



Thèse

Présentée pour obtenir le grade de docteur de l'Ecole Nationale Supérieure des Télécommunications

Spécialité : Signal et Image

Abdellatif BENJELLOUN TOUIMI

**Traitement du signal audio dans le domaine codé :
techniques et applications**

Soutenue le 21 mai 2001 devant le jury composé de

Francis Castanié	Président
Pierre Duhamel	Rapporteurs
Gang Feng	
Yannick Mahieux	Examineurs
Nicolas Moreau	
Christian Wellekens	Directeur de thèse

Ecole Nationale Supérieure des Télécommunications

A ma mère

Remerciements

Je tiens à remercier en premier lieu M. Yannick Mahieux, mon responsable au sein de France Télécom R&D, pour ses conseils et son suivi constant pendant ces trois années de thèse.

J'exprime ma profonde reconnaissance à M. Christian Wellekens de bien vouloir accepter de diriger les travaux de cette thèse et aussi pour ses encouragements et ses conseils spécialement lors de la rédaction de ce mémoire.

Je remercie M. Pierre Duhamel et M. Gang Feng pour l'intérêt qu'ils ont porté à mon travail et pour avoir accepté d'être les rapporteurs de cette thèse.

M. Francis Castanié m'a fait un grand honneur en ayant bien voulu présider le jury de ma thèse. Je tiens à remercier également M. Nicolas Moreau pour sa participation à la commission d'examen et aussi pour son acceptation de faire une première lecture critique de ce mémoire.

Certains travaux de cette thèse sont le fruit de coopération avec d'autres personnes. Je tiens à remercier en particulier Mme Claude Lamblin pour sa contribution concernant la quantification vectorielle algébrique sphérique et aussi pour sa lecture critique et ses corrections de ce mémoire. Je remercie aussi M. Cédric Boisseau avec qui j'ai développé la technique de correction de pertes paquets audio pendant son stage à FTR&D.

Je remercie M. Rolland Pattard et M. Roger Lochou pour leur support informatique et technique et Mme Janine Denmat qui nous rend la vie facile sans se soucier des problèmes administratifs.

Enfin, toute personne qui a contribué de près ou de loin à ce que le travail de cette thèse soit finalisé en soit remerciée.

Résumé

La manipulation classique de flux audio codés nécessite une opération préalable de décodage pour extraire les signaux temporels puis de recodage après traitement. Cette approche directe présente des inconvénients en terme de complexité et délai algorithmiques. Pour pallier ces problèmes, l'idée du traitement dans le domaine codé a été proposée. La mise en œuvre d'une telle démarche n'est pas immédiate, plusieurs problèmes surgissent selon le traitement et le codeur considérés. Dans cette thèse on s'intéresse à des codeurs de type fréquentiel perceptuel tels MPEG-1 et TDAC de FTR&D et précisément pour des traitements de filtrage et de mixage. Le contexte applicatif illustré concerne le traitement du son pour la téléconférence multipoint.

Le filtrage dans le domaine des sous-bandes est le premier problème abordé dans cette thèse. Une méthode générique a été développée, elle permet de transposer tout filtrage rationnel (FIR ou IIR) temporel dans ce nouveau domaine et pour tout banc de filtres à décimation maximale assurant la reconstruction parfaite. Cette méthode a été appliquée pour effectuer la spatialisation sonore par des filtres HRTF dans le domaine des sous-bandes. L'étude de la sommation sur les flux codés soulève plusieurs contraintes suivant le codeur considéré. Pour le codeur MPEG-1 Layer I et II, la problématique principale concerne la détermination des paramètres psychoacoustiques nécessaires à l'allocation de bits. L'algorithme proposé apporte donc une solution à ce problème par réestimation des seuils de masquage des signaux individuels à sommer puis recombinaison. Une méthode de réduction de débit est aussi dérivée à partir de cet algorithme. Pour le codeur TDAC de FTR&D la diminution de la complexité dans le procédé de sommation repose sur le phénomène de masquage inter-signaux en tenant compte de la structure particulière de ce codeur. Elle profite de l'imbrication des dictionnaires de la quantification vectorielle qu'il utilise.

La mise en valeur de l'intérêt du traitement dans le domaine codé a été concrétisée par la mise en œuvre dans un pont audio de téléconférence multipoints. Ce pont assure les fonctionnalités de concentration de flux, de la correction de trames effacées due au phénomène de perte de paquets sur les réseaux à qualité de service non garantie et aussi de la gestion des flux discontinus.

Mots clés : Traitement dans le domaine codé, codage audio fréquentiel perceptuel, codeur MPEG-1, codeur TDAC de FTR&D, mixage, conversion du débit, filtrage en sous-bandes, spatialisation sonore, HRTF, correction de pertes paquets, téléconférence multipoint, pont audio (MCU).

Abstract

The straightforward method to manipulate compressed audio data consists on decoding, processing and re-encoding. This method generates great algorithmic delay and complexity. In order to reduce this drawback, processing in compressed domain was proposed. However this approach is not so easy to use, some problems appears depending on the considered processing and coder. In this thesis, we are interested by perceptual frequency coders like MPEG-1 and FTR&D TDAC and processing such as filtering and mixing. The main application example considered here is audio processing in multipoint teleconferencing context.

The first problem examined in this thesis deals with filtering in subband-domain. A generic framework making possible the transposition of any temporal rational filter (FIR or IIR) to subband-domain, for any critically sampled-perfect reconstruction filter bank, was developed. This method was applied to make sound spatialisation using HRTF filters in subband-domain. The second problem considered is the summation of encoded audio signals. Lot of constraints appears depending on the considered coder. The main problem for MPEG-1 Layer I-II consists on determining psychoacoustics parameters which are required to the bit allocation. To resolve this problem the proposed algorithm makes estimation of masking thresholds of individual signal and then combines them. A new method of bit rate reduction was also derived from this algorithm. Decreasing the complexity in the summation procedure for FTR&D TDAC coder is based on the masking phenomena between different signals and using its particular structure. It takes advantage of the embedded codebooks property used by the vector quantification in this case.

Application of compressed domain processing is illustrated by the implementation in an audio bridge for multipoint teleconferencing. This audio bridge has the functionalities of mixing, recovering of erased frames due to packets loss phenomena on network with non-guaranteed QoS and also managing transmission discontinues.

Key words: Compressed domain processing, audio coding, MPEG-1 coder, FTR&D TDAC coder, subband-domain filtering, sound spatialisation, HRTF, audio packets recovering, multipoint teleconferencing, audio bridge (MCU).

Table des matières

CHAPITRE 1	PRESENTATION DU CONTEXTE ET OBJECTIFS	1
1.1.	Introduction	1
1.2.	Systèmes de codage des signaux audio et parole	1
1.2.1	Généralités	1
1.2.2	Caractéristiques	3
1.3.	Traitement du signal dans le domaine codé	4
1.3.1	Définition et généralités	4
1.3.2	Traitement dans le domaine codé des signaux images et vidéo	5
1.3.3	Traitement dans le domaine codé des signaux audio	6
1.3.4	Approche inverse : codage pour manipulation simple dans le domaine codé	7
1.4.	Traitement du signal audio et transmission sur les réseaux de communication multimédia	8
1.4.1	Réseaux multimédia : caractéristiques et contraintes	9
1.4.2	Exemples de traitements du signal audio pour les réseaux multimédia	10
1.5.	Thèmes d'étude et organisation de la thèse	11
CHAPITRE 2	CODAGE AUDIO FREQUENTIEL PERCEPTUEL	15
2.1.	Introduction	15
2.2.	Bases du codage audio fréquentiel perceptuel	15
2.3.	Banc de filtres et transformée	16
2.3.1	Bancs de filtres à décimation maximale	17
2.3.2	Banc de filtres à reconstruction parfaite ; Banc de filtres paraunitaire	19
2.3.3	Bancs de filtres FIR à cosinus modulé	20
2.3.4	Transformée MDCT	22
2.4.	Eléments sur la psychoacoustique	25
2.4.1	Phénomène de masquage	25
2.4.2	Bandes critiques	26
2.5.	Allocation des bits	26
2.6.	Quantification	28
2.6.1	Quantification vectorielle	28
2.6.2	Quantification vectorielle sphérique	30
2.6.3	Quantification vectorielle par codes à permutation	30

2.6.4 Algorithmes de codage et de décodage pour la QVS à codes à permutation	32
2.7. Codeur MPEG-1 Layer I et II	33
2.7.1 Banc de filtres	34
2.7.2 Modèles psychoacoustiques	36
2.7.3 Allocation de bits	38
2.7.4 Quantification et facteurs d'échelle	38
2.8. Codeur TDAC de FTR&D	39
2.8.1 Description générale	39
2.8.2 Détection du voisement et de la tonalité	42
2.8.3 Courbe de masquage	42
2.8.4 Enveloppe spectrale	43
2.8.5 Courbe de masquage par bandes et allocation de bits	44
2.8.6 Quantification des coefficients MDCT	45
2.8.7 Correction des trames effacées	46

CHAPITRE 3 FILTRAGE DANS LE DOMAINE FREQUENTIEL **47**

3.1. Introduction	47
3.2. Phénomène du repliement du spectre (aliasing)	48
3.3. Filtrage dans le domaine fréquentiel : aperçu de l'état de l'art	51
3.3.1 Filtrage dans le domaine transformé	51
3.3.2 Filtrage en sous-bandes du banc de filtres MPEG-1	53
3.4. Solution générale au problème du filtrage en sous-bandes	55
3.4.1 Définition du problème	55
3.4.2 Représentation bloc du filtre scalaire temporel	57
3.4.3 Formulation de la matrice du filtrage en sous-bandes	59
3.4.4 Autre formulation de la matrice du filtrage en sous-bandes	61
3.4.5 Interprétation et forme réduite de la matrice de filtrage	62
3.5. Filtrage IIR en sous-bande : génération de la matrice de filtrage et structure de l'implantation	63
3.6. Application au banc de filtres MPEG-1	64
3.6.1 Filtre retard pur	65
3.6.2 Filtre à réponse impulsionnelle finie	68
3.6.3 Filtre à réponse impulsionnelle infinie	71
3.7. Application au banc de filtres TDAC	72
3.7.1 Filtre à réponse impulsionnelle finie	73
3.7.2 Filtre à réponse impulsionnelle infinie	75
3.8. Application : spatialisation sonore dans le domaine des sous-bandes	76
3.8.1 Spatialisation sonore	76
3.8.2 Implantation dans le codeur TDAC de FTR&D	78
3.9. Conclusion	78

CHAPITRE 4 TRAITEMENT DANS LE DOMAINE COMPRESSE MPEG-1 79

4.1. Introduction	79
4.2. Etat de l'art	80
4.3. Estimation des seuils de masquage	82
4.3.1 Principe	82
4.3.2 Détermination du niveau de pression acoustique et du rapport signal à bruit	84
4.3.3 Calcul de la constante d'ancrage	85
4.3.4 Résultat	86
4.4. Algorithme de sommation	87
4.4.1 Aperçu général	87
4.4.2 Additivité des seuils de masquage	88
4.4.3 Estimation des seuils de masquage et des rapports signal à masque du signal somme	90
4.4.4 Résultats	91
4.5. Variantes de l'algorithme de sommation	92
4.5.1 Sommation à la volée	92
4.5.2 Mixage avec pondération de signaux	93
4.6. Conversion du débit	93
4.7. Evaluation de l'algorithme de sommation	95
4.7.1 Test subjectif	96
4.7.2 Mesure objective	96
4.7.3 Sensibilité de l'algorithme à la variation des modèles psychoacoustiques	97
4.8. Conclusion	98

CHAPITRE 5 TRAITEMENT DANS LE DOMAINE COMPRESSE DU CODEUR TDAC DE FTR&D 101

5.1. Introduction	101
5.2. Sommation directe dans le domaine fréquentiel	103
5.2.1 Description générale	103
5.2.2 Détermination du voisement	103
5.2.3 Détermination de la tonalité	106
5.2.4 Enveloppe spectrale et seuils de masquage	106
5.2.5 Masquage des coefficients	107
5.3. Conversion du débit sur un signal	108
5.3.1 Principe et description générale	108
5.3.2 Algorithme de trans-quantification sur une bande	109
5.3.3 A propos de la QVAS et de l'allocation de bits du codeur TDAC de FTR&D	112
5.4. Méthode de sommation avancée dans le flux audio compressé TDAC	113
5.4.1 Principe et étude préliminaire	113
5.4.2 Utilisation d'un critère de masquage par coefficient	115
5.4.3 Utilisation d'un critère de masquage basé sur l'énergie	116

5.5. Evaluation des performances des différentes méthodes de recombinaison proposées	117
5.5.1 Gain en complexité et en retard	117
5.5.2 Tests subjectifs	117
5.6. Conclusion	119

CHAPITRE 6 ARCHITECTURES ET APPLICATIONS POUR LE TRAITEMENT AUDIO DANS LE DOMAINE COMPRESSE 121

6.1. Introduction	121
6.2. Communication multimédia de groupe et téléconférence	121
6.2.1 Normes et environnement	121
6.2.2 Conférence multipoint	123
6.3. Algorithme de correction des trames TDAC effacées dans le domaine fréquentiel	124
6.3.1 Description générale	124
6.3.2 Méthode d'extrapolation sinusoïdale	125
6.4. Pont audio à traitement dans le domaine compressé	127
6.4.1 Fonctionnalités et avantages	127
6.4.2 Gestion des flux discontinus	129
6.4.3 Correction des pertes de paquets dans le pont	130
6.5. Autres exemples d'application	132
6.5.1 Terminal de téléconférence multipoint décentralisée	132
6.5.2 Conversion du débit dans les serveurs audio et les passerelles	134
6.5.3 Contexte du système MPEG-4	134
6.6. Conclusion	135

CHAPITRE 7 CONCLUSION GENERALE 137

7.1. Bilan	137
7.2. Perspectives et travaux futurs	138

ANNEXE A DECOMPOSITION POLYPHASEE DES FILTRES RECURSIFS 141

A.1. Principe	141
A.2. Forme générale des composantes polyphasées d'un filtre IIR	142
A.3. Composantes polyphasées de cellules du premier et second ordre	143

ANNEXE B ALGORITHME DE SCHALKWIJK 145

B.1. Détermination du rang d'un vecteur dans sa classe	145
B.2. Algorithme de Schalkwijk inverse	146

ANNEXE C	BASE DE DONNEES PAROLE UIT	147
BIBLIOGRAPHIE		149
BIBLIOGRAPHIE PERSONNELLE		159

Liste des figures

Fig. 1.1 : Schéma générique d'une opération de codage/décodage du signal audio.....	2
Fig. 1.2 : Schémas génériques du (a) traitement sur les données linéaires (b) traitement dans le domaine compressé.....	4
Fig. 2.1 : Schéma de principe d'un système de codage audio fréquentiel perceptuel.	15
Fig. 2.2 : Banc de filtres à décimation maximale à M bandes.	18
Fig. 2.3 : Représentation en matrices polyphasées (a) Banc de filtres d'analyse (b) Banc de filtres de synthèse.	19
Fig. 2.4 : Représentation polyphasée d'un banc de filtre à cosinus modulé (a) Banc d'analyse (b) Banc de synthèse.....	22
Fig. 2.5 : Transformée MDCT (a) Analyse : transformée directe à recouvrement (b) Synthèse : transformée inverse et addition avec recouvrement (overlap and add).	23
Fig. 2.6 : Schéma bloc d'une transformée orthogonale à recouvrement (LOT).	24
Fig. 2.7 : Méthode de détermination du plus proche voisin dans un dictionnaire à union de codes à permutation.....	31
Fig. 2.8 : Schéma de base du codeur audio MPEG-1 Layer I et II et format de la trame.	34
Fig. 2.9 : Réponse impulsionnelle du filtre prototype.....	35
Fig. 2.10 : Réponse fréquentielle du banc de filtres d'analyse du codeur MPEG-1 Layer I et II.....	35
Fig. 2.11 : Schéma bloc du codeur TDAC de FTR&D et format de la trame.....	40
Fig. 2.12 : Réponse impulsionnelle du filtre prototype du banc de filtres TDAC.	41
Fig. 2.13 : Réponse fréquentielle des seize premiers filtres du banc de filtres MDCT du codeur TDAC de FTR&D.	41
Fig. 2.14 : Fonction d'étalement de la membrane basilaire utilisée dans le codeur TDAC.	43
Fig. 2.15 : Schéma bloc de la correction des trames effacées.....	46
Fig. 3.1 : Schémas bloc du (a) filtrage dans le domaine temporel (b) filtrage dans le domaine des sous-bandes.	47
Fig. 3.2 : Réponse fréquentielle du filtre $H_k(z)$	49
Fig. 3.3 : Illustration des composantes qui se recouvrent avec la réponse fréquentielle du filtre de synthèse $F_k(z)$	49
Fig. 3.4 : Illustration de l'effet de variation de gain dans une bande sur le signal reconstruit.	50
Fig. 3.5 : Filtrage dans le domaine des sous-bandes.	56
Fig. 3.6 : Schémas blocs définissant l'équivalence entre (a) filtrage dans le domaine temporel (b) filtrage dans le domaine des sous-bandes.....	56
Fig. 3.7 : Filtrage temporel (a) fonction de transfert scalaire (b) implantation en version bloc (c) représentation de la version bloc par des opérations multiscadence.	58
Fig. 3.8 : Représentation en blocs multiscadence du schéma du filtrage temporel (a) schéma initial (b) schéma équivalent.....	60

Fig. 3.9 : Représentation en blocs multicadence du schéma du filtrage dans les sous-bandes.	60
Fig. 3.10 : Filtre délai de 64 échantillons (a) réponse impulsionnelle temporelle (b) réponses impulsionnelles des filtres de sous-bandes correspondant de la bande 0.	66
Fig. 3.11 : Filtre délai de 50 échantillons : réponse impulsionnelle (a) temporelle (b) des filtres sous- bandes correspondant de la bande 0 (c) après ajout du retard $z^{-(2m-1)}$	67
Fig. 3.12 : Illustration des opérations utilisées pour obtenir les filtres des sous bandes (a) filtre délai de 64 échantillons (b) filtre délai de 50 échantillons.	68
Fig. 3.13 : Schémas explicatifs de la forme des filtres des sous-bandes.	69
Fig. 3.14 : Réponse fréquentielle du filtre FIR à phase linéaire utilisé.	70
Fig. 3.15 : Réponses fréquentielles des filtres des sous-bandes (a) sous-bande 0 (b) sous-bande 6.	70
Fig. 3.16 : Amplitude et temps de propagation de groupe de la réponse fréquentielle du filtre IIR elliptique utilisé.	71
Fig. 3.17 : Réponses impulsionnelles de la convolution du filtre temporel IIR avec les filtres d'analyse et de synthèse pour (a) la sous-bande 0 (b) la sous-bande 5.	72
Fig. 3.18 : Réponses impulsionnelles des filtres des sous-bandes correspondant aux filtres FIR pour (a) la sous-bande 0 (b) la sous-bande 63.	73
Fig. 3.19 : Convolution et sous-échantillonnage du filtre $S(z)$ avec les filtres d'analyse et de synthèse pour la bande 0 (a) filtre initial (b) après ajout d'un retard supplémentaire de 220 échantillons.	74
Fig. 3.20 : Réponses fréquentielles des filtres des sous-bandes correspondant au filtre FIR (a) sous- bande 0 (b) sous-bande 6.	75
Fig. 3.21 : Principe de la spatialisation par synthèse binaurale.	77
Fig. 3.22 : Filtres HRTF gauche et droite correspondant à l'élévation 0° et l'azimut 30° (a) réponses impulsionnelles (b) réponses fréquentielles.	78
Fig. 4.1 : Somme des signaux audio compressés (a) méthode directe dans le domaine temporel (b) méthode dans le domaine compressé.	80
Fig. 4.2 : Schéma bloc du décodage partiel des trames MPEG-1 Layer I et II.	83
Fig. 4.3 : Exemple de rapports masque à bruit obtenus à la fin de la procédure d'allocation de bits pour une trame (débit binaire de 96 kbit/s).	84
Fig. 4.4 : Comparaison des seuils de masquage estimés et originaux pour l'exemple d'une trame.	87
Fig. 4.5 : Schéma bloc de la méthode de sommation dans le domaine codé MPEG-1.	88
Fig. 4.6 : Comparaison de la courbe de masquage obtenue directement par le modèle psychoacoustique 1, à partir du signal somme, et de la somme des deux courbes de masquage correspondant aux signaux individuels.	89
Fig. 4.7 : Illustration de l'effet de la sommation sur la détermination des composantes tonales.	90
Fig. 4.8 : Comparaison des rapports signal à masque, d'une trame, relatifs au signal somme obtenus par estimation et calcul direct.	91
Fig. 4.9 : Comparaison des allocations de bits obtenus par la méthode directe et dans le domaine compressé (débit binaire égal à 96 kbit/s).	92
Fig. 4.10 : Schéma bloc de la réduction de débit.	93
Fig. 4.11 : Illustration de l'effet de la réduction du débit binaire.	95

Fig. 4.12 : Comparaison des dégradations basilaies entre la référence et les signaux issues de (a) la méthode temporelle (b) la méthode dans le domaine compressé pour un signal somme (parole+piano).	97
Fig. 5.1 : Schéma bloc du décodage partiel des trames TDAC.....	105
Fig. 5.2 : Schéma bloc du codage partiel des trames TDAC.	105
Fig. 5.3 : Exemple d'enveloppe spectrale calculée par recombinaison en comparaison avec celle obtenue lors du codage du signal résultant d'une sommation temporelle.....	107
Fig. 5.4 : Exemple de courbe de masquage en comparaison avec les seuils de masquage par bande pour un spectre MDCT obtenu par sommation.	108
Fig. 5.5 : Schéma bloc de la conversion du débit pour une trame TDAC.....	109
Fig. 5.6 : Algorithme de trans-quantification sur une bande. Les zones ombrées ne font pas partie de l'algorithme.	111
Fig. 5.7 : Probabilités moyennes pour qu'une bande de fréquence soit totalement nulle à la sortie du codeur TDAC.	114
Fig. 5.8 : Probabilités moyennes qu'une seule bande soit active parmi N lors de la sommation de signaux codés.....	114
Fig. 5.9 : Probabilités d'utilisation d'une seule bande lors de la sommation de 2 signaux en considérant le phénomène de masquage inter-signaux dans les bandes.	115
Fig. 6.1 : Configurations de conférence multipoint (a) centralisée avec un réseau en étoile (b) décentralisée avec un réseau de type maillé.	123
Fig. 6.2 : Illustration de l'utilité de mise en œuvre d'une correction de perte de paquets dans le pont audio.	129
Fig. 6.3 : Comparaison des gains en complexité du pont à sommation dans le domaine compressé par rapport à celui réalisant la sommation en temporel avec et sans utilisation du mécanisme de gestion des flux discontinus.....	130
Fig. 6.4 : Schéma fonctionnel du pont audio de téléconférence opérant dans le domaine compressé et incluant la correction des pertes paquets.	132
Fig. 6.5 : Terminal de téléconférence en configuration multipoint décentralisée opérant dans le domaine compressé.....	133

Liste des tableaux

Tableau 1.1 : Gammes de qualités et débits.....	1
Tableau 3.1 : Comparaison des complexités par trame.....	76
Tableau 4.1 : Correspondance entre indices et niveaux de quantifications pour certains cas particuliers.	85
Tableau 4.2 : Constante de correction en fonction de l'énergie moyenne de la trame.	86
Tableau 4.3 : Résultats du test subjectif.....	96
Tableau 5.1 : Analyse de la complexité des principaux blocs des processus (a) du codage (b) et du décodage TDAC de FTR&D.	102
Tableau 5.2 : Statistiques sur le voisement du signal somme (a) de 2 signaux (b) de 3 signaux.....	104
Tableau 5.3 : Pourcentage d'erreur sur la décision de voisement.....	106
Tableau 5.4 : Gains en complexité des différentes méthodes de recombinaison dans le domaine codé par rapport à la méthode de sommation classique.	117
Tableau 5.5 : Résultats des tests pour évaluation des deux méthodes de recombinaison dans le domaine codé par rapport à la méthode de sommation classique.....	118
Tableau 5.6 : Résultats du test pour la comparaison de la qualité des deux méthodes de recombinaison.	118
Tableau 6.1 : Comparaison des configurations centralisées et décentralisées pour un système de téléconférence à N terminaux.	124
Tableau A.1 : Formules des composantes polyphasées des cellules du 1 ^{er} et 2 ^d ordre purement récursives.	144
Tableau C.1 : Statistique sur la base de donnée parole de l'UIT.....	147

Abréviations

AAC	<i>Advanced Audio Coding</i>
ADPCM	<i>Adaptive Pulse Code Modulation</i>
ATM	<i>Asynchronous Transfer Mode</i>
BIFS	<i>Binary Format for Scene description</i>
CELP	<i>Code Excited Linear Predictive</i>
CIF	<i>Common Intermediate Format</i>
CNG	<i>Comfort Noise Generation</i>
CMP	<i>Convolution-Multiplication Property</i>
DCR	<i>Degradation Category Rating</i>
DCT	<i>Discrete Cosine Transform</i>
DFT	<i>Discrete Fourier Transform</i>
DSP	<i>Digital Signal Processor</i>
DST	<i>Discrete Sine Transform</i>
DTT	<i>Discrete Trigonometric Transform</i>
FFT	<i>Fast Fourier Transform</i>
FTR&D	<i>France Télécom R&D</i>
FIR	<i>Finite Impulse Response</i>
GK	<i>Gatekeeper</i>
GW	<i>Gateway</i>
HRTF	<i>Head-Related Transfer Function</i>
QCIF	<i>Quarter Common Intermediate Format</i>
QoS	<i>Quality of Service</i>
QMF	<i>Quadrature Mirror Filters</i>
QVAS	<i>Quantification vectorielle algébrique sphérique</i>
IIR	<i>Infinite Impulse Response</i>
ITD	<i>Interaural Time Delay</i>
LAN	<i>Local Area Network</i>
LOT	<i>Lapped Orthogonal Transform</i>
LPC	<i>Linear Prediction Coding</i>
LTP	<i>Long Term Prediction</i>

MC	<i>Multipoint Controller</i>
MCU	<i>Multipoint Control Unit</i>
MDCT	<i>Modified Discrete Cosine Transform</i>
MLT	<i>Modulated Lapped Transform</i>
MMSP TC	<i>Multimedia Signal Processing Technical Comity</i>
MP	<i>Multipoint Processor</i>
MOS	<i>Mean Opinion Score</i>
MPEG	<i>Moving Picture Expert Group</i>
TDAC	<i>Time Domain Aliasing Cancellation</i>
TwinVQ	<i>Transform Weighted Interleave - Vector Quantization</i>
PCM	<i>Pulse Code Modulation</i>
PVHD	<i>Plate-forme de visiophonie haute définition</i>
RTC	<i>Réseau téléphonique commuté</i>
RNIS	<i>Réseau numérique à intégration de services</i>
UIT	<i>Union internationale des télécommunications</i>
WHT	<i>Walsh Hadamard Transform</i>

Chapitre 1 **Présentation du contexte et objectifs**

1.1. Introduction

Ce premier chapitre a pour but d'une part d'introduire et de donner un aperçu général sur les différents thèmes abordés dans cette thèse et d'autre part de situer ce travail dans son environnement technique et scientifique. Ainsi, la représentation des signaux audio et parole dans le domaine compressé est d'abord présenté de manière générique. Ensuite une définition du traitement dans le domaine compressé est donnée avant de présenter un bref aperçu sur l'état de l'art de ce qui existe pour les signaux audio et image. Le paragraphe 1.4 donne une première présentation des problématiques de transmission du signal sonore sur le réseau et précisément ceux à commutation de paquets ainsi que l'interaction que peuvent avoir avec la compression ou d'autres traitements de tel signal. Enfin, la section 1.5 résume les thèmes abordés par cette thèse, ses objectifs et l'organisation de ce document.

1.2. Systèmes de codage des signaux audio et parole

1.2.1 Généralités

Le codage et la compression des signaux audio ont connu un développement considérable pendant cette dernière décennie. De nos jours, une large panoplie d'algorithmes et de standards de compression existent permettant une adaptation pour différentes applications de communication et de stockage. Ces systèmes de codage peuvent être classés suivant la gamme de qualité à laquelle ils sont destinés. Les gammes de qualité sont définies par la largeur de bande passante et par la dynamique du signal. On distingue quatre catégories dont les caractéristiques sont résumées dans le tableau 1.1 [Noll 95, Le Guyader 00].

	Bande de fréquence	Fréquence d'échantillonnage	Nombre de bits/échantillon	Débit nominal (kbit/s)
Bande téléphonique	300-3400 Hz	8 kHz	13	104
Bande élargie	50-7000 Hz	16 kHz	16	256
Bande FM	20-1500 Hz	32 kHz	16	512 monovoie
Bande HI-FI (Qualité CD)	20-2000 Hz	44,1 kHz	16	705,6

Tableau 1.1 : Gammes de qualités et débits.

En adoptant une approche générique, on peut modéliser un système de compression de l'audio ou de la parole par le schéma de la figure 1.1. L'opération d'analyse vise principalement à réduire ou à éliminer la redondance du signal d'entrée. Le signal est alors transformé dans un autre espace où la représentation de l'information est plus compacte. Cette transformation est réversible, ainsi il est possible de reconstruire le signal initial sans aucune perte. Dans les algorithmes de codage, cette étape consiste en un banc de filtres pour les codeurs fréquentiels ou en une analyse des paramètres d'un filtre de prédiction linéaire pour les codeurs de type CELP. Elle peut aussi ne pas exister comme c'est le cas pour les codeurs de forme d'onde (algorithme ADPCM de la norme G.721 par exemple) où le codage porte directement sur le signal temporel. Pour les codeurs de type CELP, utilisé essentiellement pour le signal de la parole, la connaissance a priori des propriétés de la source est exploitée. En effet, le filtre de prédiction linéaire utilisé n'est autre qu'un modèle du conduit vocal permettant la génération de la parole. Ce sont les paramètres courants de ce dernier plus l'information sur son excitation qui sont transmis. De cette manière des taux de compression élevés sont obtenus. Dans le cas des codeurs fréquentiels, généralement utilisés pour la compression de l'audio, la variété des signaux considérés limite l'utilisation d'un modèle de source.

La seconde opération dans un schéma de compression consiste à éliminer toutes les composantes non pertinentes du signal. Cette phase est irréversible puisqu'elle est réalisée avec perte d'une partie du signal, elle est effectuée de manière à minimiser la dégradation apportée après restitution. Dans ce but un modèle de perception peut être utilisé pour ne supprimer que les composantes inutiles au système d'audition humain. Cette étape consiste essentiellement en la quantification et le codage des paramètres issus de l'analyse qui peuvent être suivi d'une opération de codage entropique pour une réduction supplémentaire du débit. La séparation dans les algorithmes de compression entre blocs d'analyse et de codage n'est pas toujours stricte. La prise en compte de l'interaction entre ces deux blocs est utilisée par exemple dans les codeurs de type CELP pour arriver à des taux de compression élevés.

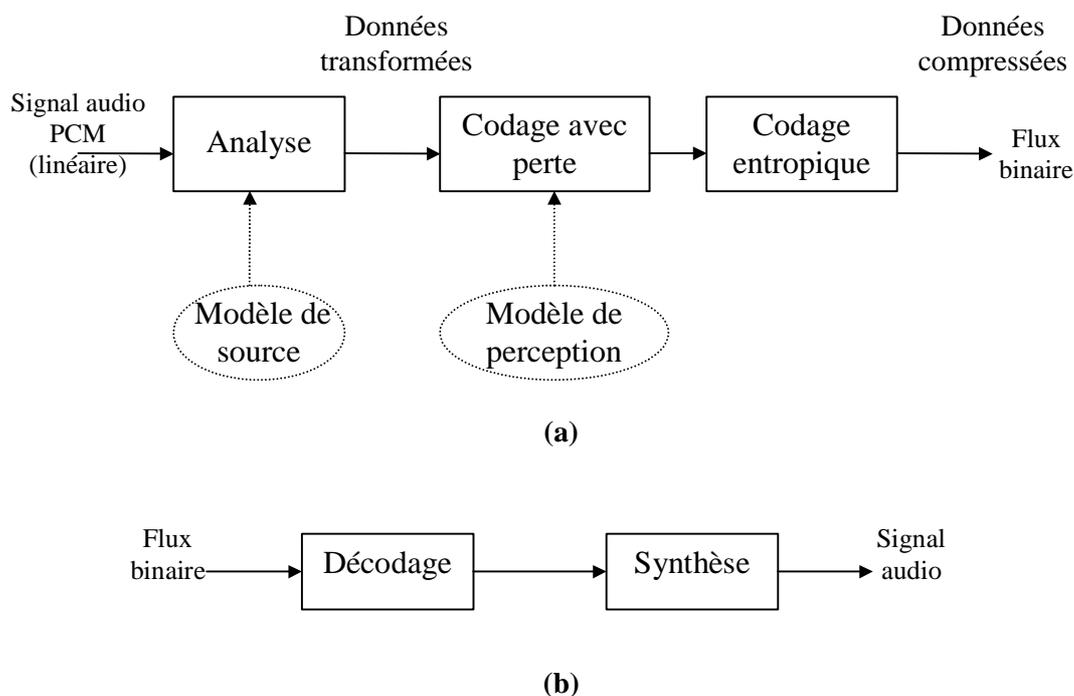


Fig. 1.1 : Schéma générique d'une opération de codage/décodage du signal audio.

La classification des codeurs en familles se base essentiellement sur la partie analyse du schéma de codage. En effet, c'est cette phase qui permet d'obtenir le type de représentation dual du signal temporel qui sera quantifié, codé et transmis en flux binaire. On peut distinguer principalement trois types de famille : codeurs de forme d'onde, codeurs fréquentiels, codeurs paramétriques par exemple de type CELP. La représentation choisie du signal est en relation directe avec le type du signal considéré et le traitement de codage ultérieur qui sera réalisé. Ainsi par exemple une représentation dans le domaine fréquentiel est faite pour pouvoir profiter des propriétés de perception de l'oreille humaine.

Dans cette thèse on s'intéresse aux codeurs de type fréquentiel perceptuel qui sont utilisés pour la parole et l'audio large bande ou haute qualité. Spécifiquement, on considérera le codeur audio MPEG-1 [Brandenburg 94] et le codeur propriétaire TDAC de FTR&D [DT&FT 98]. Une étude plus détaillée y sera consacrée dans le chapitre 2 où on donnera plus d'informations sur la structure de ce type de codeurs.

La propriété de linéarité que présentent les transformations entre l'espace temps et l'espace fréquence facilite l'application de certains traitements sur des données représentées dans ce dernier espace. Par ailleurs, d'autres problèmes surgissent du fait que les données transformées peuvent être soumises à une opération de codage ultérieur. Un autre problème posé par cette thèse consiste à analyser les possibilités d'effectuer des traitements sur les représentations codées.

1.2.2 Caractéristiques

Le cahier de charge auquel est soumis tout système de codage de la parole ou de l'audio comporte quatre attributs principaux [Noll 95, Le Guyader 00] : le débit, la complexité, le délai algorithmique et la qualité du signal reconstruit après codage/décodage. Pour certaines applications un seul de ces attributs peut être le plus déterminant, pour d'autres un certain compromis doit être fait.

Le débit est la composante la plus importante puisqu'il reflète le taux de compression fourni par l'algorithme. Il est également fonction de la gamme de qualité considérée puisque celle-ci détermine la fréquence d'échantillonnage.

La complexité algorithmique est aussi un attribut important. Pour une implantation sur des processeurs numériques du signal (DSP), elle est souvent évaluée en termes de millions d'instructions par secondes (MIPS) et par la taille de mémoire vive (RAM) et de mémoire morte (ROM) utilisées. Pour des implantations sur PC, elle est mesurée par le pourcentage du temps CPU consommé par l'algorithme. En général, les algorithmes de codage fournissant un fort taux de compression avec de faibles dégradations présentent une grande complexité. Malgré la puissance accrue des processeurs disponibles aujourd'hui, cet attribut garde toujours de l'importance pour certaines applications. Ceci est particulièrement vrai pour des terminaux portables, puisque la complexité a une influence directe sur la puissance consommée ou lorsque d'autres processus doivent partager les mêmes ressources de traitement. La répartition de complexité entre codeur et décodeur n'est pas toujours symétrique, elle dépend de l'algorithme de codage considéré. Dans certaines applications, par exemple de type conversationnel, les deux composantes de complexité doivent être prises en compte. Pour d'autres, tel que le stockage ou la diffusion sonore, la complexité du décodeur prime puisque le codage se fait hors ligne.

Les systèmes de codage introduisent aussi un retard dû à l'algorithme utilisé. Ce délai est négligeable pour des codeurs agissant échantillon par échantillon. Il devient important pour les codeurs qui traitent le signal découpé en blocs (ou trames) tels les codeurs fréquentiels. Dans ce dernier cas de figure, et si on tient compte en plus de ce délai algorithmique du temps d'acquisition et du temps de traitement, le retard de codage/décodage est généralement de l'ordre de 3 à 4 fois la durée de la trame. L'importance du délai apparaît selon l'application considérée. Ainsi, pour les communications interactives ce paramètre est très critique. Mais il l'est beaucoup moins pour le stockage ou la diffusion sonore.

La qualité est enfin fonction de ces trois premiers paramètres. Un codeur peut donc être représenté par un point dans un espace à quatre dimensions dont chacun des axes représente un des attributs précédents. L'ensemble de points à qualité égale constitue une surface dans cet espace. Le déplacement sur cette surface dans le sens de l'amélioration d'un attribut implique nécessairement une dégradation dans le sens de l'un des autres.

1.3. Traitement du signal dans le domaine codé

1.3.1 Définition et généralités

De plus en plus de données multimédia (parole, audio, image et vidéo) sont transmises, diffusées ou stockées de manière numérique. Pour des raisons d'adaptation aux capacités des réseaux de transmission ou de réduction des tailles de mémoire, ils sont tous compressés à un certain point de la chaîne. Pour effectuer un traitement sur de telles données, la méthode classique consiste à les décompresser, appliquer le traitement, puis les compresser pour revenir à la représentation codée (figure 1.2a). Les étapes supplémentaires de décompression et compression peuvent être très coûteuses en terme de complexité algorithmique, engendrer un délai additionnel et nécessite aussi beaucoup de mémoire. En plus, la mise en cascade d'opérations de décodage/codage successives peut donner lieu à une perte de qualité. Comme alternative à cette méthode et pour pallier ces inconvénients, le traitement dans le domaine compressé propose d'agir directement sur la représentation codée (figure 1.2b).

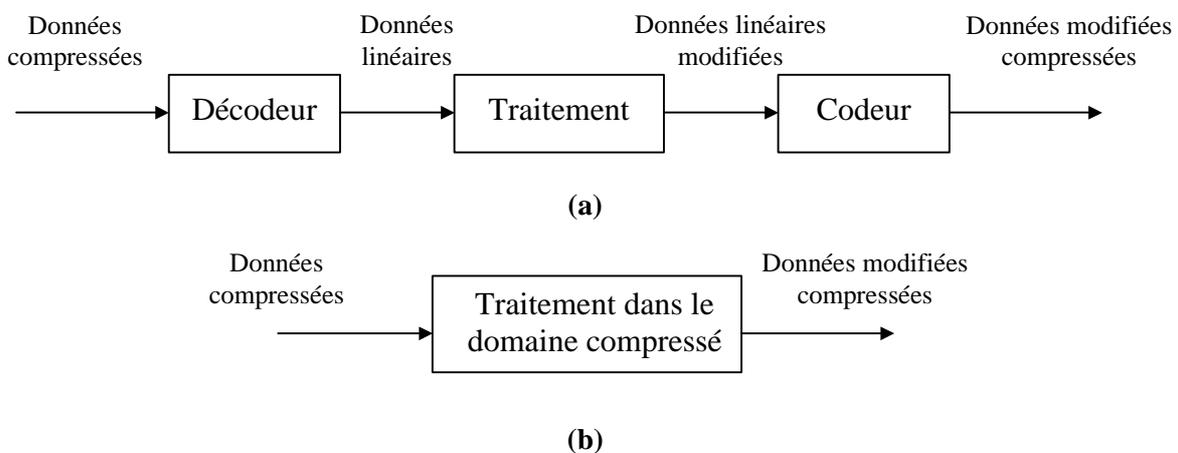


Fig. 1.2 : Schémas génériques du (a) traitement sur les données linéaires (b) traitement dans le domaine compressé.

Le traitement du signal dans le domaine compressé est un axe de recherche récent, les premiers travaux ont porté essentiellement sur les signaux images et vidéo. Cette discipline a été classée par le comité technique sur le traitement du signal multimédia de l'IEEE (MMSP TC) sous la rubrique "*Intégration de média*" [Chen 97]. A première vue cette classification est peu parlante. Par ailleurs, en examinant le problème de plus près, on peut voir que le traitement dans le domaine compressé permet de profiter des interactions entre les différentes manipulations du signal dans une chaîne de traitement et donc de réaliser une certaine intégration.

Le domaine de représentation où le signal est considéré a un impact important sur les performances du traitement à réaliser. Ainsi, à titre d'exemple, le traitement d'annulation d'écho basé sur des algorithmes de filtrage adaptatif révèle principalement des problèmes de convergence, de stabilité et de coût de calcul. L'utilisation d'algorithmes agissant dans le domaine fréquentiel ou en sous-bandes au lieu du domaine temporel améliore la vitesse de convergence et réduit la complexité [Gilloire 92]. En revanche, ils ont l'inconvénient d'introduire un délai supplémentaire dû à la transformation temps/fréquence. Pour des applications où le signal est directement disponible dans ce domaine, ce dernier problème disparaît. Le contexte d'application peut donc aussi imposer ou favoriser la représentation du signal à considérer. C'est dans ce sens que le traitement dans le domaine compressé présente son intérêt.

Contrairement au traitement dans le domaine temporel ou fréquentiel, l'appellation traitement dans le domaine compressé est plutôt générique et ne reflète pas exactement la vraie forme sous laquelle les données sont manipulées. En effet, d'une part le type de représentation en domaine compressé diffère selon le codeur initialement utilisé (codeur paramétrique, fréquentiel,...) et d'autre part, il est presque impossible de manipuler des données compressées directement sans recours à un décodage partiel. Cette phase préliminaire permet d'extraire à partir du flux binaire initial une représentation malléable qui convient au type de traitement voulu.

1.3.2 Traitement dans le domaine codé des signaux images et vidéo

Avec l'avènement des systèmes de stockage et de la diffusion numérique des signaux vidéo (DVD et DVB) et aussi de la télévision numérique haute définition (HDTV) d'énormes masses de données numériques de vidéo compressée sont utilisées actuellement. Les méthodes de compression utilisées pour l'image fixe ou la vidéo ont une complexité très élevée. Le gain en complexité que peut induire la manipulation directe dans le domaine compressé de telles données a constitué la motivation majeure de plusieurs recherches. Par ailleurs, l'élimination d'une phase de décompression pour manipulation dans le domaine spatial permet aussi d'économiser les capacités des mémoires tampons requises et de réduire le délai. Ces travaux ont porté principalement sur des flux binaires obtenus par les normes de codage MPEG, M-JPEG et H.261.

Les algorithmes de compression d'image et vidéo procèdent par une première segmentation des images en blocs de pixels de taille 8×8 . Dans la norme de codage M-JPEG, les blocs de pixels sont transformés par une DCT bi-dimensionnelle, les coefficients résultant sont quantifiés uniformément puis regroupés en suivant un balayage en zig-zag. Un codage entropique, sans distorsion, est enfin réalisé en utilisant les statistiques de l'image pour réduire le débit. Pour la vidéo et les images animées, la norme de codage MPEG utilise la même méthode pour chaque image et profite en plus de la compensation de mouvement pour réduire la redondance temporelle inter-images [Rao 96].

On trouve une littérature très abondante sur le traitement de l'image et de la vidéo dans le domaine compressé. Les premiers travaux ont porté sur des manipulations géométriques de type translation, rotation, et multiplication par un facteur d'échelle, et sur le sous-titrage [Chang 95]. La composition d'images et la conversion de format CIF en QCIF ont aussi suscité de l'intérêt surtout pour leur application dans des téléconférences multipoint incluant un pont vidéo [LeMair 97, Zhu 98, Zhu 99]. La conversion du débit binaire pour la vidéo codée en H.261 [Morrison 94] et en MPEG [Nakajima 95, Assunção 98, Werner 99] est un sujet qui a été aussi traité pour ses utilisations dans les passerelles en transmission sur des réseaux hétérogènes dans le but de s'adapter à la qualité de service (QoS) disponible. Le transcodage entre les formats de compression MPEG et M-JPEG a été abordé pour ses applications de communication entre terminaux hétérogènes [Shibata 99]. En plus, pour des applications de recherche et d'extraction rapide sur des bases de données compressées (format JPEG, JMOVIE ou MPEG) des méthodes d'indexation et de classification de scènes basées sur le contenu ont été développées [Fernando 99, Dawood 99, Yu 99]. De telles méthodes peuvent être utilisées pour la navigation rapide sur les bases de données à travers le réseau Internet, et aussi pour des applications de l'édition vidéo (sur des bases de données HDTV par exemple). L'édition vidéo dans le domaine compressé fournit des possibilités de couper, copier et coller des segments vidéo.

Le traitement des signaux vidéo dans le domaine compressé s'apparente surtout au traitement dans le domaine de la transformée DCT. Des manipulations de type composition et transformations géométriques se transposent facilement dans ce domaine vu les propriétés de linéarité et d'orthogonalité de cette transformée, d'autres posent plus de problèmes. Spécifiquement, le filtrage linéaire dans le domaine DCT nécessite une attention particulière pour faire face aux effets de blocs. Plusieurs travaux ont essayé de résoudre ce problème et des solutions exactes et efficaces n'ont été développées que récemment [Kresch 99]. Une discussion plus détaillée sur ce sujet sera donnée dans le chapitre 3 où on abordera le problème du filtrage dans le domaine fréquentiel avec une approche plus générale. Pour des algorithmes de compression utilisant la compensation de mouvement, la question de savoir si on peut aller au-delà du domaine DCT en effectuant directement des traitements sur les vecteurs de mouvement a été étudiée. La non-linéarité entre les paramètres de ces vecteurs et le domaine spatial limite les degrés de liberté de manipulation [Chang 95].

1.3.3 Traitement dans le domaine codé des signaux audio

En comparaison avec ce qui a été fait pour les signaux vidéo et images compressés, peu de travaux ont été réalisés pour les signaux audio compressés. Les premières recherches se sont concentrées sur le traitement dans le domaine audio compressé MPEG-1 Layer I-II [Brandenburg 94]. Parmi les opérations à transposer dans le domaine codé qui ont été étudiées, on peut citer le mixage et le filtrage [Broadhead 95, Lanciani 99b]. L'interaction entre codage MPEG-1 et d'autres traitements type annulation d'écho stéréophonique (SAEC) a été aussi analysée [Gänsler 98]. Cette interaction a concerné aussi les problèmes de cryptage et de tatouage de signaux dans le domaine codé du codeur MPEG-2 AAC [Allamanche 00, Neubauer 00]. Une analyse plus détaillée de certains de ces travaux sera faite au cours de cette thèse et spécialement dans les chapitres 3 et 4.

1.3.4 Approche inverse : codage pour manipulation simple dans le domaine codé

Comme il a été détaillé dans le paragraphe 1.2.2, la conception d'algorithmes de compression est objet de plusieurs contraintes. Outre la qualité du signal restitué et le débit binaire, la complexité de l'algorithme et le délai qu'il introduit peuvent être aussi des paramètres critiques selon l'application. La robustesse aux erreurs de transmission peut prendre aussi de l'importance pour des contextes particuliers tel que les radiomobiles. En plus de ces paramètres classiques, on peut ajouter lors de la construction de l'algorithme la faculté que les signaux compressés qui en résultent se prêtent à une manipulation facile pour certains types de traitements. Cette approche est inverse aux exemples précédents où les traitements effectués venaient s'adapter à des signaux compressés selon un algorithme déjà existant.

Cette idée a été adoptée dans [Levine 98] pour concevoir un codeur audio permettant aussi bien une compression haute qualité qu'une représentation dans le domaine compressé adaptée à des modifications sur l'échelle de temps et l'échelle du pitch du signal (time-scale and pitch-scale modifications). Le système conçu a pour but d'obtenir un débit de 32 kbit/s avec une qualité similaire à celle obtenue par le codeur MPEG-2 AAC au même débit. Il est basé sur une combinaison du signal audio en sinusoides, bruits et transitions. Ainsi, sur les parties stationnaires, le signal est représenté par des sinusoides et des gains du bruit résiduel dans les bandes critiques. Dans les phases de transition un codeur par transformée simplifié utilisant un banc de filtres MDCT avec fenêtre courte est utilisé pour coder le signal. La synthèse se fait alors par des générateurs sinusoidaux auxquels on ajoute du bruit et avec insertion des trames décodées correspondant aux parties transitoires. La représentation paramétrique adoptée par ce codeur facilite la manipulation directe sur le signal compressé. Ainsi la modification du pitch peut se faire directement en agissant sur les fréquences des sinusoides et l'étalement/compression du signal sur l'échelle de temps est réalisée par duplication/limitation des paramètres de la même trame.

La norme MPEG-4 Audio préconise aussi des codeurs de type paramétrique afin d'assurer ces mêmes fonctionnalités pour des débits allant de 2 à 16 kbit/s et pour divers types de signaux. Ainsi, le codeur HVXC (Harmonic Vector eXcitation Coding) assure le codage de la parole pour des débits de 2 à 4 kbit/s. Le codeur HILN (Harmonic and Individual Line plus Noise), issu de la version 2, est utilisé pour les signaux audio au-delà de 4 kbit/s [Purnahagen 99].

Ces exemples de codage paramétrique montrent qu'il est possible d'obtenir une représentation audio codée flexible pour certains traitements sans perte sur les attributs débit et qualité. Par ailleurs, la généralisation d'une telle conclusion à d'autres traitements ne peut être toujours garantie. En effet, l'augmentation du nombre de contraintes imposées sur le codeur limite nécessairement ses performances sur un axe ou un autre. D'un autre côté, le domaine d'application visé impose nécessairement certaines conditions. Pour la transmission de l'audio sur le réseau, les algorithmes existants sont optimisés en premier lieu pour avoir des taux de compression élevés avec la moindre dégradation. Ils sont soit populaires et implantés sur des milliers de machines à travers le monde (norme MPEG-1 audio par exemple) soit propriétaires et utilisés pour des applications particulières comme le cas du codeur TDAC de FTR&D intégré dans le système de téléconférence *PVHD*¹.

¹ Plate-forme de Visiophonie Haute Définition.

1.4. Traitement du signal audio et transmission sur les réseaux de communication multimédia

La prolifération des machines puissantes (PC, cartes à base de processeurs numériques du signal,...) pouvant supporter en temps réel des algorithmes très complexes de traitement du signal multimédia (parole, audio et vidéo) et l'émergence des réseaux dans le grand public a ouvert le champ à de nouveaux scénarios de communication multimédia. L'ère des réseaux de communication multimédia est donc apparue. Plusieurs applications faisant intervenir la transmission du son sur le réseau se trouvent de plus en plus répandue. On peut distinguer trois domaines où l'audio sur le réseau intervient : conférences multimédia (téléconférence, jeux sur réseau...), applications de diffusion tel que l'audio à la demande et la téléphonie sur IP.

Pendant longtemps la transmission de l'audio temps réel était réalisée sur des réseaux issus du monde des télécommunications. Ces réseaux sont tous à connexion réservée (basée sur la commutation de circuits ou de cellules) et conçus essentiellement pour des communications point à point. Au fur et à mesure de l'accroissement des besoins en communication, ce type de réseau s'est adapté par le biais de l'amélioration de la technologie des Modems, de l'apparition du RNIS et plus récemment des réseaux backbones ATM et Relais de Trame. Le traitement du signal a toujours accompagné ce développement en fournissant des algorithmes permettant d'adapter le signal aux conditions du réseau et aussi pour faire face aux problèmes et artefacts que peut engendrer ce dernier. Ainsi, dès le début et pour la transmission sur le réseau téléphonique commuté (RTC) le besoin en compression est apparu et des algorithmes de codage ont été développés : du plus élémentaire de la norme G.711 au plus sophistiqué des normes G.728 et G.729 en passant par le G.721 et G.726 basées sur l'algorithme ADPCM. L'écho électrique dû à la désadaptation d'impédance lors de la conversion hybride de quatre fils à deux fils est aussi un problème soulevé par le réseau RTC et pour lequel des algorithmes d'annulation adaptatifs ont été aussi développés.

