(0,m)$ est donc interprété comme la contribution du signal $\alpha_m \cos(2\pi n f_m)$ au signal x_n. Dans le calcul de la CQT, tout se passe comme si on prenait seulement la partie des fréquences positives du spectre, il suffit pour s'en convaincre d'écrire $\cos(2\pi n f_m) = \frac{1}{2}\left(e^{j2\pi n f_m} + e^{-j2\pi n f_m}\right)$ et de constater que la CQT revient à ne calculer que la composante correspondant à $\frac{1}{2}e^{j2\pi n f_m}$, autrement dit :

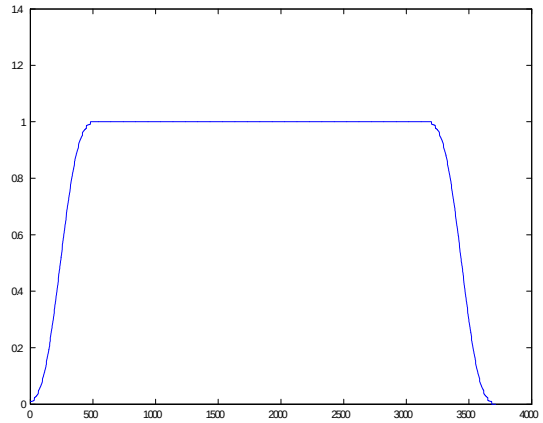
$$z(n,m) = 2\,\mathrm{real}(FFT^{-1}(Z(k,m))) \tag{12}$$

En effectuant cette opération pour chaque coefficient X_m^{cq} il suffit de faire la somme des séquences $z(n,m)$ pour obtenir une approximation de x_n par une somme de sinusoïdes pondérées :

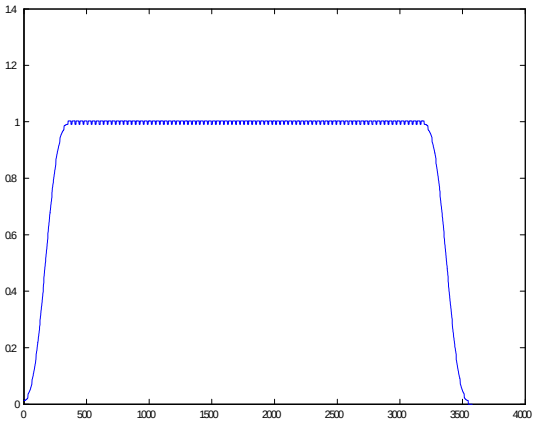
$$\hat{x}_n = \sum_m \alpha_m \cos(2\pi n f_m + \theta_m) \tag{13}$$

4 Reconstruction par addition recouvrement

Le calcul précédent permet de retrouver, pour chaque tranche analysée, l'approximation $z(n,m)$ pondérée par une fenêtre $p(n,m)$, l'analyse s'effectuant sur des fenêtres de longueur N avec un pas d'avancement de R, la reconstruction par addition recouvrement sera possible si la somme des fenêtres de pondération décalées de R donne une constante. C'est quasiment le cas pour une fenêtre de Hamming et en fait c'est toujours le cas si la fenêtre est de longueur une puissance de 2 avec un décalage R lui aussi une puissance de 2. Dans le cas de l'implémentation de la CQT par décimation (cf [2]), R est une puissance de 2, mais les fenêtres sont de longueur $N(m) \neq 2^n$, on peut cependant vérifier que la contrainte de somme constante par décalage est quasi vérifiée.

