



# INTRODUCTION D'UN BANC DE FILTRES HIERARCHIQUE DANS UN CODEUR PREDICTIF

Nicolas MOREAU † et Przemyslaw DYMARSKI ‡

† Telecom Paris, Dept Signal  
46 rue Barrault, 75634 Paris Cedex 13, FRANCE

‡ Université Technique de Varsovie  
ul. Nowowiejska 15/19, 00-665 Varsovie, POLOGNE

## RESUME

*On analyse l'intérêt et les limitations apportés par l'introduction d'un codage par transformée ou en sous-bandes, dans un codeur prédictif. On rappelle d'abord les relations existant entre codage par transformée et en sous-bandes. On donne une présentation succincte de la transformation orthogonale avec recouvrement proposée par Malvar. On montre ensuite qu'un banc de filtres hiérarchique semble un bon compromis pour analyser du signal de parole entre un banc uniforme et un banc réalisant une décomposition en octaves puisqu'il permet de s'adapter approximativement à la largeur des bandes critiques. On donne enfin des résultats de simulations montrant un comportement approximativement équivalent à celui du codeur CELP.*

## 1 Introduction

Pour compresser des signaux audio-fréquence, que cela soit de la parole ou de la musique, dans le but de les transmettre ou de les stocker, de nombreuses techniques performantes sont actuellement disponibles. Par exemple, pour compresser des signaux de parole dans la bande téléphonique, la plupart des codeurs à bas débit récemment normalisés ou en cours de développement (GSM, VSELP, LD-CELP ...) modélisent le signal par la sortie d'un filtre autorégressif excité par un vecteur recherché de façon exhaustive dans un dictionnaire pré-défini en minimisant une mesure de distorsion pondérée. Dans le domaine de l'enregistrement et de la diffusion (DAB, HDTV ...) des sons musicaux, on utilise plutôt des techniques de codage par transformée ou de codage en sous-bandes avec introduction de critères psychoacoustiques pour allouer les bits disponibles (par exemple la recommandation ISO/MPEG : Musicam). Des modèles d'audition élaborés sont alors utilisés [4]. Toutes ces techniques (liste non limitative) commencent à être bien connues et clairement exposées [1]. Les nouveaux codeurs deviennent de véritables systèmes de compression mélangeant ces différents ingrédients. On propose, dans cet article, l'introduction d'un codage par transformée ou d'un codage en sous-bandes, dans un codeur prédictif. Le but de cette étude est d'analyser la possibilité d'introduire dans un codeur à bas débit des techniques élaborées pour des codeurs à débit plus élevé.

## ABSTRACT

*In this paper, we analyse the interest and the limitations of the utilisation of transform and sub-band coding in a predictive coder. First, we point out the relationship between transform and sub-band coding. A concise presentation of the lapped orthogonal transform proposed by Malvar is presented. It is then shown that a hierarchical filter bank seems to be a good compromise between uniform and wavelet filter banks since its frequency partition approaches the critical band partition. Finally, simulation results are presented pointing out behaviour equivalent with that of a CELP coder.*

Dans une première partie, on rappelle les relations existant entre codage par transformée et codage en sous-bandes. On donne une présentation succincte de la transformation orthogonale avec recouvrement [2] qui permet d'atténuer des phénomènes parasites à la fréquence bloc dus au codage par transformée habituel. Dans une deuxième partie, on montre qu'un banc de filtres hiérarchique [2] semble un bon compromis pour du signal de parole entre un banc uniforme et un banc réalisant une décomposition en octaves puisqu'il permet de s'adapter approximativement à la largeur des bandes critiques. On montre l'intérêt et les limitations apportés par l'introduction de ce banc dans un codeur prédictif. Le problème essentiel est évidemment créé par le délai de reconstruction : il est réhibitoire pour les codeurs "full-duplex" et il rend moins efficace le prédicteur à long terme en boucle fermée. Cette étude étant exploratoire, on accepte cette contrainte et on cherche à analyser le bénéfice éventuel apporté par cette technique.