De nos jours les réseaux à commutation de paquets connaissent une grande expansion. Ces réseaux sont à connexion non réservée. Ils sont constitués de larges et petits réseaux IP interconnectés sous forme d'Internet ou de plusieurs Intranets répartis. Bien qu'ils aient été initialement conçus pour le transport de données, cette forte expansion a encouragé leur utilisation pour des applications de transmission de l'audio, et plus généralement des flux multimédia.

La diversité des supports réseaux disponibles actuellement possédant chacun des caractéristiques particulières, d'une part, et l'apparition de nouvelles architectures de communication et de nouvelles classes d'applications, d'autre part, ont fait surgir de nouveaux problèmes et contraintes. Ils ont ouvert le champ à l'ajout d'autres fonctionnalités permettant l'amélioration de la qualité totale et ainsi l'augmentation de la satisfaction de l'utilisateur. Ces différents problèmes et opportunités d'amélioration sont les principales préoccupations actuelles du traitement du signal audio pour les réseaux de communication multimédia.

Pour le développement d'applications audio sur le réseau, la bonne compréhension de l'architecture et des caractéristiques du réseau et des algorithmes de traitement de l'audio, essentiellement les algorithmes de compression, peut permettre de profiter de leurs interactions pour créer des architectures de traitement sur réseau optimisées. Le traitement du signal dans le domaine compressé peut être un des éléments-clés permettant d'aller dans ce sens.

Dans la suite nous allons expliciter les caractéristiques des réseaux multimédia et donner quelques exemples où le traitement du signal intervient pour résoudre les nouveaux problèmes. Nous montrerons aussi comment de nouvelles notions de traitement du signal sont apparues pour faire face à ces problèmes.

1.4.1 Réseaux multimédia : caractéristiques et contraintes

Les principales contraintes auxquelles doit faire face la transmission du son peuvent être résumées dans les attributs suivants [Bargar 99, Hardman 00, Kostas 98] : la bande passante, le délai, la gigue et les pertes de paquets (sur IP) ou de cellules (sur ATM). Les trois premiers définissent les paramètres de la qualité de service (QoS) pour ce type d'application de transmission.

Les flux multimédia nécessitent des débits importants. Si l'existence d'algorithmes de compression efficaces les a rendus accessibles via la plupart des réseaux, la bande passante reste toujours un problème dans certaines applications telle que les réseaux sans fils. La variabilité de largeur de bande passante est aussi un problème qui se pose essentiellement pour les réseaux à qualité de service non garantie tels ceux à protocole IP. En effet, ces réseaux sont de type "best effort service". La largeur de bande passante disponible n'est pas allouée a priori, comme c'est le cas pour le RNIS ou ATM, elle dépend des conditions du trafic. La variabilité de la bande passante intervient aussi lors de la transmission sur des architectures de communication comportant des réseaux hétérogènes. Ainsi par exemple pour une architecture H.323 [Thom 96], la largeur de bande peut varier de 28,8 kbit/s pour un accès avec un modem V.34 jusqu'à 150 Mbit/s pour des liaisons ATM. Le même type de scénario peut se présenter pour une transmission IP multicast² à un groupe de destinataires situés sur différents réseaux.

La seconde contrainte intervenant lors de la transmission du son sur réseau concerne *le retard de bout en bout* induit par le transport sur le réseau. Ce paramètre est déterminant pour l'interactivité des communications. Un retard de 250 ms est considéré comme le seuil à partir duquel l'interactivité d'une conversation commence à se dégrader [Schmitt 98]. Le retard de bout en bout comprend trois composantes : le délai des codeurs/décodeurs intervenants dans la chaîne de communication (temps d'acquisition + temps de traitement + délai algorithmique), le délai de propagation sur les supports physiques et le délai introduit par les serveurs et/ou routeurs des nœuds du réseau. Les deux premières composantes sont fixes. La dernière, due au mécanisme de temporisation (buffering) utilisé dans les routeurs pour absorber la variation des conditions du trafic, varie selon les conditions instantanées du réseau.

Si les conditions du réseau sont mauvaises en termes de qualité de service la variation du délai peut devenir importante. C'est le phénomène de *la gigue* (ou "jitter" en anglais). L'utilisation de mécanismes à base de mémoires tampons (buffer) dans les récepteurs permet de masquer ce problème mais au détriment d'un délai supplémentaire. Une adaptation dynamique des tailles de ces mémoires tampons permet d'aboutir à un compromis entre un délai minimal et une faible gigue. Par ailleurs, les paquets audio ayant un retard dépassant une certaine limite seront rejetés. La contrainte temps réel est garantie au prix de pertes possibles de paquets.

² Le "*Multicasting*" est une technique de transmission de groupe qui permet d'économiser la bande passante en évitant de transmettre N paquets d'information binaire sur la même connexion aux N différents destinataires.

Le problème de *pertes de paquets* ou de cellules n'est pas dû seulement au phénomène de la gigue. Pour les réseaux IP, la délivrance des paquets n'est pas garantie. Les exigences sur le délai pour assurer la contrainte temps réel pour les applications de transmission de l'audio sur le réseau ne permettent pas d'utiliser un protocole de transport fiable tel que TCP (Transport Control Protocol). Le protocole UDP (User Datagram Protocol) utilisé n'assure ni le contrôle de congestion ni la garantie de délivrance (comme IP). Ainsi dans le cas de forte congestion, la fluidité de transmission n'est plus possible. Un dépassement des mémoires tampons des routeurs peut donc survenir et certains paquets peuvent être rejetés.

1.4.2 Exemples de traitements du signal audio pour les réseaux multimédia

Les problèmes cités dans le paragraphe précédent, particulièrement dus à la transmission sur les réseaux à qualité de service non garantie, posent plusieurs sujets de recherche intéressants pour le traitement du signal. Ils concernent principalement l'adaptation des techniques de compression audio aux exigences imposées par le réseau. Ainsi, les aspects compression et transmission ne doivent plus être traités de manière séparée. C'est dans ce sens que de nouvelles notions sont apparues pour donner aux algorithmes classiques de compression d'autres fonctionnalités que la simple réduction du débit.

Pour faire face à la variabilité de la bande passante, rencontré lors d'une transmission sur réseau hétérogène ou à qualité de service non garantie, la notion de codeur *hiérarchique* (scalable) a été développée. Ce type de codeurs fournit un flux binaire hiérarchisé constitué de blocs emboîtés. Le bloc de base assure une qualité minimale avec un faible débit. Les autres blocs ne faisant qu'améliorer la qualité du signal synthétisé au débit faible. Cette nouvelle notion permet un fonctionnement à débit binaire variable avec différentes qualités du signal reconstruit et sans avoir recours à plusieurs codeurs à différents débits. La première réalisation de ce type de codage était faite par le codeur SB-ADPCM bande élargie de la norme UIT-T G.722 pour des débits de 48-56-64 kbit/s [G.722]. Des solutions sont aussi développées dans le cadre de la norme MPEG-4 [Grill 99]. Le cœur du codeur, fournissant le débit minimal de 24 kbit/s, utilise un codage de type CELP ou de type TwinVQ. L'étage supérieur est constitué d'un codage de type MPEG-2 AAC. La version 2 de MPEG-4 prévoit aussi un codeur AAC hiérarchique pour des débits allant de 16 kbit/s à 64 kbit/s avec une granularité plus fine de 1 kbit/s [Purnahagen 99]. L'exigence de faible délai étant une des contraintes imposées par les applications interactives, des codeurs combinant la propriété de hiérarchie avec un faible délai ont été développés récemment [Taddei 99, Jbira 99]. Le codeur hiérarchique de Taddei combine l'algorithme de codage ACELP de la norme G.729 au codeur par transformée TDAC de FTR&D [DT&FT 98] pour faire face aussi à la diversité des signaux audio (parole, musique). Il offre trois flux binaires de 8 kbit/s, 14,1 kbit/s et 24 kbit/s avec un délai de l'ordre de 92,5 ms. Cette propriété de la hiérarchie (scalability) ne se résume pas à une variation du débit binaire. Elle peut porter aussi sur la complexité, le flux binaire pouvant être décodé par des plates-formes matérielles de différentes puissances.

La seconde difficulté à laquelle s'est intéressé le traitement du signal concerne les pertes de paquets. Plusieurs techniques de récupération ont été élaborées pour remédier à ce problème [Perkins 98]. De plus en plus les nouvelles normes de codage comportent des méthodes de récupération de l'effacement de trames. C'est le cas par exemple du codeur de la norme G.723.1 adopté par l'IMTC (International Multimedia Teleconferencing Consortium) comme codeur de base pour des applications de voix sur IP, en particulier la téléphonie sur Internet. Ce codeur de type CELP utilise une méthode d'interpolation sur les paramètres du filtre prédicteur pour générer la trame perdue.

Pour les systèmes conversationnels, dans la majorité du temps une seule personne parle et les autres écoutent. Les périodes où plus de deux personnes parlent simultanément ne sont pas très fréquentes. Ainsi, dans une conversation à deux, le taux moyen des périodes de silence est de l'ordre de 60 % du temps. Cette propriété des applications de communication de groupe est utilisée pour réduire davantage le débit. Ainsi des fonctions annexes de type VAD/DTX/CNG (Voice Activity Detection / Discontinuous Transmission / Comfort Noise Generation) sont pratiquement toujours associées aussi aux nouvelles normes. Un algorithme de détection d'activité vocale est alors utilisé. Lors d'une période de silence seule l'information sur le bruit de fond (SID : Silence Insertion Descriptor) est transmise avec un débit faible. Le décodeur génère alors un bruit de confort dont les caractéristiques sont adaptées lentement. La fonction CNG peut être aussi utilisée pour lutter contre l'effacement de trame. L'annexe B de la norme G.729 offre un exemple d'algorithmes élaborés pour ces fonctionnalités [Benyassine 97].

Les recommandations G.723.1, G.729 normalisées par l'UIT-T en 1996 pour le codage de la parole, constituent de bons exemples pour un fonctionnement dans un environnement avec les contraintes de bandes passantes, pertes, faible délai et complexité. Ils sont préconisés dans la norme H.323 pour la téléphonie sur IP.

Un autre problème dû au succès commercial des réseaux multimédia est lié à la question de protection de la propriété intellectuelle. Ce point ne relève pas complètement du traitement du signal, il fait appel aussi à des éléments sur la sécurité et la protection au niveau système, y compris des méthodes de cryptographie. Par ailleurs, les techniques du traitement du signal interviennent pour introduire et dissimuler l'information de tatouage (Watermark) dans le signal audio.

Après environ trois décennies de recherche, les algorithmes de compression de la parole et de l'audio commencent à atteindre une certaine phase de maturité. Les principaux enjeux actuels et futurs concernent les fonctionnalités ci-dessus ou d'autres enrichissant le rendu sonore. En effet, il suffit de voir que la norme MPEG-4 audio version 2, finalisée en mars 2000, traite des sujets suivants : robustesse aux erreurs, hiérarchie à faible pas, codage audio à faible délai, et de la spatialisation en plus du codage paramétrique [Purnahagen 99]. Ces différents exemples montrent l'interaction entre traitement du signal et environnement global de l'application de communication y compris les caractéristiques du réseau. Le traitement dans le domaine compressé est une notion qui permet d'aller dans ce sens. L'exemple qui sera traité dans cette thèse concernant la téléconférence multipoint en architecture centralisée montrera l'apport de cette stratégie de traitement sur les paramètres de qualité (réduction de délai...). Il montrera aussi l'interaction entre le choix d'une architecture de communication réseau, les fonctionnalités de traitement améliorant la qualité et enrichissant le rendu sonore et le domaine du traitement du signal choisi.

1.5. Thèmes d'étude et organisation de la thèse

Dans le paragraphe 1.2 nous avons montré l'intérêt du traitement du signal audio dans le domaine compressé, le paragraphe 1.3 a illustré l'apport du traitement du signal aux réseaux multimédia. Cette thèse s'intéresse au traitement du signal audio dans le domaine compressé pour des signaux codés par des algorithmes de type fréquentiel perceptuel et pour des codeurs large bande et haute qualité. Les codeurs considérés sont les codeurs MPEG-1 Layer I-II et TDAC de FTR&D. L'exemple d'application considéré par cette thèse concerne les systèmes de téléconférence haute qualité multipoints. C'est dans

ce sens qu'on s'intéresse particulièrement à des traitements de type mixage audio, spatialisation sonore et recouvrement de trames effacées dues aux paquets audio perdus.

Le choix du codeur audio haute qualité MPEG-1 est dû principalement à la généralité des principes qu'il utilise pour la classe de codeurs fréquentiels perceptuels. Il se justifie aussi par sa large utilisation dans diverses applications, notamment dans les systèmes de téléconférence. En effet, MPEG-1 Layer II fait partie des codeurs obligatoires exigés par la norme UIT-T H.310 [H.310] pour les terminaux des systèmes de communication audiovisuels opérant sur des réseaux large bande (ATM).

Le codeur TDAC de FTR&D était l'un des trois codeurs en compétition pour la nouvelle norme de codage bande élargie G.722.1 de l'UIT-T [G.722.1]. Ce processus de normalisation visait à fournir un codeur pour les systèmes de téléconférence fonctionnant à des débits 16-24-32 kbit/s et avec une qualité équivalente à celle du codeur SB-ADPCM de la norme G.722 à des débits 48-56-64 kbit/s, respectivement. Cette norme, qui a été finalisée en septembre 1999, a adopté le codeur PT724 de PictureTel [PT724] pour les débits de 24 et 32 kbit/s. La qualité à 16 kbit/s n'était pas acceptable pour aucun des codeurs présentés. L'existence de points communs entre les codeurs TDAC de FTR&D et PT724 (transformée MDCT, calcul de seuils de masquage à partir de l'enveloppe spectrale au niveau codeur/décodeur...) permet d'étendre certains principes développés dans cette thèse à ce dernier codeur.

Les études menées dans le cadre de cette thèse peuvent être regroupées suivant les trois thèmes majeurs suivants :

- **Traitement du signal audio dans le domaine fréquentiel :** La transposition d'un traitement du domaine temporel au domaine fréquentiel n'est pas si directe. Des problèmes et distorsions peuvent intervenir, notamment d'aliasing, qui nécessitent une attention particulière. Ce problème d'aliasing sera détaillé dans le chapitre 3 où on étudiera le problème du filtrage dans le domaine des sous-bandes. Ainsi, après un aperçu de l'état de l'art, un algorithme innovant pour transposer tout filtrage rationnel (FIR ou IIR) dans le domaine fréquentiel sera exposé. L'illustration de cette méthode sera donnée par une application au filtrage pour la spatialisation sonore.
- **Traitement dans le domaine codé :** La figure 1.1 donne un schéma générique du codage audio et parole. Le sujet principal de ce thème se propose de répondre à la question suivante : jusqu'à quel point de ce schéma de codage peut-on aller pour un traitement d'un flux compressé? L'examen de cette question sera fait principalement pour le mixage audio et la conversion de débit. Elle sera analysée pour les structures des codeurs MPEG-1 et TDAC de FTR&D et exposée dans les chapitres 4 et 5. La diversité des méthodes de quantification et de codage ne permet pas de donner une réponse exhaustive ou générale à cette question. Par ailleurs, la variabilité entre les méthodes utilisées dans les deux codeurs examinés permet de soulever une diversité de problèmes intéressants. L'utilisation par ces codeurs de certains principes génériques permet aussi de dégager quelques règles et conclusions qui peuvent être étendues à d'autres codeurs.
- **Traitement dans le domaine compressé et architectures pour le traitement du son sur le réseau :** Ce thème a aussi pour objectif d'illustrer, à travers quelques exemples, l'intérêt du traitement audio dans le domaine compressé pour des applications de transmission du son sur le réseau et de donner aussi des schémas de réalisation fonctionnels. L'exemple principal qui

sera traité dans le chapitre 6, et qui a été mis en œuvre, concerne une architecture de téléconférence centralisée basée sur une unité de contrôle multipoint (MCU). Cet exemple illustrera l'interaction entre architecture réseau et traitement dans le domaine compressé. Ainsi, on montrera l'intérêt de l'ajout d'autres fonctionnalités de type correction des pertes paquets avec un traitement dans le domaine compressé au pont audio de téléconférence.

Chapitre 2 Codage audio fréquentiel perceptuel

2.1. Introduction

Pendant cette dernière décennie le codage des signaux audio a connu un large développement. Plusieurs codeurs ont été proposés pour fournir des débits de plus en plus faibles pour une qualité accrue et pour une large gamme d'applications. Ces codeurs sont soit propriétaires soit normalisés par des organisations internationales. Par ailleurs, ils possèdent tous une structure de base commune et reposent sur les mêmes principes.

Plusieurs références exhaustives et détaillées existent sur le codage audio [Noll 93, Brandenburg 98, Painter 00]. Dans ce chapitre nous donnerons une description des principes et outils de base utilisés en codage des signaux audio en ayant comme perspective d'aborder les problèmes liés au traitement de tels signaux dans le domaine codé. Dans le même objectif, nous donnerons une description détaillée des codeurs MPEG-1 Layer I et II et du codeur TDAC de FTR&D qui ont servi comme base d'étude dans cette thèse.

2.2. Bases du codage audio fréquentiel perceptuel

Le principe de base du codage audio fréquentiel perceptuel consiste à réduire le débit de l'information en exploitant les propriétés du système d'audition humain. Les composantes non pertinentes du signal audio, i.e. qui ne contribuent pas au mécanisme des sensations de l'oreille, sont éliminées. Cette opération utilise le phénomène de masquage qui consiste en ce que certaines composantes du signal sonore soient inaudibles à cause de la présence d'autres. Comme la description de cet effet de masquage se fait principalement dans le domaine fréquentiel, la représentation du signal se réalise dans ce domaine. Plus concrètement, le schéma de base d'un système de codage perceptuel est présenté dans la figure 2.1.

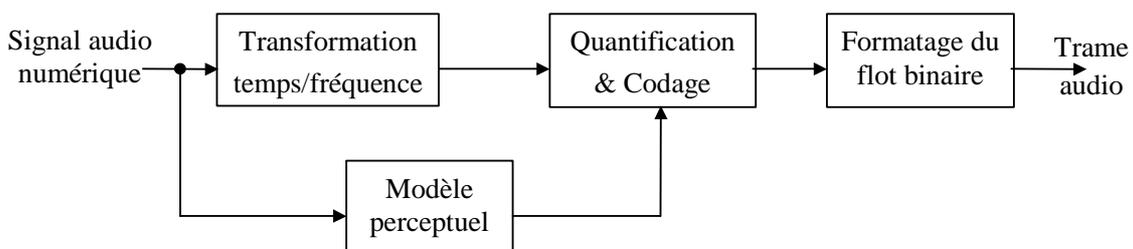


Fig. 2.1 : Schéma de principe d'un système de codage audio fréquentiel perceptuel.

Les blocs de ce schéma sont décrits comme suit :

- *Transformation temps/fréquence* : une représentation dans le domaine fréquentiel nécessite d'utiliser un banc de filtres ou une transformée pour décomposer le signal temporel d'entrée et trouver ses composantes spectrales. Cette transformation permet aussi de décorrélérer le signal et donc d'éliminer une certaine redondance statistique.
- *Modèle perceptuel* : pour estimer le masquage résultant du signal, un modèle du système d'audition humain est utilisé. Il se base sur l'analyse fréquentielle du banc de filtre d'analyse si la résolution de ce dernier est suffisante. Sinon, il utilise une transformée spécifique (DFT) appliquée directement au signal temporel d'entrée. Un modèle perceptuel parfait fournit une estimation du seuil de bruit maximal qui est à la limite de l'inaudibilité. Dans l'état de l'art actuel, les modèles existants sont loin de fournir de telles performances.
- *Quantification et codage* : pour une qualité transparente, la quantification doit se faire de manière à ce que le bruit en résultant soit en dessous des seuils de masquage. En réalité, vu le débit binaire disponible, elle est faite telle que ce bruit soit le moins perceptible possible. Cette quantification utilise les résultats d'une certaine procédure d'allocation de bits qui distribue le nombre de bits disponible aux composantes spectrales en se basant sur les paramètres de sortie du modèle psychoacoustique. Elle peut être aussi suivie par un processus de codage sans pertes (généralement codage d'Huffman) pour réduire encore plus le débit binaire.

Outre ces blocs classiquement trouvés dans tous les codeurs perceptuels, d'autres outils peuvent être utilisés. Par exemple, une prédiction adaptative peut être effectuée entre blocs successifs du signal pour réduire encore plus la redondance. Dans le cas du codage de signaux stéréo, la corrélation existante entre différentes voies est parfois exploitée dans le même but (M/S stéréo, stéréo d'intensité). Enfin, un module de mise en forme temporelle du bruit peut être utilisé pour contrôler le phénomène de pré-écho dû à l'étalement temporel du bruit.

2.3. Banc de filtres et transformée

Le premier étage de tout système de codage audio perceptuel consiste en un banc de filtres d'analyse utilisé pour la transformation temps/fréquence. Dans ce contexte, les principales caractéristiques que doit vérifier le banc de filtres consistent en la haute résolution fréquentielle et temporelle, la reconstruction parfaite, l'échantillonnage critique, la réduction des effets de bloc et la bonne séparation entre bandes. En plus, l'existence d'algorithmes rapides est une condition nécessaire pour permettre une implantation efficace. Une large variété de bancs de filtres et transformées ont été développés et exploités pour améliorer ces caractéristiques. On peut mentionner par exemple les bancs de filtres pseudo-QMF, bancs de filtres hybrides, transformée MDCT [Bosi 99]... Cette dernière s'avère actuellement la plus efficace dans ce contexte. Elle est à la base des algorithmes de codage audio les plus récents et performants : MPEG-4 AAC et TwinVQ, Dolby AC-3, TDAC de FTR&D, UIT-T G.722.1... [Shlien 97]. Bien que ces différentes transformations aient été développées séparément, elles peuvent être décrites par une approche mathématique unifiée et à partir de différents points de vue : bancs de filtres à cosinus modulé, transformées orthogonales à recouvrement (LOT : Lapped Orthogonal Transforms) et plus généralement par la théorie des bancs de filtres à décimation maximale [Vaidyanathan 93, Malvar 92].

2.3.1 Bancs de filtres à décimation maximale

2.3.1.1 Structure générale

Le schéma d'un banc de filtres uniforme à M bandes et à échantillonnage critique est donné par la figure 2.2. Il est défini par l'ensemble des filtres d'analyse, $H_k(z)$, et de synthèse, $F_k(z)$, $0 \leq k \leq M-1$. La sortie k du banc de filtres d'analyse, résultant d'un filtrage puis d'une décimation par un facteur M , est fourni par [Vaidyanathan 93] :

$$U_k(z) = \frac{1}{M} \sum_{l=0}^{M-1} H_k(z^{\frac{1}{M}} W^l) X(z^{\frac{1}{M}} W^l), \quad 0 \leq k \leq M-1, \quad (2.1)$$

avec $W = e^{-j\frac{2\pi}{M}}$.

Dans cette somme seule la composante donnée par $l=0$ correspond au signal utile sur la bande k . Les autres sont des versions translatées puis étalées du spectre du signal d'entrée filtré par $H_k(z)$ et résultent de l'opération de décimation. Ils représentent l'effet du repliement du spectre (ou d'aliasing) résultant de l'opération de décimation. En absence de tout traitement sur les composantes spectrales de sortie du banc de filtres d'analyse (quantification ou autre), on a $V_k(z) = U_k(z)$. L'interpolation puis le filtrage de chacun des signaux des sous-bandes par les filtres de synthèse donne ce qui suit :

$$\hat{X}_k(z) = V_k(z^M) F_k(z) = \frac{1}{M} \sum_{l=0}^{M-1} H_k(z W^l) X(z W^l) F_k(z), \quad 0 \leq k \leq M-1. \quad (2.2)$$

Après sommation :

$$\hat{X}(z) = \sum_{k=0}^{M-1} \hat{X}_k(z) = \frac{1}{M} \sum_{l=0}^{M-1} \left(\sum_{k=0}^{M-1} H_k(z W^l) F_k(z) \right) X(z W^l). \quad (2.3)$$

Cette formule indique que le spectre du signal reconstruit consiste en la combinaison du spectre original, $X(z)$, et de ses $(M-1)$ versions translatées, $X(z W^l)$, $1 \leq l \leq M-1$, qui représentent les termes d'aliasing. Cet effet de repliement spectral est complètement supprimé si :

$$\sum_{k=0}^{M-1} H_k(z W^l) F_k(z) = 0 \quad \text{pour } 1 \leq l \leq M-1. \quad (2.4)$$

Dans ce cas le système est linéaire invariant dans le temps, mais il existe encore des distorsions d'amplitude et de phase représentées par la fonction de transfert $T(z) = \sum_{k=0}^{M-1} H_k(z) F_k(z)$. La reconstruction parfaite est obtenue si cette fonction de transfert est réduite à un simple délai : $T(z) = cz^{-d}$.

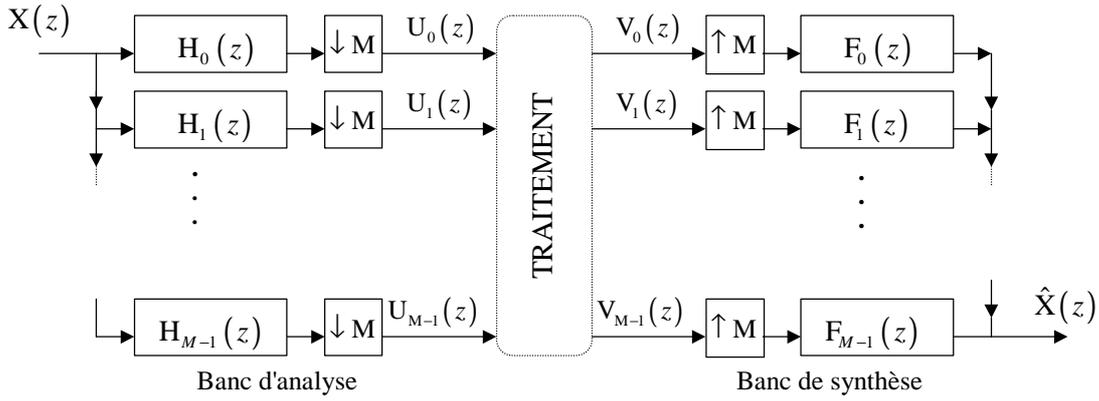


Fig. 2.2 : Banc de filtres à décimation maximale à M bandes.

2.3.1.2 Représentation polyphasée

La décomposition polyphasée de type 1 des filtres d'analyse, $H_k(z)$, et de type 2 des filtres de synthèse, $F_k(z)$, $0 \leq k \leq M-1$, s'expriment respectivement comme suit :

$$H_k(z) = \sum_{l=0}^{M-1} z^{-l} E_{kl}(z^M), \quad (2.5)$$

et

$$F_k(z) = \sum_{l=0}^{M-1} z^{-(M-1-l)} R_{lk}(z^M). \quad (2.6)$$

L'écriture matricielle de ces relations permet d'obtenir les expressions suivantes :

$$\mathbf{h}(z) = \mathbf{E}(z^M) \mathbf{e}(z), \quad (2.7)$$

$$\mathbf{f}^T(z) = z^{-(M-1)} \tilde{\mathbf{e}}(z) \mathbf{R}(z^M), \quad (2.8)$$

où $\mathbf{h}(z) = [H_0(z) \ H_1(z) \ \cdots \ H_{M-1}(z)]^T$ et $\mathbf{f}(z) = [F_0(z) \ F_1(z) \ \cdots \ F_{M-1}(z)]$ sont respectivement les vecteurs des filtres d'analyse et de synthèse, $\mathbf{e}(z) = [1 \ z^{-1} \ \cdots \ z^{-(M-1)}]^T$ est un vecteur chaîne de délai. $\mathbf{E}(z)$ est la matrice polyphasée d'analyse, elle s'écrit sous la forme suivante :

$$\mathbf{E}(z) = \begin{bmatrix} E_{00}(z) & E_{01}(z) & \cdots & E_{0,M-1}(z) \\ E_{10}(z) & E_{11}(z) & \cdots & E_{1,M-1}(z) \\ \vdots & \vdots & \ddots & \vdots \\ E_{M-1,0}(z) & E_{M-1,1}(z) & \cdots & E_{M-1,M-1}(z) \end{bmatrix}. \quad (2.9)$$

Chacune de ses lignes représente un filtre d'analyse. $\mathbf{R}(z) = [\mathbf{R}_{lk}(z)]_{0 \leq l, k \leq M-1}$ est la matrice polyphasée de synthèse. Elle est construite de la même manière avec la différence que chaque colonne représente un filtre de synthèse. La notation $\tilde{\mathbf{A}}(z)$, pour une matrice $\mathbf{A}(z)$ quelconque, correspond à $\mathbf{A}_*^T(z^{-1})$. L'indice $*$ revient à prendre le conjugué des coefficients de la matrice. Si $\mathbf{A}(z)$ est à coefficients réels alors $\tilde{\mathbf{A}}(z) = \mathbf{A}^T(z^{-1})$.

A partir des équations (2.7) et (2.8), et en utilisant les identités nobles pour faire migrer les opérations de décimation et de sur-échantillonnage à travers les matrices polyphasées, on peut obtenir la représentation en matrices polyphasées des bancs de filtres illustrée dans la figure 2.3.

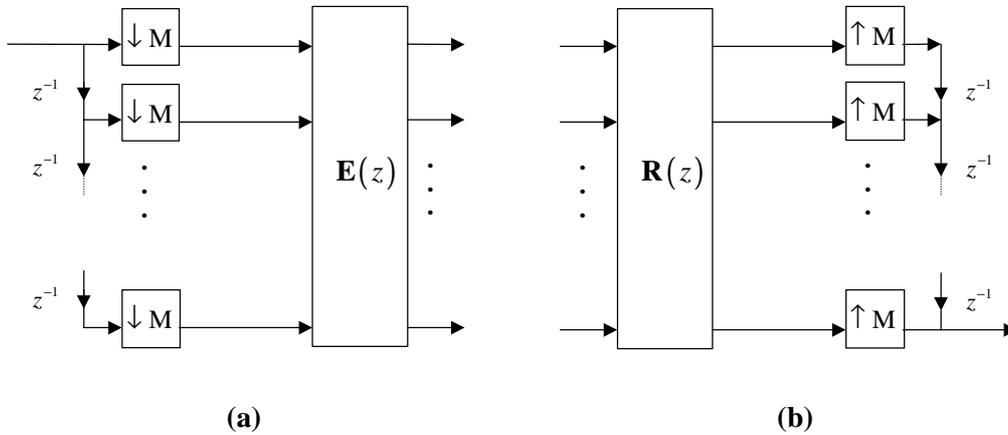


Fig. 2.3 : Représentation en matrices polyphasées (a) Banc de filtres d'analyse (b) Banc de filtres de synthèse.

2.3.2 Banc de filtres à reconstruction parfaite ; Banc de filtres paraunitaire

La mise en cascade des bancs de filtres de l'analyse et de la synthèse donne le produit des matrices polyphasées :

$$\mathbf{L}(z) = \mathbf{R}(z)\mathbf{E}(z). \quad (2.10)$$

Une condition suffisante pour obtenir la propriété de reconstruction parfaite consiste à choisir des matrices $\mathbf{E}(z)$ et $\mathbf{R}(z)$ qui vérifient la relation suivante [Vaidyanathan 93] :

$$\mathbf{R}(z)\mathbf{E}(z) = c z^{-\lambda} \mathbf{I}_M, \quad (2.11)$$

où λ est un entier positif, \mathbf{I}_M est la matrice identité de taille $M \times M$ et c une constante non nulle.

La condition nécessaire et suffisante de reconstruction parfaite s'écrit comme suit [Vaidyanathan 93] :

$$\mathbf{R}(z)\mathbf{E}(z) = c z^{-\lambda} \mathbf{\Gamma}_r(z), \quad (2.12)$$

où $\mathbf{\Gamma}_r(z) = \begin{bmatrix} \mathbf{0} & \mathbf{I}_{M-r} \\ z^{-1}\mathbf{I}_r & \mathbf{0} \end{bmatrix}$, r est un entier compris entre 0 et $M-1$.

La première relation (2.11) indique qu'il est possible de construire un banc de filtres à reconstruction parfaite, étant données les filtres d'analyse, par simple inversion de $\mathbf{E}(z)$. La recherche de solutions, dans le cas générale, par inversion de matrice peut générer des filtres de synthèse, $F_k(z)$, de réponse impulsionnelle infinie et peut être aussi instables. Une autre approche consiste à choisir la matrice polyphasée d'analyse, $\mathbf{E}(z)$, comme matrice de filtres FIR paraunitaire. Dans ce cas, le banc de filtres est dit aussi paraunitaire. La matrice $\mathbf{E}(z)$ est dite *paraunitaire* si elle vérifie :

$$\mathbf{E}(z)\tilde{\mathbf{E}}(z)=\mathbf{I}_M. \quad (2.13)$$

La matrice $\mathbf{E}(z)$ est dite *sans perte* (*lossless*) si elle est en plus stable.

Pour satisfaire à la condition de reconstruction parfaite (2.11), il suffit de choisir dans ce cas la matrice polyphasée de synthèse, $\mathbf{R}(z)$, telle que :

$$\mathbf{R}(z)=z^{-K}\tilde{\mathbf{E}}(z), \quad (2.14)$$

où K est un entier positif choisit de manière à assurer la causalité de la matrice $\mathbf{R}(z)$ et donc des filtres de synthèse, $F_k(z)$, $0 \leq k \leq M-1$.

La relation entre bancs de filtres FIR paraunitaire et à reconstruction parfaite est bien résumée par un théorème dû à Vaidyanathan [Vaidyanathan 93]. En effet, pour un banc de filtres à décimation maximale avec des filtres d'analyse FIR causaux, si deux des propriétés suivantes sont vraies alors la troisième est aussi vérifiée :

1. $\mathbf{E}(z)$ est sans perte (et donc paraunitaire).
2. Les réponses impulsionnelles des filtres de synthèse sont données par $f_k[n]=h_k^*[L-1-n]$, $0 \leq k \leq M-1$, avec L un entier positif et $c \neq 0$. Ce qui se traduit dans le domaine de la transformée en z par $F_k(z)=cz^{-(L-1)}\tilde{H}_k(z)$.
3. Le système est à reconstruction parfaite.

Notons qu'en tenant compte des relations (2.7) et (2.8), on peut montrer que la propriété 2 est équivalente au choix de matrices polyphasées vérifiant la relation (2.14), avec $L=(K+1)M$.

2.3.3 Bancs de filtres FIR à cosinus modulé

Ce type de banc de filtres est caractérisé par le fait que les filtres d'analyse et de synthèse s'obtiennent par une modulation en cosinus d'un filtre prototype passe-bas $H(z)$. Ils ont l'avantage d'avoir une complexité faible puisque l'implantation revient à un seul filtrage plus une modulation. Historiquement, les bancs de filtre pseudo-QMF ont été les premiers à être développés dans ce sens [Rothweiler 83]. Ces bancs de filtres ne vérifient pas exactement la reconstruction parfaite. Seul l'aliasing dû aux bandes adjacentes est supprimé, la distorsion de phase est complètement éliminée et l'élimination de la distorsion d'amplitude dépend de la bonne conception du filtre prototype par un procédé d'optimisation [Vaidyanathan 93]. Ces derniers peuvent être considérés comme un cas particulier des bancs de filtres à cosinus modulé et à reconstruction parfaite. L'expression des réponses impulsionnelles des filtres d'analyse et de synthèse pour ce type de bancs de filtres est donnée par :

$$h_k[l]=h[l]\cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)\left(l-\frac{L-1}{2}\right)-\theta_k\right], \quad 0 \leq k \leq M-1, \quad (2.15)$$

$$f_k[l]=h[l]\cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)\left(l-\frac{L-1}{2}\right)+\theta_k\right], \quad 0 \leq k \leq M-1, \quad (2.16)$$

avec $0 \leq l \leq L-1$ et $\theta_k = \frac{(2k+1)\pi}{4}$ et où $h[l]$ est la réponse impulsionnelle du filtre prototype de longueur L .

Pour un système issu des équations (2.15) et (2.16), la reconstruction parfaite est assurée si les conditions suivantes sont vérifiées [Vaidyanathan 93]:

1. $L = 2mM$, où m est un entier.
2. Les filtres de synthèse sont donnés par $f_k[n] = h_k[L-1-n]$.
3. Le filtre prototype est à phase linéaire : $h[n] = h[L-1-n]$.
4. Si $G_l(z)$, $0 \leq l \leq 2M-1$, sont les composantes polyphasées d'ordre $2M$ du filtre prototype $H(z)$ alors ils doivent vérifier la condition de complémentarité en puissance suivante :

$$\tilde{G}_k(z)G_k(z) + \tilde{G}_{M+k}(z)G_{M+k}(z) = \text{constante}, \quad 0 \leq k \leq M-1. \quad (2.17)$$

En utilisant la même décomposition polyphasée ci-dessus du filtre prototype on peut exprimer la matrice polyphasée d'analyse de ce type de banc de filtres sous la forme suivante [Koipillai 92] :

$$\mathbf{E}(z) = \hat{\mathbf{C}} \begin{bmatrix} \mathbf{g}_0(-z^2) \\ z^{-1} \mathbf{g}_1(-z^2) \end{bmatrix}, \quad (2.18)$$

où $\mathbf{g}_0(z)$ et $\mathbf{g}_1(z)$ sont des matrices diagonales définies comme suit :

$$\begin{cases} \mathbf{g}_0(z) = \mathbf{diag}[G_0(z) & G_1(z) & \cdots & G_{M-1}(z)], \\ \mathbf{g}_1(z) = \mathbf{diag}[G_M(z) & G_{M+1}(z) & \cdots & G_{2M-1}(z)], \end{cases} \quad (2.19)$$

et $\hat{\mathbf{C}} = [c_{kl}]$ la matrice de modulation en cosinus de taille $M \times 2M$. Ses éléments sont donnés par :

$$c_{kl} = \cos \left[\frac{\pi}{M} \left(k + \frac{1}{2} \right) \left(l - \frac{M}{2} \right) \right], \quad 0 \leq l \leq 2M-1 \text{ et } 0 \leq k \leq M-1. \quad (2.20)$$

En considérant la partition $\hat{\mathbf{C}} = [\hat{\mathbf{C}}_0 \quad \hat{\mathbf{C}}_1]$, où $\hat{\mathbf{C}}_0$ et $\hat{\mathbf{C}}_1$ sont des matrices de taille $M \times M$, on peut écrire $\mathbf{E}(z)$ sous une forme plus explicite :

$$\mathbf{E}(z) = \begin{bmatrix} c_{00}^0 G_0(-z^2) & c_{01}^0 G_1(-z^2) & \cdots & c_{0,M-1}^0 G_{M-1}(-z^2) \\ c_{10}^0 G_0(-z^2) & c_{11}^0 G_1(-z^2) & \cdots & c_{1,M-1}^0 G_{M-1}(-z^2) \\ \vdots & \vdots & \ddots & \vdots \\ c_{M-1,0}^0 G_0(-z^2) & c_{M-1,1}^0 G_1(-z^2) & \cdots & c_{M-1,M-1}^0 G_{M-1}(-z^2) \end{bmatrix} \quad (2.21)$$

$$+ z^{-1} \begin{bmatrix} c_{00}^1 G_M(-z^2) & c_{01}^1 G_{M+1}(-z^2) & \cdots & c_{0,M-1}^1 G_{2M-1}(-z^2) \\ c_{10}^1 G_M(-z^2) & c_{11}^1 G_{M+1}(-z^2) & \cdots & c_{1,M-1}^1 G_{2M-1}(-z^2) \\ \vdots & \vdots & \ddots & \vdots \\ c_{M-1,0}^1 G_M(-z^2) & c_{M-1,1}^1 G_{M+1}(-z^2) & \cdots & c_{M-1,M-1}^1 G_{2M-1}(-z^2) \end{bmatrix},$$

avec
$$G_k(-z^2) = \sum_{l=0}^{m-1} (-1)^l h(2lM+k) z^{-2l}, \quad 0 \leq k \leq 2M-1. \quad (2.22)$$

La matrice polyphasée $\mathbf{E}(z)$ est d'ordre $K = 2m - 1$. La condition 4 de reconstruction parfaite exprime en fait simplement la paraunitarité de cette matrice. Ce type de banc de filtres est donc paraunitaire et l'expression de la matrice polyphasée de synthèse, $\mathbf{R}(z)$, peut alors s'obtenir par :

$$\mathbf{R}(z) = z^{-(2m-1)} \tilde{\mathbf{E}}(z). \quad (2.23)$$

A partir de l'expression (2.18) de la matrice polyphasée d'analyse et de celle de synthèse déduite de (2.23), une représentation de ce type de banc de filtres peut être donnée par la figure 2.4.

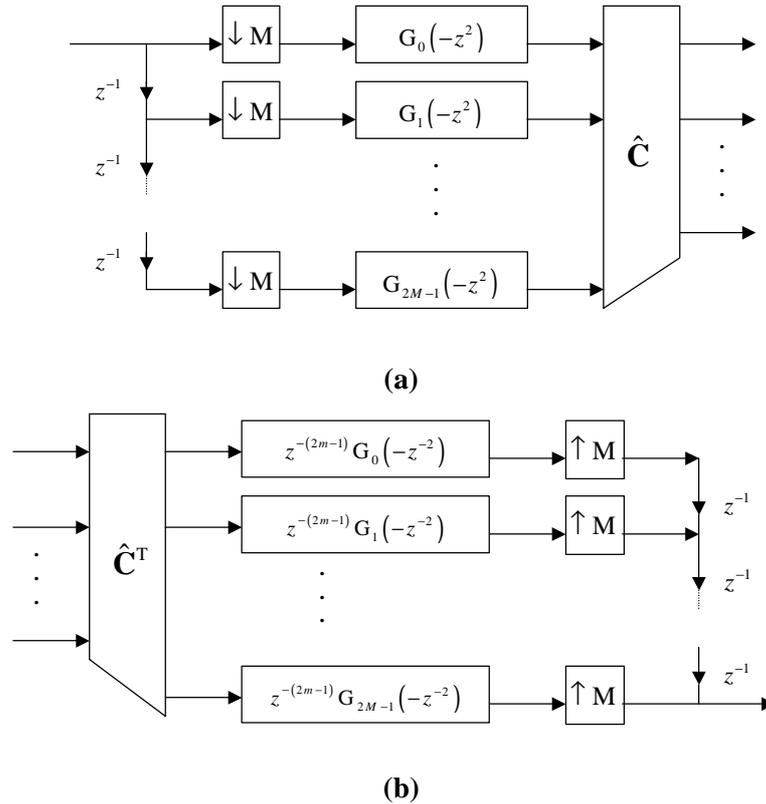


Fig. 2.4 : Représentation polyphasée d'un banc de filtre à cosinus modulé (a) Banc d'analyse (b) Banc de synthèse.

2.3.4 Transformée MDCT

Un cas particulier des bancs de filtres à cosinus modulé et à reconstruction parfaite correspond au banc de filtres TDAC (Time Domain Aliasing Cancellation) pour lequel $L = 2M$ et $m = 1$. Ce dernier a été développé pour la première fois dans [Princen 86]. Il peut être vu comme une transformée MLT (Modulated Lapped Transform) [Malvar 92] connu aussi sous le nom de MDCT (Modified DCT). La définition de la transformée MLT directe d'un signal $x[l]$ est donnée par :

$$y[k] = \sum_{l=0}^{2M-1} x[l] t_{k,l} \quad \text{pour } 0 \leq k \leq M-1, \quad (2.24)$$

avec
$$t_{k,l} = \sqrt{\frac{2}{M}} h[l] \cos \left[\frac{\pi}{M} \left(k + \frac{1}{2} \right) \left(l + \frac{M+1}{2} \right) \right], \quad 0 \leq k \leq M-1 \text{ et } 0 \leq l \leq 2M-1. \quad (2.25)$$

et $h[l]$ la fenêtre de pondération. Elle s'applique à un bloc de $2M$ échantillons avec recouvrement de M échantillons pour obtenir M coefficients de transformée. L'opération de synthèse des échantillons $x[l]$ utilise les coefficients, $y[k]$, de la trame courante et aussi ceux de la trame précédente, $y^p[k]$ (voir figure 2.5) :

$$x[l] = \sum_{k=0}^{M-1} [y[k]t_{k,l} + y^p[k]t_{k,l+M}]. \quad (2.26)$$

Pour assurer la reconstruction parfaite la fenêtre $h[l]$ doit vérifier la condition de symétrie, $h[l] = h[2M - 1 - l]$, et de complémentarité en puissance, $h^2[l] + h^2[l + M] = 1$. Cette dernière traduit la relation (2.17) pour ce cas particulier. Un choix simple pour le filtre prototype vérifiant ces conditions est donné par la fenêtre sinusoidale suivante :

$$h[l] = \sin \left[\left(l + \frac{1}{2} \right) \frac{\pi}{2M} \right], \quad 0 \leq l \leq 2M - 1. \quad (2.27)$$

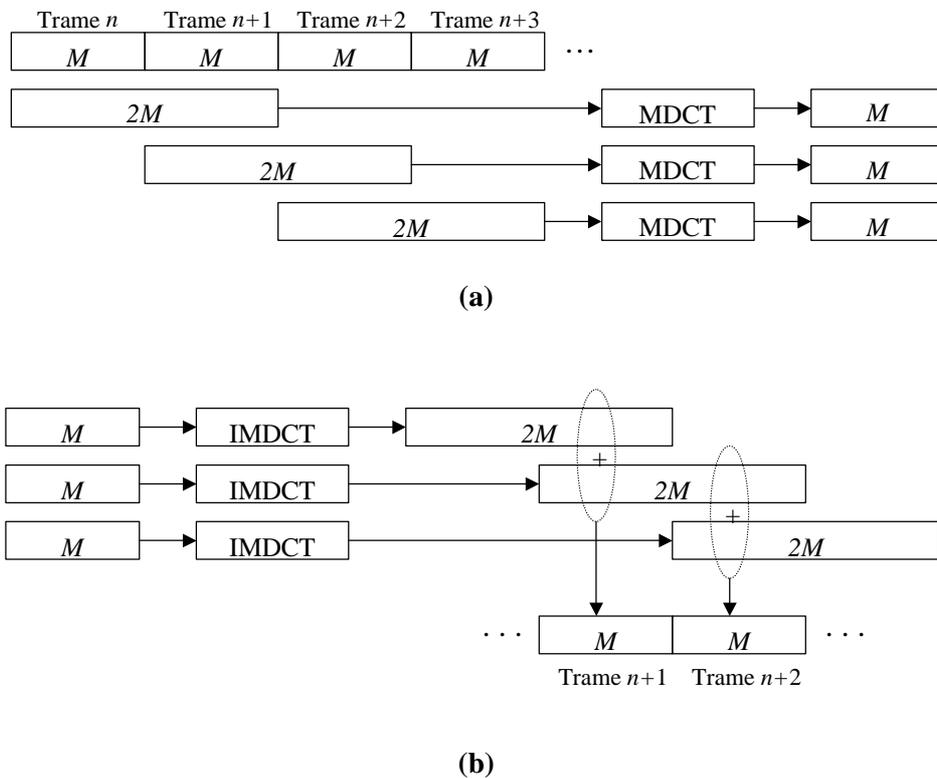


Fig. 2.5 : Transformée MDCT (a) Analyse : transformée directe à recouvrement (b) Synthèse : transformée inverse et addition avec recouvrement (overlap and add).

L'expression (2.24) donnant la transformée MLT directe peut s'écrire sous la forme matricielle suivante :

$$\mathbf{y}[n] = \mathbf{T}\mathbf{x}[n], \quad (2.28)$$

où $\mathbf{T} = [t_{k,l}]$ est une matrice de taille $M \times 2M$.

Le vecteur $\mathbf{x}[n] = [x[nM - 2M + 1] \ x[nM - 2M + 2] \ \dots \ x[nM] \ x[nM]]^T$ correspond à une partition du signal en blocs de taille $2M$ avec recouvrement de M échantillons, n étant l'indice du bloc. $\mathbf{y}[n]$ est le vecteur des coefficients transformée de taille M . Cette forme montre que la transformée MLT fait partie de la famille des transformées orthogonales à recouvrement (Lapped Orthogonal Transform : LOT). Dans le cas général des transformées LOT, les blocs du signal considérés sont de longueur $L = 2mM$, où m est le facteur de recouvrement ($m \geq 1$), avec un recouvrement de $(L - M)$ échantillons et la matrice \mathbf{T} est de taille $M \times L$.

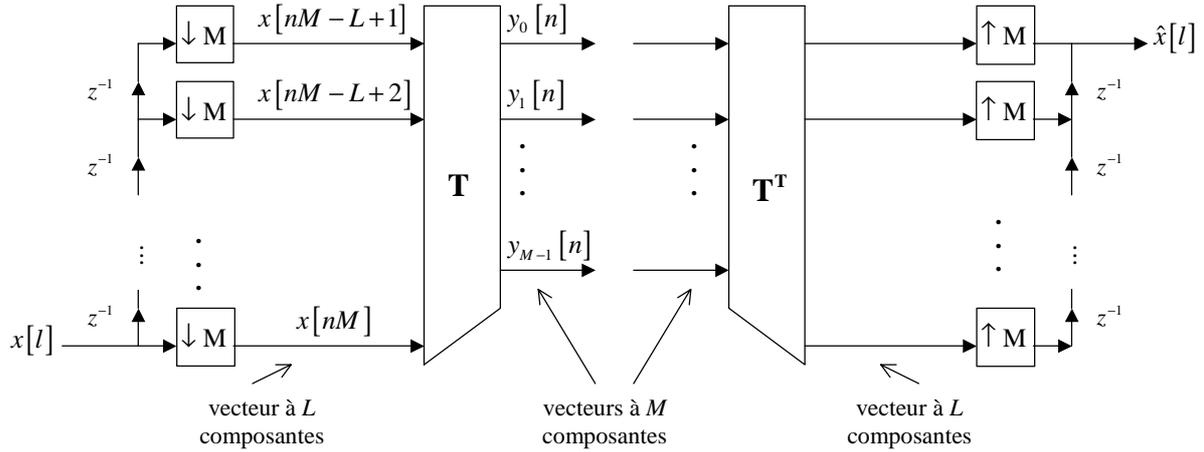


Fig. 2.6 : Schéma bloc d'une transformée orthogonale à recouvrement (LOT).

Le schéma de la figure 2.6 permet de déduire les filtres du banc d'analyse équivalent à une transformée LOT :

$$\mathbf{h}(z) = \mathbf{T}\mathbf{e}_1(z), \quad (2.29)$$

$$\text{avec } \mathbf{e}_1(z) = [z^{-(L-1)} \ \dots \ z^{-1} \ 1]^T.$$

Pour la MLT, $L = 2M$, on considère la partition suivante de la matrice \mathbf{T} :

$$\mathbf{T} = [\mathbf{T}_0 \ \mathbf{T}_1], \quad (2.30)$$

où \mathbf{T}_0 et \mathbf{T}_1 sont des matrices $M \times M$. Et on écrit $\mathbf{e}_1(z)$ sous la forme suivante :

$$\mathbf{e}_1(z) = \begin{bmatrix} z^{-M} \mathbf{J}_M \\ \mathbf{J}_M \end{bmatrix} \mathbf{e}(z), \quad (2.31)$$

avec $\mathbf{J}_M = \begin{pmatrix} 0 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 0 \end{pmatrix}$ la matrice anti-identité de taille $M \times M$, $\mathbf{e}(z)$ le vecteur chaîne de délai défini

dans le paragraphe 1.3.1.2. Notons qu'une post-multiplication d'une matrice par \mathbf{J}_M correspond à inverser l'ordre de ses colonnes, et une pré-multiplication par \mathbf{J}_M induit une inversion de ses lignes.

Ceci permet d'obtenir l'expression suivante à partir de (2.29) :

$$\mathbf{h}(z) = [\mathbf{T}_0 \mathbf{J}_M z^{-M} + \mathbf{T}_1 \mathbf{J}_M] \mathbf{e}(z). \quad (2.32)$$

En comparant (2.32) à la formule (2.7), on en déduit l'expression de la matrice polyphasée d'analyse :

$$\mathbf{E}(z) = \mathbf{T}_1 \mathbf{J}_M + \mathbf{T}_0 \mathbf{J}_M z^{-1}. \quad (2.33)$$

Pour trouver l'expression de la matrice polyphasée de synthèse, on utilise la propriété de paraunitarité du banc de filtres TDAC. En effet, l'écriture de (2.14) pour $K = 1$ donne :

$$\mathbf{R}(z) = z^{-1} \tilde{\mathbf{E}}(z). \quad (2.34)$$

Puis, remarquant que $\mathbf{J}_M^T = \mathbf{J}_M$, on obtient :

$$\mathbf{R}(z) = \mathbf{J}_M \mathbf{T}_0^T + \mathbf{J}_M \mathbf{T}_1^T z^{-1}. \quad (2.35)$$

2.4. Éléments sur la psychoacoustique

La psychoacoustique a pour objet l'étude expérimentale des relations entre les stimuli acoustiques et les réponses du système d'audition humain : sensation et perceptions auditives [Zwicker 99]. Nous détaillerons dans la suite quelques notions fondamentales qui sont couramment utilisées dans le codage audio perceptuel.

