Augmentation de la résolution temporelle du banc de filtres du codage MPEG AAC à l'aide de transformées orthogonales

Ewen Camberlein Pierrick Philippe

France Télécom R&D
4, rue du Clos Courtel
35 512 Cesson Sévigné Cedex

{ewen.camberlein,pierrick.philippe}@orange-ft.com

Résumé

Le système de référence pour le codage des signaux de musique est aujourd'hui la norme MPEG AAC [1]. Un codage par transformée est utilisé et deux tailles de transformée sont préconisées (1024 ou 128) suivant la nécessité d'avoir une bonne résolution temporelle ou fréquentielle pour coder un signal particulier. Ce changement de taille requiert l'utilisation de fenêtres de transition, qui concentrent peu l'énergie du signal. Ces fenêtres de transition imposent de plus un délai et une complexité de codage supplémentaire. La technique présentée dans cet article se base sur la combinaison des coefficients transformés par l'utilisation de transformées orthogonales afin d'améliorer à la volée la résolution temporelle sans fenêtre de transition. Dans ce contexte, l'étude présente un critère d'évaluation d'une transformée orthogonale donnée, par l'études des valeurs de résolutions temporelle et fréquentielle du banc de filtres considéré.

Mots clefs

Codage audio numérique, MDCT, Matrices orthogonales, résolution temporelle, Heisenberg.

1 Introduction

Dans une application de codage des signaux de musique, les coefficients temporels du signal sont traités et transmis au décodeur par blocs de N échantillons. Dans l'encodeur, ils sont d'abord représentés comme une combinaison linéaire des fonctions de base de la transformée et subissent ensuite une quantification suivie d'un codage entropique avant d'être transmis.
Un problème typique du codage par transformée est que le bruit de quantification introduit par la quantification des coefficients est réparti au décodage sur l'ensemble du bloc des N échantillons. Ceci est particulièrement audible pour des signaux transitoires, pour lesquels des artefacts gênants comme les effets de "pré-écho" apparaissent.
Une des solutions préconisées [1] pour répondre à ce problème consiste à changer de taille de transformée en diminuant celle-ci afin d'améliorer la résolution temporelle lorsque des signaux transitoires sont détectés. Cela implique un délai et une complexité d'encodage supplémentaire car il faut utiliser des fenêtres de transitions asymétriques, comme présenté Figure 1. Ces fenêtres de transition sont nécessaires pour conserver la propriété de reconstruction parfaite de la transformée. De plus les fenêtres de transition utilisées ne sont pas idéales en termes de résolution temps fréquence et induisent une perte en efficacité de codage.

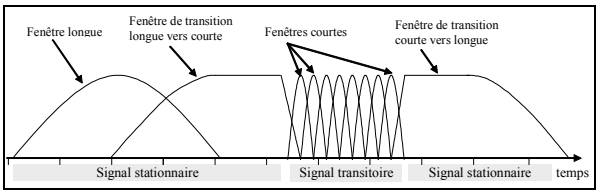

Figure 1 : *Exemple de succession de différentes fenêtres utilisées dans la norme MPEG AAC.*

Dans ce document nous présentons différentes méthodes pour effectuer la combinaison linéaire des fonctions de base de la transformée, afin d'obtenir des coefficients dans le domaine transformé ayant une meilleure localisation temporelle. Cette augmentation de la résolution temporelle sera ainsi obtenue sans recours à des fenêtres de transition.

L'opération effectuée consiste à combiner un certain nombre de coefficients successifs en sortie de la MDCT. Les coefficients transformés par une MDCT sont obtenus pour une fenêtre donnée $h[n]$ et un signal temporel $x[n]$ par :

$$X[k] = \sum_{n=0}^{2N-1} x[n] * h[n] * \cos\left[\frac{\pi}{N}\left(n + \frac{N+1}{2}\right)\left(k + \frac{1}{2}\right)\right] \qquad (1)$$

On transforme M coefficients X[k], exprimés sous forme d'un vecteur $X_k = {}^T\left[X[k],...,X[k+M-1]\right]$, à l'aide d'une

matrice carrée A_M de taille MxM et on obtient le vecteur $X'_k = {}^T\left[X'[k],...,X'[k+M-1]\right]$:

$$X'_k = A_M X_k \quad (2)$$

L'application de cette transformée A_M permet d'améliorer la résolution temporelle de la MDCT et de construire des transformées non uniformes adaptées au signal à coder à un instant donné. Différents scénarios d'utilisations de ces transformées blocs et les découpes temps fréquences en résultant sont présentés Figure 2 pour une MDCT de taille N=16.