## 2 Bancs de filtres et transformées

Considérons un banc de  $M$  filtres (non représenté ici) où l'on note  $H_k(z)$  et  $F_k(z)$  les fonctions de transfert des filtres d'analyse et de synthèse et  $y_k(m)$  les signaux sous-échantillonnés par un facteur  $M$ . On suppose que tous les filtres du banc sont des filtres de réponses impulsionnelles finies de même longueur  $N$ . On a

$$y_k(m) = \sum_{l=0}^{N-1} h_k(l)x(mM - l)$$



### 2.1 Cas particulier : $N = M$

En appelant  $x_l(m) = x(mM - l)$  pour  $l = 0 \dots M - 1$  les composantes polyphasées de  $x(n)$ ,  $X(m) = [x_0(m) \dots x_{M-1}(m)]^t$  et  $Y(m) = [y_0(m) \dots y_{M-1}(m)]^t$ , on obtient la relation  $Y(m) = TX(m)$  avec

$$T = \begin{bmatrix} h_0(0) & \dots & h_0(M-1) \\ \dots & \dots & \dots \\ h_{M-1}(0) & \dots & h_{M-1}(M-1) \end{bmatrix}$$

Un banc de filtres d'analyse est donc équivalent à un opérateur qui fabrique un vecteur à partir d'échantillons comme le montre la Figure 1 suivi d'une transformation  $T$  caractérisée par une matrice carrée  $M * M$  dont les vecteurs lignes sont les réponses impulsionnelles des filtres d'analyse. A la synthèse, les opérations inverses de transformation caractérisées par la matrice  $P$ , sur-échantillonnage, décalage et addition transforment d'abord le vecteur  $Y(m)$  en  $\tilde{X}(m) = PY(m)$  puis le vecteur  $\tilde{X}(m)$  en une séquence  $\hat{x}(n)$ . La condition nécessaire et suffisante de reconstruction

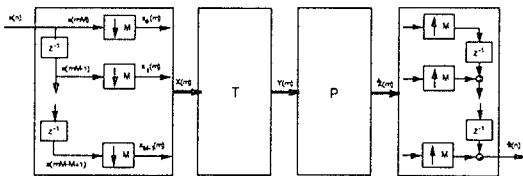


Figure 1: Equivalence entre banc de filtres et transformées lorsque  $N = M$

parfaite est donc simplement  $PT = I$ . Le signal reconstruit vérifie  $\hat{x}(n) = x(n - D)$  où  $D$  représente le délai de reconstruction avec  $D = M - 1$  si l'on suppose que les calculs entraînés par les transformations  $T$  et  $P$  sont instantanés. Pour caractériser les filtres de synthèse  $F_k(z)$  en fonction de  $P$ , il suffit d'imposer que toutes les composantes du vecteur  $Y(m)$  soient nulles excepté la  $k^{eme}$ . Le schéma habituel d'un banc de filtres montre que le signal en sortie est la réponse impulsionnelle  $f_k(n)$  du filtre  $F_k(z)$ . A partir de la Figure 1, on en déduit que  $\tilde{X}(m)$  est la  $k^{eme}$  vecteur colonne de  $P$  et comme

$$[\hat{x}(mM) \dots \hat{x}(mM + M - 1)] = [\hat{x}_{M-1}(m) \dots \hat{x}_0(m)]$$

la  $k^{eme}$  colonne renversée de la matrice  $P$  forme la réponse impulsionnelle du filtre de synthèse  $F_k(z)$ . Le signal reconstruit  $\hat{X}(m) = PY(m)$  apparaît comme une combinaison linéaire des vecteurs colonnes de  $P$  dont les coefficients  $Y(m) = TX(m)$  sont les composantes de  $X(m)$  sur les vecteurs lignes de  $T$ . Lorsque  $T = P^t$  où  $t$  dénote la transposition si  $P$  est à valeurs réelles ou la transposition hermitienne si  $P$  est à valeurs complexes, on retrouve une simple transformation orthogonale (ou unitaire). La condition précédente impose la relation  $h_k(n) = f_k(M - 1 - n)$  entre les réponses impulsionnelles des filtres d'analyse et de synthèse. Lorsque  $h_k(n) = h(n)e^{j2\pi \frac{k}{M}n}$  sont les réponses impulsionnelles d'un banc de filtres modulés, on retrouve la transformée de Fourier discrète à court terme où  $h(n)$  caractérise la fenêtre de pondération. Cette transformation est à valeurs complexes. Le banc de filtres peut être modulé par une transformation en cosinus discrète.