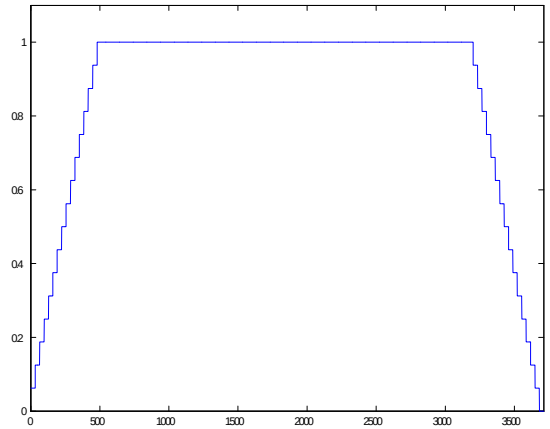


$N(m) = 512,\ R = 32.$

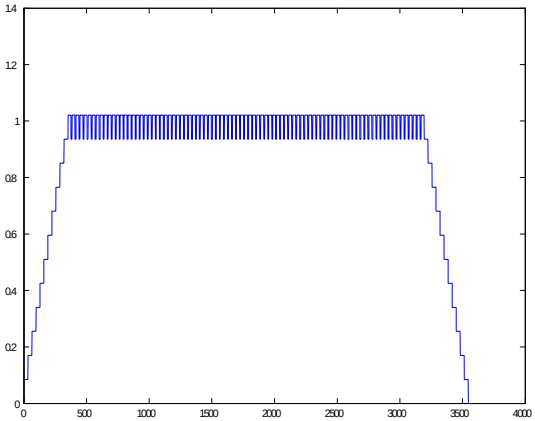


$N(m) = 376,\ R = 32.$

Mais c'est aussi le cas pour une fenêtre rectangulaire :



$N(m) = 512,\ R = 32.$



$N(m) = 376,\ R = 32.$

L'addition recouvrement est un peu moins bonne avec une fenêtre rectangulaire, mais le défaut doit être évalué sur le signal effectivement reconstruit.

5 Exemples de reconstruction

Afin d'illustrer les résultats apportés par la méthode précédemment exposée, on l'applique à différentes situations en utilisant successivement une fenêtre de Hamming et une fenêtre rectangulaire.

5.1 Une sinusoïde

La situation la plus élémentaire est l'analyse d'un signal sinusoïdal. L'analyse est effectuée sur un signal sinusoïdal de fréquence $55hz$ à l'aide d'une CQT pour laquelle la fréquence minimale est fixée à $55hz$. Pour cet exemple, comme pour les suivants, les méthodes de reconstruction sont sensiblement équivalentes, mais la reconstruction à l'aide d'une fenêtre de Hamming présente un défaut de gain systématique.

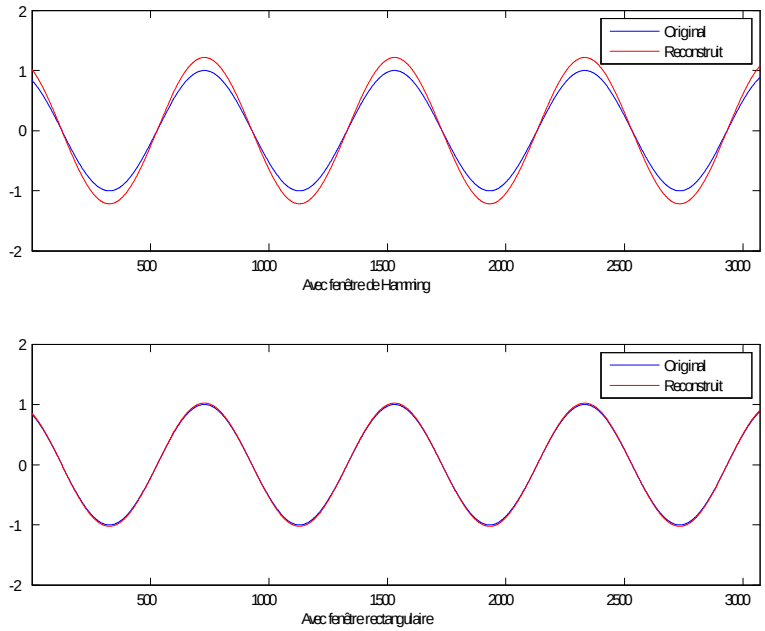


FIG. 1 – Comparaison de reconstruction pour une sinusoïde.

5.2 Somme de sinusoïdes

Ci-dessous nous examinons la somme de 3 sinusoïdes de même amplitude et dont les fréquences sont fixées à 55, 125 et 220 hertz. L'analyse s'effectue à l'aide d'une CQT sur 3 octaves avec 24 bins par octave et une fréquence de départ fixée à $55hz$.