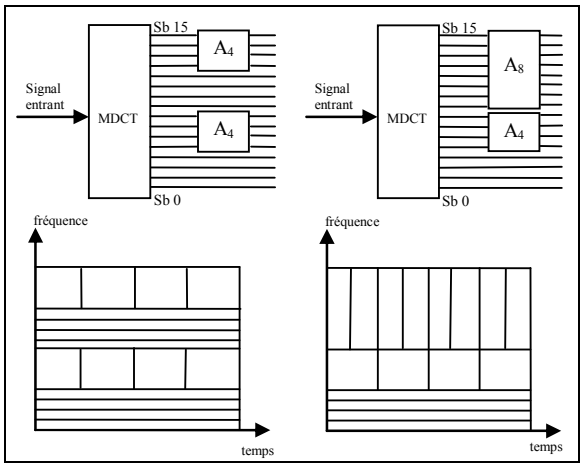

Figure 2 : *Exemples de découpes temps/ fréquence.*

Nous nous restreignons ici à l'étude des découpes uniformes, construites grâce à cette technique. Nous proposons une mesure de performance de ces transformées basé sur leur résolution temporelle et fréquentielle. Nous concluons sur les performances atteignables par ce type de structures en comparaison à celle utilisée par l'AAC.

2 Etat de l'art

Une méthode proposée par Mau [2] consiste à effectuer des opérations de somme et de différence appliquées sur deux coefficients adjacents. La transformée utilisée, équivalente à une transformée de Hadamard de taille 2 s'écrit alors :

$$A_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Nous pouvons voir Figure 3 et Figure 4 les modifications apportées aux réponses fréquentielles et impulsionnelles de la transformée initiale et après application de cette transformée de Hadamard appliquée sur les coefficients 17 et 18 d'une MDCT de taille 32.

En combinant deux coefficients, deux nouvelles sous bandes sont obtenues ayant la même localisation fréquentielles mais une localisation temporelle différente.

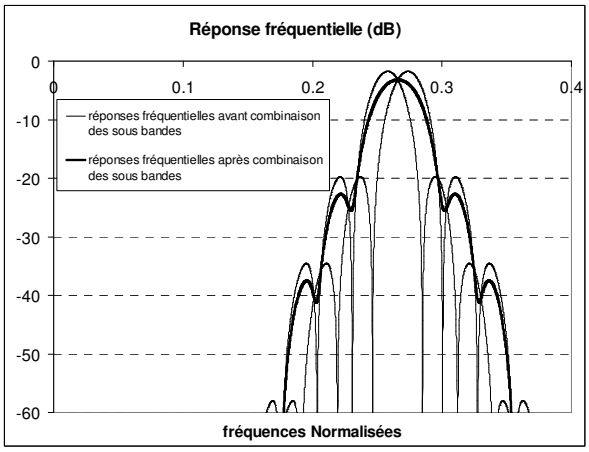

Figure 3 : *Réponses fréquentielles avant et après combinaison des coefficients.*

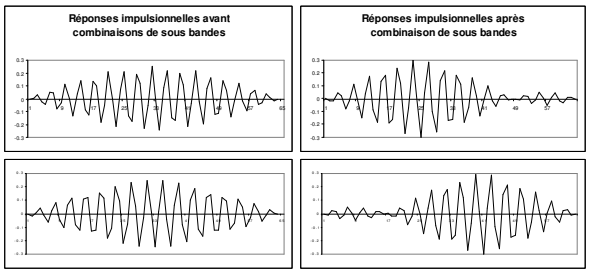

Figure 4 : *Réponses impulsionnelles avant et après combinaison des coefficients.*

Mau propose également l'utilisation de matrices de Hadamard de taille 4 appliquée sur 4 coefficients adjacents.

Malvar propose l'utilisation d'une autre transformée de taille 4 [3] afin d'augmenter la résolution temporelle en haute fréquences dans le cas d'un codage de signal de paroles.