### 2.2 Transformation orthogonale avec recouvrement : $N \geq M$

La matrice  $T$  est maintenant une matrice rectangulaire  $M * N$  contenant, comme précédemment, des vecteurs lignes correspondant aux réponses impulsionnelles des filtres d'analyse. Sans perte de généralité, on suppose que  $N$  est un multiple de  $M$  soit  $N = LM$ . On note  $T^0 \dots T^{L-1}$  les  $L$  matrices carrées  $M * M$  vérifiant  $[T^0 \dots T^{L-1}] = T$ . La matrice de synthèse  $P$  est une matrice rectangulaire  $N * M$  contenant des vecteurs colonnes correspondant aux réponses impulsionnelles renversées des filtres de synthèse. A l'analyse, le vecteur transformé  $Y(m)$  est obtenu à partir des  $L$  vecteurs  $X(m) \dots X(m - L + 1)$  par la relation

$$Y(m) = [T^0 \dots T^{L-1}] \begin{bmatrix} X(m) \\ \vdots \\ X(m - L + 1) \end{bmatrix}$$

qui s'interprète comme une opération de filtrage causal vectoriel. Cette opération est schématisée Figure 2. A la syn-

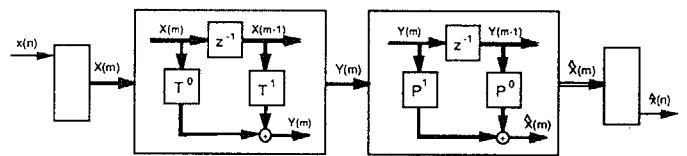


Figure 2: Transformation orthogonale avec recouvrement

thèse, la contribution de  $Y(m)$  dans le vecteur reconstruit est de la forme

$$PY(m) = \begin{bmatrix} P^0 \\ \vdots \\ P^{L-1} \end{bmatrix} Y(m)$$

Distinguons le vecteur reconstruit de façon non causale, c'est à dire à partir des vecteurs transformés futurs

$$\tilde{X}(m) = P^0 Y(m) + \dots + P^{L-1} Y(m + L - 1)$$

et le vecteur reconstruit de façon causale  $\hat{X}(m) = \tilde{X}(m - L + 1)$ . La condition de reconstruction parfaite s'obtient par simple substitution

$$\hat{X}(m) = [P^0 \dots P^{L-1}] A \begin{bmatrix} X(m + L - 1) \\ \vdots \\ X(m) \\ \vdots \\ X(m - L + 1) \end{bmatrix} = X(m)$$

avec

$$A = \begin{bmatrix} 0 & \dots & 0 & T^0 & \dots & \dots & T^{L-1} \\ & & & \dots & & & \\ T^0 & \dots & \dots & T^{L-1} & 0 & \dots & 0 \end{bmatrix}$$

La condition est donc  $[P^0 \dots P^{L-1}] A = [0 \dots 0 I 0 \dots 0]$ . Si l'on impose la contrainte  $h_k(n) = f_k(N - 1 - n)$ , on obtient la relation  $T = P^t$ . L'orthogonalité des vecteurs lignes de  $T$ , l'orthogonalité des vecteurs colonnes de  $P$  et l'orthogonalité des vecteurs lignes de  $T$  et des vecteurs colonnes de  $P$



décclés de  $M \dots (L - 1)M$  échantillons entraînent une reconstruction parfaite. Le délai de reconstruction est de  $L - 1$  blocs plus le délai nécessaire pour créer un bloc soit  $D = M - 1 + (L - 1)M$ . On obtient des conditions de reconstruction parfaite mais on désire également des propriétés de sélectivité en fréquence. Malvar propose une généralisation des bancs de filtres modulés en définissant deux types de filtres suivant que  $N = 2M$  dans le cas de la "Modulated Lapped Transform" ou que  $N = 4M$  dans le cas de "l'Extended Lapped Transform" [2]. Les réponses impulsionnelles des filtres d'analyse sont données par