2.4.1 Phénomène de masquage

On sait bien que l'oreille n'est capable de détecter un son pur que si son niveau de pression acoustique dépasse un certain seuil. Ce seuil dépend de la fréquence et la courbe exprimant cette relation s'appelle courbe de *seuil d'audition absolu*.

Lors de l'écoute de deux sons purs de fréquences suffisamment proches cette courbe augmente, il apparaît un effet de masque : le son d'intensité plus forte rend totalement ou partiellement inaudible celui qui est plus faible. On parle alors de *masquage fréquentiel* (ou simultané). Pour un son masquant de fréquence f_0 et de puissance P_0 , on peut déterminer la courbe de masquage qui lui est relative en mesurant pour toutes les fréquences f de la bande audible [20 Hz, 20 kHz] le niveau de pression acoustique limite, d'un son à cette fréquence, pour qu'il soit masqué. Cette courbe dépend de la fréquence f_0 et varie selon la nature spectrale des sons masquant et masqué : modélisables par une sinusoïde ou plutôt par un bruit à bande étroite. Dans les différents cas de figure, ces courbes ont quasiment la même allure. Ainsi on montre que si on exprime la fréquence suivant une nouvelle échelle appelée échelle des Bark, ces courbes peuvent être modélisées par une fonction de forme triangulaire. Cette dernière s'appelle *fonction d'étalement de la membrane basilaire*. Elle présente deux caractéristiques importantes : la pente du côté des fréquences supérieures est moins aiguë que du côté des fréquences inférieures et elle dépend aussi du niveau du son masquant (elle est d'autant plus raide que celui-ci est faible). La valeur, $S_m(f_0)$, de la courbe de masquage en f_0 est inférieure à P_0 , la différence $P_0 - S_m(f_0)$ est appelée *indice de masquage*.

Un autre type de masquage apparaît pour les sons non simultanés. L'expérience montre qu'un bruit de durée assez grande peut masquer une impulsion, qui le succède ou le précède, de durée plus faible. Le premier phénomène s'appelle *masquage temporel postérieur*, il traduit un effet de rémanence de l'oreille. Le second cas est le *masquage temporel antérieur*, il s'explique par le fait que le temps de

réponse mis par l'oreille pour réagir à un son dépend de l'intensité de ce dernier : un son fort peut donc générer une sensation plus rapidement qu'un son plus faible qui le précède. Le masquage antérieur est moins fort, il n'existe effectivement que si l'intervalle entre les deux sons est de 5 à 20 ms.

2.4.2 Bandes critiques

La notion de bandes critiques est liée à la manière dont l'oreille réalise la perception dans le domaine fréquentiel. Ainsi, les expériences ont montré que le système auditif se comporte comme un banc de filtres qui se chevauchent et dont les fréquences centrales s'échelonnent continûment. Il est alors modélisé par 25 filtres recouvrant la bande audible et dont la largeur de chacun correspond à une bande critique.

Cette notion de bande critique peut être illustrée en considérant le niveau de puissance juste audible pour deux sons purs de fréquence différente et de même puissance. En effet, ce niveau reste le même pourvu que l'écart en fréquence ne dépasse pas une certaine limite qui est la largeur d'une bande critique, dès que cette limite est franchie ce niveau doit être doublé : la puissance perçue par l'oreille correspond donc à la somme des puissances sur chaque bande critique. Elle apparaît aussi dans le masquage fréquentiel, ainsi lorsqu'un son pur de fréquence f_0 est masqué par un bruit à large bande centré sur la même fréquence seules les composantes du bruit contenues dans la bande critique centré sur f_0 sont effectivement masquantes et donc seule leur puissance intervient dans le calcul de la courbe de masquage [Botte 90].

La largeur des bandes critiques augmente avec la fréquence. Elle est égale à 100 Hz pour les fréquences centrales inférieures à 500 Hz, au-delà, elle vaut environ 20 % de la fréquence centrale considérée. L'échelle de Bark introduite précédemment correspond tout simplement au numéro de la bande critique.

2.5. Allocation des bits

Le principe du codage fréquentiel ou par transformée consiste à quantifier les composantes spectrales du signal de manière à minimiser l'erreur de reconstruction selon un certain critère. L'allocation des bits joue un rôle important dans ce schéma de codage puisqu'elle a pour rôle de répartir le débit binaire disponible à ces composantes spectrales. Comme le spectre du signal évolue constamment au cours du temps, l'allocation de bits est faite de manière dynamique. Elle est recalculée à chaque nouveau bloc du signal. Lorsque le nombre de composantes est faible (cas des codeurs en sous-bandes), l'allocation des bits est calculée pour chaque sous-bande. Dans le cas des codages par transformée, les coefficients sont regroupés par bandes de fréquence et l'on calcule un nombre de bits pour chacune des bandes.

Un premier critère d'allocation des bits consiste à minimiser l'erreur quadratique moyenne du signal reconstruit (LMSE : Least Mean Square Error). Dans ce cas le bruit de quantification qui en résulte est constant pour toutes les composantes du spectre (i.e. bruit blanc) [Jayant 84]. En contexte du codage perceptuel le but est de masquer totalement ce bruit ou du moins le rendre le moins audible possible. Ce critère ne permet pas donc d'atteindre cet objectif puisque pour certaines fréquences le bruit peut être largement inférieur à la courbe de masquage et être bien au-dessus de cette même courbe pour d'autres.

Le critère perceptuel intègre la connaissance a priori de l'effet de masquage du signal à coder pour minimiser l'erreur de reconstruction perceptible. L'allocation de bits doit réaliser une certaine mise en forme spectrale du bruit de quantification de manière à ce qu'il soit en dessous de la courbe de masquage. Pour cela le rapport bruit à masque, NMR , (Noise-to-Mask-Ratio) est défini pour chaque sous-bande :

$$NMR(j) = \frac{\sigma_Q^2(j)}{\tau(j)}, \quad (2.36)$$

où $\sigma_Q^2(j)$ et $\tau(j)$ sont, respectivement, la puissance de l'erreur de quantification et le seuil de masquage pour la sous-bande j . Un masquage total du bruit de quantification signifie que cette quantité, exprimée en dB, est négative pour toutes les sous-bandes. Malheureusement, la limitation en nombre de bits disponibles ne permet pas toujours d'assurer cet objectif. Elle peut être donc positive et représente dans ce cas la dégradation perceptuelle sur la sous-bande.

Pour rendre cette distorsion résiduelle minimale, l'allocation de bits suivant un critère perceptuel est exprimée comme un problème d'optimisation sous contrainte et à valeurs entières. Il consiste à minimiser la fonction de distorsion totale donnée par la somme des rapports bruit à masque (NMR total) connaissant le nombre de bits disponibles, B , et l'ensemble des quantificateurs à utiliser [Wei 97] :

$$\begin{aligned} &\text{Trouver le vecteur d'allocation de bits } \mathbf{b} = (b(j))_j \text{ qui minimise } \sum_{j=0}^{M-1} NMR(j) \\ &\text{sous la contrainte } \sum_{j=0}^{M-1} b(j) = B \text{ et } b(j) \in \mathcal{Q}_j \text{ pour } 0 \leq j \leq M-1, \end{aligned}$$

où M est le nombre de sous-bandes et \mathcal{Q}_j est l'ensemble des nombres de bits correspondant aux différents quantificateurs possibles pour la sous-bande j .

Considérons le cas où les quantificateurs utilisés sont scalaires à haute résolution (nombre élevé de niveaux de quantification) et l'ensemble des niveaux de quantification disponibles pour un nombre de bits sont autorisés. Sous cette hypothèse, la puissance du bruit de quantification dans la sous-bande j , quantifiée par $b(j)$ bits, est donnée par [Jayant 84] :

$$\sigma_Q^2(b(j)) = \varepsilon^2 2^{-2b(j)} \sigma_x^2(j), \quad 0 \leq j \leq M-1, \quad (2.37)$$

où $\sigma_x^2(j)$ est la puissance du signal dans la sous-bande j , la constante ε^2 étant le facteur de performance du quantificateur.

En utilisant la méthode des multiplieurs de Lagrange [Jayant 84] on obtient la solution optimale au problème précédent, elle s'exprime comme suit [Wei 97] :

$$b_{opt}(j) = \frac{1}{2} \log_2 \left[\frac{\sigma_x^2(j)}{\tau(j)} \right] + C = \frac{SMR(j)(dB)}{6,02} + C, \quad 0 \leq j \leq M-1, \quad (2.38)$$

où $SMR(j)$ est le rapport signal à masque exprimé en dB et $C = \frac{B}{M} - \frac{1}{2M} \sum_{l=0}^{M-1} \log_2 \left[\frac{\sigma_x^2(l)}{\tau(l)} \right]$ est une constante pour toutes les sous-bandes.

A partir des équations (2.37) et (2.38) on peut en déduire qu'à cet optimum la puissance du bruit de quantification s'exprime par la formule suivante :

$$\sigma_Q^2(b_{opt}(j)) = \tau(j) \varepsilon^2 2^{-2C}, \quad 0 \leq j \leq M-1. \quad (2.39)$$

Le rapport bruit à masque, NMR , est donc constant sur toutes les sous-bandes. Ce rapport représente aussi la puissance du bruit de quantification pondérée par le seuil de masquage. Le bruit a donc bien subi une mise en forme spectrale de manière à ce que son spectre soit parallèle à la courbe de masquage.

Les nombres de bits, $b_{opt}(j)$, obtenus par la formule (2.38) sont des nombres réels et pas forcément des entiers positifs. Cette solution n'est pas donc utilisable en pratique. Par ailleurs, elle permet de caractériser le point de convergence que doit approcher toute procédure d'allocation de bits optimale. Des solutions pratiques à ce problème développées dans les codeurs concernés par cette thèse seront exposées dans la suite.

2.6. Quantification

La quantification est la seconde étape dans tout schéma de codage audio permettant la réduction du débit. C'est un processus irréversible, il constitue la principale source de distorsion affectant le signal reconstruit. Lorsqu'on parle de quantification, on distingue toujours entre quantification scalaire qui agit sur des variables scalaires du signal et quantification vectorielle agissant sur des vecteurs de paramètres du signal. Dans les codeurs audio, c'est la quantification scalaire qui est la plus utilisée pour les paramètres résultant du banc de filtres (MPEG-1, MPEG-2/4 AAC, Dolby AC-3,...). Ceci surtout à cause de sa faible complexité. Par ailleurs, il existe certains codeurs qui utilisent la quantification vectorielle. C'est le cas par exemple du codeur MPEG-4 TwinVQ et du codeur TDAC de FTR&D. Puisque nous allons nous intéresser dans la suite à ce dernier codeur, nous nous focaliserons sur le type de quantification qu'il utilise à savoir la quantification vectorielle sphérique à unions de codes à permutation. Ce paragraphe commence par donner un bref rappel sur la quantification vectorielle avant de décrire les principes et caractéristiques de cette dernière.

2.6.1 Quantification vectorielle

De manière générique un quantificateur peut être défini comme une application Q de \mathbb{R}^N dans un sous ensemble fini $\mathcal{C} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_L\}$, appelé dictionnaire (codebook), qui affecte à chaque vecteur d'entrée \mathbf{x} une image \mathbf{y}_i dépendant de la région de l'espace où il est contenu. Un quantificateur est donc entièrement défini par la partition de l'espace \mathbb{R}^N en régions, \mathcal{R}_i , et les vecteurs, \mathbf{y}_i , $1 \leq i \leq L$, qui lui sont associées [Moreau 95] :

$$\text{Pour tout } \mathbf{x} \in \mathbb{R}^N, \quad Q(\mathbf{x}) = \mathbf{y}_i, \text{ si et seulement si } \mathbf{x} \in \mathcal{R}_i. \quad (2.40)$$

Cette définition correspond au cas général de la quantification vectorielle. Quand la dimension de l'espace est égale 1, la quantification est scalaire. Il s'agit alors de quantifier les variables scalaires du signal indépendamment. Le regroupement de ces variables en vecteurs, en cas de quantification vectorielle, n'est pas une simple généralisation. Il permet de tirer parti de la corrélation qui existe souvent entre les composantes du vecteur.

Le but de toute opération de quantification est de réduire la distorsion qui en résulte étant donné le débit binaire. Le débit dans ce cas est le nombre de bits nécessaire pour coder les indices des vecteurs

du dictionnaire, soit $\log_2(L)$ avec L la taille du dictionnaire. Une mesure de distorsion, $d(\mathbf{x}, \mathbf{y})$, doit être donc associée au quantificateur. Un choix possible consiste à prendre la distance euclidienne :

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2. \quad (2.41)$$

Un critère d'appréciation des performances du quantificateur peut être la distorsion moyenne définie comme suit :

$$D = E[d(\mathbf{x}, Q(\mathbf{x}))] = \int_{\mathbb{R}^N} d(\mathbf{x}, Q(\mathbf{x})) p(\mathbf{x}) d\mathbf{x}, \quad (2.42)$$

où $p(\mathbf{x})$ est la densité de probabilité des vecteurs d'entrée. Dans le cas de l'utilisation d'une distance euclidienne, cette distorsion moyenne correspond à l'erreur quadratique moyenne.

Le quantificateur optimal est celui qui assure le minimum de cette distorsion moyenne par la définition de son dictionnaire et des régions associées à chacun de ses vecteurs. Pour obtenir un tel quantificateur, il y a deux conditions nécessaires à satisfaire :

- Etant donné les vecteurs du dictionnaire \mathcal{C} , la partition de l'espace \mathbb{R}^N doit être obtenue par la règle du plus proche voisin :

$$Q(\mathbf{x}) = \mathbf{y}_i \text{ si et seulement si } d(\mathbf{x}, \mathbf{y}_i) \leq d(\mathbf{x}, \mathbf{y}_j) \text{ pour tout } j \neq i, 1 \leq j \leq L. \quad (2.43)$$

Les régions qui en résultent s'appellent régions de Voronoï.

- Connaissant la partition, chaque vecteur \mathbf{y}_i du dictionnaire doit être choisi de manière à minimiser la distorsion moyenne dans la région \mathcal{R}_i . Dans le cas où une distance euclidienne est utilisée, ce choix correspond au centroïde (ou centre de gravité) de la région.

Etant donnée la formule de distorsion, on voit que la construction d'un quantificateur vectoriel (et aussi scalaire) optimal dépend bien de la densité de probabilité du signal. Généralement, la connaissance de cette dernière n'est pas disponible. Des algorithmes itératifs sont plutôt utilisés, par exemple celui dit des K-moyennes, pour construire le dictionnaire par apprentissage sur un corpus. Cette quantification est alors dite statistique puisqu'elle rend bien compte de la distribution du signal.

Si ce type de quantificateurs vectoriels offre l'avantage de pouvoir approcher le quantificateur optimal, il montre par contre certains inconvénients dès qu'il s'agit d'utiliser des dictionnaires de grande taille pour mieux réduire la distorsion. En effet, dans ce cas il y a nécessité d'une capacité mémoire importante pour le stockage du dictionnaire. En plus, la recherche du vecteur quantifié d'un vecteur d'entrée par la méthode du plus proche voisin devient très complexe vu le grand nombre de vecteurs du dictionnaire.

Comme autre approche qui offre des solutions à ce problème, la quantification vectorielle algébrique (QVA) utilise des dictionnaires fortement structurés qui sont par exemple des sous-ensembles d'un réseau régulier. Un réseau régulier, dans l'espace \mathbb{R}^N , étant l'ensemble de points qui s'obtiennent par une combinaison linéaire de N vecteurs de base indépendants avec des coefficients de proportionnalité entiers [Adoul 86]. Les vecteurs du dictionnaire sont alors répartis de manière régulière dans l'espace et peuvent être générés de manière algébrique. Il n'y a pas donc nécessité de stocker le dictionnaire et aucune procédure d'apprentissage n'est requise pour sa construction. Le dictionnaire étant dans ce cas virtuel, sa taille peut être très grande. Les régions de Voronoï de ce type de quantificateur sont toutes identiques, leur forme dépend du réseau utilisé. Un quantificateur

algébrique est donc indépendant des propriétés statistiques du signal. Il ne peut être optimal que pour des signaux de distributions particulières.

2.6.2 Quantification vectorielle sphérique

La quantification vectorielle sphérique (QVS) consiste à quantifier séparément la norme et l'orientation d'un vecteur quelconque de \mathbb{R}^N . Cette séparation permet de choisir, pour la norme, un type d'échelle et aussi un type de codage particulier. En effet, cette dernière peut présenter certaines caractéristiques particulières que l'on peut exploiter. Ainsi par exemple un codage prédictif peut être utilisé pour éliminer la redondance avec une échelle logarithmique pour comprimer la dynamique. De plus le traitement séparé de la norme permet de réduire la dimension du problème (d'un volume à une surface dans \mathbb{R}^N) et donc de réduire la taille du dictionnaire [Adoul 86].

La quantification de l'orientation revient à quantifier le vecteur normé \mathbf{x} . Le dictionnaire d'orientations est constitué de vecteurs de longueur unité. Pour trouver le vecteur \mathbf{y}_i plus proche voisin, il faut minimiser la distance $d(\mathbf{x}, \mathbf{y}_i)$. Si la distance utilisée est euclidienne alors elle s'exprime dans ce cas comme suit :

$$\begin{aligned} d(\mathbf{x}, \mathbf{y}_i) &= \|\mathbf{x} - \mathbf{y}_i\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}_i\|^2 - 2\mathbf{x}^T \mathbf{y}_i, \\ &= 2 - 2\mathbf{x}^T \mathbf{y}_i. \end{aligned} \quad (2.44)$$

La recherche du plus proche voisin revient donc dans ce cas à trouver le vecteur \mathbf{y}_i qui maximise le produit scalaire $\mathbf{x}^T \mathbf{y}_i$.

2.6.3 Quantification vectorielle par codes à permutation

Un dictionnaire à code à permutation est généré à partir d'un seul vecteur \mathbf{y}_1 . Les autres éléments \mathbf{y}_i , $2 \leq i \leq L$, sont l'ensemble des vecteurs distincts qui peuvent être obtenus par toutes les permutations possibles sur les composantes du vecteur \mathbf{y}_1 . Dans cette première variante aucun changement de signe sur les composantes de \mathbf{y}_1 n'est effectué. Dans la seconde variante des codes à permutation, le vecteur \mathbf{y}_1 est choisi à composantes non négatives. Les autres éléments du dictionnaire sont formés d'abord en effectuant toutes les combinaisons possibles de changement de signes sur les composantes de \mathbf{y}_1 , puis en réalisant tous les réarrangements possibles sur ces composantes signées et de manière à avoir des vecteurs distincts [Berger 72].

Cette notion se généralise en *dictionnaire à union de codes à permutation* qui est simplement l'union de plusieurs dictionnaires à code à permutation. Considérons \mathcal{C} un dictionnaire de ce type. La relation entre les vecteurs de ce dictionnaire permet de définir une relation d'équivalence sur l'ensemble \mathcal{C} . Nous dirons que les vecteurs \mathbf{y}_i et \mathbf{y}_j de \mathcal{C} sont équivalents, $\mathbf{y}_i \sim \mathbf{y}_j$, si l'un peut se déduire de l'autre par une certaine permutation de ses composantes. On peut faire alors une partition du dictionnaire \mathcal{C} en classes d'équivalence. En fait, chaque classe représente un code à permutation. On définit en plus une relation d'ordre lexicographique dans l'ensemble \mathcal{C} . Cette relation d'ordre est totale sur chaque classe d'équivalence, elle permet d'ordonner tous ses éléments. L'élément maximal de la classe est appelé *leader*. Il s'obtient par simple ordonnancement décroissant des composantes de n'importe quel vecteur de la classe. Finalement, chaque classe d'équivalence peut être représentée par son leader et le stockage du dictionnaire peut donc se réduire au dictionnaire des leaders.

En plus de cet avantage de réduction de la mémoire de stockage, les dictionnaires à union de codes à permutation permettent d'utiliser des algorithmes rapides pour la recherche du vecteur quantifié plus proche voisin dans \mathcal{C} . En effet, pour un vecteur \mathbf{x} quelconque de \mathbb{R}^N , cette recherche n'a pas besoin de se faire de manière exhaustive sur l'ensemble des vecteurs du dictionnaire \mathcal{C} . Elle peut se restreindre au dictionnaire des leaders. Ainsi, il a été démontré [Berger 72] que si le vecteur $\tilde{\mathbf{y}}_{j_0}$, leader de la classe j_0 , est le plus proche voisin du vecteur $\tilde{\mathbf{x}}$, leader³ de \mathbf{x} dans \mathbb{R}^N , alors le plus proche voisin de \mathbf{x} est le vecteur \mathbf{y} obtenu par la permutation inverse de celle qui permet de passer de \mathbf{x} à $\tilde{\mathbf{x}}$. La figure 2.7 illustre cette méthode de détermination du vecteur quantifié \mathbf{y} .

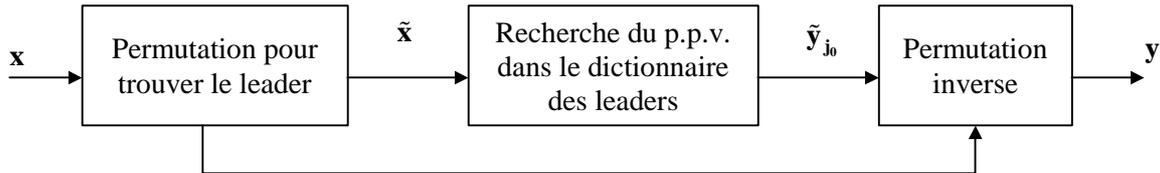


Fig. 2.7 : Méthode de détermination du plus proche voisin dans un dictionnaire à union de codes à permutation.

En quantification vectorielle c'est l'indice du vecteur quantifié, \mathbf{y} , dans le dictionnaire qui est transmis dans le mot de code. Dans ce cas de dictionnaire à union de codes à permutation, les éléments du dictionnaire des leaders peuvent être ordonnés dans l'ordre lexicographique décroissant. Ce classement induit une numérotation naturelle des classes (ou codes à permutation) du dictionnaire \mathcal{C} . L'indice j d'une classe, $j=1,2,\dots,P$, avec P le nombre de classes, correspond au numéro de son leader dans ce classement. Il est possible aussi d'utiliser un autre ordre pour les leaders, par exemple tout simplement leur indice dans une table de stockage quelconque.

Si j_0 est l'index de $\tilde{\mathbf{y}}_{j_0}$, il reste à trouver l'index lexicographique, appelé aussi *rang*, r , de \mathbf{y} dans sa classe j_0 . La détermination du rang est loin d'être une opération triviale. Elle s'effectue par la formule de Schalkwijk connaissant la permutation qui permet de passer de \mathbf{y} à $\tilde{\mathbf{y}}_{j_0}$ et par laquelle on obtient aussi $\tilde{\mathbf{x}}$ à partir de \mathbf{x} . L'expression ainsi que l'explication de cette formule sont données dans l'annexe B.1.

Finalement l'indice, t , du vecteur \mathbf{y} dans le dictionnaire \mathcal{C} peut être donné par :

$$t = H(j_0) + r(\mathbf{y}), \quad (2.45)$$

où $H(j_0) = \sum_{j=1}^{j_0-1} \kappa(j)$, $\kappa(j)$ étant le cardinal de la classe j . Cette quantité représente le cumul des cardinalités des classes précédentes $j=1,\dots,j_0-1$.

En fait, la relation d'équivalence définie précédemment correspond à la première variante des codes à permutation. Pour faire une généralisation à la deuxième variante, on définit une nouvelle relation d'équivalence : $\mathbf{y}_i \sim \mathbf{y}_j$ si et seulement si les vecteurs des valeurs absolues des composantes de \mathbf{y}_i et \mathbf{y}_j peuvent s'obtenir l'un à partir de l'autre par une certaine permutation. L'élément maximal, au sens lexicographique, de ces nouvelles classes d'équivalence est dit *leader absolu*. Il est obtenu en

³ En munissant l'ensemble \mathbb{R}^N de la relation d'équivalence \sim défini précédemment et de l'ordre lexicographique on peut étendre la notion de leader à tout vecteur \mathbf{x} de \mathbb{R}^N .

arrangeant dans l'ordre décroissant les valeurs absolues des composantes d'un vecteur de la classe. Notons que ces nouvelles classes sont des "sur-classes" de celles définies précédemment dans le sens où elles les contiennent. L'algorithme rapide de recherche du plus proche voisin dans \mathcal{C} , d'un vecteur \mathbf{x} de \mathbb{R}^N , s'effectue maintenant sur les leaders absolus.

2.6.4 Algorithmes de codage et de décodage pour la QVS à codes à permutation

Dans ce paragraphe nous résumons les algorithmes de codage et de décodage dans le cas d'une quantification vectorielle sphérique utilisant un dictionnaire à union de codes à permutation pour quantifier l'orientation des vecteurs d'entrée.

2.6.4.1 Algorithme de codage

Soit \mathbf{x} un vecteur normé de \mathbb{R}^N qu'on cherche à quantifier. L'algorithme de codage se décompose en cinq étapes : les trois premières permettent la recherche de l'orientation, \mathbf{y} , le plus proche voisin de \mathbf{x} et les deux dernières pour la recherche de l'index de \mathbf{y} dans sa classe.

Etape 1. Passage au vecteur des valeurs absolues, $|\mathbf{x}| = [|x_1| \ |x_2| \ \dots \ |x_N|]^T$, et sauvegarde de $\text{Sign}(\mathbf{x})$, vecteur des signes des composantes de \mathbf{x} .

Etape 2. Recherche du leader, $|\tilde{\mathbf{x}}|$, du vecteur $|\mathbf{x}|$ par ordonnancement décroissant de ses composantes.

Etape 3. Recherche de $|\tilde{\mathbf{y}}|_{j_0}$ plus proche voisin de $|\tilde{\mathbf{x}}|$ dans le dictionnaire des leaders absolus. j_0 étant son index dans ce dictionnaire. Il est donné par :

$$j_0 = \arg \max_{1 \leq j \leq P} [|\tilde{\mathbf{x}}|^T |\tilde{\mathbf{y}}_j|], \quad (2.46)$$

où P est le cardinal du dictionnaire des leaders.

Etape 4. Détermination du rang, r , de $|\mathbf{y}|$, vecteur quantifié de $|\mathbf{x}|$, dans la classe j_0 , par la formule de Schalkwijk (voir annexe B.1).

Etape 5. Détermination de l'indice t du vecteur quantifié $|\mathbf{y}|_{j_0}$ par la formule (2.45).

Le mot de code transmis contient donc l'indice t et le vecteur des signes $\text{Sign}(\mathbf{x})$.

2.6.4.2 Algorithme de décodage

A partir de l'indice t et du vecteur des signes $\text{Sign}(\mathbf{x})$, il faut retrouver le vecteur \mathbf{y} .

Etape 1. Recherche de la classe j_0 à laquelle appartient le vecteur $|\mathbf{y}|$. L'index j_0 est celui vérifiant :

$$H(j_0) \leq t < H(j_0 + 1). \quad (2.47)$$

Etape 2. Calcul du rang, r , de $|\mathbf{y}|$ dans sa classe j_0 :

$$r = t - H(j_0). \quad (2.48)$$

Etape 3. Détermination de $|\mathbf{y}|$, sachant que son rang dans la classe j_0 est r . Il s'agit de retrouver la permutation inverse de celle effectuée dans l'étape 2 du codage. Ceci se fait en utilisant l'algorithme de Schalkwijk inverse (voir annexe B.2).

Etape 4. Passage de $|\mathbf{y}|$ à \mathbf{y} en utilisant le vecteur des signes $\text{Sign}(\mathbf{x})$.

2.7. Codeur MPEG-1 Layer I et II

Les travaux sur la série des normes de codage audio MPEG [Brandenburg 97, Noll 97], issues de l'organisation de standardisation internationale (ISO), ont commencé depuis 1988. Le standard de compression de signaux audio haute qualité MPEG-1 [IS 11172-3, Brandenburg 94] achevé en 1992 est le résultat de la première phase de ces travaux. Ce système de codage est défini en trois versions nommées Layer I, II et III pour les fréquences d'échantillonnage 32, 44,1 et 48 kHz. Ils fournissent la possibilité d'obtenir une qualité audio quasi transparente à des débits de plus en plus faibles (192 kbits/s pour Layer I, 128 kbits/s pour Layer II et 96 kbits/s pour Layer III) mais au prix d'une complexité et d'un délai croissants. La norme de codage audio MPEG-2 BC (Backward Compatible) a été définie comme une amélioration de MPEG-1 tout en restant compatible avec cette dernière. Elle offre des extensions telles que le codage multivoies et la possibilité d'utiliser des fréquences d'échantillonnage plus faibles (16, 22,05 et 24 kHz). Le codeur MPEG-2 AAC (Advanced Audio Coding) [Bosi 97], normalisé aussi dans le cadre de MPEG-4, est le dernier de la série. Ce codeur est non compatible avec MPEG-1, il offre une très haute qualité pour un débit de 64 kbit/s. Dans la suite nous ne nous intéresserons qu'aux codeurs MPEG-1 Layer I et II.

Le codeur MPEG-1 Layer I utilise un banc de filtres à 32 sous-bandes uniformes pour réaliser la transformation temps/fréquence du signal audio d'entrée. Les échantillons en résultant sur chaque sous-bande sont regroupés en blocs de 12, puis normalisés par un facteur d'échelle commun avant d'être quantifiés. Le nombre de niveaux du quantificateur scalaire uniforme utilisé pour chaque sous-bande résulte d'une procédure d'allocation dynamique des bits. Cette procédure utilise un modèle psychoacoustique pour déterminer la répartition des bits qui rend le bruit de quantification le moins perceptible possible. La résolution fréquentielle du banc de filtres d'analyse étant faible, les modèles d'audition proposés dans la norme se basent sur l'estimation du spectre obtenue par une transformée de Fourier rapide (FFT) du signal temporel d'entrée. La trame transmise contient l'ensemble des échantillons des sous-bandes quantifiés qui représentent l'information principale sur le signal original et une information complémentaire nécessaire pour l'opération de décodage constituée par les facteurs d'échelle et l'allocation de bits (figure 2.8).

Le même schéma de codage est utilisé pour Layer II avec la différence que pour chacune des sous-bandes trois blocs de 12 échantillons sont regroupés par trame. Les facteurs d'échelle sont calculés pour chacun des blocs mais ils ne sont pas tous toujours transmis. Une procédure de compression est utilisée pour réduire le débit leur correspondant. Une information supplémentaire sur les facteurs d'échelle sélectionnés est alors introduite en plus dans la trame.

Les retards introduits par le codage/décodage MPEG-1 sont de l'ordre de 60 ms pour Layer I et 150 ms pour Layer II à une fréquence d'échantillonnage de 32 kHz. Ces retards sont en majeure partie dus à l'analyse/synthèse par les bancs de filtres.

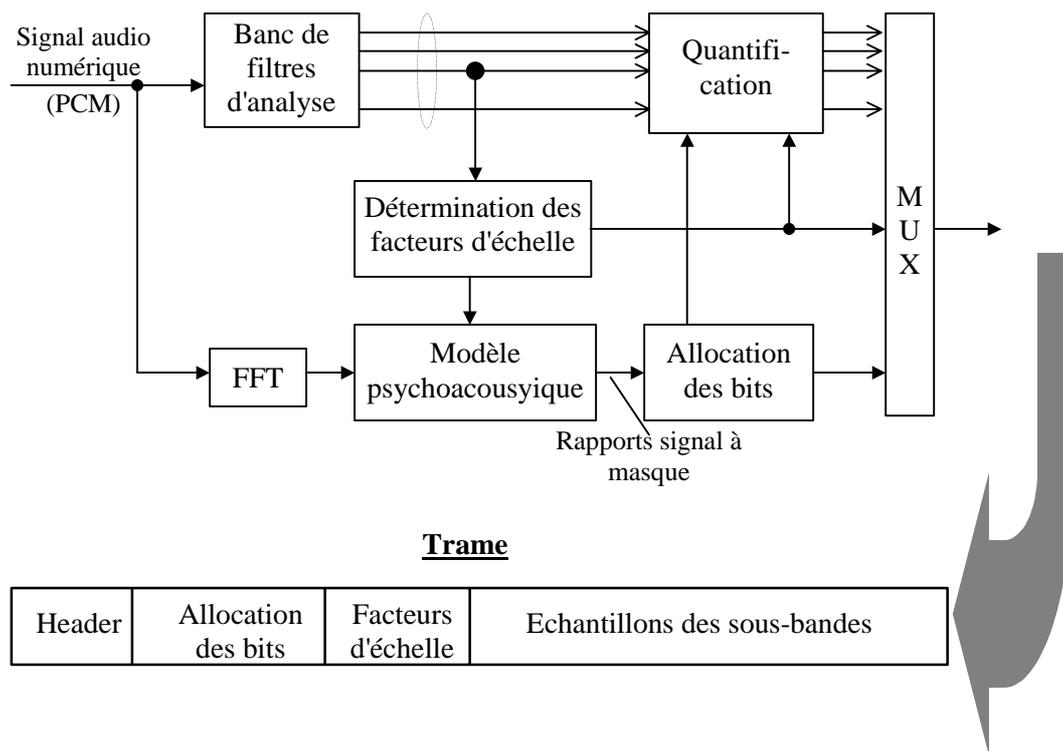


Fig. 2.8 : Schéma de base du codeur audio MPEG-1 Layer I et II et format de la trame.

2.7.1 Banc de filtres

Le banc de filtres utilisé dans Layer I et II est un pseudo-QMF. Comme il a été déjà indiqué dans le paragraphe 2.3.3, ce système est basé sur la modulation d'un filtre prototype passe-bas pour construire les différents filtres passe-bande d'analyse et de synthèse. Ce filtre prototype est à phase linéaire et de longueur 512. Sa réponse impulsionnelle, donnée dans la figure 2.9, a été optimisée de manière à approcher au mieux la condition de reconstruction parfaite. Ainsi le taux d'ondulation de la réponse en fréquence après analyse/synthèse est plus petit que 0,07 dB. Ce filtre fournit aussi une bonne résolution temporelle avec une résolution fréquentielle très acceptable (atténuation dans la bande de rejet supérieure à 96 dB).

L'implantation des opérations d'analyse et de synthèse présentée dans la norme MPEG-1 [IS 11172-3, Pan 95] se basent sur la forme polyphasée de ce type de bancs de filtres (voir figure 2.4). Malgré la réduction de complexité fournie par cette forme par rapport à une structure directe, la complexité des bancs de filtres constitue toujours la majeure partie des processus de codage et de décodage. Ainsi par exemple au décodage, la synthèse consomme 71 % du temps processeur [Hans 97]. Pour réduire encore plus cette complexité, des algorithmes rapides ont été élaborés. Ils se basent essentiellement sur la réécriture des équations de l'opération de matriçage, utilisant la matrice de modulation \hat{C} de taille 32×64 , de manière à remplacer cette dernière par une matrice de transformée en cosinus discret (DCT) de taille 32×32 [Konstantinides 94]. L'implantation se fait alors en utilisant un algorithme de DCT rapide à 32 points. D'autres méthodes ont été aussi développées pour réduire le coût de calcul résultant du filtrage par les composantes polyphasées du filtre prototype [Chan 96, Kumar 96, Hans 97].

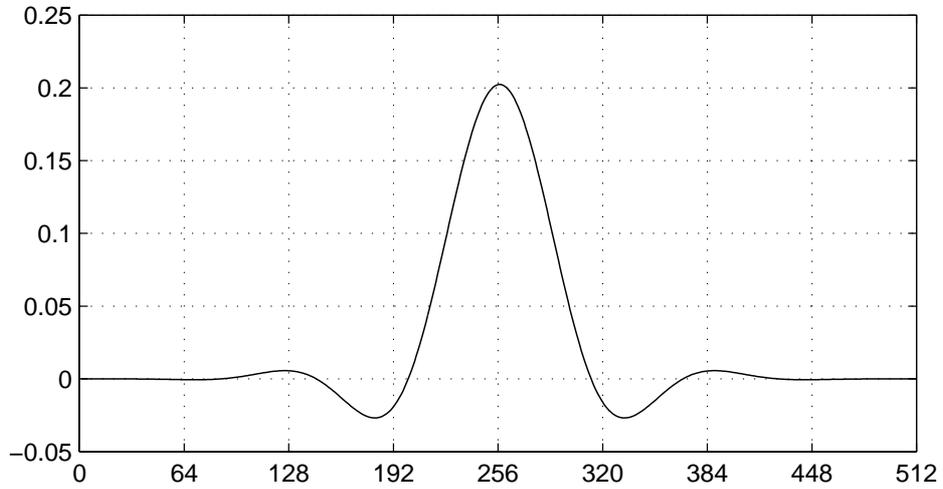


Fig. 2.9 : Réponse impulsionnelle du filtre prototype.

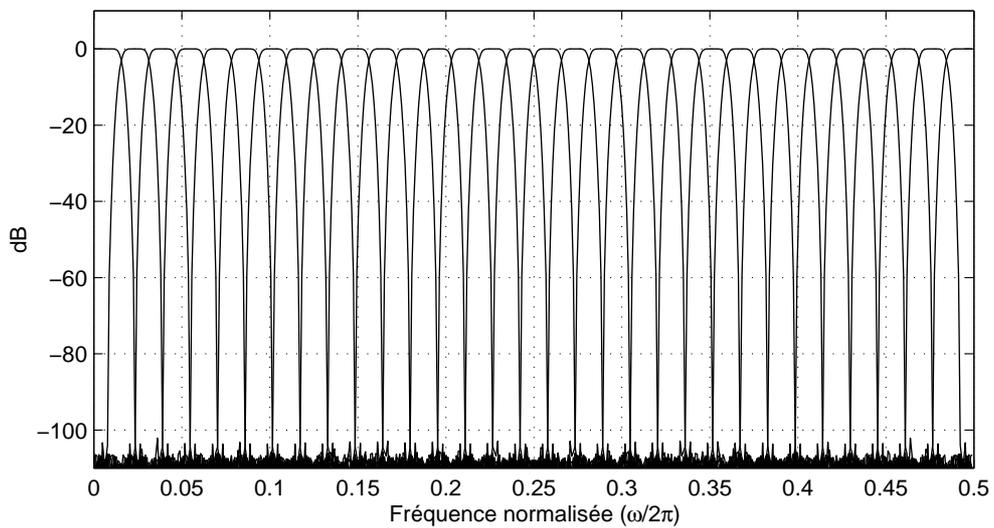


Fig. 2.10 : Réponse fréquentielle du banc de filtres d'analyse du codeur MPEG-1 Layer I et II.

2.7.2 Modèles psychoacoustiques

La norme MPEG-1 audio propose deux modèles psychoacoustiques utilisés pour la mise en forme spectrale du bruit de quantification. Ces modèles effectuent le calcul des rapports signal à masque, $SMR(j)$, pour chacune des sous-bandes. Ils serviront à alimenter la procédure d'allocation des bits. Les deux modèles peuvent être utilisés par les trois couches de MPEG-1 audio (et aussi MPEG-2 BC et AAC). Mais en pratique le modèle 1 est surtout employé dans Layer I et II et le modèle 2 dans Layer III. Nous donnerons donc dans la suite une description plus détaillée du premier modèle et un bref descriptif du second.

2.7.2.1 Modèle psychoacoustique 1

Le principe de calcul de ce modèle psychoacoustique consiste à déterminer d'abord les courbes de masquage de chaque composante du spectre d'entrée en faisant la distinction entre composantes tonales et non tonales. La courbe de masquage global est considérée comme la somme de la contribution des différentes composantes et du seuil d'audition absolu.

Ainsi, en première étape le spectre du signal temporel d'entrée, $\{X(k)\}$, est calculé par une analyse FFT, de résolution 512 points pour Layer I et 1024 pour Layer II, utilisant la fenêtre de pondération de Hann. La détermination des composantes tonales s'effectue ensuite en relevant les maxima locaux du spectre, i.e. tel que $X(k) > X(k+1)$ et $X(k) \geq X(k-1)$, vérifiant en plus la condition :

$$X(k) - X(k \pm l) \geq 7 \text{ dB}, \quad (2.49)$$

où l'intervalle de variation de l dépend de l'indice de fréquence k . En basse fréquence cet intervalle est de largeur plus petite puisque la résolution de l'oreille humaine est meilleure, sa largeur devient plus grande en haute fréquence. Le niveau de pression acoustique d'une telle composante tonale se calcule comme la somme de sa puissance et de celles des deux raies voisines de chaque côté. Les raies restantes sur chacune des 25 bandes critiques sont assimilées à un bruit. Elles sont représentées par une seule composante non tonale dont le niveau de pression acoustique est calculé par sommation de leurs puissances. Cette étape est suivie d'une phase de décimation dont le but est de réduire la complexité des traitements ultérieurs. Elle consiste d'une part à éliminer les composantes (tonales et non tonales) qui sont en dessous du seuil d'audition absolu et d'autre part à ne garder pour les composantes tonales séparées de moins de 0,5 Bark que celle de plus forte puissance.

Pour chacune des composantes obtenues précédemment, une courbe de masquage individuelle est calculée. Ainsi, le seuil de masquage, $S_m(v_j, v_i)$, résultant de l'effet d'une composante j , de niveau de pression acoustique $P(v_j)$, à l'indice de fréquence i est donnée par :

$$S_m(v_j, v_i) = P(v_j) + av(v_j) + vf(v_j, v_i). \quad (2.50)$$

Dans cette équation v_j est la fréquence exprimée en Bark correspondant à l'indice de fréquence j et toutes les valeurs sont exprimées en dB. $av(v_j)$ est l'indice de masquage, son expression dépend de la tonalité et reflète le fait qu'un son pur (tonal) masque moins qu'un bruit à bande étroite de même puissance. Elle est donnée par :

$$av(v_j) = \begin{cases} -1,525 - 0,275v_j - 4,5 & \text{si la composante est tonale,} \\ -1,525 - 0,175v_j - 0,5 & \text{sinon.} \end{cases} \quad (2.51)$$

$vf(v_j, v_i)$ est la fonction de masquage, elle s'exprime comme suit :

$$vf(v_j, v_i) = \begin{cases} 17(dv+1) - (0,4P(v_j) + 6) & \text{si } -3 \leq dv < -1 \text{ Bark,} \\ (0,4P(v_j) + 6)dv & \text{si } -1 \leq dv < 0 \text{ Bark,} \\ -17dv & \text{si } 0 \leq dv < 1 \text{ Bark,} \\ -(dv-1)(17 - 0,15P(v_j)) - 17 & \text{si } 1 \leq dv < 8 \text{ Bark,} \end{cases} \quad (2.52)$$

avec $dv = v_j - v_i$. Les pentes de cette fonction dépendent du niveau de pression acoustique de la composante considérée. L'effet de masquage est donc d'autant plus fort que $P(v_j)$ est important. Ce phénomène est plus accentué vers les fréquences supérieures.

La courbe de masquage global, $S_g(f)$, est ensuite obtenue par sommation des courbes de masquage individuelles et aussi du seuil d'audition absolu $QT(f)$.

Finalement, le rapport signal à masque pour chaque sous-bande j est donné par l'expression suivante :

$$SMR(j) = SPL(j) - MT(j), \quad (2.53)$$

où $MT(j)$ est le seuil de masquage, il est pris comme le minimum de la courbe de masquage $S_g(f)$ sur cette sous-bande. $SPL(j)$ est le niveau de pression acoustique, son calcul se base sur le facteur d'échelle maximale, $SCF(j)$, et les composantes du spectre de Fourier dans cette sous-bande j :

$$SPL(j) = \max \left\{ 10 \log_{10} \left(\sum_{k \in sb(j)} X^2(k) \right), 20 \log_{10} (SCF(j)) - 10 \right\} + 90,3. \quad (2.54)$$

2.7.2.2 Modèle psychoacoustique 2

De même que pour le modèle psychoacoustique 1, ce modèle commence par une estimation du spectre du signal d'entrée à l'aide d'une FFT. Ce spectre étant exprimé en échelle de Bark, chaque bande critique est découpée en trois segments et à chacun est affectée une composante ayant une puissance égale à la somme des celles des raies spectrales lui appartenant.

La classification des composantes en tonales ou non tonales n'est pas faite dans ce modèle. On calcul plutôt un indice de tonalité α , compris entre 0 et 1, qui mesure le degré de ressemblance d'une composante à un son pur ($\alpha = 1$) ou à un bruit ($\alpha = 0$). L'estimation de cet indice se base sur la mesure de prédictibilité du spectre du signal dans le domaine amplitude-phase.

Le spectre de puissance du signal, constitué des composantes déterminées précédemment, est convolué à la fonction de masquage élémentaire d'un bruit à bande étroite pour déterminer les courbes de masquage individuelles. L'indice de tonalité est ensuite utilisé comme indice de masquage pour apporter une correction à ces courbes due au fait que le niveau de masquage diffère selon la nature tonale ou non du son (proche d'un bruit ou d'un son pur). La courbe de masquage globale est obtenue

en prenant le niveau maximum entre la courbe de masquage individuelle et le seuil d'audition absolu. Finalement, les rapports signal à masque sont calculés pour chacune des sous-bandes.

2.7.3 Allocation de bits

Pour réaliser une allocation de bits suivant un critère perceptuel, le codeur MPEG-1 Layer I et II utilise un algorithme simple mais non-optimal connu sous le nom d'allocation de bits en "*greedy*" [Ramstad 82, Gersho 92]. Il se base sur le principe de répartir le nombre de bits disponible progressivement, un bit à la fois, de manière à faire la meilleure affectation possible. L'idée de base est que lors de chacune des itérations un bit est alloué là où il aura le plus d'effet i.e. où la diminution de la fonction de distorsion totale donnée par la somme des rapports bruit à masque sera la plus grande. Les étapes de cet algorithme se déroulent alors comme suit :

Etape 0. Initialisation à zéro des nombres de bits, $b(j)$, de chacune des sous-bandes.

Etape 1. Mise à jour de la fonction de distorsion, $NMR(j)$, sur chacune des sous-bandes :

$$NMR(j) = SMR(j) - SNR(b(j)), \quad (2.55)$$

où $SNR(b(j))$ est le rapport signal à bruit correspondant au quantificateur ayant un nombre de bits $b(j)$. $SMR(j)$ est le rapport signal à masque fourni par le modèle psychoacoustique.

Etape 2. Incrémentation du nombre de bits de la sous-bande j_0 où cette distorsion est maximale :

$$j_0 = \arg \max_j [NMR(j)]. \quad (2.56)$$

Les étapes 1 et 2 sont répétées de manière itérative jusqu'à ce que le nombre total de bits disponibles soit distribué.

2.7.4 Quantification et facteurs d'échelle

Le facteur d'échelle pour chaque sous-bande est déterminé sur chaque bloc de 12 échantillons et correspond au maximum de la valeur absolue de ces échantillons arrondi à la valeur juste au dessus prise dans une table. Pour Layer I, il est codé sur 6 bits et il n'est transmis que pour les sous-bandes dont le nombre de bits alloué est non nul. MPEG-1 Layer II utilise en plus une compression des facteurs d'échelle pour réduire le débit leur correspondant. En effet, dans ce cas, la trame contient 3 blocs de 12 échantillons par sous-bande et un facteur d'échelle pour chacun. Une classification, basée sur l'importance du changement entre facteurs d'échelle successifs, est faite pour déterminer ceux à transmettre parmi les trois. Une information sur les facteurs d'échelle sélectionnés est aussi transmise sur 2 bits.

Les quantificateurs utilisés sont tous uniformes à nombre impair de niveaux, ce qui permet une représentation exacte du zéro numérique. Layer I utilise 14 classes de quantificateurs différents, chacun a $2^b - 1$ niveaux avec $2 \leq b \leq 15$. Chaque échantillon est codé indépendamment par un mot de code. Pour Layer II, le nombre de niveaux dépend de la sous-bande et varie selon la fréquence d'échantillonnage et le débit binaire utilisé. Les nombres de niveaux permis sont 3, 5, 7, 9, 15, ..., 65535. Les quantificateurs à 3, 5 et 9 niveaux ne permettent pas une utilisation efficace des mots de code de 2, 3 ou 4 bits, un groupage de trois échantillons sous-bande en "granule" est alors utilisé avant de les coder par un même mot de code.

L'utilisation d'une procédure de compression des facteurs d'échelle et le groupage des échantillons et aussi une trame de longueur plus grande (correspondant à $3 \times 12 \times 32 = 1152$ échantillons des sous-bande dans ce cas) permettent de disposer d'un nombre de bits plus grand pour la quantification dans le cas de Layer II. Pour un même débit binaire total, la qualité audio est alors améliorée par rapport à Layer I.

2.8. Codeur TDAC de FTR&D

Le codeur TDAC de FTR&D est un codeur fréquentiel perceptuel qui a été développé initialement pour une largeur de bande de 15 kHz et pour un débit de 64 kbit/s [Mahieux 94]. Il a été ensuite adapté au codage bande élargie [0-7kHz] [Dia 93]. La version utilisée dans cette thèse est issue de cette dernière, mais modifiée et améliorée afin de participer à la campagne de normalisation de l'UIT-T visant à sélectionner un codeur bande élargie à des débits 16-24-32 kbit/s avec une qualité équivalente à celle du G.722 pour les débits 48-56-64 kbit/s [DT&FT 98, Combescure 99]. Dans la suite nous donnerons une description détaillée de cette dernière version.

2.8.1 Description générale

Le schéma de principe du codeur TDAC de FTR&D est donné dans la figure 2.11. Le signal d'entrée échantillonné à 16 kHz est découpé en blocs de 20 ms ($M = 320$ échantillons). Une transformée MDCT est appliquée sur une fenêtre à recouvrement, correspondant à deux blocs successifs, pour obtenir 320 coefficients MDCT. Le calcul de cette transformée se fait par un algorithme rapide décrit dans [Duhamel 91]. Une courbe de masquage est ensuite déterminée coefficient par coefficient à partir de ce spectre et tous les coefficients masqués sont mis à zéro. L'enveloppe spectrale de la trame est déterminée par calcul des énergies sur 32 bandes du spectre, sa quantification et son codage dépendent de la tonalité et du voisement de la trame. L'allocation dynamique des bits se base sur une courbe de masquage bande par bande calculée à partir de la version quantifiée de l'enveloppe spectrale. Ceci permet d'avoir une compatibilité entre l'affectation binaire du codeur et du décodeur. Les coefficients MDCT normalisés dans chaque bande sont ensuite quantifiés par des quantificateurs vectoriels utilisant des dictionnaires à union de codes à permutation imbriqués en dimension et en taille. Finalement, les informations sur la tonalité et le voisement ainsi que l'enveloppe spectrale et les coefficients codés sont multiplexés et transmis en trame.

Le processus de décodage se fait de manière inverse. Ainsi, après décodage de l'enveloppe spectrale, la courbe de masquage bande par bande est recalculée pour déterminer l'allocation des bits. Une quantification inverse est ensuite effectuée pour retrouver les coefficients MDCT. Finalement une opération de synthèse est réalisée par un algorithme rapide [Duhamel 91] pour déterminer les échantillons décodés de la trame temporelle. Dans le cas de trames effacées une procédure spéciale de correction basée sur la régénération des coefficients de la trame est mise en oeuvre.

Contrairement au codeur MPEG-1, on peut remarquer que la transmission de l'allocation des bits dans ce codeur se fait de manière implicite à l'aide de l'enveloppe spectrale quantifiée. Ceci permet d'économiser le nombre de bits utilisé pour transmettre les informations latérales. Par contre, il a l'inconvénient d'augmenter la complexité au niveau du décodeur puisqu'il y a nécessité de recalculer l'allocation des bits.

Ce codeur offre aussi, de manière optionnelle, la possibilité de réduire encore plus le débit en ne transmettant que les périodes actives du signal. Dans ce cas un module de détection d'activité vocale est mis en oeuvre en parallèle pour fournir une décision à partir des échantillons du signal temporel d'entrée. En plus, le premier bit de la trame, appelé bit SID, est utilisé comme fanion pour indiquer au décodeur le début d'une période de silence. La coupure de transmission ne se fait pas dès la détection de la première trame inactive. Une période de transition, dite "hangover", composée de trois trames de garde est laissée. Elle permet d'une part d'éviter les coupures brusques et d'autre part de transmettre au décodeur les trames décrivant le silence qui permettront la génération du bruit de confort. Au niveau du décodeur une procédure semblable à celle de la récupération des trames effacées est utilisée pour la génération du bruit de confort.