Dans [4] Niamut propose l'utilisation de matrices de Hadamard de tailles quelconques. Il montre que l'utilisation de telles matrices assure que le module des réponses fréquentielles des fonctions de base combinées soit égal à la somme des modules des réponses fréquentielles des sous bandes avant combinaison.

Cette contrainte forte ne nous semble pas fondamentale pour une application de codage : nous étendons ici l'étude de ces transformées au cas des matrices orthogonales.

L'orthogonalité garantie la conservation de l'énergie des coefficients par la transformée. Cette propriété permet que la variance moyenne de l'erreur de quantification introduites sur les coefficients temporels soit égale à la variance moyenne de l'erreur introduite sur les coefficients transformés. Cela permet une évaluation simple de la qualité de codage lors de l'étape d'allocation de bits et donc assure une quantification simple des coefficients transformés [5].

3 Critères d'évaluation de l'intérêt d'une transformée donnée

La localisation temporelle et fréquentielle des coefficients du signal est importante en codage audio. Ceci permet de mettre en forme le bruit de quantification en tenant compte du masquage fréquentiel pour les signaux stationnaires et du masquage temporel pour les signaux transitoires.

Nous évaluons donc ici la transformée résultante de ces deux opérations (MDCT + transformées orthogonales) en termes de résolution temporelle σ_t^2 et fréquentielle σ_f^2. Ces résolutions sont obtenues à partir des réponses impulsionnelles $h_{sb}[n]$ et fréquentielles $H_{sb}[k]$. Les formules, dérivées de [6], utilisées pour calculer ces résolutions (pour une fréquence d'échantillonnage Fe) pour chaque sous bande sb, sont les suivantes :

$$\sigma_{t,sb}^2 = E_{t,sb}^{-1} \sum_{n=0}^{2N-1} \left(\frac{n}{Fe} - \mu_{t,sb} \right)^2 h_{sb}^2[n] \tag{3}$$

$$\mu_{t,sb} = E_{t,sb}^{-1} \sum_{n=0}^{2N-1} \frac{n}{Fe} h_{sb}^2[n] \tag{4}$$

$$E_{t,sb} = \sum_{n=0}^{2N-1} h_{sb}^2[n] \tag{5}$$

$$\sigma_{f,sb}^2 = \begin{cases} E_{f,sb}^{-1} \sum_{k=-FFTSize/2}^{FFTSize/2} \left(\frac{kFe}{FFTSize} - \mu_{f,sb} \right)^2 |H_{sb}[k]|^2 & si\ sb = 0 \\ E_{f,sb}^{-1} \sum_{k=-FFTSize/2}^{FFTSize/2} \left(\left| \frac{kFe}{FFTSize} \right| - \mu_{f,sb} \right)^2 |H_{sb}[k]|^2 & si\ 0 < sb < N-1 \\ E_{f,sb}^{-1} \sum_{k=0}^{FFTSize} \left(\frac{kFe}{FFTSize} - \mu_{f,sb} \right)^2 |H_{sb}[k]|^2 & si\ sb = N-1 \end{cases} \tag{6}$$

$$\mu_{f,sb} = \begin{cases} E_{f,sb}^{-1} \sum_{k=-FFTSize/2}^{FFTSize/2} \frac{kFe}{FFTSize} |H_{sb}[k]|^2 & si\ sb = 0 \\ E_{f,sb}^{-1} \sum_{k=-FFTSize/2}^{FFTSize/2} \left| \frac{kFe}{FFTSize} \right| |H_{sb}[k]|^2 & si\ 0 < sb < N-1 \\ E_{f,sb}^{-1} \sum_{k=0}^{FFTSize} \frac{kFe}{FFTSize} |H_{sb}[k]|^2 & si\ sb = N-1 \end{cases} \tag{7}$$

$$E_{f,sb} = \begin{cases} \sum_{k=-FFTSize/2}^{FFTSize/2} |H_{sb}[k]|^2 & si\ 0 \le sb < N-1 \\ \sum_{k=0}^{FFTSize} |H_{sb}[k]|^2 & si\ sb = N-1 \end{cases} \tag{8}$$