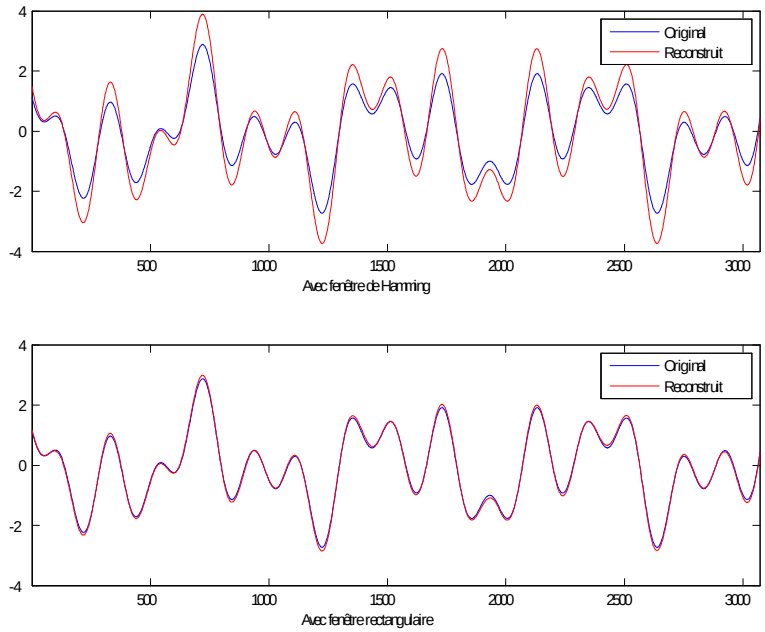


FIG. 2 – Comparaison de reconstruction pour une somme de 3 sinusoïdes.

5.3 Signal carré

Nous analysons ici un signal carré d'horloge de fréquence fondamentale $f_0 = 55hz$, avec une analyse sur 3 octaves et une analyse sur 7 octaves, chacune sur 48 bins de fréquence par octave. Comme attendu cela met en évidence le phénomène de Gibbs à la reconstruction.

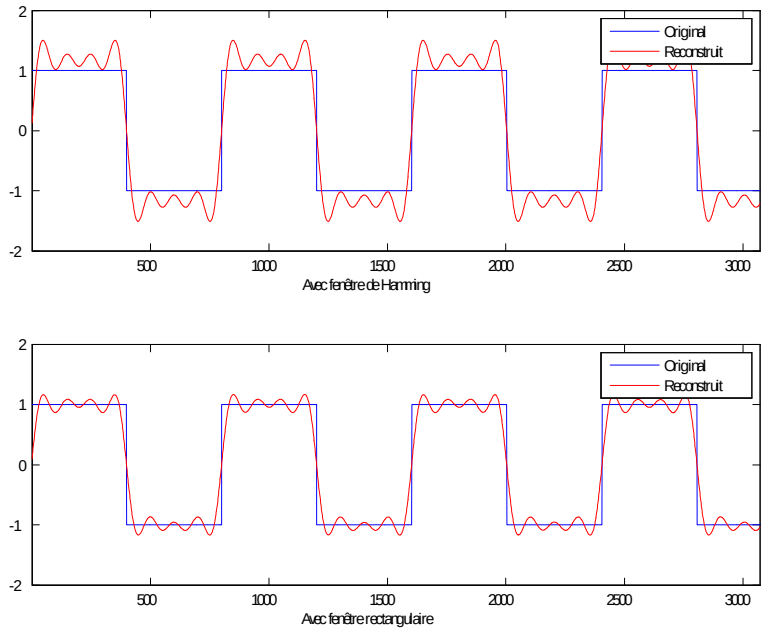


FIG. 3 – Reconstruction sur 3 octaves.

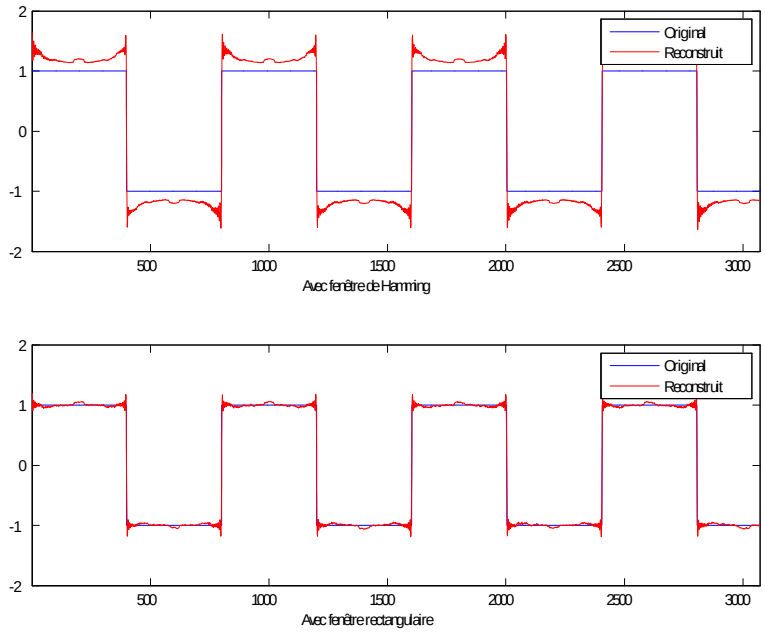


FIG. 4 – Reconstruction sur 7 octaves.