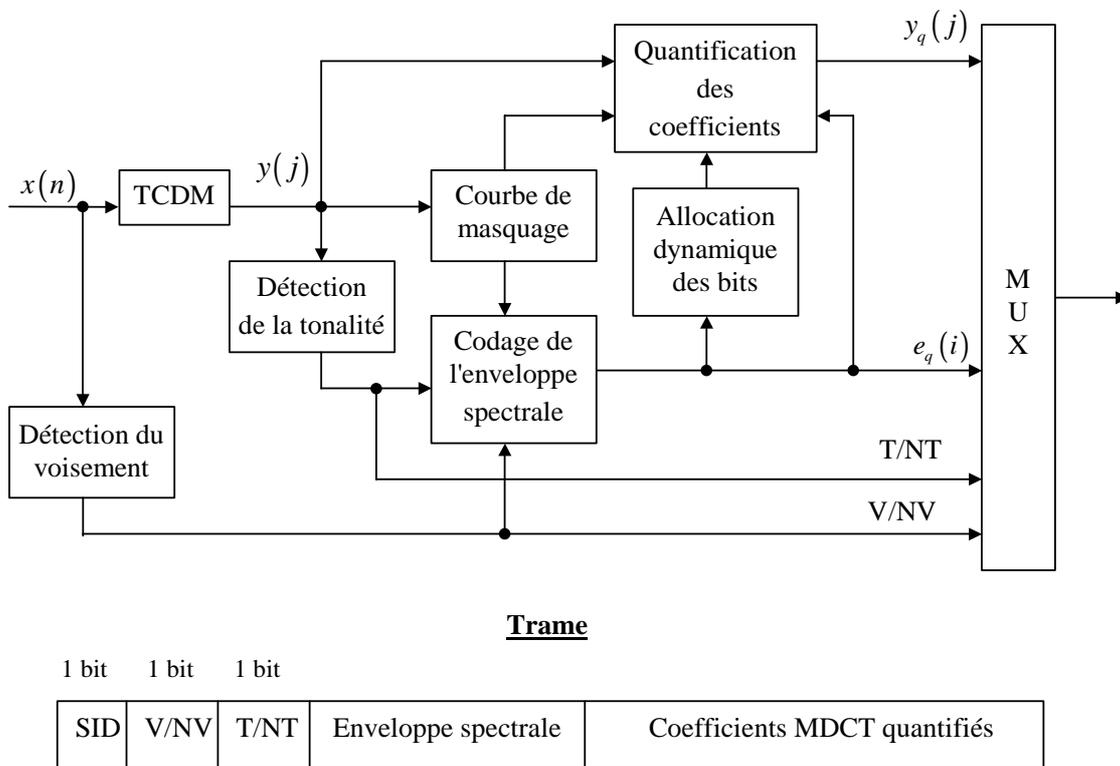


Fig. 2.11 : Schéma bloc du codeur TDAC de FTR&D et format de la trame.

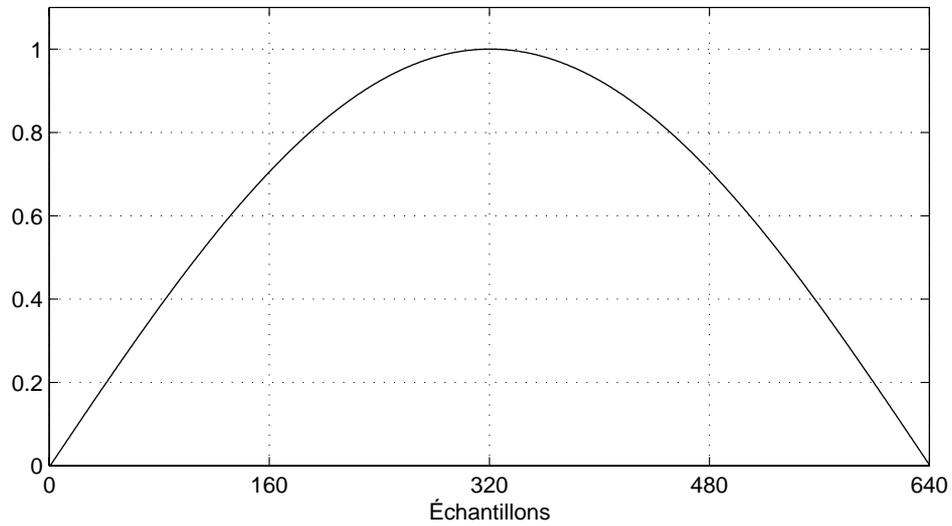


Fig. 2.12 : Réponse impulsionnelle du filtre prototype du banc de filtres TDAC.

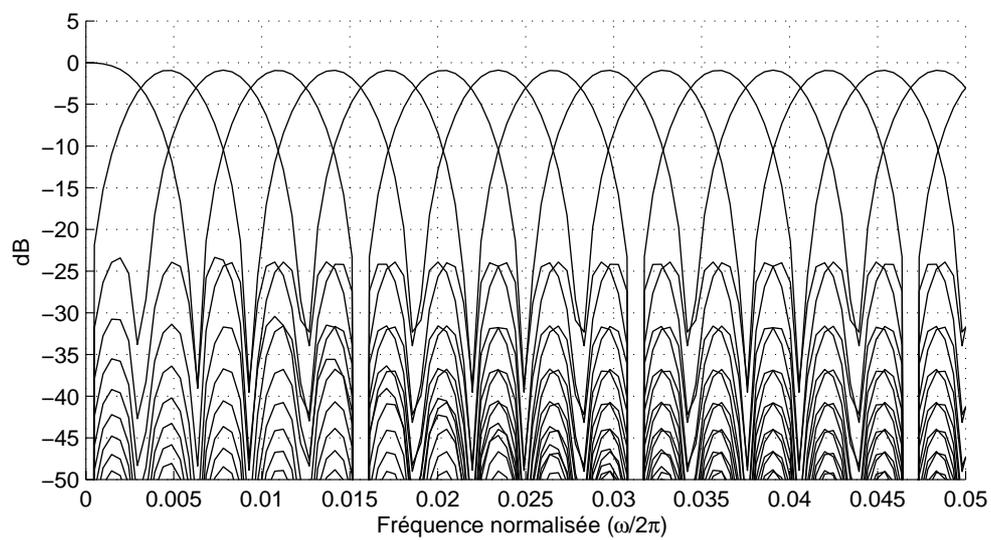


Fig. 2.13 : Réponse fréquentielle des seize premiers filtres du banc de filtres MDCT du codeur TDAC de FTR&D.

2.8.2 Détection du voisement et de la tonalité

La détection de la nature voisée/non voisée d'une trame courante du signal se base sur les trois paramètres suivants déterminés à partir des 640 échantillons temporels de cette trame :

- L'énergie moyenne de la trame :

$$E = \frac{1}{2M} \sum_{n=0}^{2M-1} x^2(n). \quad (2.57)$$

- Le premier coefficient de PARCOR :

$$\rho_1 = \frac{1}{2M} \sum_{n=0}^{2M-1} x(n-1)x(n). \quad (2.58)$$

- Le nombre de passage par zéro.

Une trame est donc considérée comme non voisée si l'une des conditions suivantes est vérifiée : faible énergie ($E(\text{dB}) < 40\text{dB}$) ou faible corrélation ($\frac{\rho_1}{E} < 0,6$) ou si le nombre de passage par zéro est grand. Dans le cas contraire la trame est voisée.

La détermination de la tonalité de la trame se fait à partir de la mesure d'étalement spectrale, sfm , calculée comme le rapport entre la moyenne géométrique et la moyenne arithmétique du carré des coefficients de transformée MDCT de la trame. Si le sfm est suffisamment faible ($sfm < -65\text{dB}$) le signal de la trame est tonal, il est non tonal si cette quantité est grande ($sfm > -50\text{dB}$). Dans les autres cas ($-65\text{dB} \leq sfm \leq -50\text{dB}$), pour éviter les changements brusques, une procédure de lissage de ce paramètre est appliquée et la valeur qui en résulte est comparée à un seuil fixe (-60dB) pour décider de la tonalité de la trame.

2.8.3 Courbe de masquage

La courbe de masquage coefficient par coefficient est déterminée pour éliminer les coefficients MDCT masqués et ainsi éliminer les composantes inutiles du signal. Son calcul se fait à partir des 320 points du spectre MDCT. Il se base sur deux hypothèses principales :

- Linéarité de l'effet de masquage : la courbe de masquage est égale à la somme des courbes masquages élémentaires dues aux différents coefficients $y(j)$, $0 \leq j \leq M - 1$.
- La forme de la courbe de masquage élémentaire due à une raie spectrale, sur l'échelle des Bark, est indépendante de la fréquence. Le niveau de masquage créé par un coefficient en une autre fréquence ne dépend donc que de la différence de leurs localisations. La forme de la fonction d'étalement de la membrane basilaire utilisée est donnée par la figure 2.14.

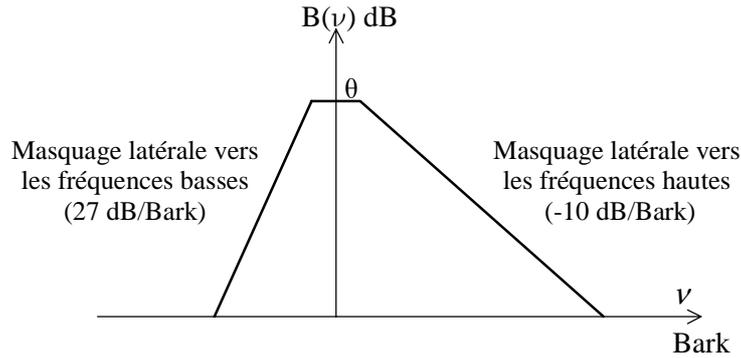


Fig. 2.14 : Fonction d'étalement de la membrane basilaire utilisée dans le codeur TDAC.

La courbe de masquage totale correspond donc à la convolution du carré des coefficients du spectre MDCT par la fonction d'étalement spectrale, elle s'exprime comme suit :

$$S(j) = \theta \sum_{k=0}^{M-1} a_0(k) B(v_j - v_k) y^2(k), \quad 0 \leq j \leq M-1. \quad (2.59)$$

où θ est l'indice de masquage, il est fixé à -20 dB indépendamment de la nature du signal (tonal ou bruit à bande étroite). Cette valeur correspond au seuil à partir duquel un bruit à bande étroite est masqué par un son pur (tonal). $a_0(k)$ est le facteur de transmission de l'oreille externe à la fréquence correspondant au coefficient $y(k)$. v_j et v_k sont les valeurs des fréquences, exprimées sur l'échelle de Bark, correspondant respectivement aux coefficients $y(j)$ et $y(k)$.

Le calcul de la courbe de masquage dans le codeur TDAC ne se fait pas directement par la formule (2.59). Un algorithme rapide est utilisé pour déterminer les points de cette dernière de manière récurrente en profitant des particularités de la fonction d'étalement de la membrane basilaire, $B(v)$, [Mahieux 94].

Un coefficient j est masqué si sa valeur au carré est inférieure à la courbe de masquage soit :

$$y^2(j) < S(j). \quad (2.60)$$

2.8.4 Enveloppe spectrale

L'enveloppe spectrale, $\{e(i), 0 \leq i \leq 31\}$, est définie comme l'ensemble des valeurs obtenues par la racine carrée de la moyenne de l'énergie sur chacune des bandes. Les coefficients masqués sur chacune des bandes ne sont pas pris en compte :

$$e(i) = \sqrt{\frac{1}{N_c(i)} \sum_{j \in \text{bande}(i)} y^2(j) (1 - \text{irm}(j))}, \quad 0 \leq i \leq 31, \quad (2.61)$$

où $\text{irm}(j)$ correspond à l'indication de masquage du coefficient d'indice j obtenue à partir de la courbe de masquage coefficient par coefficient, $\text{irm}(j)$ vaut 1 quand le coefficient est masqué, 0 sinon. $N_c(i)$ est le nombre de coefficients dans la bande i . Une bande dont tous les coefficients sont masqués est considérée comme masquée et une procédure spéciale est utilisée pour son codage.

Pour les trames de type non tonal, un codage différentiel intra-bloc dans le domaine logarithmique est utilisé pour l'enveloppe spectrale. Il permet de tenir compte de la redondance entre les valeurs successives et de réduire la dynamique de variation. Ainsi on calcule :

$$\varepsilon(i) = \log(e(i)) - \log(e(i-1)), \quad 1 \leq i \leq 31. \quad (2.62)$$

La première valeur $\log(e(0))$ sert d'ancrage, elle est quantifiée en absolu et de manière uniforme sur 7 bits. Pour les autres bandes c'est l'erreur de prédiction $\varepsilon(i)$ qui est codé et transmise. Cette erreur est quantifiée uniformément sur 32 niveaux dans son intervalle de variation [-80 dB, 80 dB]. La distribution non uniforme de cette erreur suggère d'utiliser un codage de Huffman à longueur variable. Ainsi, 16 mots de code sont utilisés pour coder les erreurs de prédiction appartenant à l'intervalle [-40 dB, 40 dB]. Les valeurs en dehors de cet intervalle étant à faible probabilité d'occurrence, le codage de Huffman leur affecte des mots de code de grandes longueurs ce qui rend le décodage plus long. Pour remédier à ce problème un seul mot de code, appelé séquence d'échappement, est affecté à tous les niveaux de quantification dans cette zone. Une indication supplémentaire, utilisant 4 bits, est ajoutée pour spécifier le niveau de quantification dans cet intervalle. Pour les bandes de fréquence entièrement masquées un mot de code spécial est utilisé. Il y a donc au total 18 mots de codes de Huffman associés à chacune des bandes.

Dans le but de réduire la mémoire nécessaire pour le stockage des mots de code de Huffman un regroupement en 4 classes a été effectué. Une méthode de K-moyenne avec la distance de Kullback été utilisée pour regrouper les densités de probabilité des bandes les plus proches en classes et de trouver ensuite leurs centroïdes [Mahieux 94]. Les mots de code de Huffman correspondant à chaque classe sont enfin générés à partir de ces derniers.

Notons aussi que compte tenu de la différence entre les statistiques des erreurs de prédiction selon la nature voisée ou non du signal, une séparation suivant ce critère est faite pour le codage de Huffman utilisé.

Pour les trames tonales, une recherche de la bande de plus forte énergie est d'abord effectuée. Son numéro est codé par 5 bits et sa valeur est quantifiée uniformément dans le domaine logarithmique sur 128 niveaux. Les autres bandes sont quantifiées de manière différentielle par rapport à ce maximum, dans le domaine logarithmique, sur 16 niveaux.

2.8.5 Courbe de masquage par bandes et allocation de bits

L'allocation dynamique des bits dans le codeur TDAC se fait suivant un critère perceptuel pour chacune des bandes de fréquence. Elle est réalisée aussi bien au codeur qu'au décodeur et se base sur les valeurs quantifiées de l'enveloppe spectrale, $e_q(i)$, $0 \leq i \leq K-1$, pour que les deux processus soient compatibles.

Pour se faire, une courbe de masquage bande par bande est d'abord déterminée. Son calcul est identique à celui de la courbe de masquage coefficients par coefficients mais en utilisant dans ce cas l'enveloppe spectrale quantifiée. En effet, au coefficient central de la bande i est associé la valeur $e_q(i)$. Les seuils de masquage bande par bande, $S_b(i)$, $0 \leq i \leq K-1$, sont ensuite obtenus par une convolution à la fonction d'étalement de la membrane basilaire de manière identique à la formule (2.59).

L'allocation de bits se réalise en deux phases. D'abord un premier calcul du nombre de bits à allouer à chaque bande est effectué suivant la formule optimale (2.38) en utilisant les rapports énergie à masque, $e_q^2(i)/S_b(i)$, $0 \leq i \leq K-1$. Chacune des valeurs obtenues est arrondie à l'entier naturel le plus proche. Si le débit total alloué n'est pas exactement égal à celui disponible, une seconde phase est utilisée pour réaliser le réajustement. Cette étape se fait par une procédure itérative basée sur critère perceptuel qui ajoute ou retire des bits aux bandes. Ainsi, si le nombre de bits total distribué est inférieur à celui disponible, l'ajout de bits se fait aux bandes où l'amélioration perceptuel serait la plus importante. Cette amélioration perceptuelle est mesurée par la variation du rapport bruit à masque entre l'allocation initiale et finale des bandes [Perkins 89]. Le débit est augmenté pour la bande où cette variation est maximale i.e. qui contribue le plus à minimiser la fonction de distorsion perceptuelle (NMR total). Dans le cas contraire où le nombre de bits total distribué est supérieure à celui disponible, l'extraction de bits sur les bandes se fait de manière duale à cette dernière procédure.

2.8.6 Quantification des coefficients MDCT

La quantification dans le codeur TDAC de FTR&D se fait de manière vectorielle sur les vecteurs de coefficients MDCT de chacune des bandes. Elle utilise des quantificateurs vectoriels sphériques avec des dictionnaires à union de codes à permutation. Ces dictionnaires correspondent aux différentes dimensions des bandes qui varient entre 3 et 15. Pour une dimension donnée, des débits binaires allant de 5 à 32 bits sont possibles. Ces différents dictionnaires ont été conçus de manière à avoir en plus la propriété d'imbrication en dimension et en taille. Ainsi on a :

$$\mathcal{C}(b,d) \subset \mathcal{C}(b',d), \text{ si } b \leq b', \quad (2.63)$$

où $\mathcal{C}(b,d)$ est le dictionnaire des leaders correspondant à la dimension d et au débit binaire b . Une telle propriété permet de réduire davantage la mémoire nécessaire pour le stockage.

Les coefficients $y(j)$ du vecteur de la bande i sont d'abord normalisés pour obtenir le vecteur normé :

$$\frac{y(j)}{\sqrt{N_c(i)} e_q(i)}. \quad (2.64)$$

Le codage se fait ensuite par l'algorithme rapide décrit dans le paragraphe 2.6.4 qui exploite la structure des dictionnaires à codes à permutation. Le dictionnaire considéré est celui correspondant à la dimension de la bande et au nombre de bits qui lui est alloué. Le mot de code transmis finalement dans la trame est celui résultant de cette opération.

La construction des dictionnaires des leaders correspondant aux différents débits et dimensions a été réalisée de manière statistique en imposant les contraintes d'imbrication. Ces dictionnaires ne sont pas donc tout à fait algébriques mais par abus de langage on continuera de parler de QVAS. Pour aboutir à la propriété d'imbrication, l'apprentissage des dictionnaires se fait de façon récurrente, en commençant du plus petit, au sens de l'inclusion, jusqu'à arriver au plus grand. Ainsi, un premier dictionnaire correspondant à la dimension et au débit binaire les plus bas est d'abord construit. Pour une dimension et un débit donnés, les vecteurs de l'union des deux dictionnaires correspondant au débit juste plus bas à la même dimension et à la dimension inférieure au même débit, qui ont été déjà construits, sont pris. Ils sont ensuite complétés par apprentissage stochastique.

L'idée d'imbrication en dimension trouve son origine dans le fait que pour une bande de dimension donnée une partie des coefficients sont mis à 0 du fait qu'ils sont masqués. Puisqu'on ne considère que les leaders des vecteurs de valeurs absolues des coefficients, les derniers coefficients nuls peuvent être tronqués et le vecteur sera considéré à dimension plus petite.

2.8.7 Correction des trames effacées

Le codeur TDAC de FTR&D met en oeuvre une procédure de régénération de trame basée sur l'extrapolation dans le cas où cette dernière est effacée. Le traitement effectué diffère suivant si c'est la première trame d'un bloc de trames effacées ou c'est les suivantes.

Ainsi, pour la première trame une analyse LPC d'ordre 14 est réalisée sur les 320 échantillons du signal temporel décodé à partir de la trame d'avant. Le signal résiduel est aussi calculé. Si cette dernière trame est de nature tonale ou voisée, une analyse LTP (Long Term Prediction) est utilisée pour estimer la période du pitch. Un bloc de 640 échantillons du signal d'excitation est généré à partir du signal résiduel en utilisant soit la périodicité du pitch dans le cas voisé ou tonal soit par simple copie des 320 échantillons précédents. Ce signal d'excitation est ensuite filtré pour obtenir les 640 échantillons du signal extrapolé auxquels une transformée MDCT est appliquée pour retrouver les coefficients MDCT de la trame effacée.

Pour les trames effacées suivantes, ce sont les paramètres LPC et LTP déjà estimés pour la première qui sont utilisés et seul 320 échantillons du signal temporel extrapolé sont calculés. Cette procédure de régénération est illustrée par le schéma de la figure 2.15.

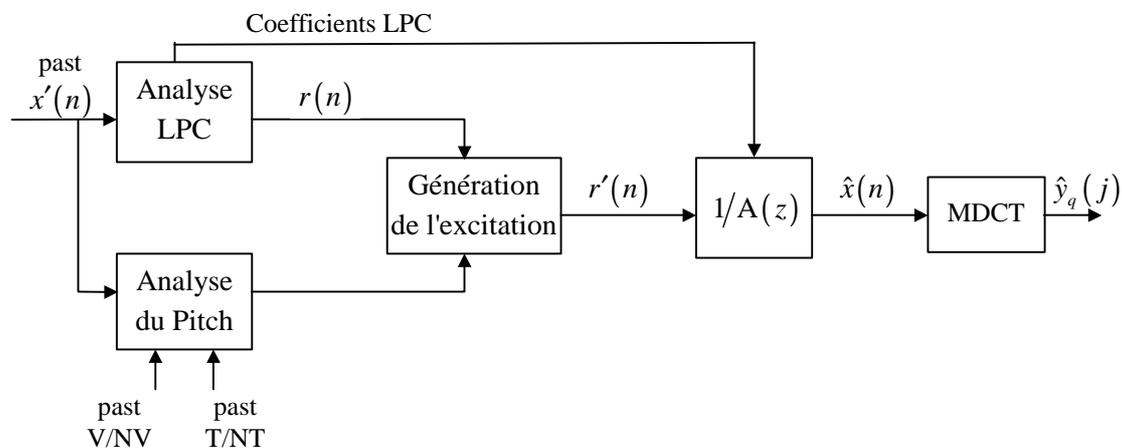


Fig. 2.15 : Schéma bloc de la correction des trames effacées.

Chapitre 3 Filtrage dans le domaine fréquentiel

3.1. Introduction

Dans le présent chapitre, nous nous intéressons au traitement du signal dans le domaine fréquentiel et particulièrement à l'opération de filtrage. C'est le premier pas vers la manipulation de signaux audio codés en fréquentiel perceptuel. Nous ferons donc abstraction de toute opération de quantification ou de codage. Seuls les blocs correspondant aux bancs de filtres d'analyse du codeur et de synthèse du décodeur sont donc considérés. Comme il a été rappelé dans le chapitre 2, un banc de filtres ou une transformée temps/fréquence sont totalement équivalents. Nous utiliserons donc de manière équivoque les appellations relatives au domaine fréquentiel et sous-bandes.

Le schéma du filtrage dans le domaine des sous-bandes est illustré par la figure 3.1. Il s'agit de remplacer la succession des blocs de synthèse, de filtrage en domaine temporel et d'analyse par un seul bloc comprenant un ensemble de filtres agissant sur les signaux des sous-bandes. L'intérêt étant d'éliminer la complexité et le retard algorithmique dus aux bancs de filtres d'analyse et de synthèse.

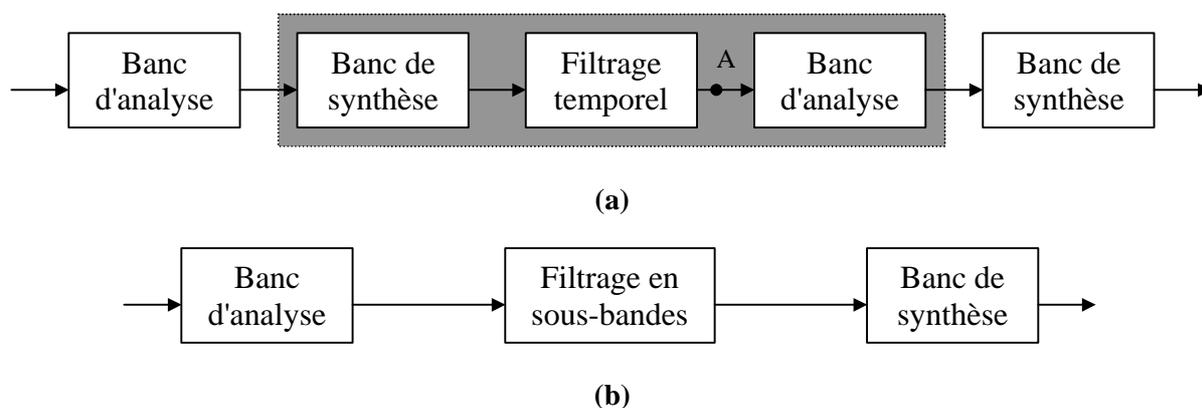


Fig. 3.1 : Schémas bloc du (a) filtrage dans le domaine temporel (b) filtrage dans le domaine des sous-bandes.

Si des opérations linéaires simples telle que la sommation ou la multiplication par une constante se transposent de manière directe du domaine temporel au domaine fréquentiel, le filtrage pose beaucoup plus de problèmes et nécessite une approche bien appropriée. La modification du spectre du signal apportée par un filtrage dans le domaine temporel ne peut être réalisé directement sur les signaux des bandes de fréquence sans tenir compte du phénomène d'aliasing introduit par le banc de filtres d'analyse. En effet, la relation de dépendance entre les composantes d'aliasing des différentes sous-bandes doit être nécessairement conservée pour que leur suppression soit assurée par le banc de filtres

de synthèse. Toute modification sur le spectre du signal doit respecter alors cette contrainte. Ce problème d'aliasing sera bien illustré dans le paragraphe qui suit.

Les applications du filtrage en sous-bandes sont diverses et multiples. Il peut être utilisé dès qu'il s'agit de filtrer des signaux représentés dans le domaine fréquentiel. Par ailleurs, dans cette thèse on s'intéresse particulièrement aux applications du filtrage audio pour la spatialisation sonore et l'ajout d'effets spéciaux (réverbération, effet de salle...).

Comme il sera détaillé dans le paragraphe 3.3, des travaux ont déjà été réalisés pour trouver des solutions au problème de filtrage en sous-bandes. Ces travaux ont porté essentiellement sur le filtrage à réponse impulsionnelle finie (FIR). L'utilisation de filtres à réponse impulsionnelle infinie (IIR) a beaucoup d'intérêt. En effet, ce type de filtres permet de modéliser un même gabarit de réponse en fréquence avec un nombre réduit de coefficients et donc une plus faible complexité par rapport aux filtres FIR. Ils ont aussi plusieurs applications spécifiques telle que la réverbération sonore utilisant des filtres passe-tout ou en peigne (Comb filter) [Gardner 98, Jot 95] ou encore la modification de la phase sans altération de l'amplitude du spectre du signal par des filtres passe-tout. L'utilisation de filtres HRTF (Head Related Transfer Functions) à réponse impulsionnelle infinie pour la spatialisation sonore, dans le but d'une meilleure réduction de la complexité, a constitué la motivation majeure du développement de la méthode présentée dans ce chapitre. Cette méthode donne une solution générale au problème du filtrage en sous-bandes [Benjelloun 00b]. Elle est applicable pour n'importe quel filtre rationnel (FIR ou IIR) et pour tout banc de filtres à échantillonnage critique vérifiant la condition de reconstruction parfaite.

3.2. Phénomène du repliement du spectre (aliasing)

Dans un banc de filtres à M bandes uniformes les filtres d'analyse ont une largeur de bande passante égale à $\frac{\pi}{M}$. Par ailleurs, les bandes de transition de ces filtres dépassent cette largeur et se recouvrent. Après opération de décimation, les signaux de chacune des sous-bandes sont alors échantillonnés à la fréquence $\frac{2\pi}{M}$ qui est inférieure à la fréquence d'échantillonnage de Nyquist. Il y a donc un phénomène de repliement du spectre (ou aliasing). Dans ce paragraphe nous allons illustrer ce phénomène pour le cas du banc de filtres pseudo-QMF du codeur MPEG-1.

Rappelons d'abord l'expression du signal à la sortie du filtre de synthèse sur la branche k :

$$\hat{X}_k(z) = \frac{1}{M} \sum_{l=0}^{M-1} H_k(zW^l) X(zW^l) F_k(z), \quad 0 \leq k \leq M-1. \quad (3.1)$$

avec $W = e^{-j\frac{2\pi}{M}}$.

Dans la somme ci-dessus seule la composante donnée par $l=0$ correspond au signal utile sur la bande k . Les autres, résultant de l'opération de décimation, sont des images du spectre du signal d'entrée filtré par $H_k(z)$. A cause du recouvrement du spectre entre bandes, le filtrage passe-bande par $F_k(z)$ ne permet pas d'isoler que le signal utile, il y reste une partie des composantes images.

Considérons la décomposition de chacun des filtres d'analyse et de synthèse en filtres passe-bandes complexes (figure 3.1) :

$$H_k(z) = H_k^+(z) + H_k^-(z), \quad (3.2)$$

$$F_k(z) = F_k^+(z) + F_k^-(z), \quad (3.3)$$

Ces filtres résultent de la modulation du filtre prototype, $H(z)$, par la fréquence porteuse $\frac{(2k+1)\pi}{2M}$. Ils représentent les spectres obtenus sur les fréquences positives et négatives et diffèrent par un terme de phase.

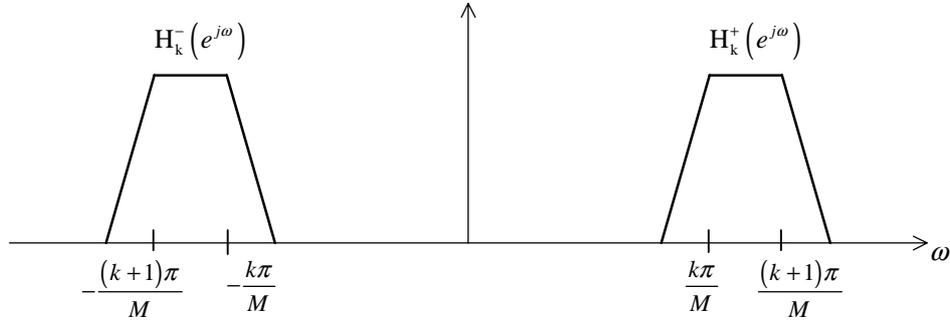


Fig. 3.2 : Réponse fréquentielle du filtre $H_k(z)$.

Pour le banc de filtres pseudo-QMF, tous les filtres ont une bande de transition suffisamment raide pour que seuls ceux qui sont adjacents se recouvrent (voir figure 2.10). En plus, ils ont une forte atténuation dans la bande de rejet. Ces propriétés s'obtiennent en imposant la contrainte $H(\omega) = 0$ pour $\omega \geq \pi/M$, lors de l'optimisation du filtre prototype. Dans la formule (3.1), les composantes images $H_k^+(zW^l)X(zW^l)$ et $H_k^-(zW^l)X(zW^l)$, $1 \leq l \leq M-1$, ne se recouvrent pas donc tous avec les bandes passantes du filtre de synthèse $F_k(z)$. Seules celles correspondant aux deux bandes adjacentes sont à considérer. Comme le montre la figure 3.3, ces dernières sont relatives à :

$$l = k, k+1 \text{ et } (M-k), (M-k-1) \text{ pour } 1 \leq k \leq M-1.$$

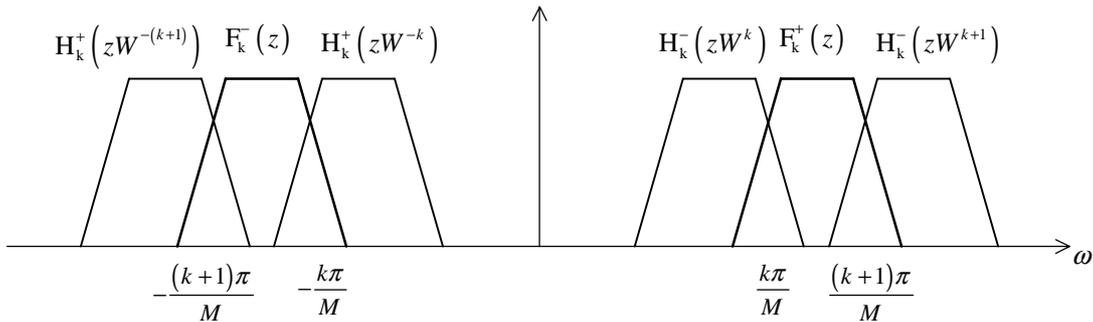


Fig. 3.3 : Illustration des composantes qui se recouvrent avec la réponse fréquentielle du filtre de synthèse $F_k(z)$.

La formule (3.1) qui donne le signal de sortie du filtre de synthèse se réduit donc à la relation suivante :

$$\begin{aligned} \hat{X}_k(z) = \frac{1}{M} \{ & H_k(z) F_k(z) X(z) \\ & + H_k^+(zW^{-(k+1)}) F_k^-(z) X(zW^{-(k+1)}) + H_k^-(zW^{k+1}) F_k^+(z) X(zW^{k+1}) \\ & + H_k^+(zW^{-k}) F_k^-(z) X(zW^{-k}) + H_k^-(zW^k) F_k^+(z) X(zW^k) \}. \end{aligned} \quad (3.4)$$

Chacune des composantes d'aliasing $X(zW^{\pm k})$ est commune aux branches $(k-1)$ et k . Pour que de telles composantes soient supprimées après sommation sur toutes les branches, il suffit que :

$$H_k^+(zW^{-k}) F_k^-(z) + H_{k-1}^+(zW^{-k}) F_{k-1}^-(z) = 0, \quad (3.5)$$

et

$$H_k^-(zW^k) F_k^+(z) + H_{k-1}^-(zW^k) F_{k-1}^+(z) = 0. \quad (3.6)$$

La conception des filtres $H_k^+(z)$, $H_k^-(z)$, $F_k^+(z)$ et $F_k^-(z)$ qui vérifient (3.5) et (3.6) se fait par choix approprié du terme de la phase qui les différencie [Vaidyanathan 93].

Tout traitement sur les signaux des sous-bandes ne permettant pas de conserver les relations (3.5) et (3.6) peut donner lieu à un résidu d'aliasing dans le signal obtenu après synthèse. Pour illustrer ce phénomène, considérons le cas simple d'une variation de gain par un facteur α sur les échantillons du signal de sortie de la sous-bande k_0 ($1 \leq k_0 \leq M-2$) du banc de filtres d'analyse. Dans ce cas, l'expression (3.5) par exemple devient :

$$\alpha H_{k_0}^+(zW^{-k_0}) F_{k_0}^-(z) + H_{k_0-1}^+(zW^{-k_0}) F_{k_0-1}^-(z). \quad (3.7)$$

Cette expression, qui est le facteur multiplicatif de la composante d'aliasing $X(zW^{k_0})$, n'est plus égale à zéro. Une telle composante ne sera donc pas totalement supprimée après restitution du signal par le banc de filtres de synthèse. La figure 3.4 illustre le résidu d'aliasing qui subsiste dans le cas de l'utilisation d'un signal sinusoïdale.

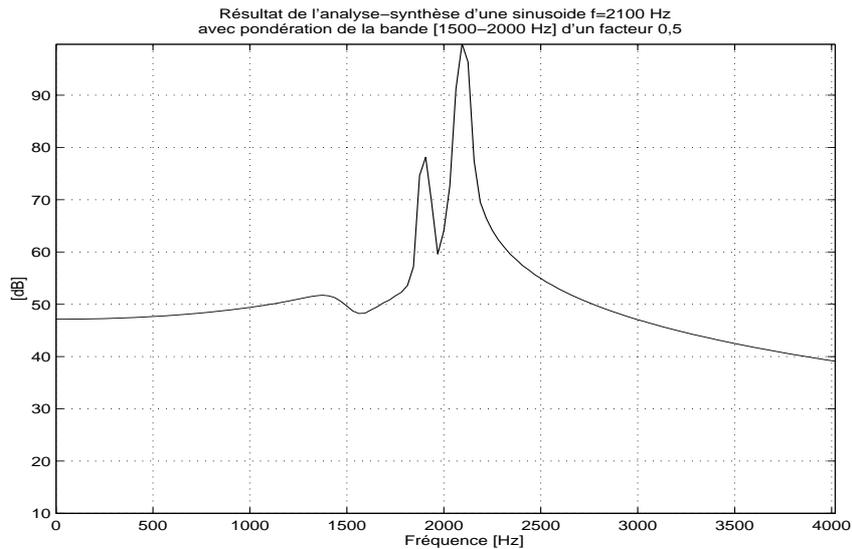


Fig. 3.4 : Illustration de l'effet de variation de gain dans une bande sur le signal reconstruit.

3.3. Filtrage dans le domaine fréquentiel : aperçu de l'état de l'art

3.3.1 Filtrage dans le domaine transformé

Le problème de filtrage dans le domaine transformé a été largement abordé et spécialement pour des applications en traitement d'image. Cet intérêt particulier est dû essentiellement à l'utilisation du filtrage dans plusieurs opérations de traitement du signal appliquées à l'image telle que la suppression du bruit, le lissage ou encore le filtrage anti-repliement associé au changement de la fréquence d'échantillonnage. La plupart des travaux déjà effectués dans ce sens concernent surtout le filtrage dans le domaine de la transformée DCT puisque cette dernière est à la base de différents standards de compression d'image et vidéo (JPEG, MPEG, H.261 et H.263). L'analyse de ces différents travaux permet de distinguer deux types d'approches principales.

3.3.1.1 Approche basée sur les propriétés de convolution-multiplication

L'extension de la propriété de convolution-multiplication (CMP : Convolution-Multiplication Property) de la transformée de Fourier discrète (DFT) à la transformée DCT est le but visé par cette première classe d'approche. Cette propriété consiste en ce qu'une convolution circulaire dans le domaine temporel équivaut à une multiplication dans le domaine de la DFT [Oppenheim 99]. Elle est classiquement exploitée pour réaliser le filtrage FIR rapide par blocs. En effet, les blocs du signal filtré peuvent s'obtenir par multiplication des coefficients de la DFT du filtre et de la FFT du bloc du signal, suivie d'une FFT inverse puis d'une opération d'addition avec recouvrement (Overlap-Add).

La première tentative pour trouver une propriété similaire pour la DCT a été faite dans [Chen 76]. Il a été démontré que la multiplication coefficient par coefficient dans le domaine DCT correspond à une convolution circulaire de trois séquences dans le domaine temporel. Une de ces séquences est un noyau fixe indésirable et qui est un facteur de distorsion. Le travail de [Chitprasert 90] a apporté une simplification en éliminant ce dernier facteur. Il a montré que si la réponse impulsionnelle du filtre FIR est réelle et de symétrie paire alors il suffit d'utiliser sa DFT, à la place de la DCT, dans l'opération de multiplication pour réaliser la bonne convolution circulaire. Malgré cette simplification, cette méthode a l'inconvénient d'introduire des artefacts de type effet de bloc à cause de l'utilisation d'une convolution circulaire au lieu d'une convolution linéaire.

Une généralisation de la propriété CMP a été faite dans [Martucci 94] pour toute la famille des transformées trigonométriques discrètes (DTT : Discrete Trigonometric Transform). Il a ainsi dérivé un ensemble de formules, basées sur les convolutions symétriques (circulaire et "skew-circular"⁴), permettant d'obtenir dans le domaine temporel l'équivalent d'une multiplication coefficient par coefficient d'un couple de transformées appartenant à l'ensemble des DCT et DST types I-IV. Elles peuvent être résumées par la formule suivante :

$$w_n = \mathcal{E}_a \{x_n\} \otimes \mathcal{E}_b \{y_n\} = \mathcal{T}_c^{-1} \{ \mathcal{T}_a \{x_n\} \times \mathcal{T}_b \{y_n\} \}, \quad (3.8)$$

où $\{x_n\}$ et $\{y_n\}$ sont les deux séquences d'entrée de longueur finie et $\{w_n\}$ est la séquence de sortie convoluée. \mathcal{E}_a et \mathcal{E}_b sont des opérateurs d'extension symétrique et \otimes désigne une convolution

⁴ Lors de la permutation circulaire les éléments passant au début de la séquence sont munis d'un signe moins.

symétrique. \mathcal{T}_a et \mathcal{T}_b sont respectivement les transformées DTT des séquences $\{x_n\}$ et $\{y_n\}$, \mathcal{T}_c^{-1} est la transformée inverse appropriée. L'opérateur \times désigne la multiplication élément par élément de ses opérandes.

L'utilisation de ces formules pour réaliser des convolutions linéaires est possible mais nécessite de faire préalablement une opération de "zero-padding" appropriée dans le domaine temporel. Elles ne sont donc pas applicables quand des données sont fournies directement dans le domaine DCT sans traitement a priori par cette opération. En plus, elle est limitée aux seuls cas de filtres symétriques ou antisymétriques.

La solution proposée récemment dans [Kresh 99] est basée sur ces propriétés mais en palliant leurs problèmes. Elle donne une méthode rapide pour effectuer le filtrage linéaire dans le domaine de la transformé DCT type IIe. Le développement de cette solution est basé essentiellement sur la conversion des propriétés de convolution-multiplication en une forme matricielle [Sánchez 95]. Le gain en complexité obtenu par rapport au filtrage dans le domaine spatial est de 35 à 64 %, il dépend de la taille du filtre FIR utilisé. Le meilleur résultat s'obtient pour des filtres causals et symétriques de longue taille.

3.3.1.2 Approche algébrique directe

Les travaux présentés dans [Lee 92, Chang 95] fournissent une méthode algébrique directe comme alternative à l'approche par convolution-multiplication. Ils proposent de trouver l'opérateur équivalent à la mise en cascade de la DCT inverse (IDCT), du filtrage FIR temporel (ou spatial) et de la transformée DCT directe (ces trois blocs correspondent à la partie en gris de la figure 3.1a où le banc de filtres est une DCT dans ce cas). Pour effectuer cette combinaison, la DCT et la DCT inverse sont considérées comme des opérateurs matriciels et le filtre FIR est représenté par sa matrice de convolution. L'opérateur trouvé permet donc de réaliser directement sur les blocs du signal, représentés dans le domaine DCT, l'opération équivalente à la convolution linéaire dans le domaine temporel. Comme elle est décrite dans [Lee 92], cette méthode est applicable à n'importe quelle transformée par bloc (DFT, DCT, WHT,...). Elle a été aussi généralisée pour le cas du filtrage IIR [Kim 95].

Une amélioration de cette approche est donnée dans [Merhav 95a] pour le cas particulier d'une DCT bi-dimensionnelle de taille 8×8 , généralement utilisée dans les algorithmes de compression d'images. L'algorithme rapide dérivé exploite d'une part le recouvrement entre les calculs intervenant pour des blocs adjacents et d'autre part une structure en papillon des données d'entrée permettant d'obtenir des matrices creuses.

Il est envisageable d'utiliser la même méthodologie de l'approche décrite dans [Lee 92], en cas du filtrage FIR, pour l'étendre à des transformées orthogonales à recouvrement (LOT) telle que la MDCT. Par ailleurs, la méthode développée dans le paragraphe 3.4 permet de donner une solution plus générale pour ce problème de filtrage dans le domaine des sous-bandes. Son application pour ces cas particuliers permettra d'arriver aux mêmes résultats.

3.3.1.3 Autre approche

En alternative à ces différentes approches, visant à élaborer des algorithmes exacts, [Merhav 95b] propose d'approximer l'opération de convolution par des coefficients de multiplication aux

coefficients de la DCT. La synthèse de tels coefficients se traduit sous forme d'un problème de minimisation de l'erreur entre le signal filtré de manière directe dans le domaine temporel et celui obtenu par cette méthode. L'auteur a proposé deux solutions. La première se base sur le critère de minimisation de l'erreur quadratique moyenne (MMSE), elle nécessite une connaissance a priori sur les statistiques du signal. La seconde utilise le critère "MinMax". Ces solutions restent par ailleurs valables seulement si le filtre est symétrique. Des résultats satisfaisants ont été obtenus surtout dans le cas où les coefficients centraux du filtre sont relativement dominants. Cette approche est la plus simple en terme de complexité de traitement, complexité qui peut être réduite davantage en intégrant les coefficients synthétisés directement dans les tables de quantification.

3.3.2 Filtrage en sous-bandes du banc de filtres MPEG-1

L'ajout d'effets sonores sur des signaux MPEG-1 dans le domaine compressé a déjà motivé certains travaux auparavant pour réaliser du filtrage dans le domaine des sous-bandes. Le phénomène d'aliasing est le principal problème qui pose des contraintes à cette réalisation. S'il n'est pas convenablement traité, des artefacts peuvent être entendus après synthèse du signal. Les premières tentatives pour résoudre ce problème ont été plutôt approximatives et restrictives [Broadhead 95, Levine 96], le travail de [Lanciani 99a] a donné une solution plus générale.

3.3.2.1 Approches approximatives

Dans [Broadhead 95] une opération d'égalisation est effectuée par application de filtres sur les signaux des sous-bandes. Pour faire face au problème d'aliasing, l'auteur suggère de contraindre les réponses fréquentielles des filtres des sous-bandes à avoir approximativement les mêmes amplitudes aux voisinages des frontières des sous-bandes. Le choix d'une telle solution s'explique par le fait que d'une part le recouvrement entre bandes du banc de filtres MPEG-1 est restreint aux bandes adjacentes, et d'autre part la distorsion d'aliasing est plus importante proche des frontières. Cette stratégie permet donc d'apporter la même modification aux composantes d'aliasing les plus significatives, ainsi, lors de la synthèse leur suppression peut être assurée. Si cette approche élimine les distorsions, elle impose par ailleurs une contrainte limitative sur l'ensemble des filtres qui peuvent être utilisés.

Dans [Levine 96] une implémentation des effets sonores standards (réverbération, effet de chœur, écho) est réalisée par le biais de filtres. Pour l'exemple de la réverbération, l'auteur utilise des filtres IIR passe-tout sur les signaux des dix premières sous-bandes. Cette structure de filtres est classiquement utilisée dans le domaine temporel pour générer de la réverbération [Gardner 98]. La longueur de la ligne à retard utilisée et le gain de la boucle de retour sont directement liés au temps de réverbération. L'auteur choisit d'utiliser un temps de réverbération qui décroît linéairement à partir de la première bande. Le gain de la boucle de retour est le même pour toutes les bandes ce qui permet de déterminer le délai du filtre sur chacune. Dans ce cas l'auteur n'impose aucune contrainte supplémentaire aux filtres pour éviter le problème d'aliasing. Il espère tout simplement que les effets de masquage permettent de rendre ces distorsions inaudibles. La méthode proposée dans ce chapitre permet de donner une solution à ce problème de filtrage IIR.

3.3.2.2 Travaux de Lanciani

Le travail réalisé par Lanciani a consisté à trouver une solution générale au problème du filtrage à réponse impulsionnelle finie dans le domaine des sous-bandes du banc de filtres pseudo-QMF des

codeurs MPEG-1 et MPEG-2 BC Layer I et II [Lanciani 99a]. Sa méthode est par ailleurs utilisable pour n'importe quel banc de filtres FIR à échantillonnage critique et vérifiant la condition de reconstruction parfaite. Dans son principe, la méthode adoptée est similaire à l'approche algébrique directe utilisée pour le filtrage dans le domaine transformé [Lee 92] (voir paragraphe 3.2.1.2). La résolution du problème se base sur une représentation temporelle des bancs de filtres d'analyse et de synthèse tirée des travaux de [Nayebi 90]. Elle utilise les matrices de convolution de ces deux bancs de filtres et aussi du filtre temporel à transposer dans le domaine des sous-bandes.

En effet, pour un filtre, $S(z)$, à réponse impulsionnelle finie de longueur N , sa matrice de convolution est de taille $(2N-1) \times N$ et s'écrit :

$$\mathbf{S} = \begin{bmatrix} s[0] & 0 & \cdots & 0 \\ s[1] & s[0] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ s[N-1] & s[N-2] & \cdots & s[0] \\ 0 & s[N-1] & \cdots & s[1] \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s[N-1] \end{bmatrix}. \quad (3.9)$$

Les bancs de filtres d'analyse et de synthèse sont représentés respectivement par les matrices de convolution \mathcal{H} et \mathcal{F} , de taille $2N \times N$, et qui ont les formes suivantes :

$$\mathcal{H} = [\mathcal{H}_0 \quad \mathcal{H}_1 \quad \cdots \quad \mathcal{H}_{M-1}]^T, \quad (3.10)$$

$$\mathcal{F} = [\mathcal{F}_0 \quad \mathcal{F}_1 \quad \cdots \quad \mathcal{F}_{M-1}], \quad (3.11)$$

où \mathcal{H}_k et \mathcal{F}_k , $0 \leq k \leq M-1$, correspondent respectivement aux matrices de convolution des filtres d'analyse, $H_k(z)$, et de synthèse, $F_k(z)$, qui ont été décimées par un facteur M sur les lignes pour les premiers et sur les colonnes pour les seconds.

La matrice de convolution du filtrage dans le domaine des sous-bandes s'exprime alors comme suit :

$$\mathcal{S}_{\text{sb}} = \mathcal{H}'_{2N \times 2N} \mathcal{S}'_{2N \times 2N} \mathcal{F}_{2N \times N}, \quad (3.12)$$

où \mathcal{S}' est formée par extension de la matrice \mathcal{S} suivant l'équation (3.13). \mathcal{H}' s'obtient de manière similaire à partir de \mathcal{H} en utilisant la même extension pour chacune de ces sous-matrices \mathcal{H}_k . Ces extensions sont réalisées pour avoir des matrices ayant des dimensions compatibles avec le produit matriciel utilisé.

$$\mathcal{S}' = \begin{bmatrix} \left[\begin{array}{c} \mathbf{S} \\ \mathbf{0}_{1 \times N} \end{array} \right] & \begin{array}{cccc} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \\ s[0] & 0 & \cdots & 0 \\ s[1] & s[0] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ s[N-2] & s[N-3] & \cdots & 0 \\ s[N-1] & s[N-2] & \cdots & s[0] \end{array} \end{bmatrix}. \quad (3.13)$$

Les mêmes représentations des bancs de filtres ont été utilisées pour réaliser la conversion de la fréquence d'échantillonnage dans le domaine des sous-bandes [Lanciani 99b]. Les opérations telles que la décimation, interpolation et ré-échantillonnage par un facteur rationnel [Crochiere 81] sont transposées dans ce domaine.

Si cette solution de Lanciani peut être considérée comme l'état de l'art sur le filtrage dans le domaine des sous-bandes, elle reste par ailleurs limitée au seul cas des filtres temporels à réponse impulsionnelle finie. L'utilisation de la représentation en matrices de convolution pour la résolution du problème ne permet pas de l'étendre au cas des filtres IIR.

3.4. Solution générale au problème du filtrage en sous-bandes

3.4.1 Définition du problème

Considérons un banc de filtres à M bandes et à échantillonnage critique donné respectivement par ses filtres d'analyse et de synthèse, $H_k(z)$ et $F_k(z)$, $0 \leq k \leq M-1$. On supposera que ce banc de filtres vérifie la propriété de reconstruction parfaite. Soit $S(z)$ la fonction de transfert scalaire d'un filtre rationnel (FIR ou IIR) quelconque. Cette fonction de transfert représentera le filtrage temporel de la figure 3.1a.

Comme il a été mentionné précédemment, le problème essentiel qui se pose pour le filtrage en sous-bandes concerne la distorsion due aux composantes d'aliasing. Pour assurer leur suppression par le banc de filtres de synthèse, la relation de dépendance qui les relie doit être conservée par l'opération de filtrage en sous-bandes. C'est dans ce but que cette opération est représentée par une matrice de filtres, $\mathbf{S}_{sb}(z)$, (figure 3.5). Le vecteur, $\mathbf{V}(z)$, des signaux des sous-bandes filtrés s'obtient par la relation suivante :

$$\mathbf{V}(z) = \mathbf{S}_{sb}(z) \mathbf{U}(z), \quad (3.14)$$

avec $\mathbf{U}(z) = [U_0(z) \ U_1(z) \ \cdots \ U_{M-1}(z)]^T$ et $\mathbf{V}(z) = [V_0(z) \ V_1(z) \ \cdots \ V_{M-1}(z)]^T$.

Le signal de sortie, $V_n(z)$, de la sous-bande n résulte donc d'une combinaison linéaire de tous les signaux des sous-bandes, $U_k(z)$, filtrés par $S_{sb,nk}(z)$, $0 \leq k \leq M-1$. Il contient, a priori, la contribution de toutes les bandes adjacentes y compris des composantes d'aliasing qui y sont contenues.