Pour mesurer l'intérêt d'une transformée bloc donnée nous calculons la moyenne des résolutions temporelles et fréquentielles :

$$\sigma_f^2 = \frac{1}{N} \sum_{sb=0}^{N-1} \sigma_{f,sb}^2 \tag{9}$$

$$\sigma_t^2 = \frac{1}{N} \sum_{sb=0}^{N-1} \sigma_{t,sb}^2 \tag{10}$$

Le principe d'incertitude d'Heisenberg nous permet d'évaluer si le compromis résolution temporel/ résolution fréquentielle est proche de la borne théorique :

$$\sigma_t^2 \sigma_f^2 \ge \frac{1}{(4\pi)^2} \tag{11}$$

4 Résultats

Les transformées blocs optimales A* présentées Figure 6 résultent de l'optimisation du paramètre de résolution temporelle moyenne σ_t^2 sur le sous ensemble des transformées blocs orthogonales constitué par l'ensemble des matrices de rotations {A}.

La matrice A* est donc recherchée tel que :

$$A^* = \arg\min_{\{A\}} \left(\sigma_t^2 \right) \tag{12}$$

En utilisant le fait qu'une matrice de rotation de dimension M*M peut être définie par seulement $M/2 (M-1)$ angles, cette optimisation devient possible en un temps raisonnable.

Nous avons testé les différents types de fenêtres utilisés par l'AAC (Figure 5). Il s'agit des fenêtres de Kaiser

Bessel Dérivées (KBD) et sinusoïdales de taille 1024 [1]. Les formes de ces fenêtres permettent de concentrer différemment l'énergie temporelle du signal, et ont une influence directe sur les résolutions temporelles obtenues.

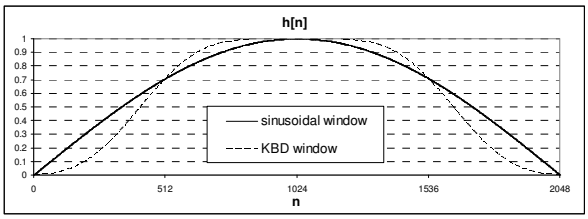

Figure 5 : *Différents types de fenêtres utilisés dans la norme MPEG AAC.*

Sur la Figure 6 sont représentées un certain nombre de transformées caractérisées par leurs résolutions temporelles et fréquentielles moyennes (pour une fréquence d'échantillonnage de 48 kHz).

A titre de références, nous avons tracé les caractéristiques de Heisenberg et celle obtenue par la MDCT seule. Sont également présentés les résultats obtenus après application des matrices orthogonales optimales, obtenues après optimisation sur les fenêtres définies par la norme AAC.

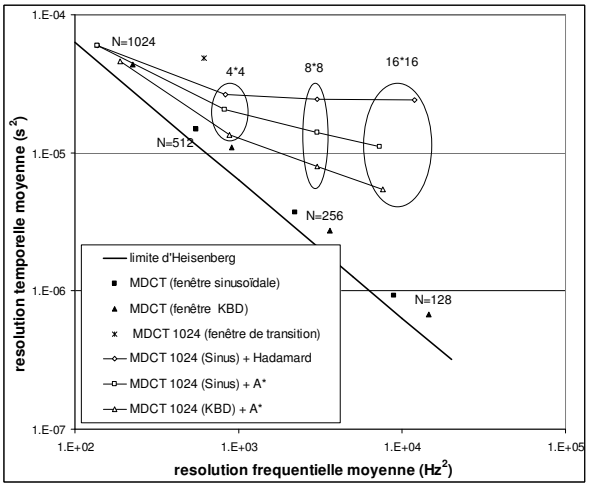

Figure 6 : *Comparaison des concentrations énergétiques obtenues par différentes transformées.*

L'utilisation de transformées orthogonales en sortie d'une MDCT utilisant une fenêtre de Kaiser Bessel dérivée (KBD) permet une meilleure efficacité par rapport à une configuration utilisant des fenêtres sinusoïdales.

Il apparaît clairement sur cette figure que l'utilisation de transformées de Hadamard ne permet pas d'obtenir les meilleurs résultats. Les matrices optimales obtenues par notre algorithme offrent de meilleures concentrations temps fréquence.

En revanche, on observe que l'application de transformées orthogonales ne permet pas de retrouver une concentration énergétique aussi bonne que celle de la MDCT : on s'éloigne de la limite d'Heisenberg à mesure que la taille de la transformée orthogonale croît.