$$h_k(n) = h(n) \sqrt{\frac{2}{M}} \cos\left[\left(n + \frac{M+1}{2}\right)\left(k + \frac{1}{2}\right)\frac{\pi}{M}\right]$$

avec

$$h(n) = -\sin\left[\left(n + \frac{1}{2}\right)\frac{\pi}{2M}\right] \text{ si } N = 2M$$

$$h(n) = -\frac{1}{2\sqrt{2}} + \frac{1}{2}\cos\left[\left(n + \frac{1}{2}\right)\frac{\pi}{2M}\right] \text{ si } N = 4M$$

### 3 Banc de filtres hiérarchique

#### 3.1 Justification

Le modèle d'audition habituellement utilisé [4] consiste à admettre que l'oreille se comporte comme un banc de filtres réalisant une partition non-régulière de l'axe des fréquences, la largeur de chaque partition augmentant avec la fréquence. Les 18 premières bandes critiques correspondant à la bande téléphonique sont indiquées sur le premier tracé de la Figure 3. Les bancs de filtres modulés précédents divisent l'axe

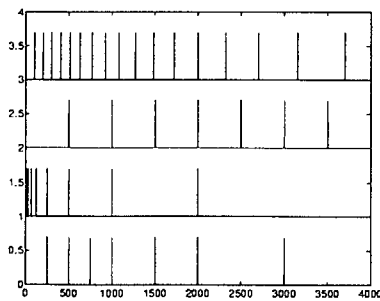


Figure 3: Bandes critiques, Banc uniforme, Banc itéré en octaves, Banc hiérarchique

des fréquences en  $M$  bandes de largeur approximativement égales. Le deuxième tracé de la Figure 3 correspondant à un banc avec  $M = 8$ , montre que cette solution n'est pas satisfaisante. Une solution élégante pourrait être fournie par l'utilisation de transformées en ondelettes mais la découpe en octaves de l'axe des fréquences imposée par cette méthode n'est pas non plus satisfaisante comme le montre le troisième tracé de la Figure 3. Malvar définit un banc hiérarchique [2] consistant à filtrer d'abord un signal par un premier banc de filtres puis à définir de nouveaux bancs adaptés dans chacune des sous-bandes. Cette technique semble être un meilleur candidat. On adopte le compromis suivant. La bande  $[0, 4 \text{ kHz}]$  est d'abord divisée en 4 sous-bandes. Ensuite, la bande  $[0, 1 \text{ kHz}]$  est encore divisée en 4 sous-bandes par un banc de filtres identique. Enfin, la bande

$[1, 2 \text{ kHz}]$  est divisée en 2 sous-bandes. Ainsi, chaque sous-bande correspond approximativement à 2 bandes critiques comme on peut l'observer sur le quatrième tracé de la Figure 3. Les 4 premières sous-bandes sont sous-échantillonnées par un facteur 16, les 2 sous-bandes suivantes par un facteur 8 et les 2 dernières sous-bandes par un facteur 4. La répartition temps-fréquence des 16 échantillons filtrés est donnée Figure 4. La Figure 5 donne le schéma fonctionnel complet du banc hiérarchique d'analyse et de synthèse.

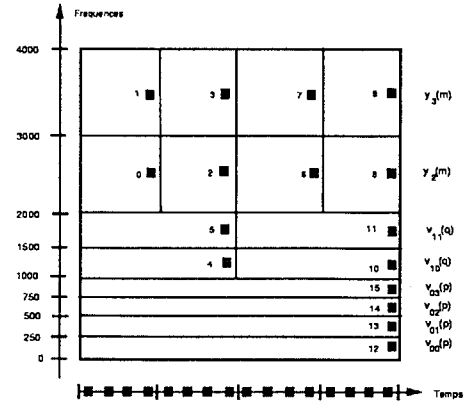


Figure 4: Répartition temps-fréquence du banc hiérarchique

#### 3.2 Introduction du banc dans un codeur prédictif

Considérons le schéma d'un codeur CELP standard avec un prédicteur à long terme en boucle fermée montré Figure 6. Le fait de disposer d'un modèle de production pour