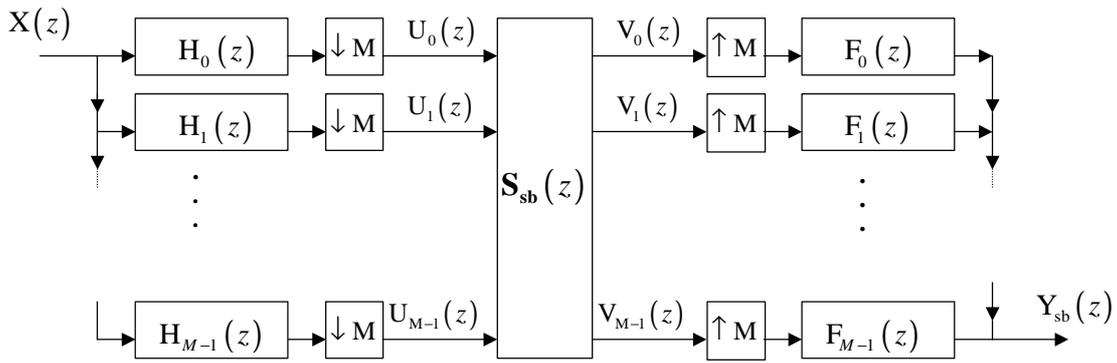


Fig. 3.5 : Filtrage dans le domaine des sous-bandes.

Les schémas de la figure 3.1 ont déjà présenté le problème à résoudre pour le filtrage dans le domaine des sous-bandes. Par ailleurs, pour simplifier la résolution on peut le reformuler comme dans la figure 3.6. Ainsi, dans ce cas, on cherche à trouver la matrice de filtrage dans les sous-bandes, $S_{sb}(z)$, permettant d'obtenir le même signal, après synthèse, que celui obtenu par filtrage temporel direct, par $S(z)$, sur le signal synthétisé.

Il est aisé de voir que cette dernière formulation implique la première. En effet, si les signaux en sortie des figures 3.6a et 3.6b sont équivalents alors il sera de même pour le signal au point A de la figure 3.1a et du signal de sortie de la figure 3.1b. Le banc de filtres étant à reconstruction parfaite, la succession des deux dernières opérations d'analyse et de synthèse de la figure 3.1a est équivalente à un simple retard. D'où l'équivalence des signaux de sortie des figures 3.1a et 3.1b.

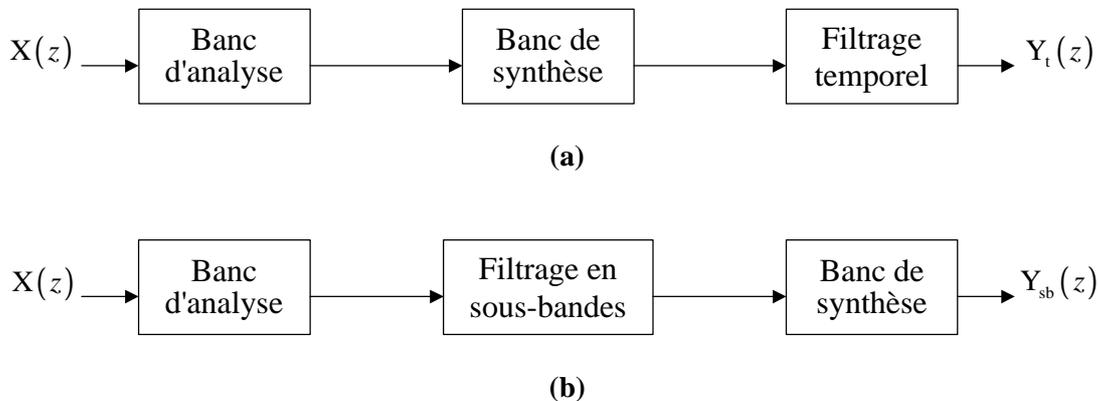


Fig. 3.6 : Schémas blocs définissant l'équivalence entre (a) filtrage dans le domaine temporel (b) filtrage dans le domaine des sous-bandes.

La représentation des bancs de filtres donnée dans la figure 3.5 n'est pas très adéquate pour trouver une solution mathématique au problème de filtrage dans le domaine des sous-bandes. Nous utiliserons des matrices de transformée en z pour représenter les différents blocs du problème décrits dans la figure 3.6. Les bancs de filtres seront alors présentés par leurs matrices polyphasées (figure 2.3) et le filtre temporel par sa version bloc. Comme il sera démontré dans la suite, cette représentation permettra de trouver une solution générale à ce problème, valable pour n'importe quel type de filtres (FIR ou IIR).

3.4.2 Représentation bloc du filtre scalaire temporel

Soient $x[i]$ et $y[i]$ les signaux d'entrée et de sortie du filtre temporel représenté par sa fonction de transfert scalaire $S(z)$. Considérons $\mathbf{x}[j]$ et $\mathbf{y}[j]$ les vecteurs correspondants à la mise en blocs, de taille M , de ces signaux :

$$\mathbf{x}[j] = [x_{M-1}[j] \quad \cdots \quad x_1[j] \quad x_0[j]]^T, \quad (3.15)$$

$$\mathbf{y}[j] = [y_{M-1}[j] \quad \cdots \quad y_1[j] \quad y_0[j]]^T, \quad (3.16)$$

avec $x_n[j] = x[jM + n]$ et $y_n[j] = y[jM + n]$, pour $0 \leq n \leq M - 1$. Les éléments des vecteurs $\mathbf{x}[j]$ et $\mathbf{y}[j]$ sont respectivement les composantes polyphasées des signaux $x[i]$ et $y[i]$.

La version bloc du filtre scalaire $S(z)$ peut être représentée par une matrice de transfert $\mathbf{S}(z)$ qui caractérise un système à M entrées - M sorties (figure 3.7b). On a :

$$\mathbf{Y}(z) = \mathbf{S}(z)\mathbf{X}(z), \quad (3.17)$$

où $\mathbf{Y}(z) = \sum_{j=0}^{+\infty} \mathbf{y}[j]z^{-j}$ et $\mathbf{X}(z) = \sum_{j=0}^{+\infty} \mathbf{x}[j]z^{-j}$ sont respectivement les vecteurs transformés en z des vecteurs $\mathbf{x}[j]$ et $\mathbf{y}[j]$.

Les opérations de conversion série-parallèle puis de conversion parallèle-série après filtrage peuvent être facilement représentées en utilisant une chaîne de délais et des décimations / expansions. Ainsi, on obtient le schéma de filtrage par bloc de la figure 3.7c utilisant des opérations multicaudence.

La matrice de transfert $\mathbf{S}(z)$ s'obtient directement à partir des composantes polyphasées $S_k(z)$, $0 \leq k \leq M - 1$, du filtre scalaire $S(z)$ [Vaidyanathan 88]. Ainsi, si :

$$S(z) = \sum_{k=0}^{M-1} z^{-k} S_k(z^M), \quad (3.18)$$

$$\text{alors } \mathbf{S}(z) = \begin{bmatrix} S_0(z) & S_1(z) & \cdots & & S_{M-1}(z) \\ z^{-1} S_{M-1}(z) & S_0(z) & S_1(z) & \cdots & S_{M-2}(z) \\ z^{-1} S_{M-2}(z) & z^{-1} S_{M-1}(z) & S_0(z) & S_1(z) & \cdots & S_{M-3}(z) \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ z^{-1} S_1(z) & \cdots & & z^{-1} S_{M-1}(z) & S_1(z) & S_0(z) \end{bmatrix}. \quad (3.19)$$

Cette matrice possède une structure *pseudo-circulante*. Chaque ligne n , $1 \leq n \leq M - 1$, de $\mathbf{S}(z)$ s'obtient par décalage circulaire, de n unités, des éléments de la ligne 0 et de plus les éléments sous la diagonale sont multipliés par z^{-1} .

On peut remarquer la similarité entre la représentation bloc d'un filtre scalaire (figure 3.7c) et la représentation polyphasée d'un banc de filtres (figure 2.3). En fait, pour un banc de filtres QMF à M bandes et à échantillonnage critique la structure pseudo-circulante de la matrice de transfert

$\mathbf{L}(z) = \mathbf{R}(z)\mathbf{E}(z)$, produit des matrices d'analyse et de synthèse, est une condition nécessaire et suffisante pour la suppression d'aliasing [Vaidyanathan 88]. Si cette condition est vérifiée, le banc de filtres d'analyse/synthèse se réduit à une fonction de transfert qui s'écrit sous la forme suivante :

$$T(z) = z^{-(M-1)} V(z), \quad (3.20)$$

où $V(z)$ est un filtre scalaire. L'élimination de la distorsion d'amplitude est assurée, i.e. le filtre $V(z)$ est passe-tout, si et seulement si la matrice $\mathbf{L}(z)$ est en plus sans pertes (i.e. paraunitaire et stable).

La relation (3.19) permet d'obtenir la matrice de transfert $\mathbf{S}(z)$ connaissant les composantes polyphasées du filtre scalaire $S(z)$. Si ce filtre est à réponse impulsionnelle finie (FIR), ces composantes s'obtiennent de manière directe à partir de l'expression de sa transformée en z . Le cas des filtres à réponse impulsionnelle infinie (IIR) n'est pas si direct. L'annexe A explicite une méthode pour déterminer leurs composantes polyphasées et ainsi la matrice de transfert.

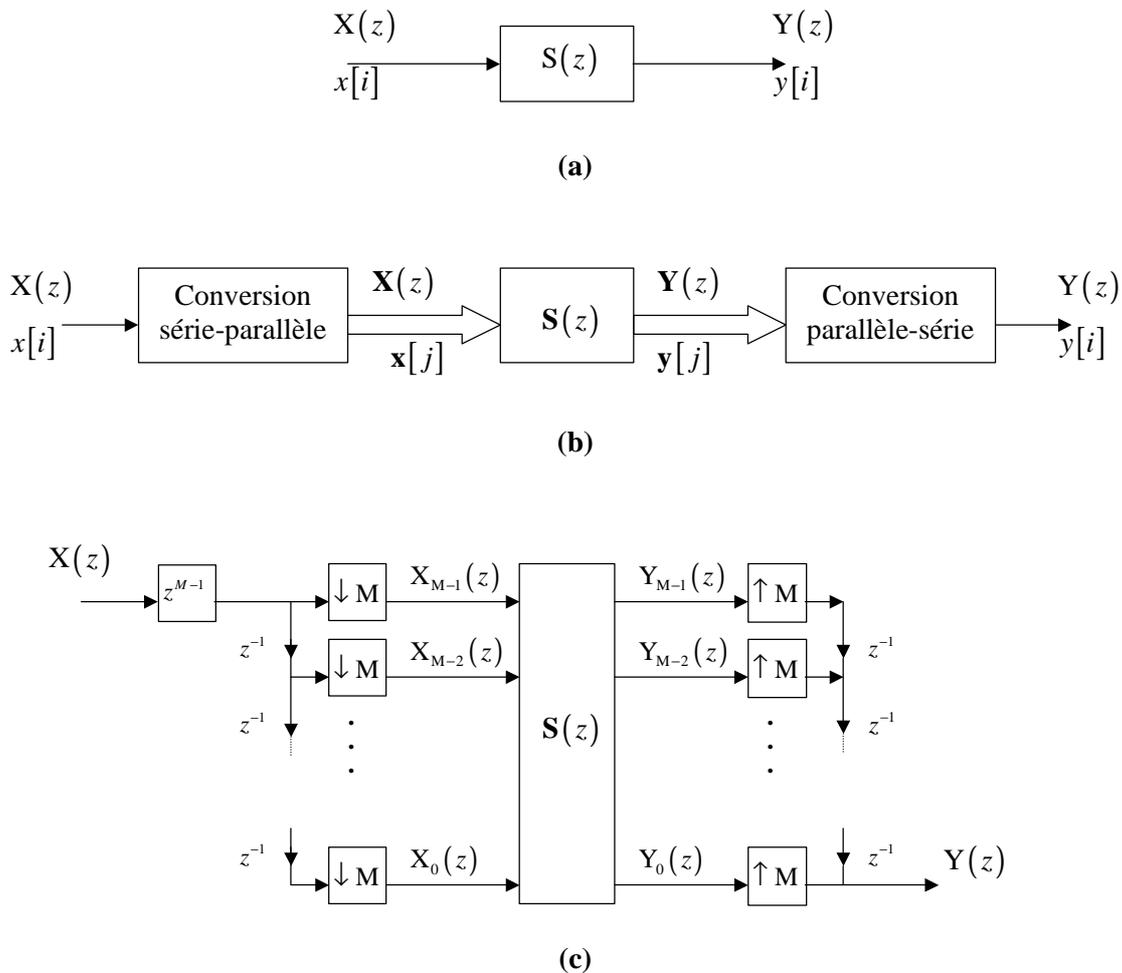


Fig. 3.7 : Filtrage temporel (a) fonction de transfert scalaire (b) implantation en version bloc (c) représentation de la version bloc par des opérations multicaudence.

3.4.3 Formulation de la matrice du filtrage en sous-bandes

La représentation polyphasée des bancs de filtres ainsi que la représentation bloc du filtre temporel permettent de décrire le problème du filtrage en sous-bandes par des blocs multicadence en utilisant des matrices de transfert en z . Ainsi, les figures 3.6a et 3.6b peuvent être remplacées, respectivement, par les figures 3.8a et 3.9. Dans la figure 3.8a, il est aisé de voir que la mise en cascade des deux blocs de conversion parallèle-série puis série-parallèle équivaut à l'identité ce qui permet d'obtenir le schéma équivalent de la figure 3.8b.

En se basant sur les figures 3.8b et 3.9, on peut écrire les expressions des signaux de sorties obtenues respectivement par filtrage dans le domaine temporel et en sous-bandes :

$$\mathbf{Y}_t(z) = \mathbf{S}(z)\mathbf{R}(z)\mathbf{E}(z)\mathbf{X}(z), \quad (3.21)$$

$$\mathbf{Y}_{sb}(z) = \mathbf{R}(z)\mathbf{S}_{sb}(z)\mathbf{E}(z)\mathbf{X}(z), \quad (3.22)$$

avec $\mathbf{X}(z) = [X_0(z) \ X_{M-1}(z) \ \cdots \ X_1(z)]^T$, où $X_k(z)$ sont les composantes polyphasées de $\mathbf{X}(z)$, et $\mathbf{Y}_t(z) = [Y_{t,M-1}(z) \ \cdots \ Y_{t,1}(z) \ Y_{t,0}(z)]^T$, $\mathbf{Y}_{sb}(z) = [Y_{sb,M-1}(z) \ \cdots \ Y_{sb,1}(z) \ Y_{sb,0}(z)]^T$.

Le banc de filtres est supposé à reconstruction parfaite, les matrices $\mathbf{E}(z)$ et $\mathbf{R}(z)$ vérifient alors la condition générale (2.12). Elles sont donc inversibles et on a la relation suivante :

$$\mathbf{R}^{-1}(z) = c^{-1}z^\lambda \mathbf{E}(z)\mathbf{\Lambda}_r(z), \quad (3.23)$$

$$\text{avec } \mathbf{\Lambda}_r(z) = \mathbf{\Gamma}_r^{-1}(z) = \begin{bmatrix} \mathbf{0} & z\mathbf{I}_r \\ \mathbf{I}_{M-r} & \mathbf{0} \end{bmatrix}.$$

En égalisant les équations (3.21) et (3.22) puis en utilisant la relation (3.23) on en déduit l'expression de la matrice de filtrage en sous-bandes :

$$\mathbf{S}_{sb}(z) = \mathbf{R}^{-1}(z)\mathbf{S}(z)\mathbf{R}(z), \quad (3.24)$$

puis

$$\mathbf{S}_{sb}(z) = c^{-1}z^\lambda \mathbf{E}(z)\mathbf{\Lambda}_r(z)\mathbf{S}(z)\mathbf{R}(z). \quad (3.25)$$

Cette dernière relation permet donc de déterminer la matrice $\mathbf{S}_{sb}(z)$, connaissant les matrices polyphasées d'analyse et de synthèse du banc de filtres et la version bloc du filtre temporel. Notons que la condition de reconstruction parfaite du banc de filtres est la seule hypothèse effectuée pour aboutir à cette solution générale. Si on se restreint à l'hypothèse où le banc de filtres vérifie la condition suffisante $\mathbf{R}(z)\mathbf{E}(z) = z^{-K}\mathbf{I}_M$ (i.e. cas particulier où $r = 0$) alors la relation (3.25) se réduit à la forme suivante :

$$\mathbf{S}_{sb}(z) = z^K \mathbf{E}(z)\mathbf{S}(z)\mathbf{R}(z). \quad (3.26)$$

Dans la suite de ce chapitre on se placera sous cette hypothèse et on utilisera donc la solution donnée par cette dernière formulation.

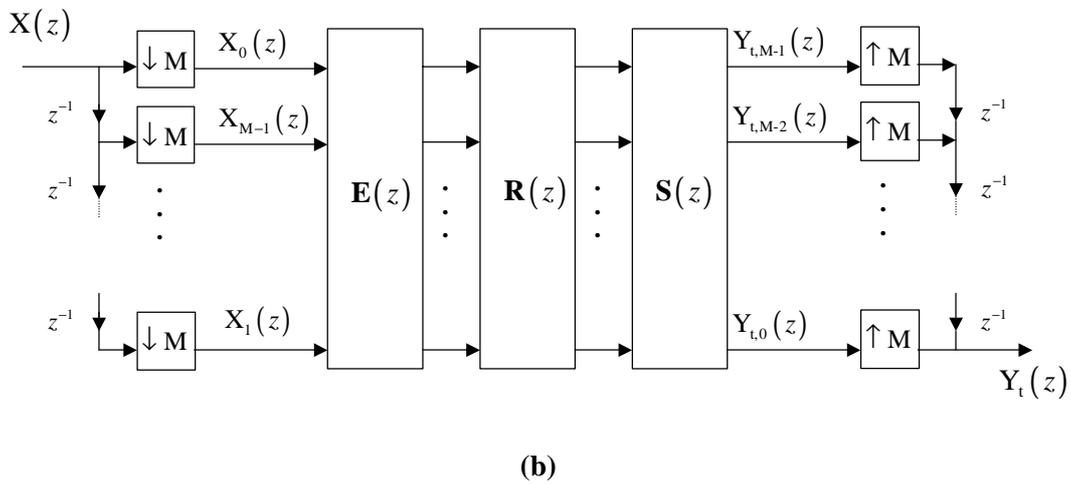
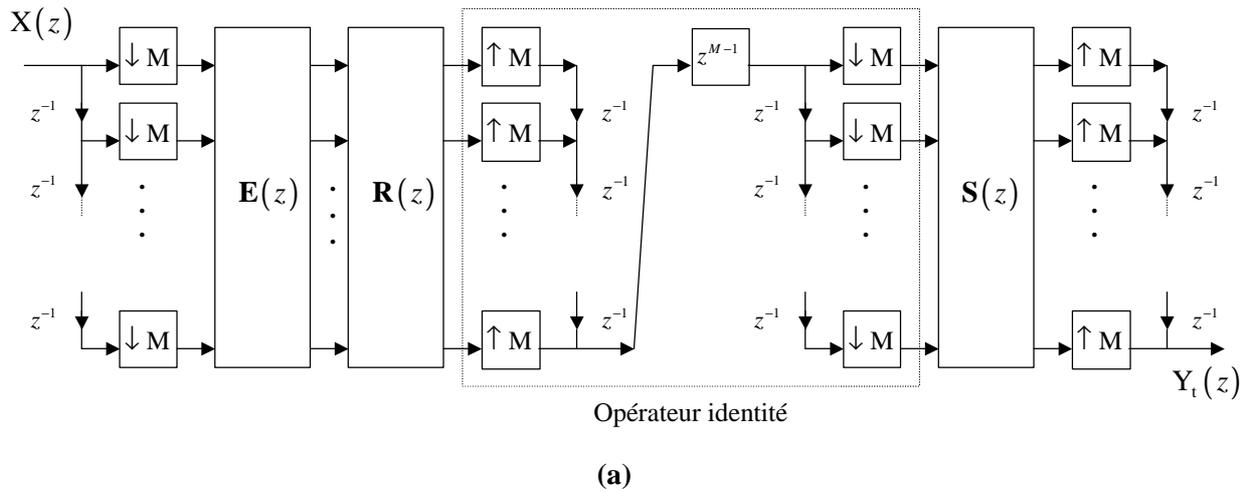


Fig. 3.8 : Représentation en blocs multiscalaire du schéma du filtrage temporel (a) schéma initial (b) schéma équivalent.

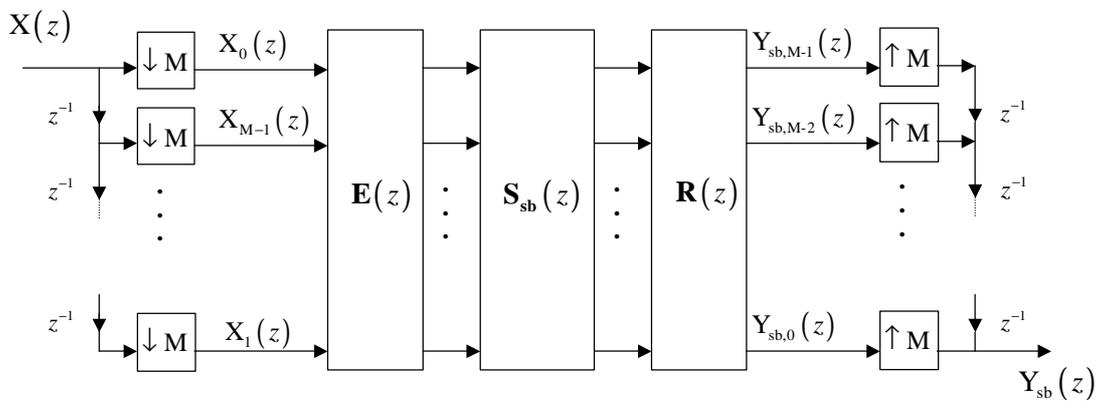


Fig. 3.9 : Représentation en blocs multiscalaire du schéma du filtrage dans les sous-bandes.

3.4.4 Autre formulation de la matrice du filtrage en sous-bandes

Les filtres des sous-bandes peuvent être exprimés directement à partir des filtres d'analyse/synthèse et de la fonction de transfert scalaire du filtre temporel, par les formules suivantes :

$$\mathbf{S}_{\text{sb},nk}(z) = \left[z^{L-1} \mathbf{H}_n(z) \mathbf{S}(z) \mathbf{F}_k(z) \right] \Big|_{\downarrow M}, \quad 0 \leq n, k \leq M-1, \quad (3.27)$$

où $\downarrow M$ désigne l'opération de décimation par un facteur M .

Ou de manière équivalente, en utilisant une écriture matricielle :

$$\mathbf{S}_{\text{sb}}(z) = \left[z^{L-1} \mathbf{h}(z) \mathbf{S}(z) \mathbf{f}(z) \right] \Big|_{\downarrow M}, \quad (3.28)$$

où $\mathbf{h}(z) = [\mathbf{H}_0(z) \quad \mathbf{H}_1(z) \quad \dots \quad \mathbf{H}_{M-1}(z)]^T$ et $\mathbf{f}(z) = [\mathbf{F}_0(z) \quad \mathbf{F}_1(z) \quad \dots \quad \mathbf{F}_{M-1}(z)]$ sont respectivement les vecteurs des filtres d'analyse et de synthèse.

Preuve : Considérons les matrices $\mathbf{Q}(z)$ et $\mathbf{W}(z)$, intervenant dans le calcul intermédiaire de la matrice $\mathbf{S}_{\text{sb}}(z)$, par la formule (3.26), et qui sont définies comme suit :

$$\mathbf{Q}(z) = \mathbf{E}(z) \mathbf{S}(z), \quad (3.29)$$

$$\mathbf{W}(z) = \mathbf{Q}(z) \mathbf{R}(z). \quad (3.30)$$

En utilisant la structure de la matrice $\mathbf{S}(z)$, donnée par la formule (3.19), on peut voir que les éléments de $\mathbf{Q}(z)$ s'écrivent comme suit :

$$\begin{cases} \mathbf{Q}_{nk}(z) = \sum_{l=0}^k \mathbf{E}_{nl}(z) \mathbf{S}_{(k-l)}(z) + \sum_{l=k+1}^{M-1} \mathbf{E}_{nl}(z) \mathbf{S}_{(M+k-l)}(z) z^{-1}, & 0 \leq k \leq M-2, \\ \mathbf{Q}_{n,(M-1)}(z) = \sum_{l=0}^{M-1} \mathbf{E}_{nl}(z) \mathbf{S}_{(M-1-l)}(z), \end{cases} \quad (3.31)$$

pour $0 \leq n \leq M-1$.

La formule (A.3) de l'annexe A donne la relation entre les composantes polyphasées du produit de deux filtres et de celles de chacun d'eux. La comparaison des équations données dans (3.31), pour n fixé, avec l'équation (A.3) permet de déduire que les fonctions $\mathbf{Q}_{nk}(z)$, $0 \leq k \leq M-1$, éléments de la $n^{\text{ème}}$ ligne de la matrice $\mathbf{Q}(z)$, ne sont autres que les composantes polyphasées du filtre produit $\mathbf{H}_n(z) \times \mathbf{S}(z)$:

$$\mathbf{H}_n(z) \mathbf{S}(z) = \sum_{k=0}^{M-1} z^{-k} \mathbf{Q}_{nk}(z^M), \quad 0 \leq n \leq M-1. \quad (3.32)$$

D'un autre côté, les éléments de la matrice $\mathbf{W}(z)$ s'écrivent comme suit :

$$\mathbf{W}_{nk}(z) = \sum_{l=0}^{M-1} \mathbf{Q}_{nl}(z) \mathbf{R}_{lk}(z), \quad 0 \leq n, k \leq M-1. \quad (3.33)$$

Les fonctions $R_{lk}(z)$, $0 \leq l \leq M-1$, sont les composantes polyphasées de type 2 du filtre de synthèse $F_k(z)$. Elles sont reliées aux composantes polyphasées de type 1, $R'_{lk}(z)$, comme suit :

$$R_{lk}(z) = R'_{(M-1-l)k}(z), \quad 0 \leq l \leq M-1. \quad (3.34)$$

L'équation (3.33) devient donc :

$$W_{nk}(z) = \sum_{l=0}^{M-1} Q_{n1}(z) R'_{(M-1-l)k}(z), \quad 0 \leq n, k \leq M-1. \quad (3.35)$$

Cette dernière écriture permet de déduire, en regardant encore une fois les formules (A.3), que la fonction $W_{nk}(z)$ correspond à la $(M-1)^{\text{ème}}$ composante polyphasée du filtre produit $[H_n(z)S(z)] \times F_k(z)$.

Pour un filtre quelconque, $A(z)$, sa $l^{\text{ème}}$ composante polyphasée de type 1, $A_l(z)$, peut être exprimée par [Vaidyanathan 93] :

$$A_l(z) = \left[z^l A(z) \right] \Big|_{\downarrow M}, \quad 0 \leq l \leq M-1. \quad (3.36)$$

D'après ce qui précède, on peut conclure que les éléments de la matrice $\mathbf{W}(z)$ peuvent s'écrire finalement comme suit :

$$W_{nk}(z) = \left[z^{M-1} H_n(z) S(z) F_k(z) \right] \Big|_{\downarrow M}, \quad 0 \leq n, k \leq M-1. \quad (3.37)$$

Si on tient compte de l'avance z^K de la formule (3.26), et sachant que $z^K = \left[z^{KM} \right] \Big|_{\downarrow M}$, on en déduit que les éléments de la matrice du filtrage en sous-bandes, $\mathbf{S}_{sb}(z)$, sont donnés par :

$$S_{sb,nk}(z) = \left[z^{(K+1)M-1} H_n(z) S(z) F_k(z) \right] \Big|_{\downarrow M}, \quad 0 \leq n, k \leq M-1. \quad (3.38)$$

or, d'après le paragraphe 2.3.2, on sait que la longueur, L , des filtres d'analyse et de synthèse FIR s'écrit :

$$L = (K+1)M. \quad (3.39)$$

Ce qui permet de déduire finalement l'expression de la formule (3.27).

Cette dernière formulation de la matrice de transfert du filtrage en sous-bandes, donnée par l'équation (3.28), peut être considérée comme la forme équivalente, dans le domaine de la transformée en z , de l'expression (3.12) développée dans [Lanciani 99a] et basée sur des matrices de convolution. Elle généralise donc cette dernière pour l'ensemble des filtres rationnels.

3.4.5 Interprétation et forme réduite de la matrice de filtrage

La relation (3.27) montre que les filtres des sous-bandes, $S_{sb,nk}(z)$, $0 \leq k \leq M-1$, permettant d'obtenir le signal filtré dans la sous-bande n , s'obtiennent en multipliant le filtre temporel $S(z)$ par les filtres d'analyse $H_n(z)$ et de synthèse $F_k(z)$. Cette multiplication est suivie d'une avance z^{L-1} qui élimine le retard introduit par ces deux derniers. Le filtre qui en résulte est finalement sous-

échantillonné par un facteur M . Les filtres d'analyse et de synthèse ayant les mêmes gabarits d'amplitude de la réponse fréquentielle, la relation (3.27) permet de voir que pour une sous-bande n , parmi les filtres des sous bandes, $S_{sb,nk}(z)$, seules les bandes k qui ont un recouvrement suffisant avec cette bande ont une énergie significative. Ce résultat est bien attendu puisque seuls les filtres d'analyse de ces sous-bandes adjacentes produisent des composantes d'aliasing importantes dans la sous-bande n considérée. Pour remédier à ces distorsions, les filtres des sous-bandes apportent un effet proportionnel à leurs niveaux d'énergie.

Dans la matrice $\mathbf{S}_{sb}(z)$, il suffit donc de retenir les filtres $S_{sb,nk}(z)$ tel que $(n \ominus \delta) \leq k \leq (n \oplus \delta)$, avec δ est le nombre de bandes qui se recouvrent suffisamment d'un côté avec la bande n et \oplus, \ominus désignent respectivement l'opération d'addition et de soustraction modulo M . Ce nombre δ dépend du banc de filtres utilisé. Chaque ligne n , $0 \leq n \leq M - 1$, de cette matrice se réduit alors à la forme suivante :

$$\left[0 \quad \cdots \quad S_{sb,n(n \ominus \delta)}(z) \quad \cdots \quad S_{sb,nn}(z) \quad \cdots \quad S_{sb,n(n \oplus \delta)}(z) \quad \cdots \quad 0 \right], \quad (3.40)$$

3.5. Filtrage IIR en sous-bande : génération de la matrice de filtrage et structure de l'implantation

Pour le filtrage à réponse impulsionnelle finie, la formule (3.26) ou (3.27) permettent de générer directement les filtres des sous-bandes et en tenant compte de la réduction donnée par la forme (3.40). L'implantation dans ce cas de figure a été déjà étudiée dans [Lanciani 99a]. Dans le cas du banc de filtres MPEG-1, le signal de chaque sous-bande se présente sous forme d'un bloc d'échantillons dans chaque trame (de taille 12 pour Layer I et 36 pour Layer II). L'auteur a donc utilisé la technique de convolution par bloc suivie d'une addition avec recouvrement (overlap-add) [Oppenheim 99] pour réaliser une implantation efficace.

Dans ce paragraphe, on suppose que le filtre temporel $S(z)$ est à réponse impulsionnelle infinie. Pour générer la matrice de filtrage en sous-bandes lui correspondant il faut d'abord déterminer sa matrice de transfert basée sur ses composantes polyphasées. L'annexe A donne une méthode pour décomposer ce type de filtres. Les composantes polyphasées dans ce cas ont la forme suivante :

$$S_k(z) = \frac{N_k(z)}{D(z)}, \quad 0 \leq k \leq M - 1, \quad (3.41)$$

où $D(z)$ est un dénominateur commun à toutes les composantes.

La matrice de filtrage en sous-bandes s'écrit donc comme suit :

$$\mathbf{S}_{sb}(z) = \frac{1}{D(z)} \mathbf{N}_{sb}(z), \quad (3.42)$$

avec $\mathbf{N}_{sb}(z) = z^K \mathbf{E}(z) \mathbf{N}(z) \mathbf{R}(z)$, où $\mathbf{N}(z)$ est une matrice de transfert de filtres FIR contenant les numérateurs de $\mathbf{S}(z)$.

Cette dernière expression suggère d'implanter les filtres des sous-bandes sous une forme canonique directe [Oppenheim 99] et en factorisant la boucle de retour entre les filtres de la même ligne de la matrice $\mathbf{S}_{\text{sb}}(z)$. En effet, si :

$$\mathbf{V}(z) = \mathbf{S}_{\text{sb}}(z) \mathbf{U}(z), \quad (3.43)$$

on définit le vecteur $\mathbf{G}(z) = [G_0(z) \ G_1(z) \ \dots \ G_{M-1}(z)]^T$, pour le calcul intermédiaire, comme suit :

$$\mathbf{G}(z) = \frac{1}{D(z)} \mathbf{U}(z). \quad (3.44)$$

Ainsi on a :

$$\mathbf{V}(z) = \mathbf{N}_{\text{sb}}(z) \mathbf{G}(z). \quad (3.45)$$

Si le dénominateur, $D(z)$, et les numérateurs, $N_{\text{sb},nk}(z)$, s'écrivent respectivement comme suit :

$$D(z) = 1 - \sum_{j=1}^{L_d-1} b[j] z^{-j}, \quad (3.46)$$

$$N_{\text{sb},nk}(z) = \sum_{j=0}^{L_{nk}-1} a_{nk}[j] z^{-j}, \quad 0 \leq n, k \leq M-1. \quad (3.47)$$

Alors les équations de l'implantation sont données par :

$$\begin{cases} g_n[i] = u_n[i] + \sum_{j=1}^{L_d-1} b[j] g_n[i-j], \\ v_n[i] = \sum_{k \in \mathfrak{S}_n} \sum_{j=0}^{L_{nk}-1} a_{nk}[j] g_k[i-j], \end{cases} \quad (3.48)$$

où $0 \leq n \leq M-1$ et \mathfrak{S}_n désigne l'ensemble des indices des filtres de sous-bandes pris en compte pour obtenir le signal filtré de la sous-bande n . Sous cette forme d'implantation, une seule ligne à délais d'une longueur égale à $\max\{L_d, \max_{n,k}\{L_{nk}\}\}$, est utilisée pour chaque composante du vecteur à calculer.

3.6. Application au banc de filtres MPEG-1

Le banc de filtres utilisé dans MPEG-1 est un Pseudo-QMF obtenu par modulation d'un filtre prototype, $H(z)$, FIR à phase linéaire de longueur $L = 2mM$, avec $M = 32$ et $m = 8$. Ce banc de filtres ne vérifie pas exactement la propriété de reconstruction parfaite mais le filtre prototype est suffisamment bien optimisé pour supposer que cette condition est réalisée. On peut donc l'assimiler à un banc de filtres FIR à cosinus modulé et à reconstruction parfaite. Les expressions des matrices polyphasées d'analyse, $\mathbf{E}(z)$, et de synthèse, $\mathbf{R}(z)$, de ce type de banc de filtres ont déjà été données dans le paragraphe 2.3.3. Ces deux matrices sont d'ordre $K = 2m-1$ et la solution au problème de filtrage en sous-bandes s'écrit dans ce cas comme suit :

$$\mathbf{S}_{\text{sb}}(z) = z^{2m-1} \mathbf{E}(z) \mathbf{S}(z) \mathbf{R}(z). \quad (3.49)$$

Pour ce banc de filtres le recouvrement entre bandes de la réponse fréquentielle ne dépasse pas les deux bandes adjacentes (voir figure 2.10). En plus il y a une atténuation de 96 dB dans la bande de rejet. On choisit donc de prendre $\delta = 1$ et chaque ligne de la matrice $\mathbf{S}_{\text{sb}}(z)$ se réduit alors au filtre de la diagonale plus les deux filtres adjacents.

Dans la suite nous donnerons quelques exemples du filtrage en sous-bandes pour ce banc de filtres. Ils permettront d'illustrer certains problèmes qui se manifestent dans ce cas.

3.6.1 Filtre retard pur

Considérons un filtre retard $S(z) = z^{-d}$, $d \in \mathbb{N}^*$. Si on écrit $d = rM + l$, avec $r \in \mathbb{N}$ et l un entier compris entre 0 et $M - 1$, alors ses composantes polyphasées sont données par :

$$S_l(z) = z^{-r} \quad \text{et} \quad S_k(z) = 0 \quad \text{pour } k \neq l, \quad k = 0, \dots, M - 1. \quad (3.50)$$

La matrice de transfert correspondant à la version bloc s'écrit donc comme suit :

$$\mathbf{S}(z) = \begin{bmatrix} \mathbf{0} & \mathbf{I}_{M-1} \\ z^{-1} \mathbf{I}_1 & \mathbf{0} \end{bmatrix} z^{-r}. \quad (3.51)$$

Cas d'un retard multiple de M :

Dans ce cas $\mathbf{S}(z) = z^{-r} \mathbf{I}_M$. En tenant compte de la relation entre $\mathbf{E}(z)$ et $\mathbf{R}(z)$, donnée par (2.11), on obtient la matrice de transfert du filtrage en sous-bandes :

$$\mathbf{S}_{\text{sb}}(z) = z^{-r} \mathbf{I}_M. \quad (3.52)$$

Cette matrice est réduite donc à un simple délai r sur les signaux de chaque sous-bande. Ce résultat est bien évident puisque effectuer un retard de r échantillons dans le domaine des sous-bandes, est équivalent à réaliser un retard de rM échantillons dans le domaine temporel. La figure 3.10 illustre les réponses impulsionnelles pour ce cas.

Cas d'un retard non multiple de M :

Dans ce cas, il n'est pas possible d'éliminer l'avance z^{2m-1} dans la formule (3.49). Pour tenir compte de cette avance, il faut tronquer les réponses impulsionnelles des filtres des sous-bandes (figure 3.11b), ce qui donne des artefacts lors de l'écoute du signal synthétisé. On peut envisager aussi de remplacer (3.49) par :

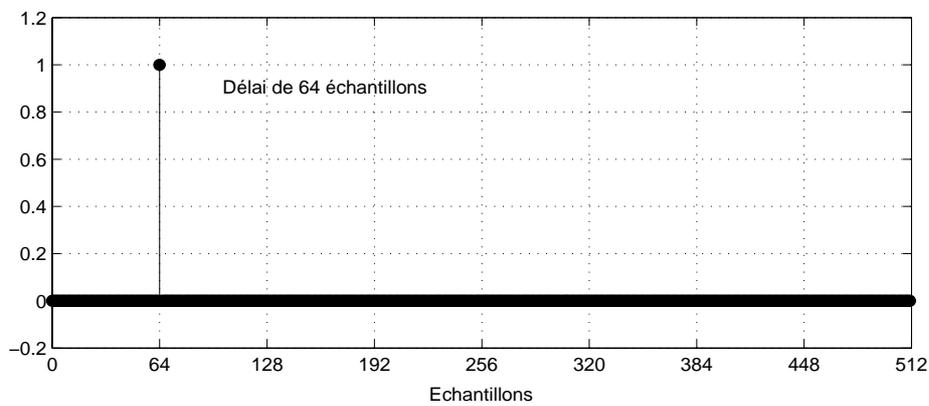
$$\mathbf{S}_{\text{sb}}(z) = \mathbf{E}(z) \mathbf{S}(z) \mathbf{R}(z). \quad (3.53)$$

Par ailleurs, cette nouvelle solution introduit un retard supplémentaire de $KM = 480$ échantillons. Aucune réduction de délai ne sera donc réalisée pour le signal de sortie synthétisé de la figure 3.1b par rapport à celui de la figure 3.1a obtenu par un schéma de filtrage dans le domaine temporel. La figure 3.11 illustre les réponses impulsionnelles des filtres des sous-bandes utilisés pour obtenir le signal filtré de la bandes 0, avec et sans troncature.

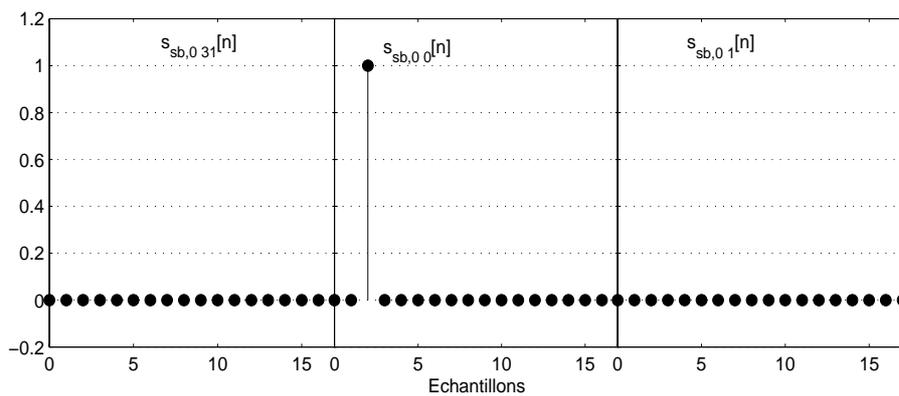
Pour comprendre ce problème, considérons les filtres des sous-bandes sous leur seconde forme donnée par la formule (3.27). La convolution des filtres d'analyse, $H_n(z)$, et de synthèse, $F_k(z)$,

donne une réponse impulsionnelle qui s'annule périodiquement sur tous les points multiples de $M = 32$ sauf sur son point de symétrie où elle est maximale. L'élimination de leur délai se fait par z^{L-1} et permet de centrer cette réponse impulsionnelle autour de zéro. Après convolution avec le filtre $S(z)$, si le délai apporté par ce dernier est multiple de M , les points obtenus par sous-échantillonnage, d'un facteur M , de la réponse impulsionnelle résultante seront tous nuls sauf le point de symétrie. Dans le cas contraire, cette propriété ne sera plus vérifiée et les points de la partie non causale ne seront pas nuls. Ces deux cas de figure sont bien illustrés dans la figure 3.12 pour des filtres retard de 64 et 50 échantillons.

Pour limiter le retard additionnel, il est possible de ne considérer que les points significatifs des réponses impulsionnelles des filtres des sous-bandes. Pour l'exemple de la figure 3.11c, on peut éliminer les 10 premiers points. Le retard total additionnel sera donc de $(480 - 10 \times 32) = 160$ échantillons, soit une réduction de délai de la figure 3.1b par rapport à la figure 3.1a de 320 échantillons.

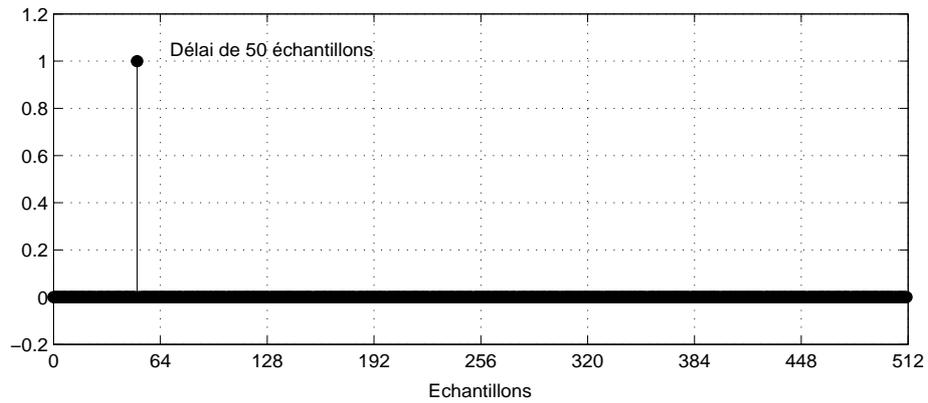


(a)

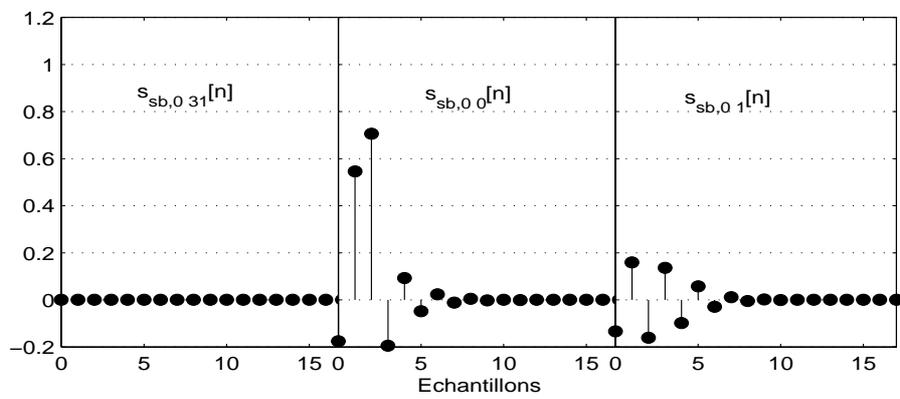


(b)

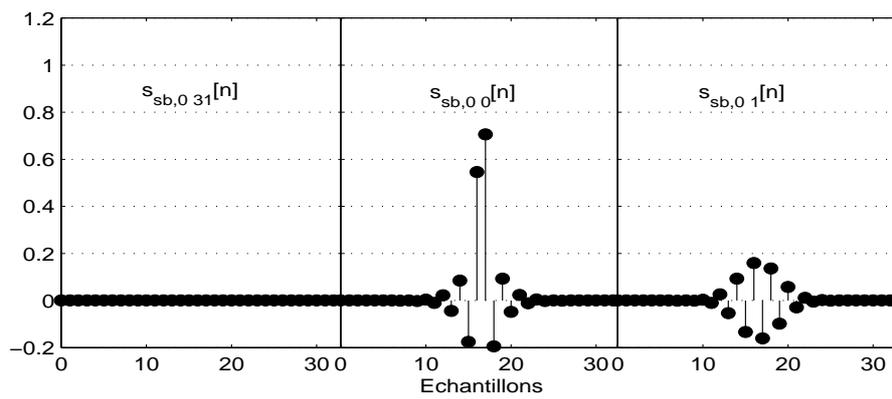
Fig. 3.10 : Filtre délai de 64 échantillons (a) réponse impulsionnelle temporelle (b) réponses impulsionnelles des filtres de sous-bandes correspondant de la bande 0.



(a)



(b)



(c)

Fig. 3.11 : Filtre délai de 50 échantillons : réponse impulsionnelle (a) temporelle (b) des filtres sous-bandes correspondant de la bande 0 (c) après ajout du retard $z^{-(2m-1)}$.

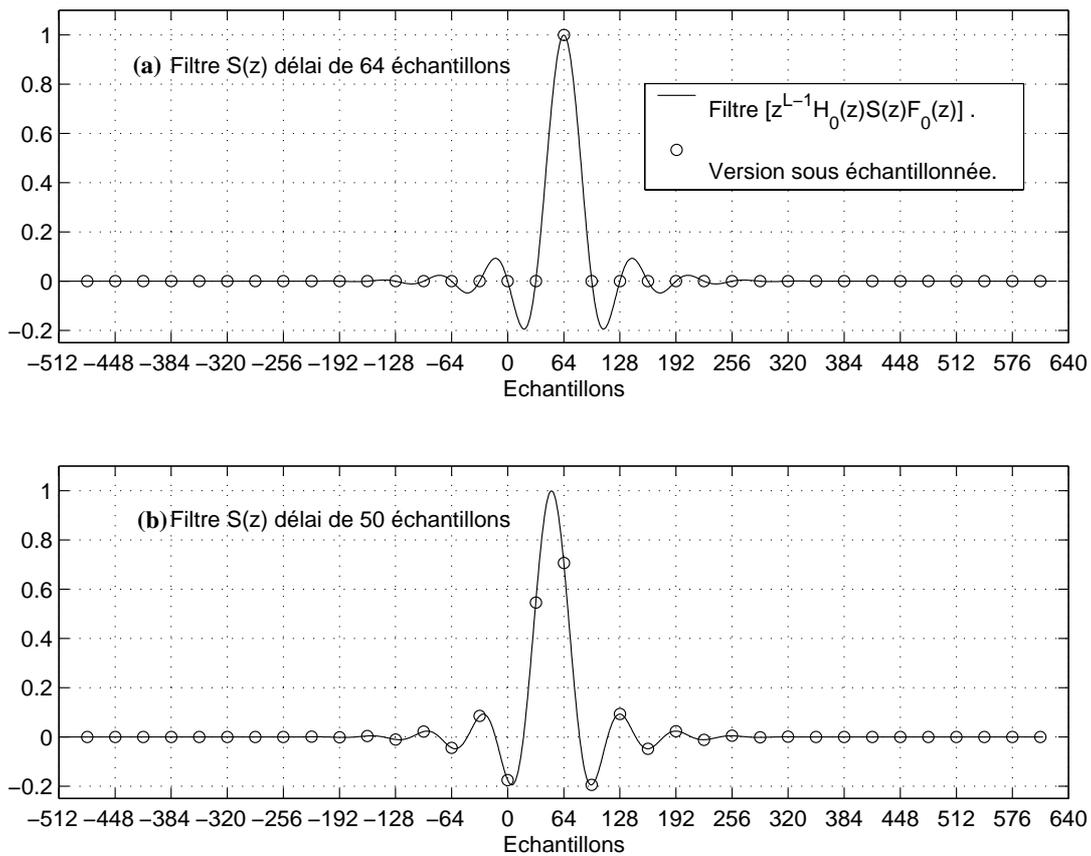


Fig. 3.12 : Illustration des opérations utilisées pour obtenir les filtres des sous bandes (a) filtre délai de 64 échantillons (b) filtre délai de 50 échantillons.

3.6.2 Filtre à réponse impulsionnelle finie

Pour illustrer le filtrage FIR en sous-bandes, nous utilisons un filtre $S(z)$ passe-bas à phase linéaire de longueur $N = 200$ et de fréquence de coupure $\omega_c = 0,2\pi$. La réponse fréquentielle de ce filtre est donnée dans la figure 3.14. Sa fréquence de coupure est située dans la bande 6 du banc de filtres (entre $\omega_1 = 0,1875\pi$ et $\omega_2 = 0,21875\pi$). Vu la forme (3.27) des filtres des sous-bandes, on s'attend à ce que les filtres $S_{sb,n}(z)$ soient passe-tout pour $n \leq 5$. La figure 3.15a illustre ce phénomène pour la bande 0. Les réponses fréquentielles des filtres de sous-bandes pour la bande 6, contenant la zone de transition de $S(z)$, sont illustrées dans la figure 3.15b.

Pour expliquer la forme de l'amplitude de la réponse fréquentielle des filtres des sous-bandes latéraux, il faut revenir à la forme (3.27) de ces filtres et aux caractéristiques du banc de filtres MPEG-1. En effet, la sous-bande k de ce banc de filtres est centrée sur $\frac{(2k+1)\pi}{2M}$ et son support (en incluant les bandes de transition) varie entre $\frac{(2k-1)\pi}{2M}$ et $\frac{(2k+3)\pi}{2M}$. Le filtre produit $H_k(z)F_{(k+1)}(z)$ est donc centré sur $\frac{(k+1)\pi}{M}$ et son support est $\left[\frac{(2k+1)\pi}{2M}, \frac{(2k+3)\pi}{2M} \right]$. Si les bandes k et $k+1$ sont à l'intérieur d'une zone passe-tout du filtre $S(z)$, après produit et décimation dans la formule (3.27), le support du filtre $S_{sb,k(k+1)}(z)$ est donc centré sur les points $(k+1)\pi + p2\pi$, $p \in \mathbb{N}$, et varie entre $\frac{(2k+1)\pi}{2} + p2\pi$ et $\frac{(2k+3)\pi}{2} + p2\pi$.

Ce filtre est donc un demi-bande, pour k pair, il est passe-haut et pour k impair il est passe-bas. Le même raisonnement est valable pour le filtre, $S_{sb,k(k-1)}(z)$ de la sous-bande latérale gauche, sauf que dans ce cas ce filtre est passe-haut pour k impair et passe-bas pour k pair (voir figure 3.13). Ceci explique donc les formes des amplitudes des réponses fréquentielles de $S_{sb,01}(e^{j\omega})$ et $S_{sb,65}(e^{j\omega})$ données dans la figure 3.15.

Le retard additionnel évoqué dans le paragraphe précédent pose toujours problème dans ce cas. Dans le cas d'un filtre à phase linéaire, le temps de propagation de groupe apporté est constant. Après convolution de $S(z)$ avec $H_n(z)$ et $F_k(z)$ la réponse impulsionnelle résultante vérifiera toujours la propriété citée précédemment. Pour avoir des points non causals nuls après décimation, il suffit alors que ce délai soit multiple de M . Il faut donc ajouter au filtre $S(z)$ un retard supplémentaire de manière à ce qu'il vérifie cette condition. Dans le cas du filtre considéré ici on doit ajouter un délai de 28 échantillons ainsi le délai total est de 128. Dans le cas général le délai à apporter, lorsque N est pair, peut être calculé par $M - \left\lfloor \left(\frac{N}{2} \right) \right\rfloor_M$, où $\lfloor x \rfloor_M$ désigne x modulo M .