Ces résultats montrent également qu'il semble difficile de retrouver une résolution temporelle équivalente à celle d'une MDCT de taille 128 à partir d'une MDCT de taille 1024 sur laquelle est appliquée une matrice orthogonale.

A titre d'illustration, nous présentons Figure 7 la concentration temporelle obtenue par les fonctions de bases avant et après transformée orthogonale, dans le cadre d'une transition. Ces résultats sont présentés dans le cadre des fenêtres sinusoïdales et de Kaiser Bessel dérivées. On observe la meilleure concentration d'énergie temporelle obtenue grâce aux fenêtres KBD.

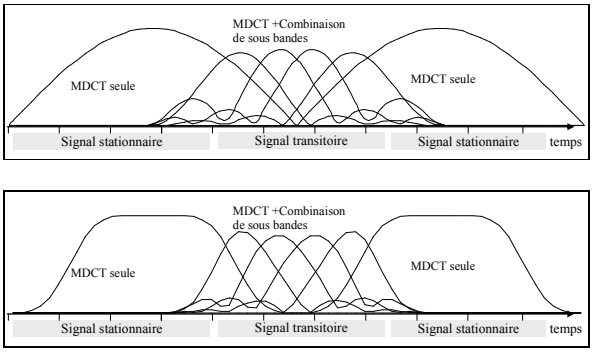

Figure 7 : Exemple de fenêtres temporelles obtenues par l'utilisation d'une transformée A_4^*.

Les matrices A_4^* utilisées Figure 7 sont les suivantes :

$$A^*_{4,Sinus} = \begin{bmatrix} 0.378621 & 0.597329 & 0.597190 & 0.378429 \\ -0.597205 & -0.378430 & 0.378656 & 0.597291 \\ 0.597254 & -0.378561 & -0.378489 & 0.597265 \\ -0.378524 & 0.597227 & -0.597269 & 0.378562 \end{bmatrix}$$

$$A^*_{4,KBD} = \begin{bmatrix} 0.386597 & 0.592050 & 0.592072 & 0.386614 \\ -0.592050 & -0.386643 & 0.386618 & 0.592040 \\ 0.592102 & -0.386662 & -0.386548 & 0.592020 \\ -0.386569 & 0.592010 & -0.592079 & 0.386692 \end{bmatrix}$$

5 Conclusion

Nous avons défini une méthode permettant de sélectionner une transformée orthogonale afin d'augmenter la résolution temporelle d'une MDCT de taille donnée. Cette technique permet l'utilisation de différents compromis résolution temporelle/ résolution fréquentielle qui n'étaient pas envisageables avec les transformées utilisées dans la norme MPEG AAC, et ceci sans utilisation de fenêtres de transition. Par contre obtenir une résolution temporelle équivalente à celle d'une MDCT 128 semble difficile.

La suite de cette étude consistera à définir un sous ensemble des transformées présentées ainsi qu'un critère permettant de décider comment construire un banc de filtre non uniforme s'adaptant à un signal donné. La quantité d'information à transmettre au décodeur pour la construction de la transformée inverse sera également prise en considération dans cette étude. L'insertion de cette technique au sein d'un schéma de codage nous permettra enfin d'évaluer les performances envisageables par l'utilisation de ces transformées.

Références

[1] MPEG-2 Advanced Audio Coding, Norme internationale *AAC.*, ISO/IEC 13818-7, Edition 2003.

[2] J. Mau. et al. Time-varying orthogonal filter banks without transient filters. *ICASSP,* vol. 2, pages 1328 – 1331. mai 1995

[3] H. S. Malvar. Enhancing the performance of subbands audio coders for speech signals. Dans *Proc. Int. Symp. Circuits and Systems'98* : 90-101, juin 1998.

[4] O.A. Niamut et R. Heusdens. Subband merging in cosine-modulated filter banks. *IEEE Signal processing letter*. vol. 10,num. 4, avril 2003.

[5] N.S. Jayant et Peter Noll, *Digital Coding of waveforms*, pages 517- 525, Prentice-Hall, 1984.

[6] C. Taswell. Empirical tests for the evaluation of multirate filter bank parameters. Rapport technique. Computational Toolsmiths. Stanford, février 1998.