5.4 Signal musical

L'analyse sur 8 octaves, d'un signal musical échantillonné à $44.1khz$, avec 72 bins par octave et une fréquence initiale de $55hz$ donne le résultat suivant :

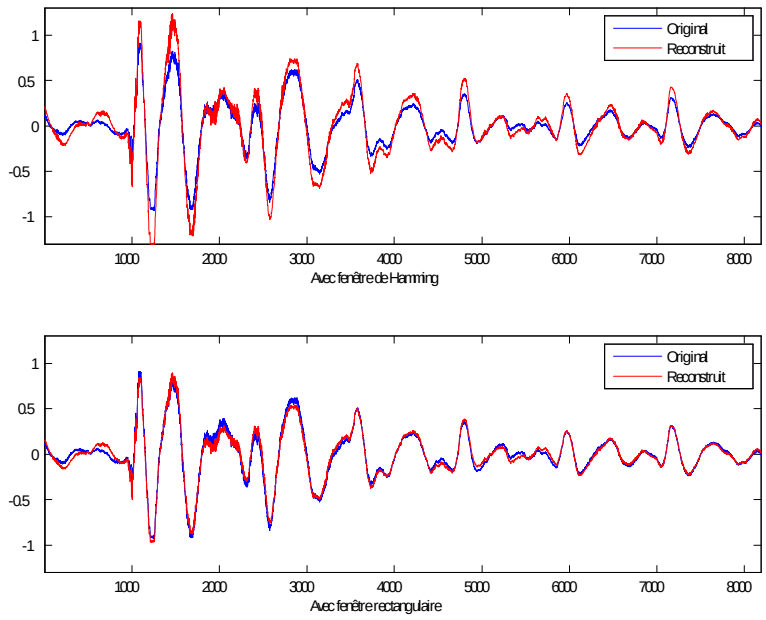


FIG. 5 – Reconstruction sur 8 octaves d'un signal musical.

6 Conclusion

On peut constater que les défauts de recouvrements des fenêtres rectangulaires ne sont pas rédhibitoires, dans la mesure où le signal reconstitué est suffisamment proche du signal original. Par contre dans le cas d'utilisation d'une fenêtre de Hamming, le gain du signal reconstruit est variable, il dépend en fait de la largeur de la fenêtre utilisée, largeur qui dépend de la fréquence analysée. Il faudrait donc tenir compte de ce paramètre pour pondérer en conséquence les différentes composantes servant à la reconstruction. Cette complication de

l'algorithme, qui n'est pas nécessaire dans le cas d'une fenêtre rectangulaire, ne semble pas a priori utile pour obtenir un signal reconstruit d'une qualité subjectivement meilleure.

Nous avons donné une interprétation simple de l'inversion de la CQT qui permet d'obtenir une reconstruction du signal initial par un polynôme trigonométrique qui ne prend en compte que les fréquences définies par l'analyse. Les résultats expérimentaux montrent que pour peu que les signaux analysés puissent être modélisés par des sommes de sinusoïdes dont les fréquences sont celles des notes de la gamme, les signaux reconstitués peuvent être aussi proche que possible des signaux originaux, au sens de la convergence des séries de Fourier.

Références

[1] J. C. Brown, "Calculation of a constant q spectral transform," *J. Acoust. Soc. Am.*, pp. 425–434, 1991.

[2] J. Prado, "Transformée à q constant," Telecom-ParisTech, Rapport interne 2010D004, 2010.

Dépôt légal : 2011 – 2ème trimestre
Imprimé à Télécom ParisTech – Paris
ISSN 0751-1345 ENST D (Paris) (France 1983-9999)

Télécom ParisTech

Institut TELECOM - membre de ParisTech

46, rue Barrault - 75634 Paris Cedex 13 - Tél. + 33 (0)1 45 81 77 77 - www.telecom-paristech.frfr

Département TSI