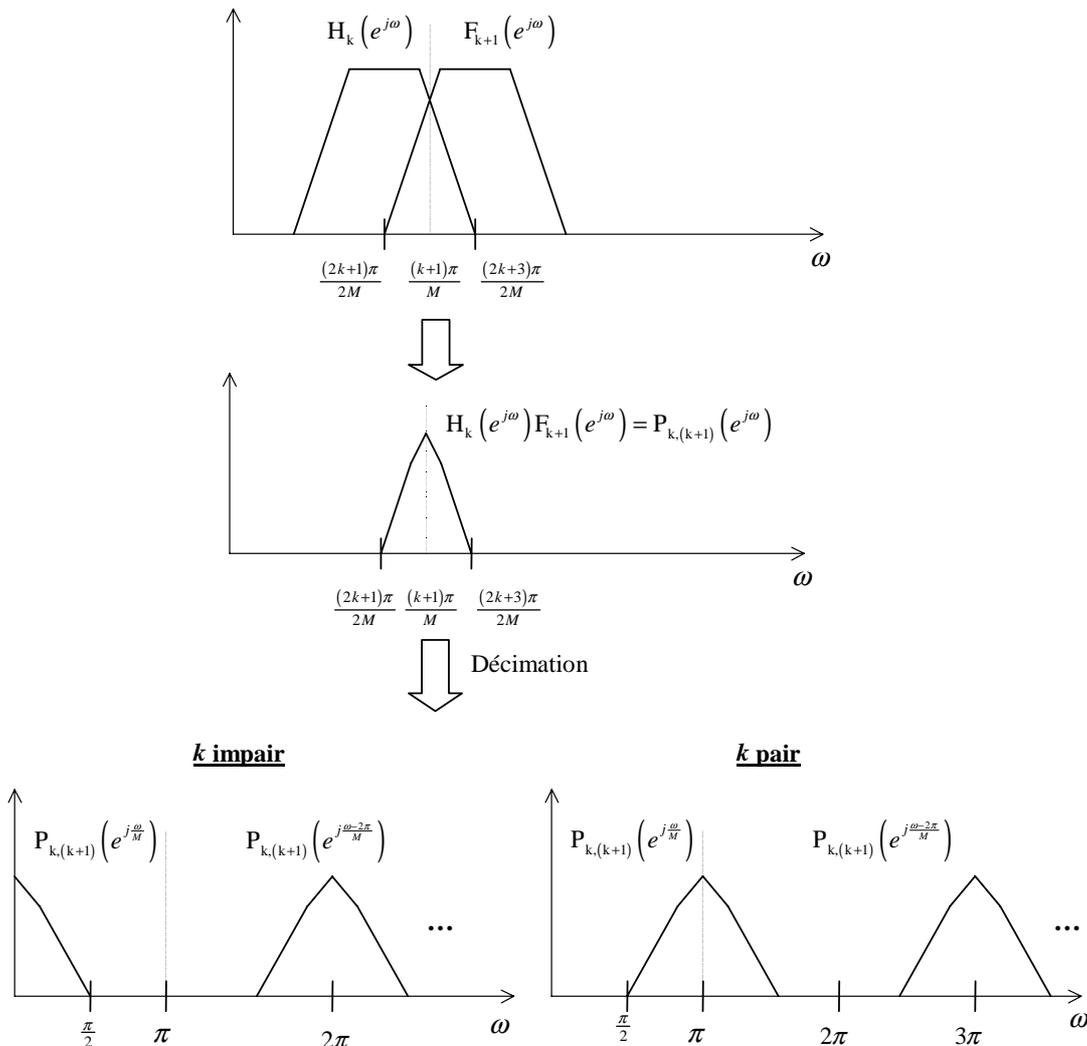


Fig. 3.13 : Schémas explicatifs de la forme des filtres des sous-bandes.

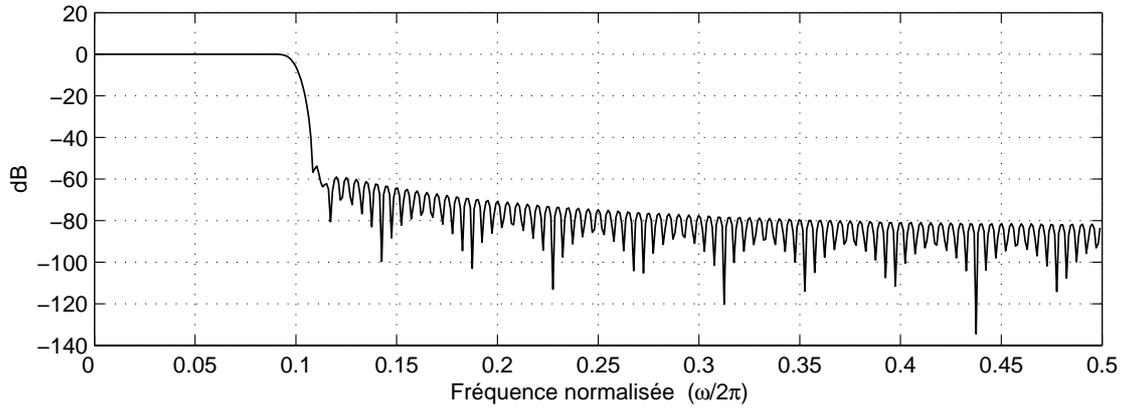
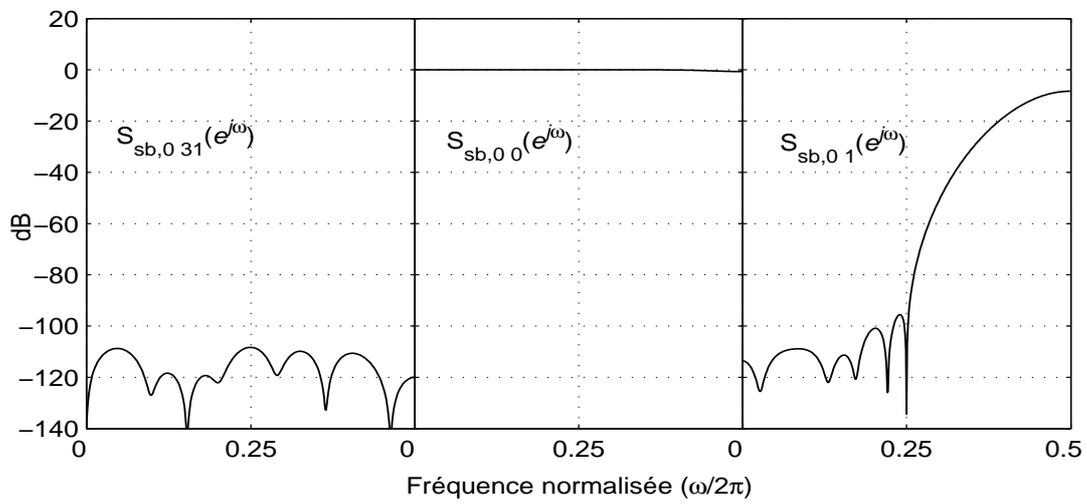
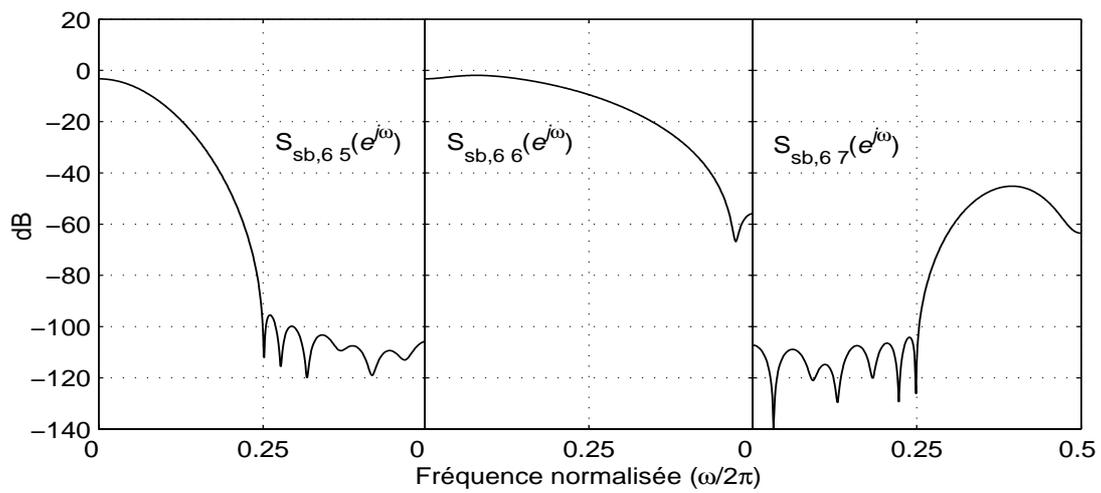


Fig. 3.14 : Réponse fréquentielle du filtre FIR à phase linéaire utilisé.



(a)



(b)

Fig. 3.15 : Réponses fréquentielles des filtres des sous-bandes (a) sous-bande 0 (b) sous-bande 6.

3.6.3 Filtre à réponse impulsionnelle infinie

Dans cet exemple nous utilisons un filtre IIR elliptique passe-bas d'ordre 12, de fréquence de coupure $\omega_c = 0,2\pi$ avec un taux d'ondulation dans la bande passante de $R_p = 0,1\text{dB}$ et un taux de d'atténuation dans la bande de rejet de $R_s = 50\text{dB}$. La figure 3.16 illustre son amplitude et son temps de propagation de groupe en fonction de la fréquence.

Le problème du retard additionnel ne peut être résolu comme dans le cas du filtre à phase linéaire. En effet, dans ce cas le temps de propagation de groupe varie en fonction de la fréquence. La figure 3.17 montre les réponses impulsionnelles de la convolution des filtres $S(z)$, $H_n(z)$ et $F_n(z)$ pour les sous-bandes $n=0$ et $n=5$ qui sont dans la bande passante du filtre. Ces deux réponses impulsionnelles ne sont plus symétriques et ne vérifient plus la propriété du filtre prototype de MPEG-1. En plus, la première est centrée autour de l'échantillon 8 et la seconde autour de l'échantillon 27. D'une part la variation du temps de propagation de groupe à l'intérieur de la même bande déplace les points de la réponse impulsionnelle différemment, d'autre part ce temps moyen entre bandes est différent. Il n'est donc plus possible de synchroniser toutes les bandes autour d'un point multiple de M pour que l'opération de décimation donne des points non causaux nuls. La seule solution pour limiter le retard additionnel dans ce cas est d'ajouter un même délai à toutes les bandes de manière à inclure dans la zone causale tous les points significatifs. Pour le filtre considéré, il suffit alors d'ajouter par exemple un délai de 192 échantillons.

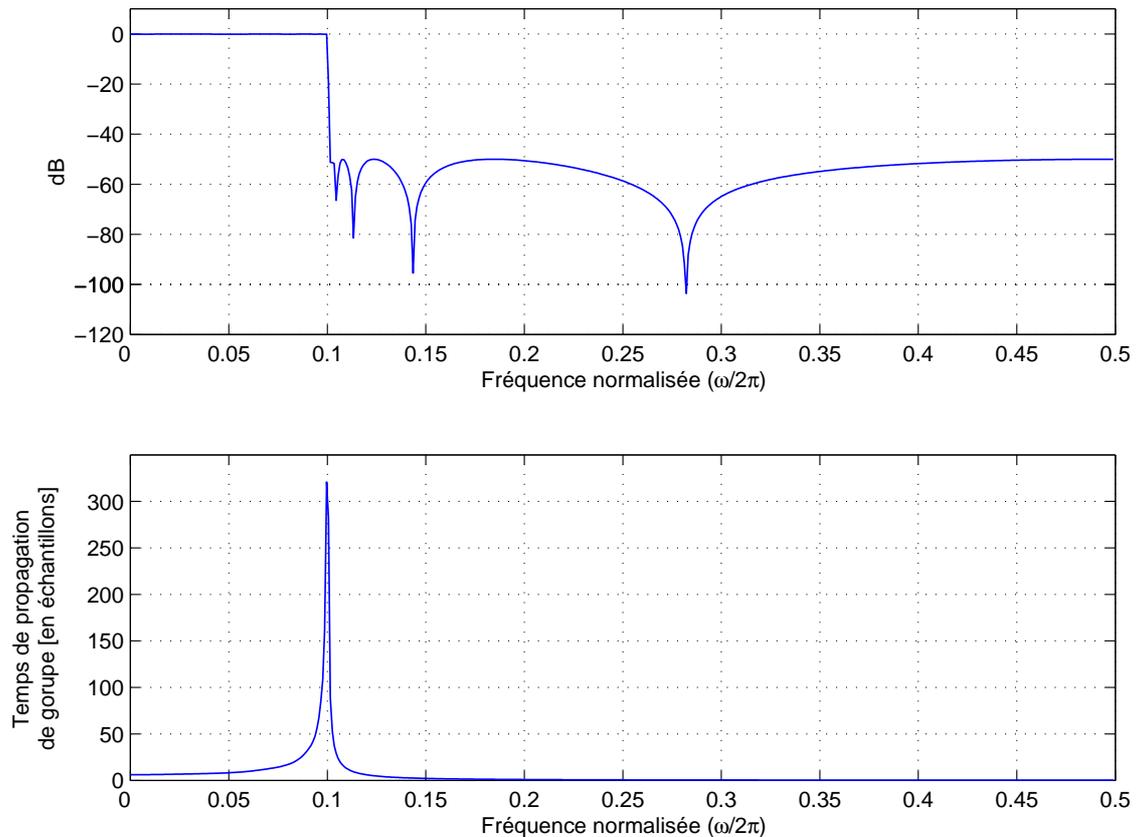
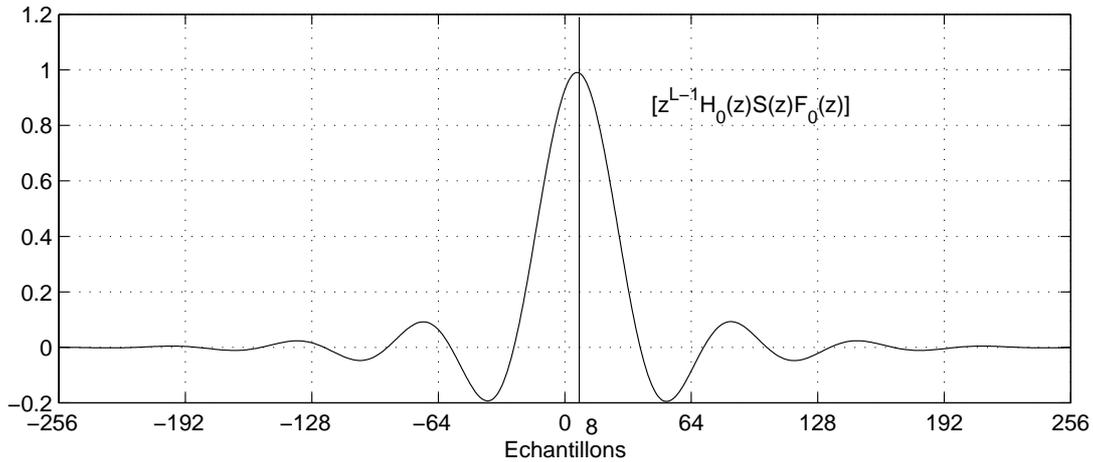
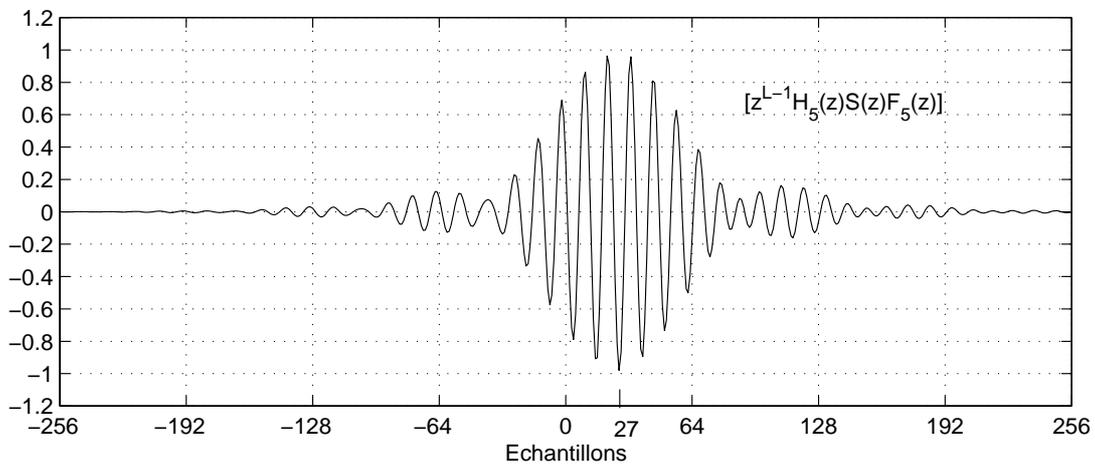


Fig. 3.16 : Amplitude et temps de propagation de groupe de la réponse fréquentielle du filtre IIR elliptique utilisé.



(a)



(b)

Fig. 3.17 : Réponses impulsionnelles de la convolution du filtre temporel IIR avec les filtres d'analyse et de synthèse pour (a) la sous-bande 0 (b) la sous-bande 5.

3.7. Application au banc de filtres TDAC

Nous utilisons dans ce cas le banc de filtres du codeur TDAC de FTR&D. Rappelons que ce banc de filtres est à $M = 320$ sous-bandes. Le filtre prototype, $H(z)$, à une réponse impulsionnelle correspondant à une fenêtre sinusoidale de longueur $L = 2M = 640$. Les expressions des matrices d'analyse et de synthèse sont données par les formules (2.33) et (2.35). Ces matrices sont d'ordre $K = 1$, l'équation (3.26) solution au problème du filtrage en sous-bande s'écrit dans ce cas comme suit :

$$\mathbf{S}_{sb}(z) = z\mathbf{E}(z)\mathbf{S}(z)\mathbf{R}(z). \quad (3.54)$$

Pour rendre la matrice $\mathbf{S}_{sb}(z)$ causale il faut éliminer l'avance z dans (3.54), ce qui introduit un retard supplémentaire de $M = 320$ échantillons. Comme le montre la figure 2.13, chaque bande de ce banc de filtres se recouvre avec les deux adjacentes de chaque côté, mais vu que le taux de rejet hors bande

passante de ces filtres est faible, de l'ordre de -24 dB, on prend $\delta = 3$. Chaque ligne de la matrice $\mathbf{S}_{sb}(z)$ réduite est donc constituée du filtre de la diagonale et de trois filtres latéraux de chaque côté. Nous illustrerons dans la suite la forme de ces filtres pour les cas d'un filtre FIR à phase linéaire et un filtre IIR.

3.7.1 Filtre à réponse impulsionnelle finie

Nous utilisons le même filtre FIR à phase linéaire du paragraphe 3.5.2. Dans ce cas la fréquence de coupure $\omega_c = 0,2\pi$ coïncide avec l'intersection des bandes numéro 63 et 64 du banc de filtres TDAC. La figure 3.18 illustre les réponses impulsionnelles pour les filtres des sous-bandes de la bande 0 et 63. Ces filtres sont d'ordre 3. Le problème du retard additionnel peut être résolu de la même manière que celle utilisée dans le paragraphe 3.5.2. Ainsi pour diminuer ce retard on ajoute un délai de 220 échantillons au filtre FIR. Le délai global est donc de 320 et l'opération de sous-échantillonnage de la réponse impulsionnelle de la convolution des trois filtres $S(z)$, $H_n(z)$ et $F_n(z)$ donnera donc des points non causals quasiment nuls (figure 3.19).

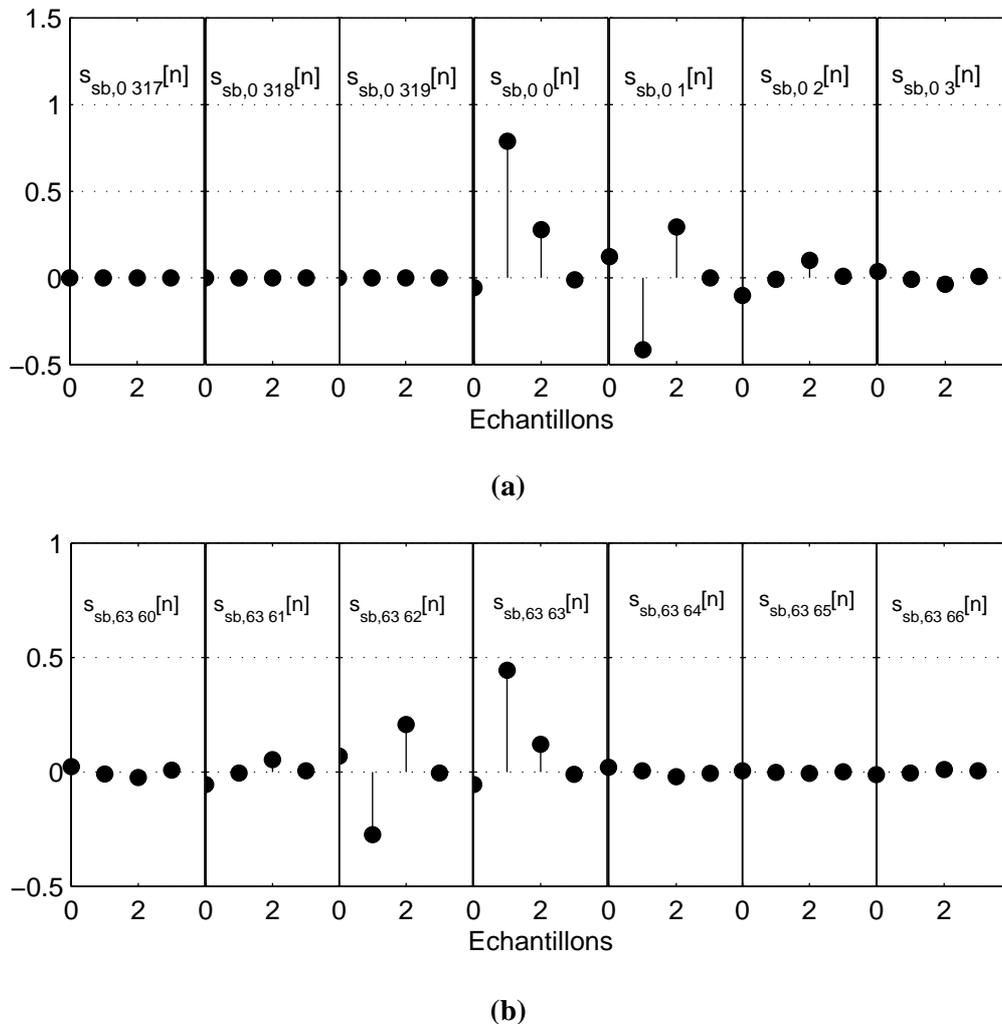


Fig. 3.18 : Réponses impulsionnelles des filtres des sous-bandes correspondant aux filtres FIR pour (a) la sous-bande 0 (b) la sous-bande 63.

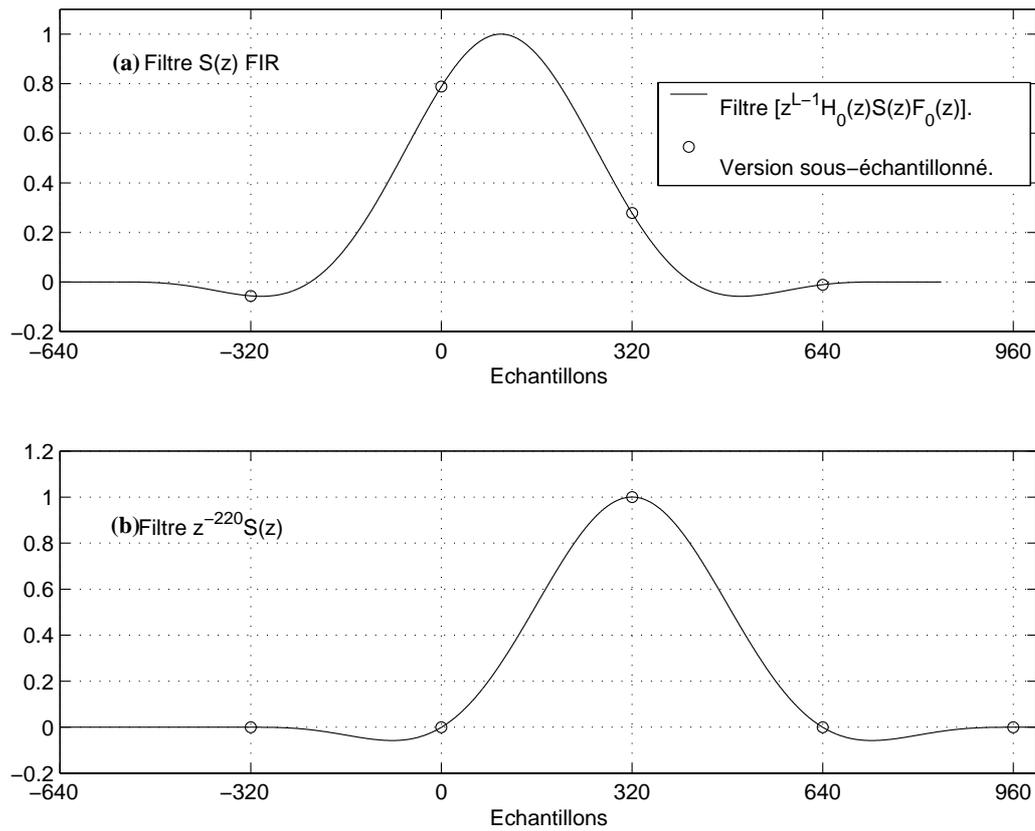
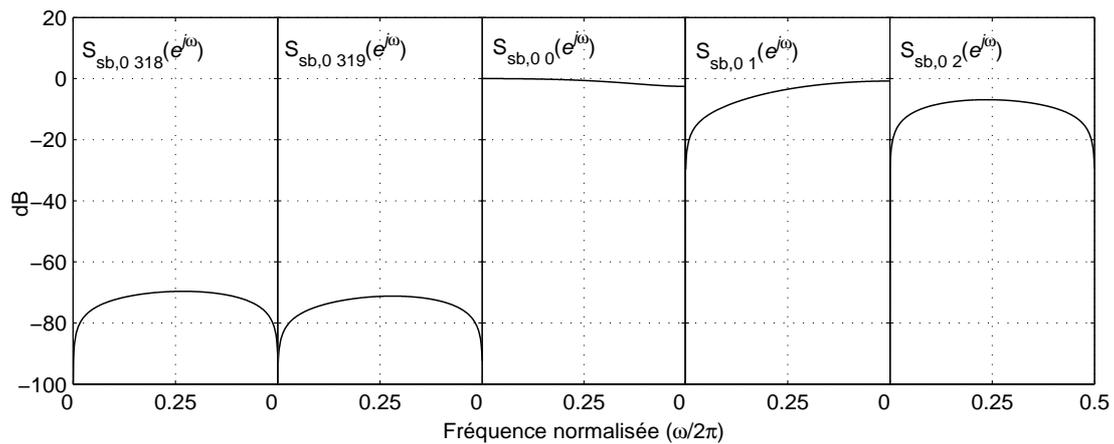
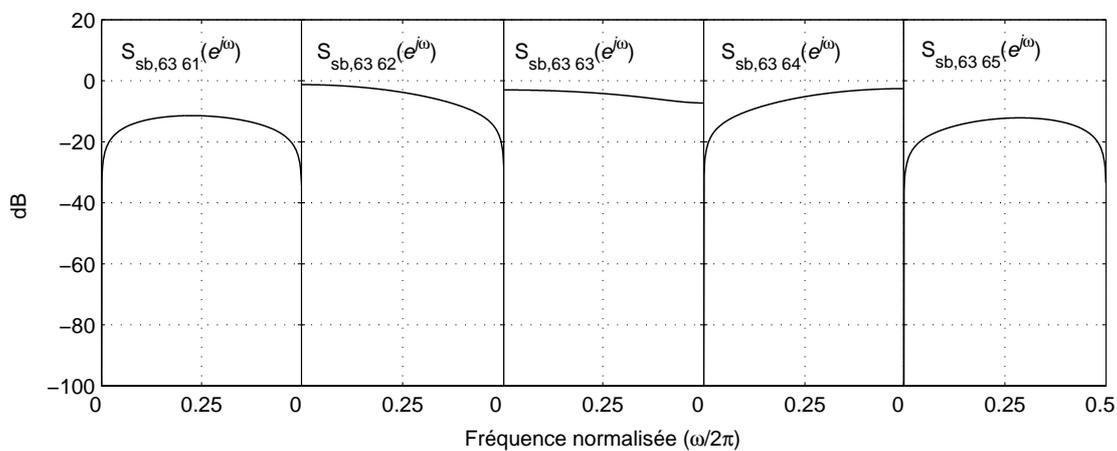


Fig. 3.19 : Convolution et sous-échantillonnage du filtre $S(z)$ avec les filtres d'analyse et de synthèse pour la bande 0 (a) filtre initial (b) après ajout d'un retard supplémentaire de 220 échantillons.



(a)



(b)

Fig. 3.20 : Réponses fréquentielles des filtres des sous-bandes correspondant au filtre FIR (a) sous-bande 0 (b) sous-bande 6.

3.7.2 Filtre à réponse impulsionnelle infinie

La solution du paragraphe 3.5.3 pour limiter le retard additionnel pour ce type de filtres n'est plus possible pour le banc de filtre TDAC. Il faut donc ajouter nécessairement un retard de 320 échantillons dans ce cas.

Si on considère un filtre IIR d'ordre N , en regardant la forme des composantes polyphasées lui correspondant (annexe A), sa matrice de transfert sera aussi d'ordre N . D'un autre côté, les matrices polyphasés d'analyse et de synthèse du banc de filtres TDAC sont d'ordre 1. Les numérateurs des filtres des sous-bandes sont donc d'ordre $N+2$ et le dénominateur d'ordre N . La complexité d'implantation par trame dans ce cas est indiquée dans le tableau 3.1, elle est comparée à la complexité obtenue par filtrage dans le domaine temporel.

	Analyse/Synthèse ⁵	Filtrage temporel	Filtrage en sous-bandes
Additions	$M(3\log_2 M - 1)/2$	$M(2N - 1)$	$M(13N + 24)$
Multiplications	$M(\log_2 M + 1)/2$	$M(2N + 1)$	$M(15N + 42)$

Tableau 3.1 : Comparaison des complexités par trame.

La complexité dans ce cas de figure est comparable au filtrage dans le domaine temporel. Ce résultat s'explique par le fait que la décomposition d'un filtre IIR donne des composantes polyphasées du même ordre. Contrairement au cas des filtres FIR, aucune réduction de la taille des filtres n'est donc réalisée. Par ailleurs, on peut voir que tous les coefficients ne sont pas nécessaires. En effet, considérons à titre d'exemple le cas d'une cellule du premier ordre, $S(z) = \frac{1}{1 - az^{-1}}$, $|a| < 1$, ses composantes polyphasées sont données par (tableau A.1) :

$$S_k(z) = \frac{a^k}{1 - a^M z^{-1}}, \quad 0 \leq k \leq M - 1. \quad (3.55)$$

Pour M grand, le coefficient du dénominateur, a^M , tend vers zéro ainsi que les coefficients des numérateurs des $S_k(z)$ pour k élevé. Cette remarque peut être utilisée pour réduire la complexité en ne conservant que les coefficients significatifs des filtres des sous-bandes. Un tel résultat est bien cohérent vu que pour M grand les bandes sont de faible largeur (haute résolution fréquentielle). Le spectre du filtre initial ne varie pas alors trop sur une bande. La modélisation de cette partie du spectre ne nécessite donc pas un filtre d'ordre élevé.

3.8. Application : spatialisation sonore dans le domaine des sous-bandes

3.8.1 Spatialisation sonore

La possibilité de localisation des sources sonores qui nous entourent dans l'espace 3D est l'une des facultés importantes du système d'audition humain. En effet, cette propriété accroît l'intelligibilité et le confort en cas de discussion avec plusieurs personnes puisque l'oreille associe une position dans l'espace à chaque évènement sonore correspondant à un locuteur distant. Elle permet aussi de focaliser son attention sur le contenu d'une source sonore en milieu bruyé : effet "Cocktail Party".

Le mécanisme de perception par le cerveau de la localisation d'une source sonore repose sur les différences entre les signaux arrivant à chacune des deux oreilles. En effet, le son atteignant les oreilles gauche et droite arrive avec des retards différents selon la proximité de la source du côté gauche ou droit de l'auditeur. Cette différence est appelée différence interaurale de temps (ITD : Interaural Time Difference). En plus, l'onde sonore subit des diffractions sur la tête et le torse de l'auditeur avant d'atteindre le tympan. L'atténuation due à ce phénomène donne lieu à une différence d'intensité interaurale (IID : Interaural Intensity Difference). Les réflexions sur le pavillon de l'oreille

⁵ Complexité évaluée dans [Duhamel 91].

externe ainsi que les diffractions citées précédemment engendrent aussi une modification spectrale du signal qui dépend de la direction d'incidence. La localisation d'un son dans l'espace, donnée par l'azimut et l'élévation, se base essentiellement sur ces différents indices.

Pour reproduire l'effet de perception 3D, il est possible d'utiliser ces indices psychoacoustiques sur lesquels se base le cerveau. Une autre approche considère les fonctions de transfert HRTF (Head Related Transfer Function) qui modélisent la transformation subit par le son le long de son chemin de propagation en champ libre, à partir de la source jusqu'aux canaux des oreilles de l'auditeur. Ces filtres HRTF contiennent toutes les informations sur les différences interaurales de temps et d'intensité ainsi que les modifications spectrales.

La spatialisation sonore consiste à reproduire le champ sonore 3D à partir de sources monophoniques donnant ainsi l'impression à l'auditeur que le son qu'il perçoit provient d'une position donnée dans l'espace. La technique de la synthèse binaurale recrée le champ sonore présent à l'entrée du canal de l'oreille par filtrage du signal monophonique par les deux filtres HRTF gauche et droit correspondant à une direction donnée (figure 3.21). Elle vise à reproduire les effets spatiaux au niveau des oreilles de l'auditeur et non pas sur un champ d'écoute étendu. Une écoute sur casque est donc nécessaire pour réaliser une simulation parfaite dans ce cas.

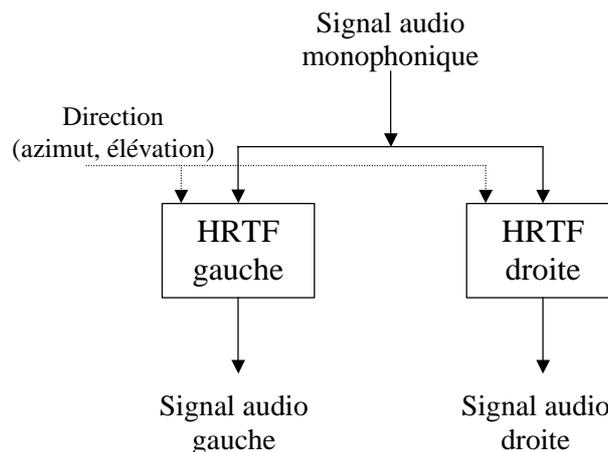


Fig. 3.21 : Principe de la spatialisation par synthèse binaurale.

La conception des filtres HRTF se fait à partir des réponses impulsionnelles mesurées à l'entrée du canal de l'oreille pour chaque direction. Généralement un échantillonnage de 5° est utilisé puisque cette valeur correspond à la sensibilité de l'oreille pour la perception de la localisation [Zwicker 99]. La figure 3.22 montre un exemple de réponses impulsionnelles et fréquentielles des HRTF correspondant à l'azimut 30° dans le plan horizontal. Afin de donner une représentation compacte des HRTF et réaliser aussi une implantation efficace, ces filtres sont décomposés en le produit d'un filtre à phase minimale et un filtre passe-tout qui représente l'excès de phase. Ce dernier s'avère quasiment linéaire en fonction de la fréquence et correspond au retard ITD [Jot 95]. A la fréquence d'échantillonnage 32 kHz, les HRTF peuvent être modélisés par un filtre FIR de longueur 200 échantillons. L'utilisation de filtres IIR réduit cet ordre à 12.

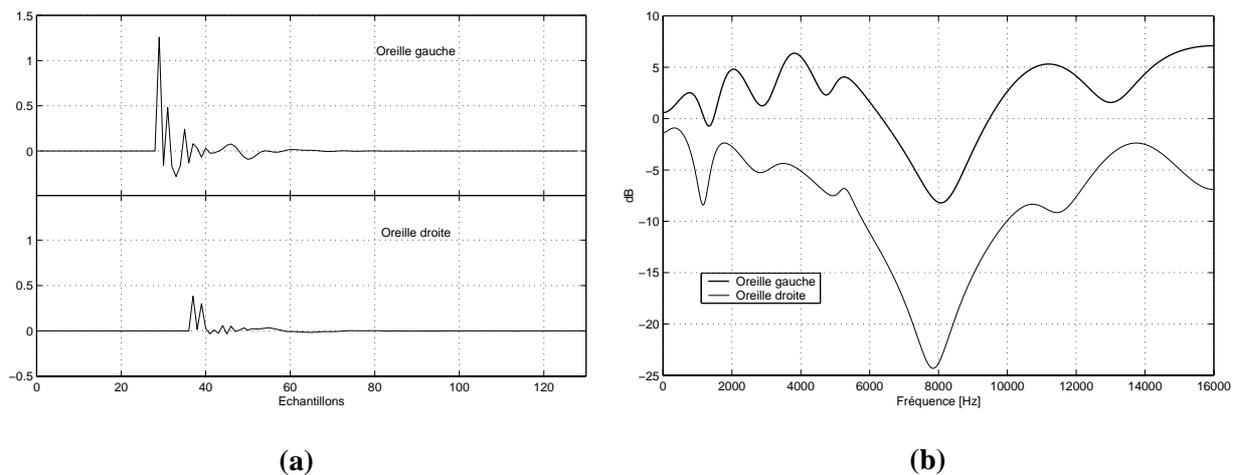


Fig. 3.22 : Filtres HRTF gauche et droite correspondant à l'élévation 0° et l'azimut 30°
(a) réponses impulsionnelles (b) réponses fréquentielles.

3.8.2 Implantation dans le codeur TDAC de FTR&D

La synthèse binaurale a été implantée dans le domaine des sous-bandes du codeur TDAC de FTR&D. Les filtres HRTF modélisés initialement dans le domaine temporel par un IIR d'ordre 12 suivi du retard pur (ITD) sont transposés dans le domaine des sous-bandes. Les filtres de la matrice de filtrage obtenus ont un numérateur d'ordre 15 et un dénominateur d'ordre 12. L'implantation se base sur la structure donnée dans le paragraphe 3.4. L'écoute a montré que pour avoir une transparence par rapport à la qualité obtenue lors d'une synthèse binaurale dans le domaine temporel, il suffit d'utiliser un nombre de bandes adjacentes, δ , égal à 2, au lieu de 3 comme c'est indiqué dans le paragraphe 3.7. Cette restriction à seulement deux bandes adjacentes de la diagonale résulte du fait que le bruit de quantification masque dans ce cas certains résidus d'aliasing.

L'implantation des filtres des sous-bandes donne une complexité comparable à la méthode de filtrage direct dans le domaine temporel. Mais l'observation des coefficients de ces filtres montre que seuls quelques coefficients du numérateur et le coefficient d'ordre zéro du dénominateur sont significatifs. En considérant alors les filtres des sous-bandes comme des FIR d'ordre 2 et en réalisant une implantation directe on obtient un gain en complexité de 40 %.

3.9. Conclusion

Le développement d'une solution générique pour le filtrage rationnel dans le domaine des sous-bandes constitue le résultat principal de ce chapitre. L'application de cette solution aux exemples particuliers des bancs de filtres pseudo-QMF et TDAC a permis de faire surgir des éléments intéressants sur les réductions espérées en complexité et en délai. Une modélisation du filtre temporel initial par un IIR ne permet pas toujours une réduction de complexité après transposition dans le domaine des sous-bandes. Le choix d'un IIR ou un FIR dépend du nombre des sous-bandes du banc de filtres utilisé et de la longueur de ses filtres. Il dépend aussi de la longueur du filtre temporel.

Chapitre 4 **Traitement dans le domaine compressé MPEG-1**

4.1. Introduction

Dans le chapitre précédent nous avons abordé les problèmes liés au traitement dans le domaine fréquentiel. Dans le présent chapitre, on s'intéresse aux problèmes relatifs à la manipulation des flux codés obtenus par le codeur MPEG-1 Layer I et II en exploitant les spécificités de ce codeur. Comme il a été évoqué dans le chapitre 1 une telle manipulation est nécessairement précédée par une phase de décodage partiel du flux codé permettant d'obtenir les paramètres nécessaires pour le traitement. Dans ce chapitre, et aussi dans le suivant concernant le codeur TDAC de FTR&D, nous examinerons les possibilités, d'une part, de minimiser l'étape de décodage partiel, et d'autre part, d'exploiter au maximum les paramètres de la trame afin de minimiser la complexité du traitement. Le type de manipulation considéré concerne essentiellement la sommation directe de flux audio codés. Le but est alors de réaliser une recombinaison directe des informations contenues dans les trames d'entrée pour construire la trame de sortie correspondant au signal somme (figure 4.1b). Une telle trame doit fournir, après décodage, un signal sonore de qualité perceptuelle au moins égale à celle d'un signal obtenu par transcodage direct.

Dans l'algorithme de codage MPEG-1 l'allocation des bits est faite suivant un critère perceptuel. Elle est basée sur un modèle psychoacoustique dont le rôle est de déterminer une courbe de masquage et de fournir des rapports signal à masque à la procédure itérative d'allocation des bits. La résolution fréquentielle du banc de filtres d'analyse étant faible, les modèles psychoacoustiques proposés sont dérivés du spectre du signal de la trame obtenu par une transformée de Fourier discrète du signal temporel. Si un traitement est réalisé dans le domaine des sous-bandes sur le signal d'une trame qui sera soumis à une phase ultérieure de codage, une nouvelle allocation de bits doit être effectuée. Le problème essentiel est donc relatif à la détermination de cette allocation de bits. En effet, aucune information précise sur le spectre du signal ne peut être obtenue explicitement, sans recours au signal temporel, pour alimenter le modèle psychoacoustique.

Pour les manipulations visées dans ce chapitre, la résolution de ce problème, comme il sera détaillé dans la suite, nécessite de déterminer les seuils de masquage correspondant aux trames d'entrée. Après rappel de certains travaux sur le traitement dans le domaine codé MPEG-1, ce chapitre commence par établir une méthode d'estimation de ces paramètres à partir des simples informations contenues dans ces trames. Un algorithme de sommation sera ensuite décrit puis évalué. Un algorithme de conversion de débit sera aussi proposé dans le paragraphe 4.6. Il illustrera un exemple de traitement avancé sur le flux audio codé. Enfin, une discussion sur les conclusions à tirer des manipulations considérées dans le domaine codé MPEG-1 sera donnée.

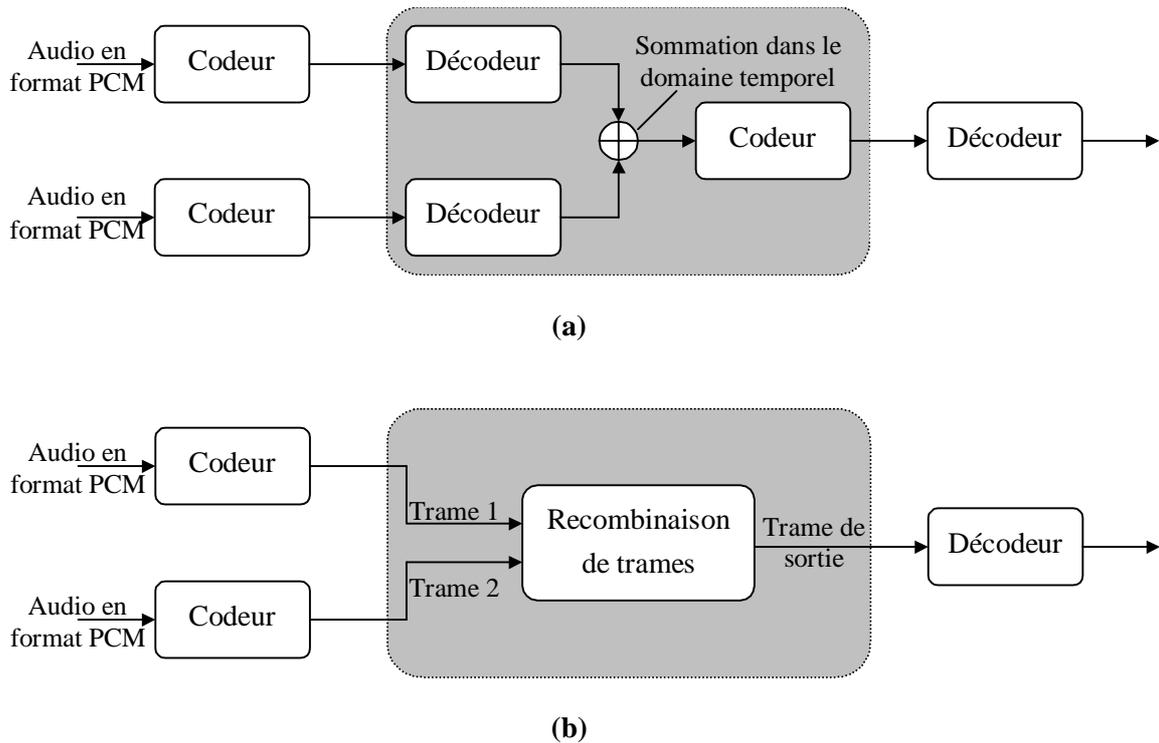


Fig. 4.1 : Somme des signaux audio compressés (a) méthode directe dans le domaine temporel (b) méthode dans le domaine compressé.

4.2. Etat de l'art

La généralité du codeur MPEG-1 et sa large utilisation ont suscité l'intérêt de plusieurs recherches pour réaliser des traitements sur des flux obtenus par un tel codeur. Des travaux concernant la manipulation du spectre de type filtrage ont été déjà abordés dans le paragraphe 3.3.2. Dans ce qui suit nous discuterons brièvement ce qui a été fait pour d'autres traitements de type mixage, conversion de débit et analyse et classification des signaux audio MPEG-1.

Le mixage de flux audio MPEG-1 a été abordé dans [Broadhead 95]. L'approche utilisée dans ce cas pour résoudre le problème de détermination de l'allocation de bits est approximative. Elle utilise la formule suivante pour retrouver la nouvelle allocation de bits à partir des nombres de bits alloués aux signaux à mixer :

$$b_0 = \left\lceil \log_2 \left(\frac{g_1 SCF_1 2^{b_1} + g_2 SCF_2 2^{b_2}}{g_1 SCF_1 + g_2 SCF_2} \right) \right\rceil, \quad (4.1)$$

où SCF_1 et SCF_2 sont les facteurs d'échelle, g_1 et g_2 sont les facteurs de gain utilisés pour les deux signaux, b_1 et b_2 les nombres de bits alloués initialement à chaque signal individuel. $\lceil x \rceil$ désigne l'entier juste plus grand que le réel x .

Cette équation se base sur le principe que le nombre de bits se déduit de la dynamique du signal. La dynamique est calculée comme le rapport entre l'amplitude maximale du signal et le niveau du bruit de

quantification. Pour le signal mixé, la somme pondérée des maxima des deux signaux et la somme de niveaux de bruit est effectuée. Le nombre de bits nécessaire est ensuite retrouvé par le logarithme à base 2 de leur rapport. Cette méthode ne prend aucune considération des critères perceptuels utilisés dans le codeur MPEG-1. Les travaux donnés dans [Lanciani 99b] ont apporté une solution à ce problème de mixage en tenant compte de ce critère. Cette solution est proche de la méthode qui sera présentée dans la suite et qui a été développée en parallèle et indépendamment.

La conversion de débit sur les flux audio codés MPEG-1 a été discuté dans [Nakajima 98]. Trois algorithmes sont proposés pour réduire le débit d'un flux binaire initial. Le premier est le plus simple, il se base sur la limitation de la largeur de bande. Ainsi, une élimination des bandes hautes est effectuée jusqu'à obtention d'un nombre total de bits alloué compatible avec le débit binaire voulu. Le second algorithme réalise une nouvelle allocation de bits et effectue ensuite une requantification des échantillons des sous-bandes. La réallocation de bits se fait de manière itérative en diminuant de un le nombre de bits alloué à chaque sous-bande en commençant par les bandes hautes et jusqu'à adaptation du débit. Dans le troisième algorithme, cette réallocation de bits se fait suivant un critère perceptuel pour respecter le principe utilisé dans MPEG-1. Ainsi, un bit à la fois est retiré itérativement à la sous-bande où l'incrément du rapport bruit à masque, *NMR*, résultant de cette opération est minimale.

Une conversion de débit a été aussi utilisée dans le transcodeur hiérarchique à base de MPEG-1 Layer II conçu dans [Hans 98] pour la transmission sur Internet. Ce dernier extrait à partir d'une trame codée initiale une autre à plus faible débit qui sera transmise à tous les utilisateurs, dite couche fondamentale, et d'autres couches d'amélioration pour les connexions à plus haut débit. La couche fondamentale s'obtient par une réduction du débit du signal original dans le domaine codé. L'algorithme retenu effectue une requantification après réallocation de bits. Contrairement à la troisième approche de Nakajima, cette dernière est refaite par la procédure classique de la norme MPEG-1 en utilisant comme paramètre d'entrée les rapports signal à masque, *SNR*, correspondant à l'allocation de bits initial. Les couches d'amélioration codent le signal de différence entre échantillons sous-bande du signal original et ceux requantifiés de la couche fondamentale.

L'accès aux bases de données audio demande des méthodes d'indexation efficaces et d'extraction rapide d'information. Quand ces données sont compressées, leur utilisation directe sous cette forme assure ces propriétés. Des travaux récents ont été réalisés dans ce sens pour des flux codés MPEG-1. chacun d'eux propose un traitement direct sur les échantillons des sous-bandes pour extraire un certain nombre de paramètres. Ainsi, une analyse de scène sonore et détection des changements de scène a été proposé dans [Venugopal 99]. La segmentation des séquences est basée sur le changement musique/parole, changement du locuteur ou changement du genre du locuteur (homme/femme). La distinction musique/parole se base sur la tonalité, la largeur de bande, la forme d'onde de l'excitation, la durée des tonals et l'énergie. Une mesure de vraisemblance de chacun de ces paramètres à l'hypothèse que le signal correspond à la parole est calculée, et la vraisemblance totale est prise comme leur somme pondérée. L'identification du locuteur utilise un mélange de gaussiennes (GMM) et se base sur les coefficients cepstraux linéaires (LFCC : Linear Frequency Cepstral Coeff.) extraits des échantillons des sous-bandes. La distinction voix d'homme/voix de femme utilise la combinaison de deux décisions : d'abord la détection du pitch sur les segments voisés basée sur la méthode d'analyse par synthèse du codeur MBE (Multiband Excitation Vocoder) puis l'identification basée sur un modèle GMM.

Le travail présenté dans [Nakajima 99] effectue la classification de données audio en 4 classes : silence, parole, musique et segments d'applaudissement. La détection de silence se fait par simple seuillage sur l'énergie. Pour caractériser la musique/parole et les applaudissements les paramètres suivants sont calculés : densité d'énergie des sous-bandes, nombre moyen de sous-bandes ayant un niveau d'énergie significatif, moyenne et variance des centroïdes d'énergie de chaque trame. Ces paramètres sont regroupés en vecteurs et la classification musique, parole, applaudissement se fait par application d'une fonction de discrimination Bayésienne pour une distribution gaussienne multivariable.

Un algorithme de classification musique/parole et un autre de segmentation générale a été aussi proposé dans [Tzanetakis 00]. Les paramètres utilisés se basent sur le vecteur des énergies des sous-bandes de chaque trame. Ainsi, on extrait son centroïde, la limite haute des sous-bandes contenant 85% d'énergie de la trame (Rolloff), la racine carrée de leur moyenne au carré (RMS) et le flux spectral correspondant à la norme de la différence entre deux vecteurs normalisés successifs. Ces paramètres sont regroupés en vecteurs et une distance est mesurée entre vecteurs successifs. Pour réaliser la segmentation, la dérivée de l'évolution temporelle de cette distance est effectuée et une méthode heuristique est utilisée pour déterminer les pics et ainsi délimiter les régions du signal.