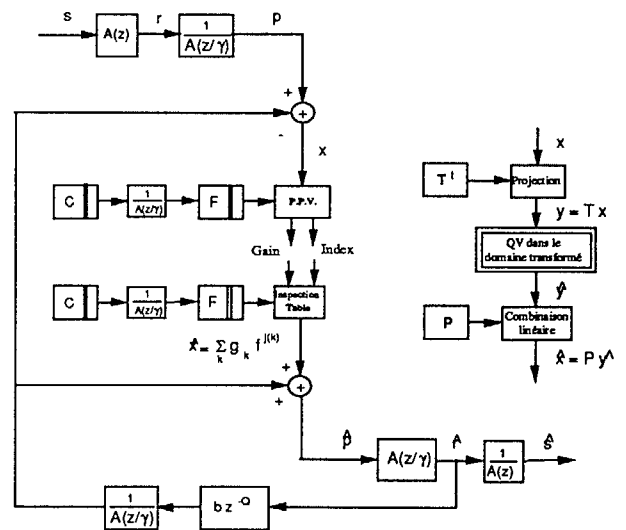


Figure 6: Forme générale d'un codeur CELP ou d'un codeur par transformée avec un prédicteur à long terme

le signal de parole sous la forme excitation-filtre permet de modéliser le vecteur  $x$ , que l'on supposera de dimension  $M$ , par une combinaison linéaire de  $K$  vecteurs  $f^j$  appartenant au dictionnaire filtré  $F$ , le dictionnaire d'excitation  $C$  étant pré-déterminé, par exemple par apprentissage. Généralement, on choisit  $K = 1$  ou  $2$  vecteurs dans ce dictionnaire. La modélisation du vecteur  $x$  peut être réalisée aussi par

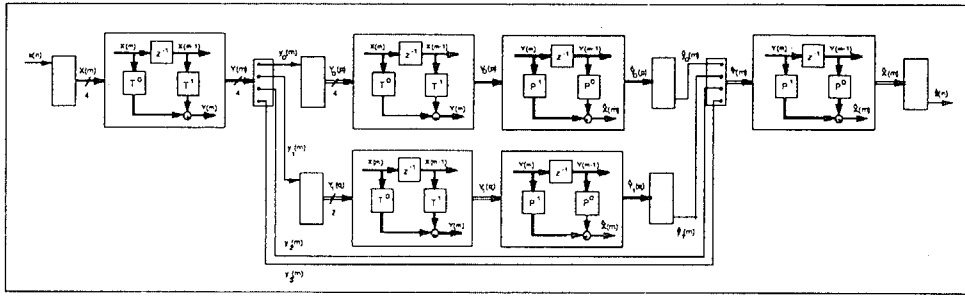


Figure 5: Schéma fonctionnel du banc hiérarchique d'analyse et de synthèse

utilisation d'un banc de filtres. On projette alors le vecteur  $x$  sur les vecteurs lignes de la matrice  $T$ . Le vecteur reconstruit  $\hat{x}$  est une combinaison linéaire des  $M$  vecteurs colonnes de  $P$ . Ces deux approches deviennent équivalentes si l'on sélectionne un grand nombre de vecteurs ( $K \rightarrow M$ ) dans le codeur CELP et si l'on réalise une orthogonalisation progressive du dictionnaire filtré  $F$  comme expliqué dans [3]. Ce que l'on appelait les gains relatifs aux vecteurs orthogonalisés devient les composantes du vecteur  $x$  relativement à la base définie par les vecteurs de la matrice  $T$ . On a affaire alors à une transformation très particulière puisque, dans ce cas, la matrice  $T$  est construite à partir de la matrice  $F$  qui dépend elle-même du signal de parole dans l'intervalle d'analyse. Lorsque l'on utilise un banc de filtres, la matrice  $T$  est pré-déterminée.