Les propriétés de masquage utilisées par le codeur MPEG-1 ont été exploitées pour résoudre le problème de l'annulation d'écho acoustique stéréophonique (SAEC). En effet, l'annulation d'écho repose sur des algorithmes de filtrage adaptatif pour l'identification de la réponse impulsionnelle modélisant le chemin de l'écho [Murano 90]. Dans le cas multivoies, des problèmes de convergence peuvent apparaître pour les algorithmes utilisés pour cette identification. Ces problèmes ont pour cause la dépendance linéaire entre les signaux sur les deux voies. La solution examinée dans [Gänsler 98] propose d'introduire un bruit supplémentaire dans les sous-bandes des deux signaux stéréophoniques lors du codage. Cet ajout se fait de manière à ce que l'énergie du bruit total, y compris celui dû à la quantification, ne dépasse pas les seuils de masquage et donc reste imperceptible. Par cette méthode les signaux sur les deux voies sont décorrélés du fait que les bruits ajoutés sur chacun d'eux sont indépendants.

4.3. Estimation des seuils de masquage

4.3.1 Principe

Le calcul des seuils de masquage correspondant au signal d'une trame d'entrée doit se baser sur les échantillons des sous-bandes, les facteurs d'échelle, et l'allocation binaire. Ces différentes informations sont obtenues par une phase de décodage partiel de la trame (figure 4.2). L'idée de base consiste à suivre la démarche inverse du codage à savoir retrouver les seuils de masquage, $MT(j)$, à partir de l'allocation de bits.

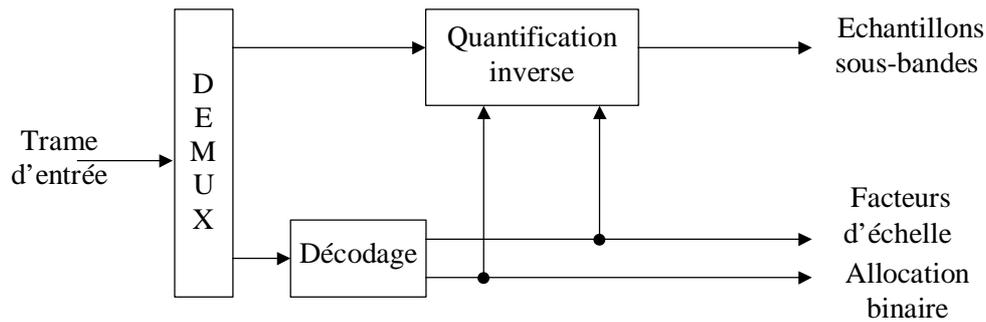


Fig. 4.2 : Schéma bloc du décodage partiel des trames MPEG-1 Layer I et II.

En effet, l'allocation des bits effectuée initialement pour le codage de la trame suit un critère perceptuel. Les seuils de masquage à estimer sont donc en relation directe avec cette allocation binaire. On supposera qu'elle a été réalisée de manière optimale lors du codage initial. Comme il a été démontré dans le paragraphe 2.5, cette hypothèse implique que le rapport bruit à masque est constant pour toutes les bandes :

$$NMR(j) = C, \quad (4.2)$$

où C est une constante.

En réalité cette hypothèse n'est pas exactement vérifiée. L'algorithme en "greedy" utilisé pour minimiser la fonction de distorsion, $D(\mathbf{b}) = \sum_j NMR(j)$, est sous-optimale. Il ne garantit pas d'atteindre le minimum global de cette fonction [Gersho 92]. Par ailleurs, l'observation des valeurs des $NMR(j)$ obtenues à la fin de la procédure itérative d'allocation des bits montre qu'elles sont très proches pour les bandes à nombre de bits non nul (figure 4.3). Le cas simple de la Layer I permet d'expliquer ce phénomène. En effet, dans ce cas 14 classes de quantificateurs sont utilisées pour toutes les sous-bandes. Ils correspondent aux nombres de bits variant entre 2 et 15 bits. Lors de chaque itération, la précision du quantificateur de la sous-bande à NMR maximale est incrémentée par ajout d'un bit [Ramstad 82]. Une telle bande voit alors son rapport signal à bruit augmenter, et donc son rapport bruit à masque diminuer, de l'ordre⁶ de 6 dB [Jayant 84]. Pour un nombre total de bits suffisant, à la fin des itérations toutes les valeurs des $NMR(j)$ seront confinées dans un intervalle de largeur 6 dB.

⁶ La variation sur ces quantités, après ajout d'un bit, n'est pas exactement de 6,02 dB car les quantificateurs scalaires uniformes utilisés sont à $2^b - 1$ niveaux de quantifications mais pas 2^b .

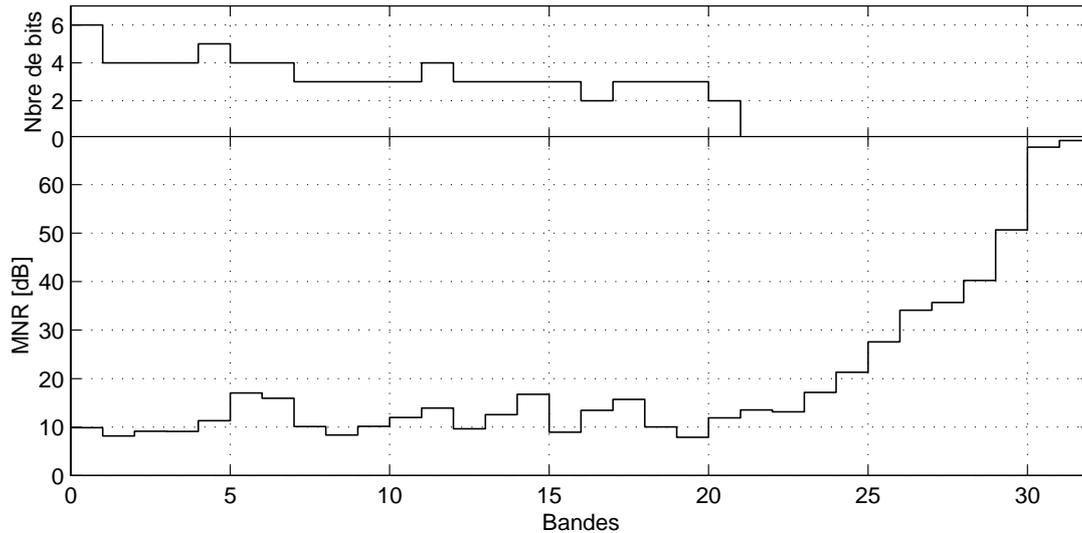


Fig. 4.3 : Exemple de rapports masque à bruit obtenus à la fin de la procédure d'allocation de bits pour une trame (débit binaire de 96 kbit/s).

Sous l'hypothèse d'optimalité et à partir des relations (2.53) et (2.55) on peut en déduire, pour les sous-bandes j où le nombre de bits alloué est non nul, une expression du seuil de masquage :

$$MT(j) = SPL(j) - SNR(b(j)) - C, \quad (4.3)$$

où $SPL(j)$ est le niveau de pression acoustique sur la bande j exprimé en dB.

4.3.2 Détermination du niveau de pression acoustique et du rapport signal à bruit

Le niveau de pression acoustique, $SPL(j)$, est déterminé de manière similaire à son expression originale (2.54) :

$$SPL(j) = \max\{10\log_{10}(\sigma_x^2(j)), 20\log_{10}(SCF(j)) - 10\} + 90,3 \text{ dB}, \quad (4.4)$$

où $\sigma_x^2(j) = \frac{1}{K} \sum_{k=0}^{K-1} x^2(k, j)$ est une estimée de la puissance du signal sur la bande j à partir de ses échantillons, $\{x(k, j)\}$. $SCF(j)$ est le facteur d'échelle de cette bande lu sur la trame d'entrée.

Le rapport signal à bruit, $SNR(b(j))$, dépend du quantificateur utilisé (i.e. du nombre de bits) sur la bande. Il se déduit directement de l'information sur l'affectation binaire lue à partir de la trame. Pour Layer I, cette correspondance est fournie directement par un tableau comme celui utilisé dans MPEG-1 [IS 11172-3].

Pour Layer II, l'indice contenu dans la trame permet de connaître le nombre de niveaux de quantification utilisé à partir des tables de quantification du codeur MPEG-1 dépendant de la fréquence d'échantillonnage, f_e , et du débit binaire utilisé. Lors de l'allocation de bits initiale la précision des quantificateurs utilisés pour cette couche ne varie pas de manière continue. Ainsi par exemple pour la bande 30, à la fréquence d'échantillonnage $f_e = 32 \text{ kHz}$ et au débit binaire 96 kbit/s, les nombres de niveaux des quantificateurs autorisés sont 3, 5 et 65535. Si le quantificateur à nombre

de niveaux maximal est atteint pour cette sous-bande, le SNR et par suite le NMR va subir une variation brusque de 78 dB. Si on prend le SNR correspondant à ce nombre de niveaux de quantification, le NMR relative à cette bande sera loin de la constante caractérisant le point de convergence optimal. Pour de telles sous-bandes, le nombre de niveaux de quantification qui est considéré est celui du quantificateur à précision juste au-dessus de l'avant dernier quantificateur autorisé. Les tableaux 4.1a et 4.1b indiquent les bandes figurant dans ce cas. Si l'indice lu pour de telles sous-bandes correspond à celui indiqué sur ces tableaux alors il faut prendre le nombre de niveaux de quantification qui y est donné. La déduction des $SNR(b(j))$ se fait finalement de manière directe à partir du nombre de niveaux de quantification en utilisant le tableau indiqué dans la norme MPEG-1 [IS 11172-3].

Numéro de la sous-bande	Indice	Nombre de niveaux de quantification
3-10	15	32767
11-22	7	63
23-26	3	7

- (a) $f_e = 48\text{kHz}$ débits binaires = 56, 54, 80, 96, 112, 160, 192 kbits/s et format libre
 $f_e = 44,1\text{kHz}$ débits binaires = 56, 64, 80 kbits/s
 $f_e = 32\text{kHz}$ débits binaires = 56, 64, 80 kbits/s

Numéro de la sous-bande	Indice	Nombre de niveaux de quantification
3-10	14	32767
11-22	7	63
23-29	3	7

- (b) $f_e = 44,1\text{kHz}$ débits binaires = 96, 112, 128, 160, 192 kbits/s et format libre
 $f_e = 32\text{kHz}$ débits binaires = 96, 112, 128, 160, 192 kbits/s et format libre

Tableau 4.1 : Correspondance entre indices et niveaux de quantifications pour certains cas particuliers.

4.3.3 Calcul de la constante d'ancrage

Le problème résiduel pour déterminer les seuils de masquage relève du calcul de la constante d'ancrage C . A partir de la relation (4.3), on effectue une moyenne sur les M' sous-bandes où le nombre de bits affecté est non nul :

$$C = \frac{1}{M'} \left[\sum_j SPL(j) - \sum_j SNR(b(j)) - \sum_j MT(j) \right]. \quad (4.5)$$

Dans cette expression le terme $\frac{1}{M'} \sum_j MT(j)$ est indéterminé. Il représente la valeur en dB de la moyenne géométrique des seuils de masquage. L'observation de cette quantité montre une forte dépendance de la moyenne géométrique des puissances des sous-bandes (coefficient de corrélation de 0,96). L'explication de ce phénomène est liée au principe de calcul de la courbe de masquage. En effet, ce calcul est basé sur la convolution du spectre du signal avec la fonction d'étalement de la membrane basilaire suivi d'une correction par les indices de masquage [Zwicker 99]. Si on fait abstraction de cette dernière correction, une relation directe peut être déterminée entre l'énergie totale de la trame dans le domaine fréquentiel et son équivalente dans le domaine basilaire. Bien qu'une relation déterministe entre les moyennes géométrique des énergies des bandes dans les deux domaines ne soit pas simple à obtenir, ceci explique l'observation expérimentale. On remplace donc la relation (4.5) par l'expression suivante :

$$C = \frac{1}{M'} \left[\sum_j SPL(j) - \sum_j SNR(b(j)) - \sum_j 10 \log_{10}(\sigma_x^2(j)) \right] - \Delta, \quad (4.6)$$

où Δ est une constante de correction.

Les valeurs de Δ sont tabulées en fonction de l'énergie totale, E , de la trame (tableau 4.2). Ces valeurs sont déterminées par apprentissage sur un corpus de séquences audio. Pour chaque intervalle d'énergie, on calcule la différence entre la moyenne géométrique des énergies des bandes et celle des seuils de masquage des trames d'apprentissage, connus a priori. La valeur retenue pour la constante Δ sur un intervalle est celle qui maximise la distribution de probabilité des valeurs de cette différence, soit $P(\Delta/E)$.

Energie de la trame (dB)	≤ 4]4,8]]8,12]]12,16]]16,22]]22,32]]32,48]
Constante Δ (dB)	0	1,5	3,5	4,5	6,5	9,5	10,5
Energie de la trame (dB)]48,56]]56,60]]60,62]]62,66]]66,68]]68,70]]70,91]
Constante Δ (dB)	11,5	12,5	13,5	12,5	14,5	15,5	11,5

Tableau 4.2 : Constante de correction en fonction de l'énergie moyenne de la trame.

4.3.4 Résultat

Des seuils de masquage obtenus par cette méthode d'estimation, pour une trame, sont donnés dans la figure 4.4 en comparaison avec ceux calculés par le modèle psychoacoustique 1 à partir du signal temporel original. Le signal de cette trame est initialement codé par le codeur MPEG-1 Layer II à la fréquence d'échantillonnage 32 kHz et avec un débit binaire de 96 kbit/s. Notons que pour les sous-bandes ayant un nombre de bits initial nul, on affecte la valeur minimale du seuil d'audition absolu sur la bande. Dans cet exemple, ils correspondent aux bandes $j = 26 \dots 30$.

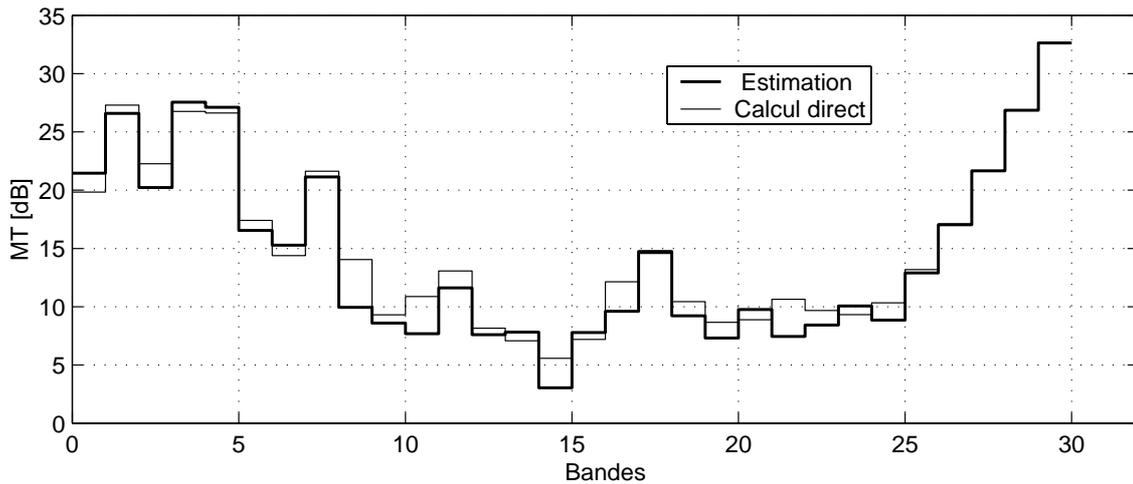


Fig. 4.4 : Comparaison des seuils de masquage estimés et originaux pour l'exemple d'une trame.

4.4. Algorithme de sommation

4.4.1 Aperçu général

Considérons le cas de la sommation de flux codés en MPEG-1 correspondant à N signaux d'entrée. L'algorithme de recombinaison a pour but de construire le train binaire du signal somme de sortie à partir des simples informations contenues dans les trames des signaux individuels d'entrée. Le décodage partiel qui permet de lire ces informations constitue alors la première étape de l'algorithme (figure 4.2). Les bancs de filtres étant linéaires, la sommation peut s'effectuer directement sur les échantillons des sous-bandes pour obtenir ceux correspondant au signal de la trame de sortie :

$$x_T(k, j) = \sum_{n=1}^N x_n(k, j), \quad 0 \leq k \leq K-1, \quad 0 \leq j \leq M-1. \quad (4.7)$$

où K est la taille d'un bloc d'échantillons sur une sous-bande (égale à 12 pour Layer I et 36 pour Layer II).

Une fois ces échantillons calculés, les facteurs d'échelle des sous-bandes peuvent être déterminés par la même méthode utilisée dans le codeur MPEG-1. Comme il a été mentionné dans le début de ce chapitre, la problématique essentielle à laquelle doit répondre l'algorithme concerne la détermination de l'allocation des bits. Le cœur de l'algorithme se propose donc d'extraire d'abord l'information sur les seuils de masquages correspondants à chacun des signaux individuels d'entrée. Il utilise pour cela la méthode d'estimation du paragraphe 4.3. Ensuite, une recombinaison de ces paramètres est faite pour construire ceux correspondant à la trame du signal somme. Les puissances des sous-bandes étant simples à estimer à partir des échantillons correspondants, les rapports signal à masque, $\{SMR_T(j)\}$, de la trame de sortie peuvent être donc déterminés et l'allocation binaire déduite par simple utilisation de la procédure itérative classique. La figure 4.5 illustre le schéma de base de cet algorithme de sommation.

Le point essentiel qui reste à étudier dans ce schéma de sommation concerne l'estimation des seuils de masquage globaux. Il faut donc trouver une relation entre les seuils de masquage des signaux à

recombinaison et celle du signal somme. Nous discuterons d'abord du problème d'additivité des courbes de masquage.

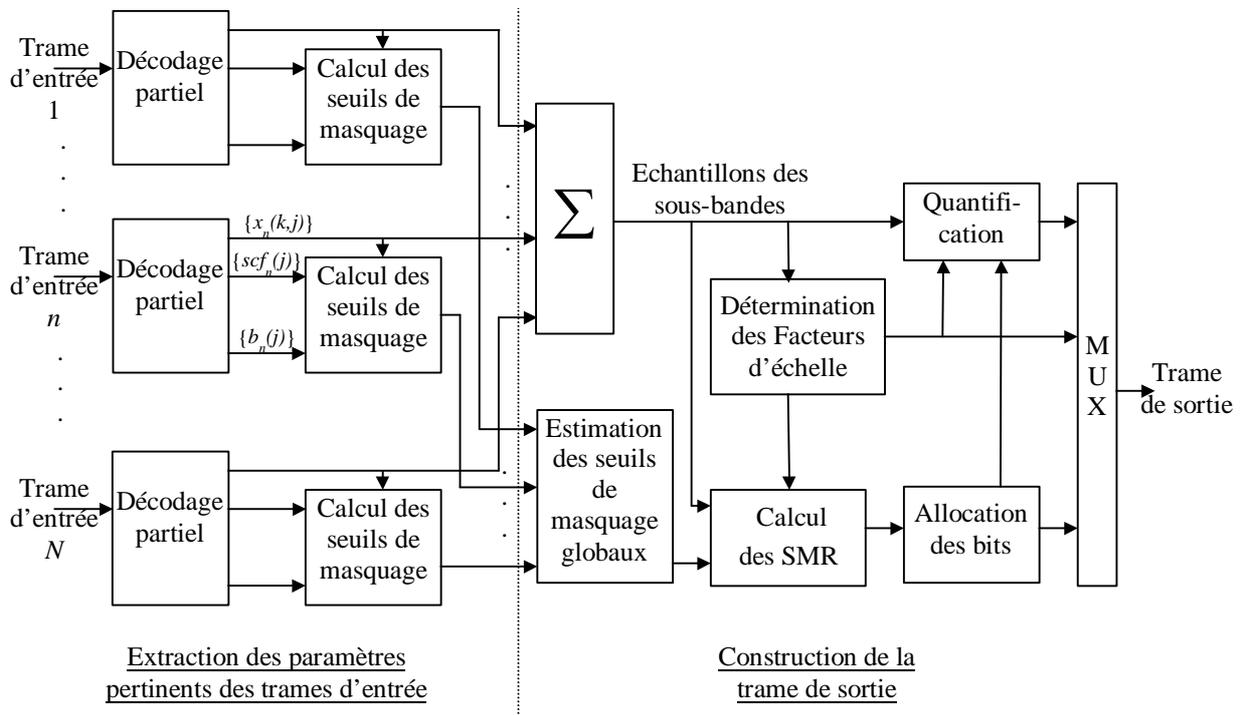


Fig. 4.5 : Schéma bloc de la méthode de sommation dans le domaine codé MPEG-1.

4.4.2 Additivité des seuils de masquage

Le principe de combinaison des courbes de masquage dues aux sons élémentaires (sinusoïde ou bruit à bande étroite), dans les modèles psychoacoustiques 1 et 2 de la norme de codage MPEG-1, est basé sur une loi d'addition linéaire [IS 11172-3]. Ceci suppose une linéarité de la courbe de masquage en fonction des effets de masquage individuels. En fait, cette hypothèse n'est pas tout à fait exacte. En effet, les études menées dans [Humes 89, Humes 92] ont fait la comparaison de plusieurs modèles d'addition. Elles ont montré que le système d'audition a un comportement non linéaire. L'effet de masquage résultant de deux stimuli est généralement plus important que le masquage global calculé à partir de la somme des deux courbes de masquage individuelles. La validité de la loi d'addition linéaire est restreinte au cas où les spectres des deux signaux masquant sont à l'intérieur d'une même bande critique. Le modèle basé sur la loi de puissance modifiée est apparu le plus adapté aux données expérimentales. Pour ce modèle la transformation qui donne le masquage résultant de la combinaison de deux sons masquant élémentaires A et B, à partir de leurs masquages individuels, est donnée par :

$$J(M_{A+B}) = J(M_A) + J(M_B), \quad (4.8)$$

avec $J(M_x) = \left(10^{M_x/10}\right)^p - \left(10^{QT/10}\right)^p$, où QT est le seuil d'audition absolu et p varie entre 0,1 et

0,5. Ce modèle a été déjà utilisé dans le système de mesure objective de la qualité audio PAQM (Perceptual Audio Quality Measure) [Beerends 92].

Malgré ces résultats, on adoptera l'hypothèse de linéarité des effets de masquage. Ce qui correspond à utiliser la formule (4.8) avec $p=1$. Cette hypothèse est faite surtout pour rester en cohérence avec le traitement réalisé dans le cas d'une sommation directe dans le domaine temporel (figure 4.1a) utilisant les modèles psychoacoustiques de la norme MPEG-1.

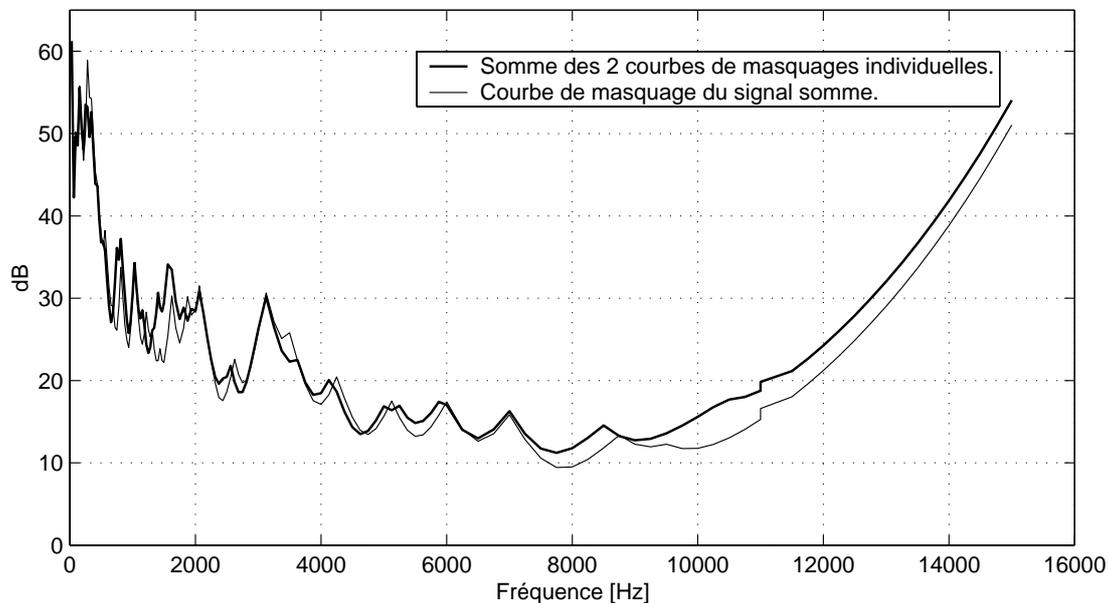


Fig. 4.6 : Comparaison de la courbe de masquage obtenue directement par le modèle psychoacoustique 1, à partir du signal somme, et de la somme des deux courbes de masquage correspondant aux signaux individuels.

La figure 4.6 donne une comparaison de la courbe obtenue directement et celle par sommation des courbes de masquage individuelles. Elle montre une légère différence qui est due au fait que le modèle psychoacoustique 1 fait intervenir quelques opérations non linéaires lors de la détermination du spectre basilaire pour le calcul de la courbe de masquage. Ainsi les deux étapes suivantes peuvent être source de non-linéarité :

- La phase d'extraction des composantes tonales à partir du spectre de Fourier du signal, $\{X(k)\}$, se fait d'abord par détermination des maxima locaux puis seules les raies qui dépassent de 7 dB un groupe de raies voisines sont sauvegardées. Cette étape d'étiquetage peut générer une différence entre les deux méthodes de calcul de la courbe de masquage. En effet, si on considère à titre d'illustration le cas de la figure 4.7, on peut voir que le calcul individuel des deux courbes de masquages 1 et 2 fait intervenir des composantes masquantes tonales. Après sommation des deux signaux, la composante X_1 étant suffisamment proche de X_2 en amplitude (différence inférieure à 7dB) et en fréquence (faisant partie des composantes voisines) aucune des deux n'est considérée comme tonale et le calcul se basera sur une courbe de masquage élémentaire non tonale. Ce phénomène intervient peu en basses fréquences (entre 0 et 2 kHz) puisque le groupe de raies voisines, auxquelles est comparé chaque maxima local, est limité à deux raies. En revanche, dans les hautes fréquences la zone de comparaison étant plus grande, le calcul direct peut éliminer plusieurs raies de la liste des composantes tonales et qui sont gardées lors d'une sommation de courbes de masquage.

- Une opération de décimation est effectuée sur les composantes de masquage tonales et non tonales pour réduire le nombre de masqueurs pris en compte pour le calcul de la courbe de masquage globale. En effet, d'une part, toutes les composantes en dessous du seuil de masquage absolu sont éliminées, d'autre part, seule la composante tonale de puissance la plus élevée est sauvegardée sur une fenêtre glissante de largeur 0,5 Bark. Il est donc possible que certaines composantes faibles, et correspondant à la même fréquence, soient éliminées lors du calcul des courbes de masquage individuelles alors qu'elles auraient été gardées si le calcul se faisait sur le signal somme. De même certaines composantes tonales sont gardées lors d'un calcul individuel mais seule la plus forte est prise en compte dans la détermination de la courbe de masquage du signal somme.

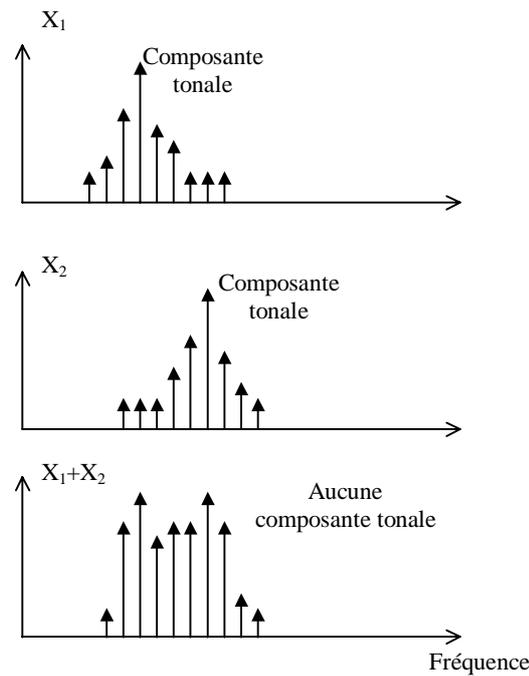


Fig. 4.7 : Illustration de l'effet de la sommation sur la détermination des composantes tonales.

4.4.3 Estimation des seuils de masquage et des rapports signal à masque du signal somme

Quoique les seuils de masquage estimés pour les signaux individuels correspondent aux minima des courbes de masquage, on utilise la loi d'addition linéaire pour estimer les seuils globaux du signal somme à partir de ces derniers paramètres. Le calcul s'effectue comme suit :

$$J'(MT_r(j)) = \sum_n J'(MT_n(j)), \quad (4.9)$$

où $J'(MT(j)) = 10^{\frac{MT(j)}{10}} - 10^{\frac{QT_m(j)}{10}}$, $QT_m(j)$ étant le minimum du seuil d'audition absolu sur la bande j . La somme dans cette expression porte sur les trames d'entrée d'indice n où le nombre de bits alloué sur la bande j n'est pas nul.

Puis

$$MT_T(j) = 10 \log_{10} \left[J'(MT_T(j)) + 10^{\frac{Q_{T_m}(j)}{10}} \right]. \quad (4.10)$$

Finalement, pour les sous-bandes où au moins un signal de voie d'entrée n'est pas nul, le rapport signal à masque de la trame totale, s'obtient par :

$$SMR_T(j) = SPL_T(j) - MT_T(j), \quad (4.11)$$

où $SPL_T(j)$ est le niveau de pression acoustique du signal somme. Il est estimé, comme dans l'expression (4.4), à partir des échantillons du signal somme, $x_T(k, j)$, $0 \leq k \leq K-1$, et du facteur d'échelle, $SCF_T(j)$, de la sous-bande j déjà déterminés dans les premières étapes de l'algorithme de sommation.

Pour les sous-bandes où toutes les trames d'entrée sont nulles, on affecte aux SMR_T une valeur très faible (-200 dB). Ceci permet de s'assurer qu'elles se verront allouer zéro bit lors de la procédure d'allocation des bits.

4.4.4 Résultats

Le noyau de l'algorithme de sommation consiste en la détermination de l'allocation des bits en estimant les nouveaux rapports signal à masque, $\{SMR_T(j)\}$. La première comparaison à faire concerne donc ces paramètres estimés avec ceux obtenus par le modèle psychoacoustique 1 lors de la phase de codage du signal somme dans le schéma de sommation directe dans le domaine temporel (figure 4.1a). La figure 4.8 illustre une telle comparaison et montre des différences relatives très légères. Ceci est dû aux approximations faites en utilisant les hypothèses sur l'optimalité de l'allocation binaire et la linéarité du masquage.

La figure 4.9 montre des statistiques sur la différence en nombre de bits alloués entre les deux méthodes sur chaque ensemble de bandes de fréquence. Ces statistiques ont été réalisées sur une base de données audio de 13 minutes de longueur, contenant 32 % de séquences musical et 68 % de séquences de parole. On peut remarquer que cette différence dépasse rarement un bit.

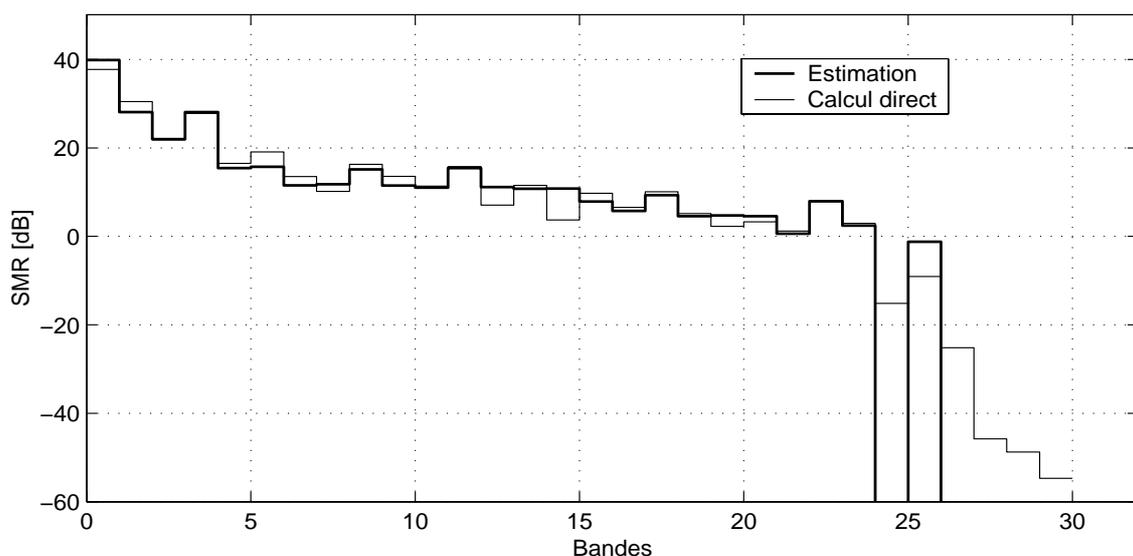


Fig. 4.8 : Comparaison des rapports signal à masque, d'une trame, relatifs au signal somme obtenus par estimation et calcul direct.

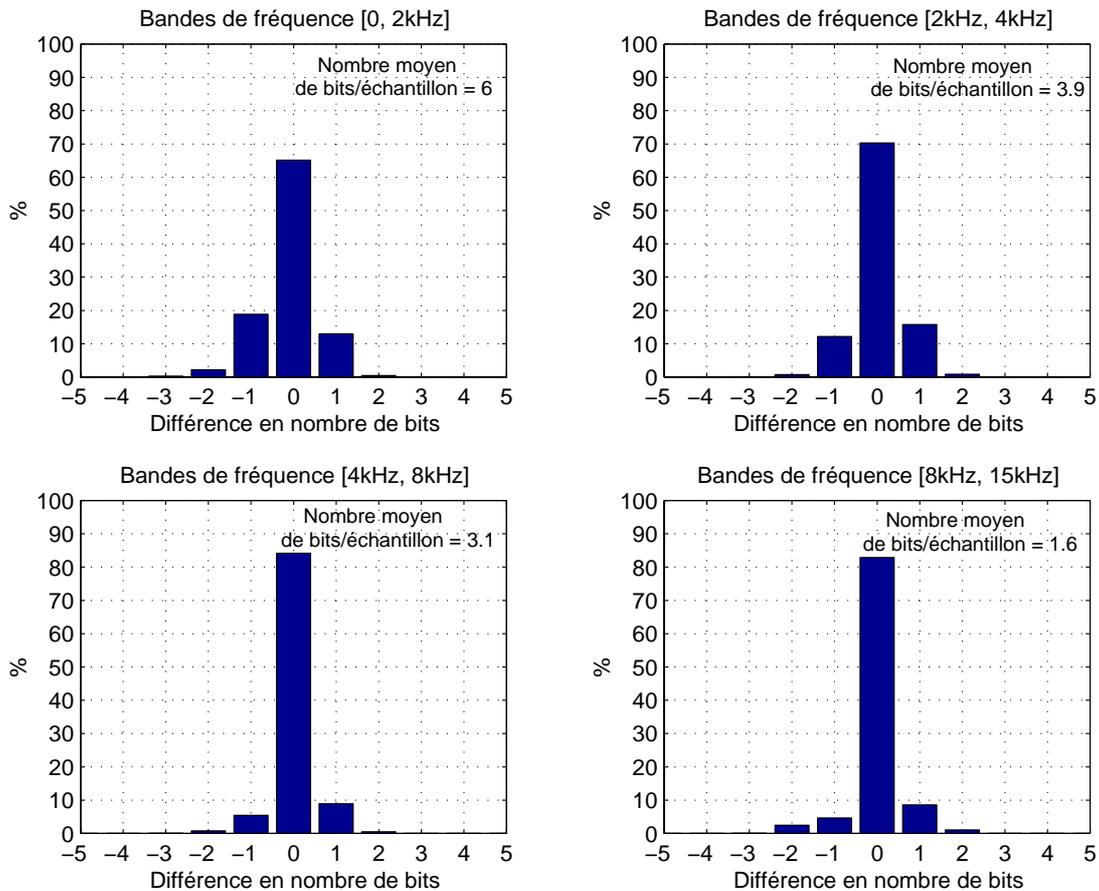


Fig. 4.9 : Comparaison des allocations de bits obtenus par la méthode directe et dans le domaine compressé (débit binaire égal à 96 kbit/s).

4.5. Variantes de l'algorithme de sommation

4.5.1 Sommation à la volée

L'algorithme de sommation proposé dans le paragraphe 4.4 nécessite la lecture de la totalité de chacune des différentes trames qui se présentent à l'entrée avant de commencer la construction de la trame totale de sortie. Cette stratégie fait accumuler le délai de lecture au délai de traitement algorithmique. L'algorithme de sommation à la volée produit les informations latérales (allocation des bits et facteurs d'échelle) de la trame somme de sortie à partir de la seule connaissance de ceux des trames individuelles d'entrée. Il calcule les échantillons des sous-bandes au fur et à mesure de la lecture des échantillons individuels. Combiné avec un mécanisme adéquat de gestion des processus de lecture et de traitement, le délai total est alors réduit au minimum.

La détermination des facteurs d'échelle dans cet algorithme se fait par simple sommation de ceux des trames d'entrée. L'allocation des bits utilise les rapports signal à masque estimés par la méthode du paragraphe 4.4.2. La seule différence réside dans l'estimation des niveaux de pression acoustique.

Comme les échantillons des sous-bandes ne sont pas disponibles, l'estimation de la puissance du signal sur chacune des sous-bandes se base sur le facteur d'échelle. Ainsi :

$$SPL(j) = 20 \log_{10} (c(j) \times SCF(j)) - 10 + 90,3 \text{ dB}, \quad (4.12)$$

où $c(j)$ est une constante tabulée qui dépend de la bande. Elle a été déterminée par estimation de l'écart type des échantillons normalisés des sous-bandes sur un corpus de séquences audio.

4.5.2 Mixage avec pondération de signaux

Considérons le cas de mixage de N signaux avec pondération par des facteurs de gains α_n , $0 \leq n \leq N$. Le schéma de sommation donné par la figure 4.5 reste toujours valable sauf quelques changements mineurs. Ainsi, lors de la sommation des coefficients des sous-bandes il faut considérer les facteurs de pondération. Pour l'algorithme d'estimation des rapports signal à masque du signal mixé le seul changement affecte la formule de sommation des seuils de masquage individuels. En tenant compte de la linéarité des seuils de masquage en fonction de la puissance du signal, cette formule devient alors :

$$J'(MT_r(j)) = \sum_n \alpha_n^2 \cdot J'(MT_n(j)). \quad (4.13)$$

4.6. Conversion du débit

La conversion du débit binaire dans le domaine compressé MPEG-1 a pour objectif de changer le nombre de bits alloué par trame. Cette opération peut s'effectuer par le même algorithme utilisé pour la sommation en considérant le cas particulier d'un seul signal ($n=1$) et en spécifiant le nouveau débit de sortie lors de la nouvelle allocation des bits. Par ailleurs, plusieurs simplifications supplémentaires peuvent être faites pour réduire davantage la complexité.

Les facteurs d'échelle des bandes de la trame de sortie sont les mêmes que ceux de la trame originale, à l'exception des bandes où la nouvelle allocation de bits donne zéro bit. Pour cette raison, lors du décodage partiel, les échantillons des sous-bandes ne sont pas dénormalisés après quantification inverse (figure 4.10).

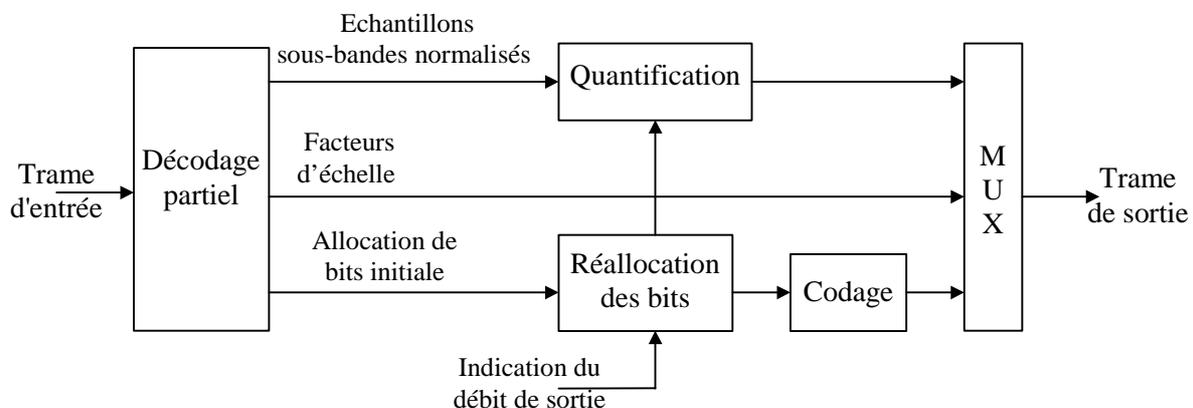


Fig. 4.10 : Schéma bloc de la réduction de débit.

Pour la réallocation de bits, il est possible d'utiliser la procédure itérative classique en tenant compte du nouveau nombre total de bits disponible et en se basant sur les rapports signal à masque estimés à partir des relations (4.3) et (4.4). L'utilisation d'une méthode qui tiendrait compte de la distribution de bits initiale pour arriver à la nouvelle serait plus avantageuse en terme de la diminution de complexité. Pour cela, examinons la fonction objective du problème de minimisation relatif à l'allocation des bits. Initialement, cette fonction s'écrit :

$$D(\mathbf{b}) = \sum_j NMR(j), \quad (4.14)$$

où $NMR(j) = SMR(j) - SNR(j)$.

Le développement du paragraphe 4.3 permet d'estimer les seuils de masquage, pour déterminer les $SMR(j)$, à partir de l'équation (4.3). Dans ce cas, il n'est pas nécessaire de déterminer la constante d'ancrage C puisqu'elle est égale pour toutes les bandes. Ainsi, la nouvelle fonction objective peut être exprimée comme suit :

$$D'(\mathbf{b}) = \sum_j [SNR(b_{init}(j)) - SNR(j)] = -\sum_j \Delta SNR(j), \quad (4.15)$$

où les $SNR(b_{init}(j))$ correspondent à l'allocation de bits initiale.

Le problème d'optimisation peut être alors reformulé comme suit :

Trouver le vecteur d'allocation de bits $\mathbf{b}' = (b'(j))_j$ qui maximise

$$\Delta SNR = \sum_j \Delta SNR(j) \text{ sous la contrainte } \sum_j b'(j) = B',$$

où B' est le nouveau nombre de bits disponibles.

La figure 4.11 donne une illustration graphique de ce problème. Il est équivalent à minimiser la distance de déplacement vers le haut que va subir le bruit de quantification, initialement mis en forme, à cause de la réduction du débit binaire. Cette distance n'est autre que la valeur absolue du $\Delta SNR(j)$.

L'algorithme en "*greedy*" utilisé dans l'allocation de bits de la norme MPEG-1 se base sur l'idée de réaliser à chaque itération la meilleure affectation qui contribue à l'optimisation de la fonction objective. Ainsi à chaque étape cette fonction, qui représente la distorsion totale, est minimale étant donnée l'allocation partielle courante [Gersho 92]. En se basant sur la même idée, on peut développer un algorithme itératif permettant de trouver une solution au problème de maximisation décrit ci-dessus et en tenant compte de l'allocation de bits initiale. Cet algorithme se déroule comme suit :

Etape 0. Initialisation :

- Les valeurs des $SNR(b_{init}(j))$ sont relevées à partir des tables données dans la norme MPEG-1 [IS 11172-3], en se basant sur l'allocation de bits initiale.
- Les valeurs des $SNR(j)$ correspondent au quantificateur juste en dessous de celui qui est utilisé initialement.

Etape 1. Décrémentement du nombre de bits alloué à la sous-bande j_0 donnée par :

$$j_0 = \arg \max_j [\Delta SNR(j)], \quad (4.16)$$

où $\Delta SNR(j) = SNR(j) - SNR(b_{init}(j))$.

Etape 2. Mise à jour du $SNR(j_0)$ en prenant dans les tables la valeur correspondant au quantificateur juste en dessous de celui utilisé actuellement.

Les étapes 1 et 2 sont répétées de manière itérative jusqu'à ce que le nouveau nombre total de bits disponible, B' , soit atteint. Notons qu'à chaque itération de l'algorithme, seules les sous-bandes avec un nombre de bits non nul sont prises en compte.

Une autre interprétation de ce problème d'optimisation peut être considérée. En effet, en remarquant que les niveaux de pression acoustique et les seuils de masquage ne sont pas affectés par le changement du débit binaire, on arrive à écrire :

$$\Delta SNR(j) = -\Delta NMR(j). \quad (4.17)$$

Le critère précédent est donc équivalent à minimiser l'augmentation de la fonction de distorsion totale, NMR , sous la contrainte du nouveau nombre de bits disponible. Ce qui revient à déplacer le point de convergence initial vers le point optimal le plus proche. L'approche développée ci-dessus rejoint donc le troisième algorithme présenté dans [Nakajima 98]. Notons aussi que ce dernier peut s'appliquer pour retrouver la couche fondamentale dans le transcodeur hiérarchique développé dans [Hans 98] (voir paragraphe 4.2).

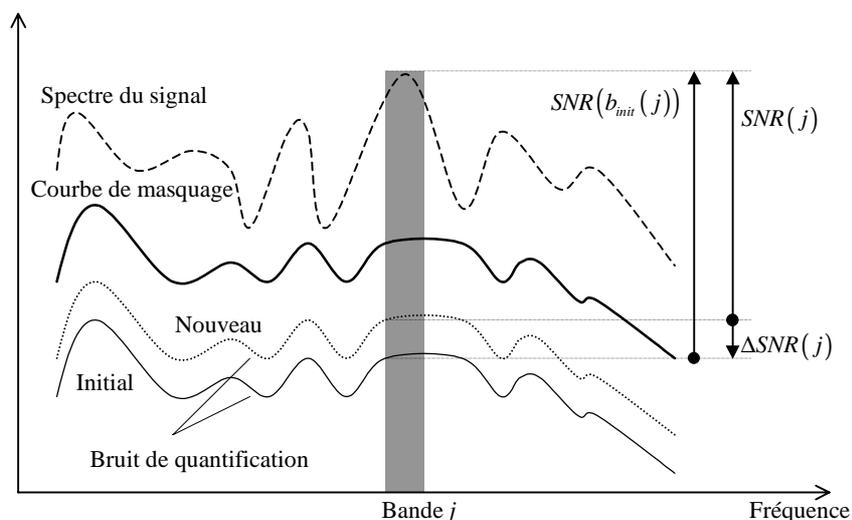


Fig. 4.11 : Illustration de l'effet de la réduction du débit binaire.

4.7. Evaluation de l'algorithme de sommation

L'implantation de l'algorithme de sommation dans le domaine compressé a montré de bonnes performances en termes de réduction du retard et de la complexité par rapport à la méthode de traitement directe dans le domaine temporel. Ainsi, pour Layer II, le retard est réduit par un facteur 3 (de 150 ms à 50 ms pour une fréquence d'échantillonnage à 32 kHz) et la complexité est diminuée d'environ 70 %. Afin d'évaluer aussi la performance en termes de qualité des signaux audio obtenus par cette méthode, nous avons effectué des tests subjectif et objectif que nous présentons dans la suite.

4.7.1 Test subjectif

Une première évaluation de la qualité a été menée à l'aide d'un test subjectif formel à triple stimulus A-B-C. Dans ce test, les différents auditeurs effectuent l'écoute respective des trois séquences suivantes :

- Séquence A : correspond au signal original obtenu par sommation de deux signaux originaux.
- Séquence B : correspond au signal résultant de la méthode de sommation dans le domaine temporel (figure 4.1a).
- Séquence C : correspond au signal résultant de la méthode de sommation dans le domaine compressé (figure 4.1b).

Les séquences B et C sont présentées dans un ordre aléatoire et les deux possibilités sont fournies. Chaque auditeur évalue la différence de qualité de chacune des séquences B et C par rapport à la séquence de référence A. Les écoutes ont été effectuées sur casque. L'échelle de notation adoptée est celle utilisée dans la méthode DCR (Degradation Category Rate) de la recommandation ITU-T P.800-Annexe D [P.800]. Cette échelle comporte cinq niveaux de dégradation :

- dégradation imperceptible,
- dégradation perceptible mais non gênante,
- dégradation légèrement gênante,
- dégradation gênante,
- dégradation très gênante.

Dix auditeurs ont participé à ce test et six séquences de musique et de parole ont été utilisées. Le codeur considéré pour obtenir les différentes séquences compressées initiales est le MPEG-1 Layer II à la fréquence d'échantillonnage $f_e = 32$ kHz, au débit binaire 96 kbit/s et utilisant le modèle psychoacoustique 1. Les résultats obtenus sont résumés dans le tableau 4.3. Ils montrent la transparence de l'algorithme de recombinaison par rapport au schéma classique de sommation dans le domaine temporel.

Méthode de sommation	Moyenne	Ecart-type	Intervalle de confiance
<i>Domaine temporel</i>	4,63	0,48	0,088
<i>Domaine compressé</i>	4,66	0,48	0,089

Tableau 4.3 : Résultats du test subjectif.

4.7.2 Mesure objective

La qualité des signaux audio a été aussi évaluée à l'aide du système de mesure de la qualité objective POM (Perceptual Objective Measurement) [Colomes 94]. Ce système utilise un modèle d'audition pour déterminer une représentation interne du signal dans le domaine basilaire. Ainsi, pour des périodes du signal de 20 ms, il procède d'abord par une transformée de Fourier discrète pour

obtenir une représentation fréquentielle. L'effet d'atténuation sur le signal de l'oreille externe et moyenne est intégré par une fonction de transfert d'un filtre passe-bas. Pour modéliser le signal au niveau de l'oreille interne un passage de l'échelle des Hertz à l'échelle des Bark est d'abord effectué puis l'énergie du signal est regroupée sur cette échelle avec une résolution de $1/25$ Bark (soit 625 raies basilaire sur l'ensemble des 25 bandes critiques). Une fonction d'étalement est utilisée ensuite pour calculer l'excitation de chacune des composantes puis l'excitation globale est déterminée en utilisant une loi d'addition linéaire. La représentation du signal est donc donnée par un vecteur d'excitation basilaire de 625 points.

Pour comparer un signal sonore à un autre de référence, une mesure de distance est effectuée à partir de leurs vecteurs d'excitation basilaire. Cette distance représente la dégradation basilaire qui est fournie à la sortie du système POM. La figure 4.12 montre les dégradations basilaire des signaux résultant respectivement des méthodes de sommation dans le domaine temporel et dans le domaine compressé, comparés au signal somme original. Elle confirme l'équivalence entre les deux méthodes de sommation.

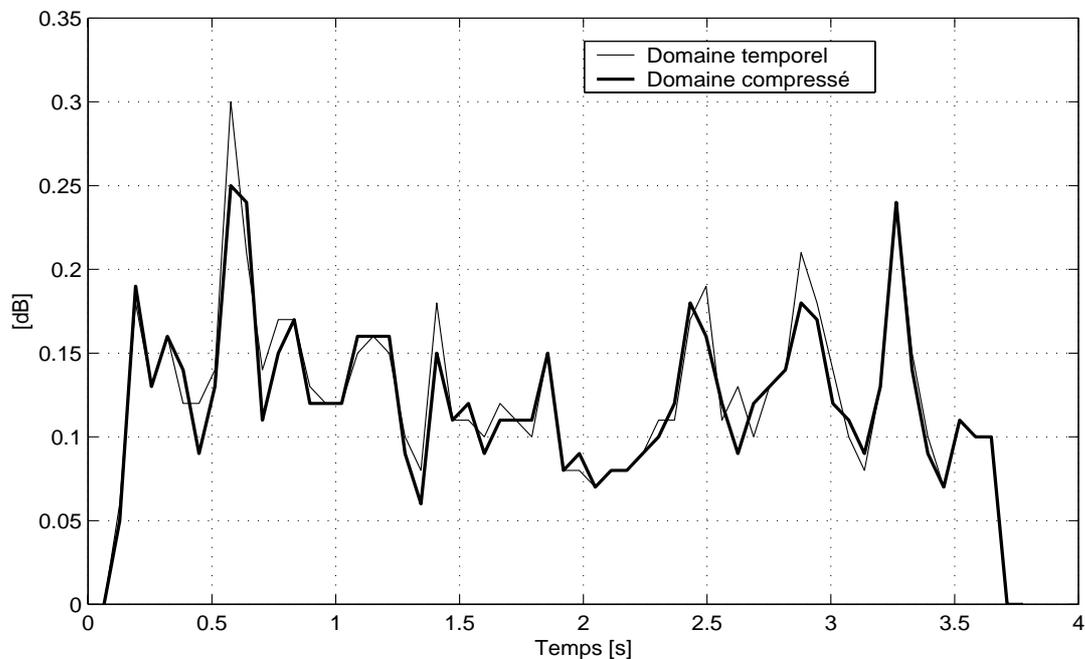


Fig. 4.12 : Comparaison des dégradations basilaire entre la référence et les signaux issus de (a) la méthode temporelle (b) la méthode dans le domaine compressé pour un signal somme (parole+piano).