Le problème essentiel concerne le délai de reconstruction. Il est d'abord dû, dans tous les cas, au fait de regrouper des échantillons pour former un vecteur soit  $M - 1$  échantillons. Il est dû aussi au recouvrement des vecteurs dans le cas du banc hiérarchique. Comme pour reconstruire un vecteur  $\hat{X}(m)$  de dimension  $M = 4$ , il faut connaître le vecteur  $Y(m + L - 1)$  et que, pour reconstruire la première composante de ce vecteur, c'est à dire le vecteur  $\hat{Y}_0(p)$  en adoptant les notations de la Figure 5, il faut connaître  $V_0(p + L - 1)$ , on constate que le délai supplémentaire dû au banc hiérarchique est, dans le cas le plus défavorable, de  $L * 16$  échantillons, c'est à dire de 32 ou 64 échantillons suivant que l'on emploie la "Modulated Lapped Transform" ou l'"Extended Lapped Transform". Le prédicteur à long terme en boucle fermée pose également un problème. On sait qu'il est très efficace mais qu'il impose que la plus petite valeur de  $Q$  permise soit égale à la dimension du vecteur  $x$ . Lorsque l'on utilise un banc de filtres hiérarchique, il faut rajouter le délai de reconstruction.

### 3.3 Résultats de simulations et conclusion

On compare les deux solutions schématisées Figure 6. Le compromis habituel pour un codeur CELP standard est de réaliser une prédiction à court terme tous les 160 échantillons et une prédiction à long terme tous les 40 échantillons. Pour que tous les paramètres du codeur soient identiques dans les deux cas excepté le type de modélisation du résiduel et comme le banc sélectionné impose des vecteurs de dimension 16, on choisit les paramètres suivants. On définit des sous-sous-fenêtres de 16 échantillons où l'on réalise une quantification vectorielle du signal résiduel  $x$  dans le cas du codeur CELP ou du signal résiduel transformé  $y$  dans le cas du banc hiérarchique. Dans les deux cas, on effectue une

quantification vectorielle de type "gain-forme" en sélectionnant un vecteur parmi 128 et un gain (quantification du gain non réalisée). Toutes les sous-fenêtres de  $3 * 16 = 48$  échantillons, on calcule un prédicteur à long terme. Ce calcul est habituellement réalisé en boucle fermée. La contrainte sur  $Q$  rend irréaliste le cas  $L = 4$ . On se limite donc à l'emploi de la "Modulated Lapped Transform". Même dans ce cas, la contrainte  $Q_{min} = 48 + 32 = 80$ , reste trop importante, ce qui nous oblige à calculer les paramètres du prédicteur à long terme en boucle ouverte. La contrainte devient alors  $Q_{min} = 16 + 32 = 48$ , valeur plus raisonnable. Toutes les fenêtres de  $4 * 48 = 192$  échantillons, on actualise les coefficients du filtre de synthèse. Le dictionnaire de quantification vectorielle codant le vecteur transformé  $y$  a été déterminé par apprentissage. Les résultats de simulation montre une différence peu significative (moins de 0.5 dB) entre les deux cas examinés. On a cherché alors à introduire des quantificateurs scalaires prédictifs "backward" dans les 4 premières sous-bandes du banc hiérarchique. Des estimations des fonctions d'auto-corrélation ont montré la décorrélation (coefficients inférieurs à 0.5) des signaux dans les différentes sous-bandes et donc le peu d'efficacité de ces prédicteurs.

Il est normal que ces premières simulations fournissent des résultats comparables dans les deux cas examinés, la quantification vectorielle s'adaptant à la distribution de la source. L'introduction d'un banc hiérarchique dans un codeur prédictif présente l'intérêt de réaliser un pré-traitement réalisant une mise en forme du signal adaptée à l'utilisation de critères psychoacoustiques. Ceci fera l'objet d'une étude ultérieure.

## 4 Références

1. A. Gersho, R. Gray "Vector Quantization and Signal Compression" Kluwer Academic Publishers 1992
2. H. Malvar "Signal Processing with Lapped transforms" Artech House 1992
3. N. Moreau, P. Dymarski "Successive Orthogonalizations in the Multistage CELP coder" Proc. Int. Conf. Acoust., Speech, Signal Processing, 1992
4. E. Zwicker, E. Feldtkeller "Psychoacoustique, L'oreille récepteur d'information" Collection CNET/ENST, Masson 1981