4.7.3 Sensibilité de l'algorithme à la variation des modèles psychoacoustiques

Dans le standard de codage audio MPEG-1 seul le processus de décodage et la syntaxe du flux binaire codé sont normalisés. Pour donner une certaine flexibilité et fournir des possibilités d'amélioration, les parties du codeur concernant l'allocation binaire et les modèles psychoacoustiques sont données seulement à titre informatif [IS 11172-3]. L'algorithme de sommation doit donc tenir compte de cette variabilité des codeurs MPEG-1 utilisés initialement lors de la construction des trames d'entrée. En effet, la méthode d'estimation des rapports signal à masque de l'algorithme décrite dans le

paragraphe 4.4 s'inspire essentiellement du modèle psychoacoustique 1. Pour des trames d'entrée initialement codées par utilisation du modèle psychoacoustique 2, l'algorithme doit toujours assurer la même qualité sonore.

Dans cet algorithme la phase liée au codage initial concerne l'estimation des seuils de masquage individuels basée sur la formule (4.3). Seule la détermination du $SPL(j)$ et de la constante C dépendent du modèle psychoacoustique 1 dans cette formule. Mais le calcul du premier paramètre ne change pas trop dans le modèle psychoacoustique 2, en effet les rapports signal à masque sont fournis par :

$$SMR(j) = 10 \log_{10} \left[\frac{e(j)}{MT(j)} \right], \quad (4.18)$$

où $e(j)$ est l'énergie dans la bande j calculée à partir des composantes du spectre de Fourier du signal.

Un test d'écoute informel a été mené en utilisant des séquences initialement codées par un codeur utilisant le modèle psychoacoustique 2. Aucune différence perceptuelle n'a été trouvée, ce qui assure la robustesse de l'algorithme face à la variation des codeurs.

4.8. Conclusion

Ce chapitre s'est intéressé au traitement dans le domaine compressé MPEG-1. Nous avons considéré principalement le problème de sommation pour lequel un algorithme a été développé. Cet algorithme permet une réduction considérable du retard et de la complexité sans perte de qualité par rapport au schéma classique opérant dans le domaine temporel. Un algorithme de conversion du débit a été aussi dérivé pour le cas particulier d'un seul signal. Il illustre un exemple de possibilité d'effectuer un traitement minimal sur le flux codé et avec un décodage partiel minimal.

L'algorithme proposé pour la sommation concerne essentiellement la combinaison des allocations des bits et est basé sur deux hypothèses principales : optimalité de l'allocation des bits lors du codage initial et linéarité du masquage. Quoique la validité théorique de ces hypothèses reste objet à discussion, de bons résultats qualitatifs ont été obtenus. En fait, ceci n'est pas surprenant. Dans cet exemple du codeur MPEG-1, le calcul des seuils de masquage ne correspond pas exactement aux vrais seuils théoriques. Les modèles développés pour le système d'audition humaine sont encore approximatifs et très simplifiés. Ils ne correspondent pas de manière exacte aux vrais phénomènes qui se passent dans l'oreille humaine. Dans l'état de l'art des connaissances actuelles, aucun modèle psychoacoustique ne peut être qualifié d'exact [Brandenburg 98]. D'autre part, le problème d'allocation de bits n'est pas exactement résolu comme l'exige la théorie de l'optimisation psychoacoustique [Guimares 98]. En résumé, le codeur MPEG-1 ne correspond pas au codeur perceptuel optimal. Un tel exemple de codeur montre qu'il y a un certain degré de liberté dans la conception des algorithmes de codage perceptuel. Ce degré de liberté explique les résultats obtenus pour l'algorithme proposé.

Certains principes développés dans ce chapitre peuvent trouver d'autres contextes applicatifs. Ainsi, la méthode d'estimation des seuils de masquage à partir des données de la trame peut être utilisée dans les algorithmes de tatouage de signaux audio dans le domaine compressé. Le principe de base de tels algorithmes consiste à utiliser une modulation à spectre étalé des données de tatouage (Watermark), le

signal résultant est ensuite transposé dans le domaine des sous-bandes par le banc de filtre d'analyse. Les signaux des sous-bandes du signal de tatouage sont pondérés par les seuils de masquage pour que ce dernier soit dissimulé dans le signal audio utile. Un exemple d'un tel schéma est présenté dans [Neubauer 00] pour le codeur MPEG-4 AAC. Il peut être transposé pour le codeur MPEG-1 et 2 BC Layer I et II directement en utilisant les seuils de masquage estimés par la méthode du paragraphe 4.3.

Chapitre 5 **Traitement dans le domaine compressé du codeur TDAC de FTR&D**

5.1. Introduction

Dans le chapitre précédent nous avons étudié le cas du traitement de flux audio codés MPEG-1, où nous avons discuté les possibilités de traitements avancés sur le flux codé. Dans la même optique, nous considérons dans ce chapitre le cas du codeur TDAC développé par FTR&D. Le traitement principalement étudié concerne toujours la sommation.

Pour le cas du codeur MPEG-1, la problématique essentielle soulevée concerne l'allocation de bits. Si ce problème existe pour ce codeur c'est principalement à cause de l'utilisation d'une transformée, ou banc de filtres, à faible résolution fréquentielle. Une telle transformée ne permet pas d'utiliser le résultat de son analyse pour calculer directement les paramètres psychoacoustiques nécessaires pour la détermination de l'allocation des bits. Dans la méthode élaborée dans le chapitre précédent, pour résoudre ce problème, on peut dissocier deux étapes : d'abord l'estimation des paramètres de masquage puis leur recombinaison en vue de l'utilisation ultérieure. Cette première étape est nécessaire dans le cas de ce codeur à cause de la transmission explicite de l'information d'allocation des bits.

Le codeur TDAC de FTR&D constitue un autre type de codeur fréquentiel perceptuel qui n'a pas ces mêmes propriétés. D'une part la transmission de l'information sur l'allocation des bits est faite de manière implicite en transmettant, dans la trame, l'enveloppe spectrale qui permettra le calcul de l'allocation de bits au niveau du décodeur. D'autre part, ce codeur utilise une transformée à forte résolution fréquentielle. Après une phase de décodage partiel permettant l'extraction des informations de la trame (coefficients MDCT, enveloppe spectrale...), une sommation directe peut être effectuée. La recombinaison des informations latérales qui se fait pour le cas de ce codeur ne nécessite donc aucune phase préalable d'estimation. Une autre particularité de ce codeur est l'utilisation de la quantification vectorielle algébrique sphérique (QVAS). En effet, si dans le cas du codeur MPEG-1 la quantification scalaire utilisée a une complexité négligeable par rapport aux autres blocs du codage et du décodage ce n'est plus le cas pour la QVAS (voir tableaux 5.1a et 5.1b). Examiner les possibilités d'aller le plus loin possible dans le flux codé en essayant de faire une intégration et une factorisation entre les opérations de la quantification inverse et de la requantification par exploitation de leurs propriétés spécifiques est aussi l'un des problèmes que l'on cherche à résoudre.

Ce chapitre suit l'organisation suivante : d'abord nous résolvons différents problèmes dus à la recombinaison des informations latérales lors d'une sommation directe sur les coefficients MDCT entièrement déquantifiés. Le problème de factorisation entre quantification inverse et directe sera

ensuite examiné pour le cas simple de la conversion du débit où un seul signal est mis en jeu. L'extension des principes issus de ce cas particulier sera faite dans le cas de la sommation de plusieurs signaux. Finalement, une évaluation de la qualité audio et aussi des performances (retard et complexité) des différentes méthodes élaborées dans ce chapitre sera donnée dans un dernier paragraphe.

Blocs du codeur		Pourcentage du temps de calcul total
Transformée MDCT directe		10,2 %
Détection de la tonalité		12,7 %
Détection du voisement		2,7 %
Courbe de masquage		12,8 %
Enveloppe spectrale	Calcul à partir du spectre MDCT	2,5 %
	Quantification et codage entropique	3,0 %
Seuils de masquage par bandes		0,8 %
Allocation dynamique des bits		10,8 %
QVAS	Etapes préliminaires	16,4 %
	Algorithme de Schalkwijk	11,3 %

(a)

Blocs du décodeur		Pourcentage du temps de calcul total
Transformée MDCT inverse		19,7 %
Décodage de l'enveloppe spectrale		4,5 %
Seuils de masquage par bande		1,4 %
Allocation dynamique des bits		44,4 %
QVAS inverse	Etapes préliminaires	2,2 %
	Algorithme de Schalkwijk inverse	21,3 %

(b)

Tableau 5.1 : Analyse de la complexité des principaux blocs des processus (a) du codage (b) et du décodage TDAC de FTR&D.

5.2. Sommation directe dans le domaine fréquentiel

5.2.1 Description générale

La sommation des trames audio codées TDAC dans le domaine fréquentiel commence d'abord par une phase de décodage partiel des différentes trames d'entrées. Dans l'étude considérée dans ce paragraphe, cette étape du décodage partiel se réalise de manière similaire au schéma du décodage TDAC classique en excluant simplement le bloc de synthèse permettant d'obtenir le signal temporel à partir des coefficients MDCT. Cette première version de méthode de sommation, élaborée dans ce paragraphe, sera appelée méthode V1. La figure 5.1 explicite le schéma de cette phase de décodage partiel à l'issue de laquelle on dispose de l'ensemble des coefficients MDCT déquantifiés, $y_q(j)$, $0 \leq j \leq M-1$, des valeurs des coefficients de l'enveloppe spectrale déquantifiés, $e_q(i)$, $0 \leq i \leq K-1$, et des informations sur le voisement et la tonalité du signal de la trame.

La première étape de la phase de recombinaison, pour obtenir la trame de sortie, consiste à calculer les coefficients MDCT de la trame de sortie par sommation de ceux des trames d'entrée :

$$y_T(j) = \sum_{n=1}^N y_{q,n}(j), \quad 0 \leq j \leq M-1. \quad (5.1)$$

L'étape suivante vise à déterminer les informations latérales consistant d'une part en l'enveloppe spectrale utilisée pour la détermination de l'allocation des bits et d'autre part les paramètres utilisés pour le choix du type du codage de l'enveloppe spectrale et des caractéristiques de la quantification. Ainsi, la nature voisée / non voisée de cette trame de sortie est déterminée en se basant sur le voisement et l'énergie des trames d'entrée utilisées pour la construire, et une prédétermination de la tonalité de la trame est aussi effectuée. Ces différents paramètres serviront à alimenter le codeur partiel dont le schéma bloc est donnée par la figure 5.2. Nous expliciterons avec plus de détails certains blocs de ce schéma ainsi que la recombinaison des informations latérales indiquées ci-dessus dans les paragraphes suivants.

5.2.2 Détermination du voisement

Dans le schéma du codage TDAC, l'enveloppe spectrale est codée en utilisant une quantification de l'erreur de prédiction fréquentielle dans le domaine logarithmique suivi d'un codage entropique. Les mots de code utilisés dans le codage de Huffman dépendent de la nature voisée ou non voisée de la trame. Cette classification permet d'exploiter la différence entre les densités de probabilité de l'erreur de prédiction des deux cas. Elle permet ainsi de mieux optimiser les tables de codage de Huffman. Comme il a été indiqué dans le chapitre 2, la procédure de détection voisée / non voisée utilise une méthode temporelle qui se base sur trois paramètres : l'énergie du signal de la trame, le nombre de passages par zéro et le premier coefficient PARCOR.

Dans le domaine fréquentiel de la transformée MDCT, l'utilisation de cette méthode n'est plus possible. En effet, la détermination des deux derniers paramètres ne peut être faite. Par contre, il est possible de profiter de l'information déjà disponible sur le voisement des trames d'entrée contribuant à la trame somme. Pour aboutir à une stratégie de décision, nous avons d'abord effectué des statistiques sur la nature de la trame obtenue par sommation dans le domaine temporel en fonction de celles des trames d'entrée. Elles ont été effectuées sur l'ensemble des trames de la base de données parole UIT

décrite dans l'annexe C. Les résultats indiqués dans les tableaux 5.2a et 5.2b montrent que dans le cas où toutes les trames d'entrée sont de même nature (voisé ou non voisé) alors la trame correspondant au signal somme peut être considérée dans le même état. Ceci s'explique par le fait que dans le cas de sommation de plusieurs signaux voisés le spectre résultant a toujours une accentuation de l'énergie autour d'une zone réduite en basse fréquence. De même lors de la sommation de signaux non voisés, le spectre du signal résultant a toujours la propriété d'être lisse.

Nature des trames d'entrée	Nombre de cas	% des trames de sortie voisées
V / V	1207	99,2 %
V / NV	1453	93,3 %
NV / NV	2152	0,3 %

(a)

Nature des trames d'entrée	Nombre de cas	% des trames de sortie voisées
V / V / V	748	100 %
V / V / NV	1212	98,3 %
V / NV / NV	1016	90 %
NV / NV / NV	1836	0,2 %

(b)

Tableau 5.2 : Statistiques sur le voisement du signal somme (a) de 2 signaux (b) de 3 signaux.

Dans le cas de mélanges de trames voisées et non voisées la prise de décision n'est pas immédiate. Le critère additionnel utilisé consiste à comparer le rapport entre l'énergie totale des trames voisées et celle des trames non voisées à un seuil S . Ce seuil, déterminé pour le cas de la sommation de deux trames, est celui qui minimise la probabilité d'erreur de décision calculée comme suit :

$$P_{\text{err}} = P(\text{décision voisé} | \text{non voisé})P(\text{non voisé}) + P(\text{décision non voisé} | \text{voisé})P(\text{voisé}), \quad (5.2)$$

$$P_{\text{err}} = P(E_{\text{voisé}}/E_{\text{non voisé}} > S)P(\text{non voisé}) + P(E_{\text{voisé}}/E_{\text{non voisé}} \leq S)P(\text{voisé}). \quad (5.3)$$

où $E_{\text{voisé}}$ et $E_{\text{non voisé}}$ sont respectivement les énergies de la trame voisé et de la trame non voisé.

Les mesures de cette probabilité montrent que la valeur optimale du seuil S est de 0 dB. Ce qui revient à faire la décision selon le groupe de trames de plus forte énergie.

En utilisant les deux critères ci-dessus, nous avons calculé le pourcentage de fausses décisions sur le voisement en prenant comme référence celle obtenue par la procédure utilisée dans le cas de sommation dans le domaine temporel. Les résultats sont résumés dans le tableau 5.3 et sont obtenus sur un corpus de 11407 trames voisées et 3029 trames non voisées.

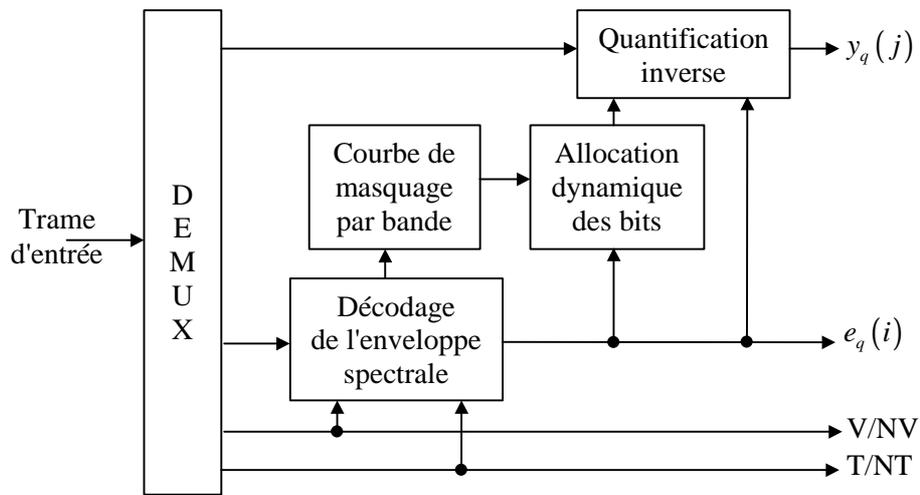


Fig. 5.1 : Schéma bloc du décodage partiel des trames TDAC.

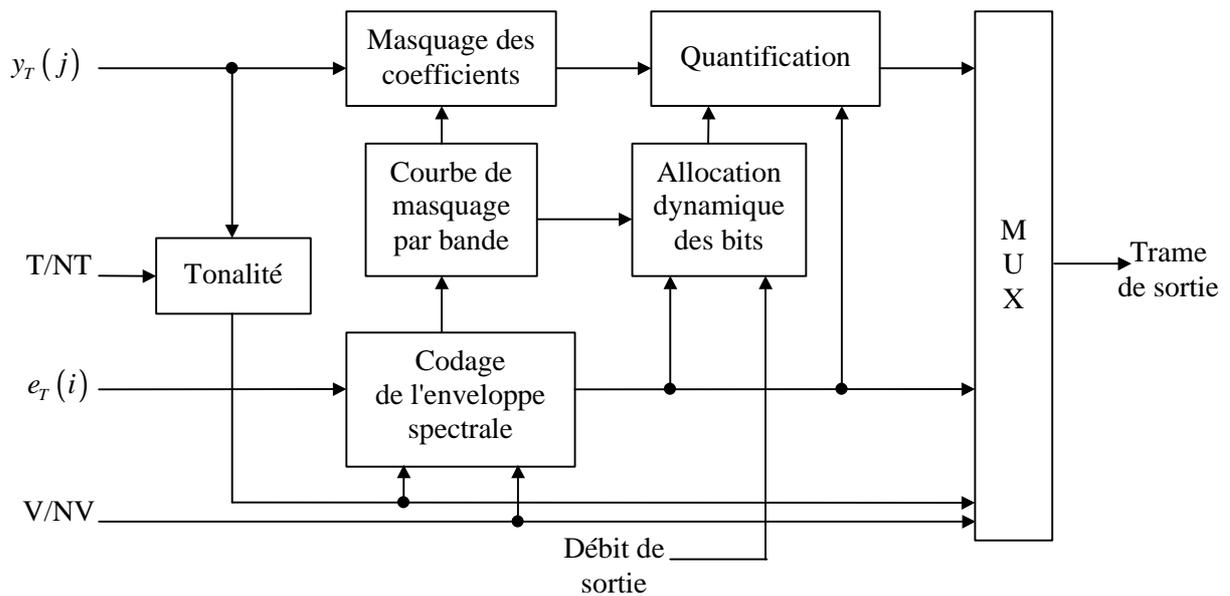


Fig. 5.2 : Schéma bloc du codage partiel des trames TDAC.

Erreur sur les trames voisées	1 %
Erreur sur les trames non voisées	8,6 %
Erreur totale sur le voisement	2,6 %

Tableau 5.3 : Pourcentage d'erreur sur la décision de voisement.

5.2.3 Détermination de la tonalité

La détermination de la tonalité dans le codeur TDAC utilise la mesure de l'étalement spectral, sfm , évaluée par le logarithme du rapport entre les moyennes géométrique et arithmétique des énergies des coefficients MDCT. Cette même méthode peut être utilisée dans ce cas d'une sommation dans le domaine fréquentiel sans problème particulier. Par ailleurs, on utilisera l'information disponible sur la tonalité des trames à sommer dans le but de réduire la complexité due à ce bloc (voir tableau 5.1a). En effet, la sommation de trames non tonales donne nécessairement une trame non tonale. La procédure de décision consiste donc à considérer la trame de sortie non tonale si toutes les trames d'entrée le sont. Dans le cas contraire, où il existe une trame tonale, la détermination de la tonalité est reportée à la phase de codage partiel (figure 5.2). Elle est réalisée, de manière classique, comme dans le codeur TDAC. Le faible pourcentage de trames tonales dans des signaux de parole ou de musique fait que le premier cas est le plus courant. Le recours à un calcul postérieur d'une mesure d'étalement spectral, consommant le plus de complexité, est très rare.

5.2.4 Enveloppe spectrale et seuils de masquage

Dans le codeur TDAC, les coefficients de l'enveloppe spectrale correspondent à la racine carrée de l'énergie moyenne des coefficients MDCT dans les bandes, pour lesquelles les coefficients masqués ne sont pas pris en compte. Pour profiter des paramètres individuels correspondant aux trames d'entrée, l'enveloppe spectrale de la trame somme est calculée comme suit :

$$e_T(i) = \sqrt{\sum_{n=1}^N (e_{q,n}(i))^2}, \quad 0 \leq i \leq K-1. \quad (5.4)$$

La figure 5.3 illustre la comparaison de l'enveloppe spectrale calculée directement à partir des coefficients MDCT du signal somme et celle estimée par la formule (5.4).

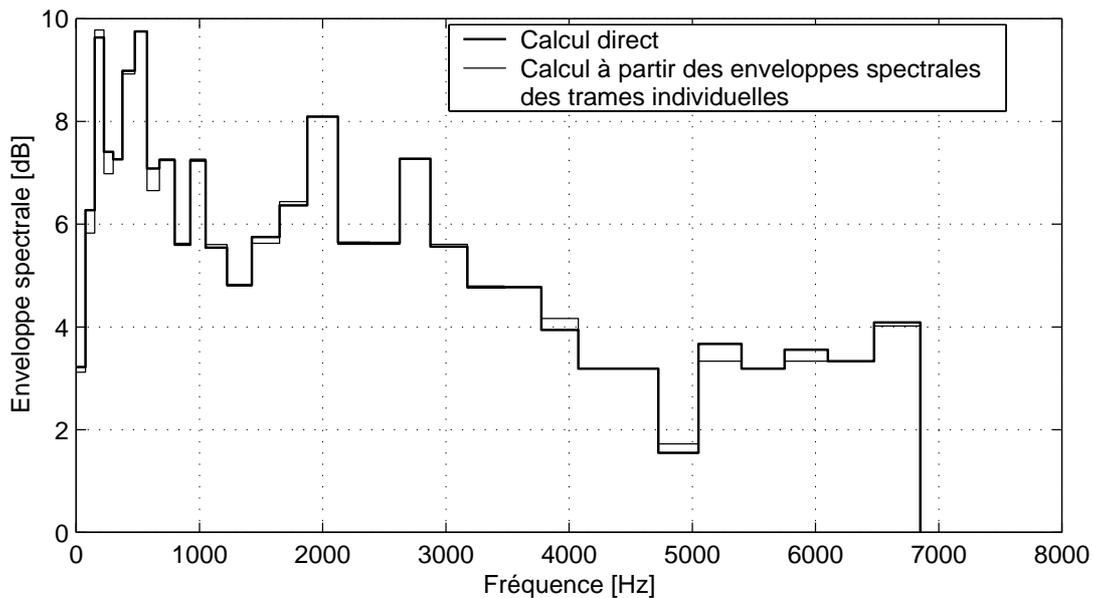


Fig. 5.3 : Exemple d'enveloppe spectrale calculée par recombinaison en comparaison avec celle obtenue lors du codage du signal résultant d'une sommation temporelle.

Cette méthode de calcul de l'enveloppe spectrale suggère de déterminer les seuils de masquage par bande, du signal somme, par sommation directe sur ceux correspondant aux trames d'entrée, obtenues lors du décodage partiel. En effet, le calcul de ces seuils de masquage, dans le codeur TDAC, est basé sur une convolution directe de la fonction d'étalement basilaire et des énergies en bandes. Il y a donc une parfaite linéarité entre l'énergie et le masquage.

Mais, les quantités, $e_T(i)$, calculées par la formule (5.4) ne sont pas directement transmises dans la trame. Ce sont leurs valeurs quantifiées, $e_{q,T}(i)$, $0 \leq i \leq K-1$, qui sont codées puis transmises. En effet, ceci permet d'avoir des facteurs d'échelle identiques dans les procédés de codage à la recombinaison et du décodage distant. L'utilisation de seuils de masquage par bandes de la trame somme obtenus par sommation de ceux des trames d'entrée entraînerait une allocation de bits au codeur différente de celle qui serait effectuée dans le décodeur distant. Cette solution de sommation directe des seuils de masquage par bande ne peut être donc utilisée.

5.2.5 Masquage des coefficients

Dans le codeur TDAC, la détection des coefficients MDCT masqués et leur mise à zéro se fait pour réduire le débit et aussi pour éliminer l'erreur de quantification sur ces raies. Des statistiques, menées sur un ensemble de 24960 trames, ont montré que lors du codage initial d'un signal en moyenne 47% de coefficients MDCT sont détectés masqués. Après décodage partiel, une bonne partie non pertinente des différents signaux à sommer a été donc déjà mise à zéro. Le pourcentage de complexité pris par le calcul d'une courbe de masquage (voir tableau 5.1a) suggère donc de l'éliminer lors de la phase du codage partiel et d'utiliser plutôt les seuils de masquage par bande qui représentent une approximation grossière de la courbe de masquage. La figure 5.4 illustre un exemple des deux types de masquages et du spectre MDCT correspondant au signal après sommation.

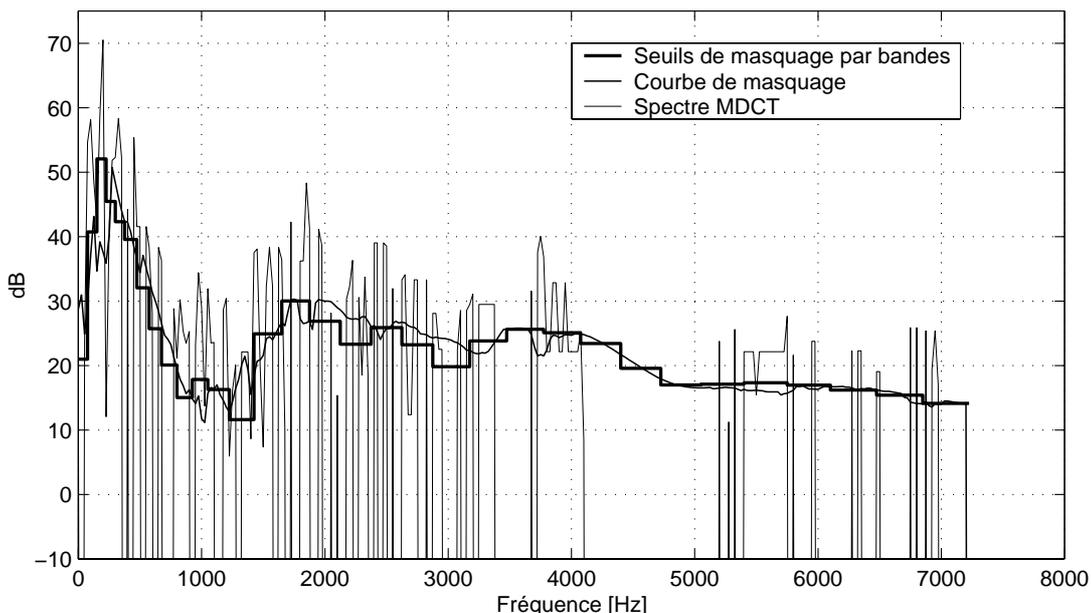


Fig. 5.4 : Exemple de courbe de masquage en comparaison avec les seuils de masquage par bande pour un spectre MDCT obtenu par sommation.

5.3. Conversion du débit sur un signal

5.3.1 Principe et description générale

Dans le but d'effectuer un traitement avancé sur les flux codés TDAC, i.e. avec un décodage partiel minimal, nous examinons dans ce paragraphe le cas simple de la conversion du débit sur un seul flux audio. La conversion, ou plus précisément la réduction, du débit (passage de 24 à 16 kbit/s par exemple) peut se réaliser de manière directe dans le domaine MDCT. Le processus d'un tel schéma préliminaire se résume en une phase de décodage partiel (figure 5.1) suivie d'une réallocation de bits basée sur le nouveau débit binaire disponible. Cette nouvelle allocation binaire utilise les mêmes coefficients d'enveloppe spectrale et seuils de masquage par bande déterminés lors du décodage partiel. Elle servira à requantifier les coefficients MDCT des bandes. Les informations latérales concernant la tonalité, le voisement et l'enveloppe spectrale codée restent les mêmes dans la nouvelle trame de sortie.

Le tableau 5.1 indique que les processus de quantification vectorielle directe et inverse prennent respectivement 27,7 % du temps de calcul total du codage et 23,5 % de celui du décodage. Une bonne partie de la complexité algorithmique de ces deux processus est due à l'algorithme de Schalkwijk permettant de déterminer le rang d'un vecteur quantifié de coefficients MDCT sur une bande dans sa classe et inversement. Dans le but de réduire cette complexité, nous proposons un schéma de conversion du débit plus évolué qui utilise une trans-quantification directe sur les mots de code des bandes quantifiées. Cette opération vise à intégrer et à éliminer certains blocs de la QVAS directe et inverse en profitant de leurs propriétés, notamment celle concernant l'imbrication des dictionnaires algébriques utilisés. Les détails de l'algorithme de trans-quantification seront donnés dans le paragraphe suivant. Le schéma de conversion complet est illustré par la figure 5.5.

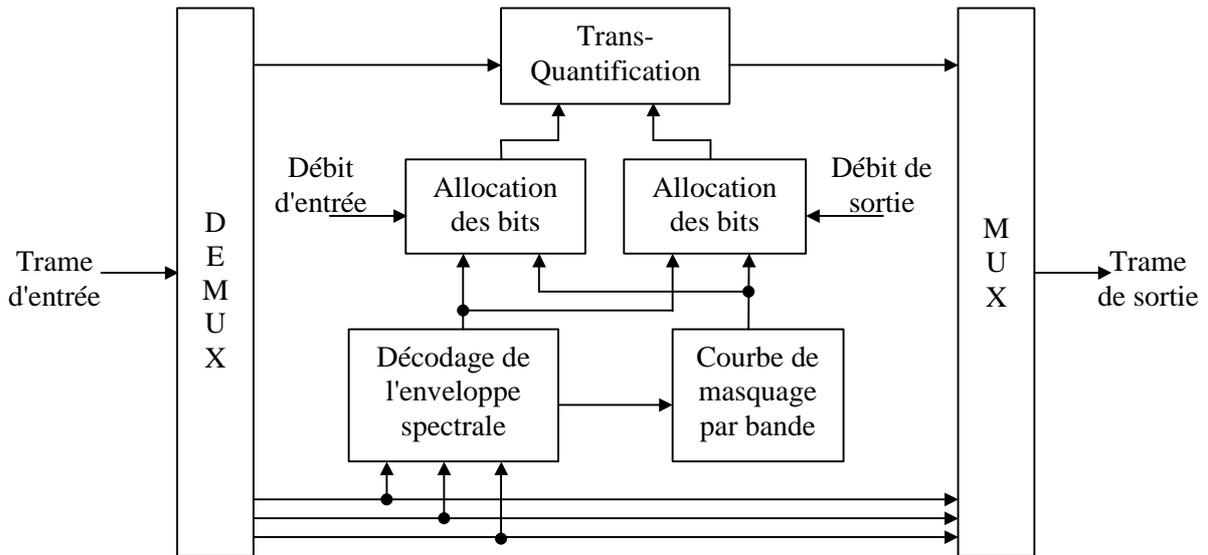


Fig. 5.5 : Schéma bloc de la conversion du débit pour une trame TDAC.

Dans ce schéma de conversion du débit deux procédures d'allocation de bits sont effectuées. Comme pour le codeur MPEG-1 (voir paragraphe 4.6), il est possible de suggérer un algorithme profitant de l'allocation de bits initiale pour passer à la finale. Un tel algorithme peut utiliser une procédure d'extraction de bits basée sur un critère perceptuel. Par ailleurs, l'allocation de bits utilisée dans le codeur TDAC se fait par une affectation préliminaire selon la formule optimale suivie d'un réajustement par ajout ou extraction de bits suivant un critère perceptuel [Perkins 89]. La seconde phase de ce processus est la plus complexe. Elle consomme 70 % du temps de calcul utilisé pour l'allocation des bits. L'utilisation d'une procédure basée seulement sur une méthode d'extraction perceptuelle des bits pour passer du débit initial à un autre plus bas (de 24 à 16 kbit/s par exemple) sera alors plus complexe. En effet, utiliser préalablement la formule optimale limite le nombre de bits à redistribuer et par conséquent la complexité. Dans ce cas du codeur TDAC une telle procédure ne présente donc aucun intérêt.

5.3.2 Algorithme de trans-quantification sur une bande

Considérons le cas de la bande i de dimension d_i . Soient b_i et b'_i , respectivement, les nombres de bits initial et final qui lui sont alloués. Le mot de code résultant de la QVAS initiale de cette bande, reçu dans la trame d'entrée, contient les informations suivantes :

- Le vecteur des signes, $\text{Sign}(i)$, des coefficients normalisés de la bande.
- L'index lexicographique, L_i , du vecteur leader quantifié, $\tilde{Y}_q(i)$, dans le dictionnaire $\mathcal{C}(b_i, d_i)$.
- Le rang, r_i , du vecteur absolu quantifié, $Y_q(i)$, dans la classe du leader $\tilde{Y}_q(i)$.

Soient $Y'_q(i)$ le nouveau vecteur absolu quantifié correspondant au nombre de bits b'_i et $\tilde{Y}'_q(i)$ son leader. On notera par L'_i l'index lexicographique de ce nouveau vecteur leader quantifié dans le dictionnaire $\mathcal{C}(b'_i, d_i)$ et par r'_i le rang de $Y'_q(i)$ dans la classe de $\tilde{Y}'_q(i)$.

Par exploitation des propriétés de la QVAS, on peut faire une recherche directe du vecteur leader quantifié $\tilde{Y}'_q(i)$ sans avoir à déterminer $Y_q(i)$ et son vecteur requantifié $Y'_q(i)$. En effet, $\tilde{Y}'_q(i)$ n'est autre que le plus proche voisin de $\tilde{Y}_q(i)$ dans le nouveau dictionnaire $\mathcal{C}(b'_i, d_i)$. Dans la QVAS utilisée dans le codeur TDAC, les dictionnaires correspondant aux différents débits sont imbriqués. Cette dernière propriété conduit à distinguer les cas suivants :

- Si $b'_i \geq b_i$ alors $\mathcal{C}(b_i, d_i) \subset \mathcal{C}(b'_i, d_i)$ donc $\tilde{Y}_q(i) \in \mathcal{C}(b'_i, d_i)$. Le plus proche voisin du vecteur $\tilde{Y}_q(i)$ dans le nouveau dictionnaire $\mathcal{C}(b'_i, d_i)$ n'est autre que lui-même :

$$\tilde{Y}'_q(i) = \tilde{Y}_q(i) \quad (5.5)$$

- Sinon, il faut distinguer deux cas selon la position du vecteur $\tilde{Y}_q(i)$ par rapport au dictionnaire $\mathcal{C}(b'_i, d_i)$:
 - Si $\tilde{Y}_q(i) \in \mathcal{C}(b'_i, d_i)$, i.e. $L_i \leq \kappa(b'_i, d_i)$ où $\kappa(b'_i, d_i)$ est le cardinal du dictionnaire $\mathcal{C}(b'_i, d_i)$, alors $\tilde{Y}'_q(i) = \tilde{Y}_q(i)$.
 - Sinon $\tilde{Y}_q(i) \in \mathcal{C}(b_i, d_i) \setminus \mathcal{C}(b'_i, d_i)$. Dans ce cas il faut faire la recherche de $\tilde{Y}'_q(i)$ puis déterminer le rang r'_i de $Y'_q(i)$ dans la classe de ce dernier.

L'algorithme complet de la trans-quantification sur la bande i , pour obtenir le mot de code de sortie à partir de celui d'entrée, se déroule donc comme suit :

1. Décodage du mot de code de la bande et extraction de $\text{Sign}(i)$, L_i et r_i .
2. On distingue deux cas :
 - i. Si l'une des conditions $(b'_i \geq b_i)$ ou $((b'_i < b_i) \text{ et } (L_i \leq \kappa(b'_i, d_i)))$ est vérifiée alors le mot de code de la trame de sortie est le même correspondant à la trame d'entrée.
 - ii. Sinon, il y a cinq étapes à effectuer dont les deux premières font parties de la quantification inverse et les trois dernières de la requantification :
 - a. Détermination du leader quantifié $\tilde{Y}_q(i)$ dans le dictionnaire $\mathcal{C}(b_i, d_i)$ à partir de son index lexicographique L_i .
 - b. Décodage par l'algorithme de Schalkwijk inverse, en utilisant L_i et r_i , pour retrouver la permutation entre $Y(i)$ et son leader $\tilde{Y}(i)$. Le vecteur $Y'_q(i)$ peut être donc obtenu en effectuant l'inverse de cette permutation sur $\tilde{Y}_q(i)$.
 - c. Recherche dans le dictionnaire $\mathcal{C}(b'_i, d_i)$ du vecteur $\tilde{Y}'_q(i)$ le plus proche voisin de $\tilde{Y}_q(i)$. L'index L'_i est ainsi déterminé.
 - d. Détermination du rang r'_i par l'algorithme de Schalkwijk en utilisant la permutation trouvée dans l'étape b.
 - e. Construction du mot de code de sortie de la bande à partir de $\text{Sign}(i)$, L'_i et r'_i .

La figure 5.6 donne une illustration de cet algorithme. Elle indique les étapes intégrées entre la quantification inverse et la requantification ainsi que celles qui sont éliminées.

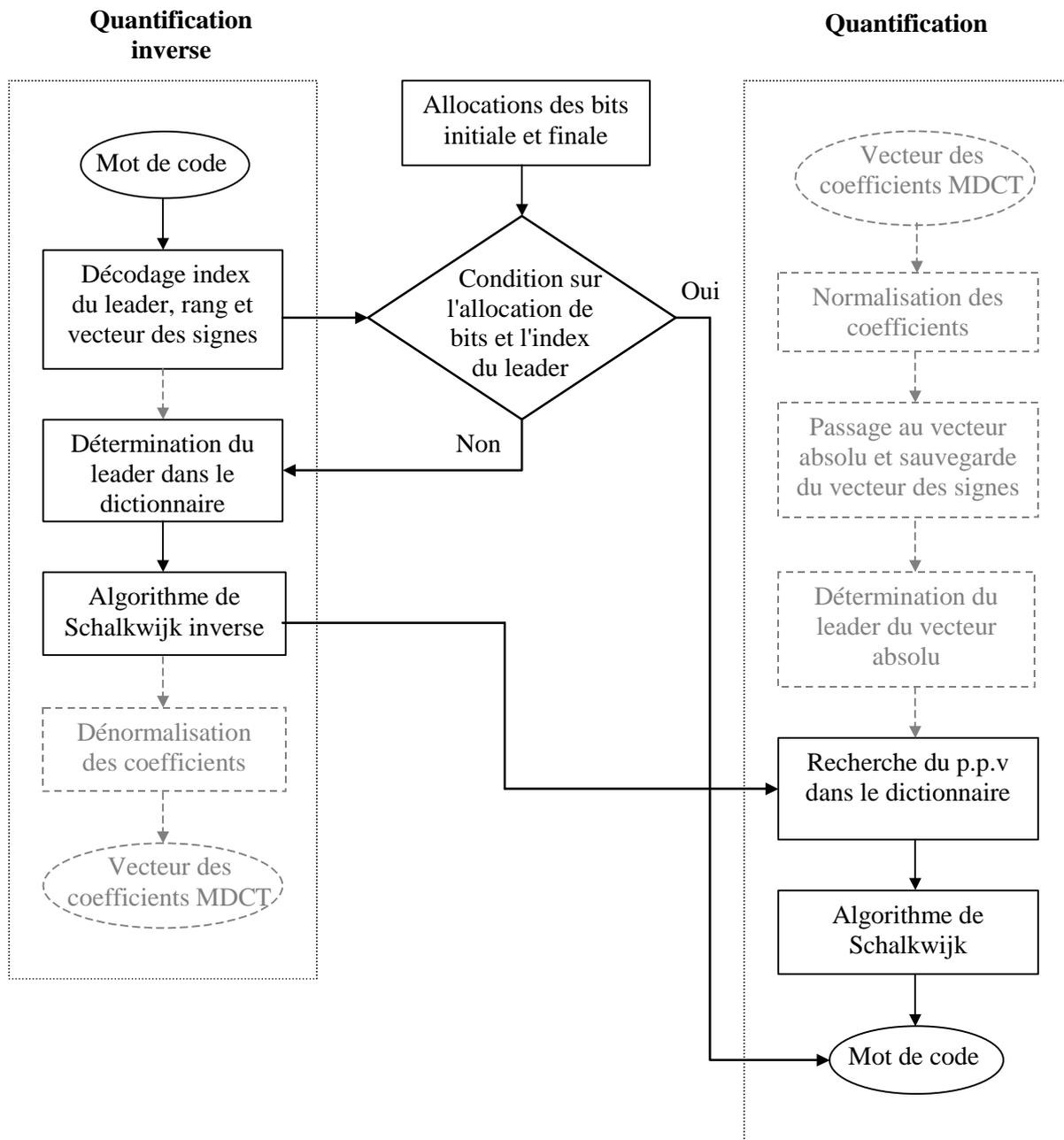


Fig. 5.6 : Algorithme de trans-quantification sur une bande. Les zones ombrées ne font pas partie de l'algorithme.

5.3.3 A propos de la QVAS et de l'allocation de bits du codeur TDAC de FTR&D

Dans le cas d'une réduction de débit, il est hautement improbable que la condition $b'_i \geq b_i$ soit réalisée lors d'une trans-quantification, puisqu'en principe toutes les bandes auront des bits retirés. La seconde condition $((b'_i < b_i) \text{ et } (L_i \leq \kappa(b'_i, d_i)))$ permet de relever une remarque importante à propos de la quantification et de l'allocation de bits du codeur TDAC. En effet, cette condition indique que $\tilde{Y}_q(i)$, le vecteur leader quantifié initial, appartient à un dictionnaire, $\mathcal{C}(b'_i, d_i)$, correspondant à un nombre de bits, b'_i , inférieur à celui alloué à la bande, à savoir b_i . Cela ne signifie pas que c'est une situation impossible qui ne peut exister, mais révèle une faiblesse du codeur TDAC de FTR&D. En effet, dans ce cas de figure il y a un "gaspillage" de bits puisqu'il suffisait à cette bande d'avoir b'_i bits ou peut-être même moins.

Pour remédier au problème qui vient d'être décrit, il faudrait réaliser conjointement l'allocation des bits et la quantification vectorielle. Une solution consiste à utiliser une méthode d'analyse par synthèse. Il s'agit de faire l'allocation de bits et la quantification vectorielle puis d'extraire aux bandes les bits en excès. Le nombre de bits retenu pour une bande est celui correspondant au plus petit dictionnaire, au sens de l'inclusion, contenant le vecteur leader quantifié de la bande. Une boucle peut être effectuée ensuite en redistribuant les bits prélevés aux bandes suivant un critère perceptuel. Cette dernière opération de redistribution de bits n'est pas si évidente puisqu'il se peut qu'on augmente le nombre de bits d'une bande qui avait déjà un excès de bits. Nous pourrions envisager l'utilisation d'un algorithme dont le principe de déroulement est le suivant :

1. Réalisation d'une allocation de bits initiale suivant la méthode classique utilisée dans le codeur TDAC.
2. Recherche de la bande où l'ajout de bits apporte l'amélioration perceptuel minimale par rapport aux autres. Pour cette bande, on effectuera la recherche du leader quantifié plus proche voisin dans le dictionnaire correspondant à l'allocation de bits courante. On retiendra pour cette bande le nombre de bits correspondant au plus petit dictionnaire, au sens de l'inclusion, contenant le vecteur leader quantifié déterminé.
3. Redistribution de l'excès de bits courant aux autres bandes restantes suivant un critère perceptuel i.e. en favorisant les bandes où l'amélioration perceptuel est la plus grande.
4. Elimination de la bande traitée et retour à l'étape 2 s'il reste encore des bandes non traitées.

Par ailleurs, l'application d'une telle méthode ne peut être faite sans changement de la structure de la trame du codeur TDAC de FTR&D. En effet, le décodage se base sur l'allocation de bits initiale faite à partir de l'enveloppe spectrale contenue dans la trame. Si cette allocation de bits est modifiée dans le codeur, le décodeur n'est plus compatible. Une solution possible est d'ajouter dans la trame une information sur le dictionnaire utilisée pour chaque bande. En tenant compte du nombre de dictionnaires utilisés, cela nécessiterait 132 bits supplémentaires comme information latérale. Ce nombre de bits à soustraire à celui réserver à la quantification des bandes ne peut compenser le nombre de bits retirés. Cette méthode n'a pas donc d'intérêt.

5.4. Méthode de sommation avancée dans le flux audio compressé TDAC

5.4.1 Principe et étude préliminaire

L'étude de la conversion du débit sur un seul signal a montré des possibilités intéressantes d'intégration entre le décodage et le recodage pour une réduction de la complexité. L'algorithme de trans-quantification sur le vecteur d'une bande est à la base d'une telle intégration. Nous nous intéressons donc à appliquer cet algorithme dans le cas de la sommation de plusieurs signaux. Pour cela, il faut utiliser un critère permettant de garder sur une bande un seul vecteur de signal parmi ceux à sommer. On se base alors sur deux phénomènes de masquage : le masquage *intra-signal* et le masquage *inter-signaux* par bandes.

Dans le codeur TDAC, une courbe de masquage est utilisée pour détecter les coefficients masqués et les mettre à zéro. Certaines bandes peuvent être totalement masquées. En plus la procédure d'allocation de bits peut fournir zéro bit à d'autres bandes. Ce phénomène de bandes entièrement nulles est intéressant lors de la sommation puisqu'il permet de ne considérer que les signaux *actifs* sur la bande. Pour évaluer la pertinence d'un tel phénomène, nous avons évalué la probabilité d'être totalement nul, après un codage TDAC, pour chacune des 32 bandes de fréquence (figure 5.7). Ces statistiques ont été effectuées sur l'ensemble des trames de parole de la base de données UIT (annexe C). La figure 5.7 montre un taux plus fort dans les bandes hautes ainsi que sur la bande 5 correspondant à la zone de fréquence [300 Hz, 375 Hz]. Le pic sur cette bande est dû à l'effet de masquage par l'énergie du pitch concentrée vers les bandes 2 à 4 ([150 Hz, 300 Hz]). Les bandes autour de la fréquence 2 kHz sont pratiquement toujours non nulles. Cette dernière constatation s'explique par le croisement de deux phénomènes :

- Les bandes sont de largeur plus grande quand la fréquence augmente. Ceci diminue la probabilité que la totalité de leurs coefficients soient masqués.
- L'effet de masquage est plus fort dans les bandes hautes d'une part à cause de la faible énergie dans cette zone par rapport aux basses fréquences et d'autre part du fait que le masquage d'une excitation élémentaire vers les hautes fréquences est plus fort. En effet, la fonction d'étalement de la membrane basilaire décroît moins vite vers les fréquences hautes (-27 dB/Bark) qu'en fréquences basses (10 dB/Bark) (voir figure 2.14).

Notons en plus que l'allocation des bits favorise généralement plus les bandes basses au détriment des bandes hautes ce qui explique aussi que les bandes dans cette zone ont une probabilité plus élevée d'être nulle.

Soit p_i la probabilité que la bande i soit totalement nulle. La probabilité qu'une seule bande soit active lors de la sommation de N signaux est alors donnée par $Np_i^{N-1}(1-p_i)$. A partir des résultats de la figure 5.7, nous illustrons dans la figure 5.8 cette probabilité pour chacune des bandes dans les cas de N égal à 2 et 3. Nous pouvons constater que ce phénomène de masquage intra-signal ne suffit pas pour arriver à des situations fréquentes où un seul signal est actif dans la bande et donc à une réduction plus forte de complexité.

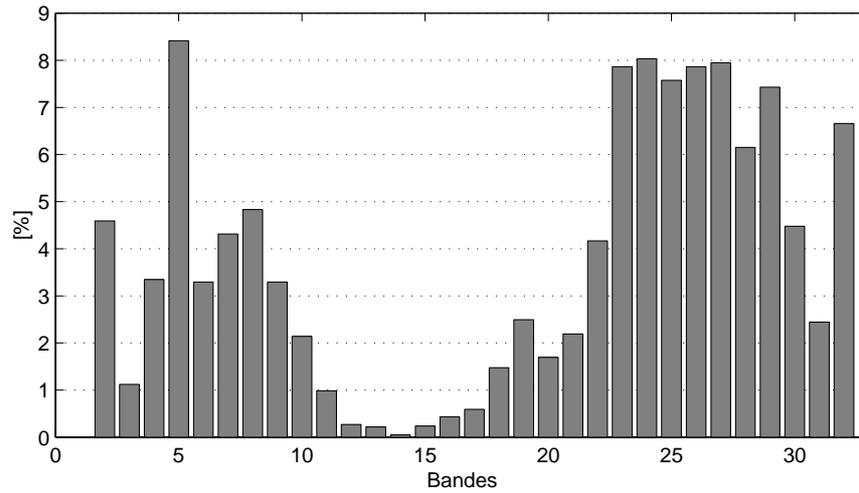


Fig. 5.7 : Probabilités moyennes pour qu'une bande de fréquence soit totalement nulle à la sortie du codeur TDAC⁷.

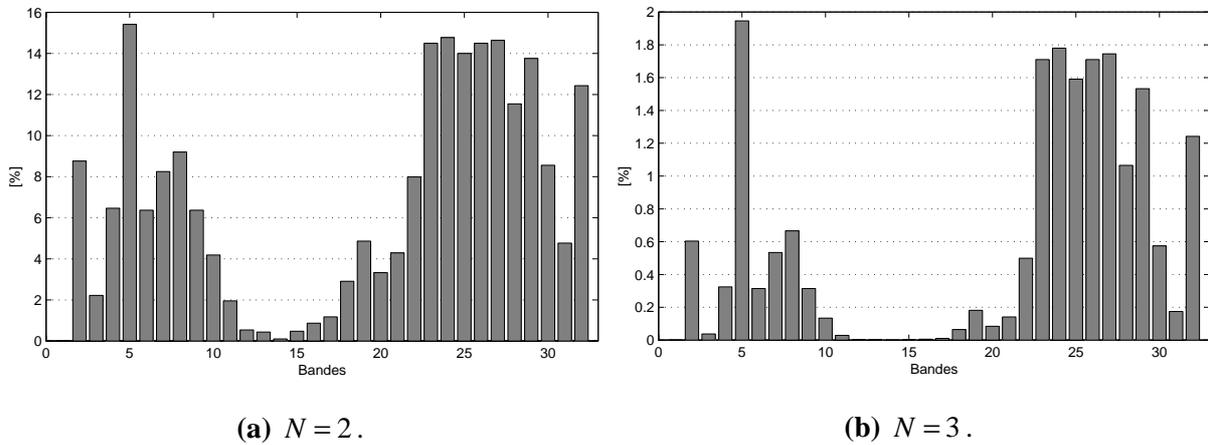


Fig. 5.8 : Probabilités moyennes qu'une seule bande soit active parmi N lors de la sommation de signaux codés.

Pour augmenter cette probabilité d'avoir un seul signal sur chaque bande, nous utiliserons en plus le masquage inter-signaux dans les bandes. Le principe consiste à chercher sur chacune des bandes le signal de plus forte énergie et à tester si les autres signaux sont masqués par ce dernier. Tout signal totalement masqué n'est pas pris en compte dans l'opération de sommation sur cette bande. Ainsi toutes les composantes perceptuellement inutiles des signaux utilisés sont éliminées. L'utilisation d'un tel critère se base essentiellement sur le fait que la découpe en bandes utilisée dans le codeur TDAC s'approche de l'échelle des bandes critiques. En effet, tout le phénomène de masquage se passe dans ces bandes [Zwicker 99].

En utilisant ce critère de masquage inter-signaux par bande, nous avons effectué des statistiques sur la base de données parole de l'UIT pour déterminer le taux moyen où une seule bande intervient lors de la sommation. Les résultats illustrés dans la figure 5.9 montrent des taux intéressants. Le taux moyen sur toutes les bandes est de 63%. Notons que les séquences utilisées sont bien synchronisées de

⁷ Dans cette figure ainsi que dans les figures 5.8 et 5.9 les bandes sont numérotées de 1 à 32.

manière à ne pas superposer une période de silence d'un signal à la période d'activité de l'autre. Les résultats obtenus n'incluent donc pas le cas où un seul signal est actif. Par ailleurs, ils englobent le cas où un masquage inter-signaux est dû au fait que la bande sur un des deux signaux était initialement nulle.

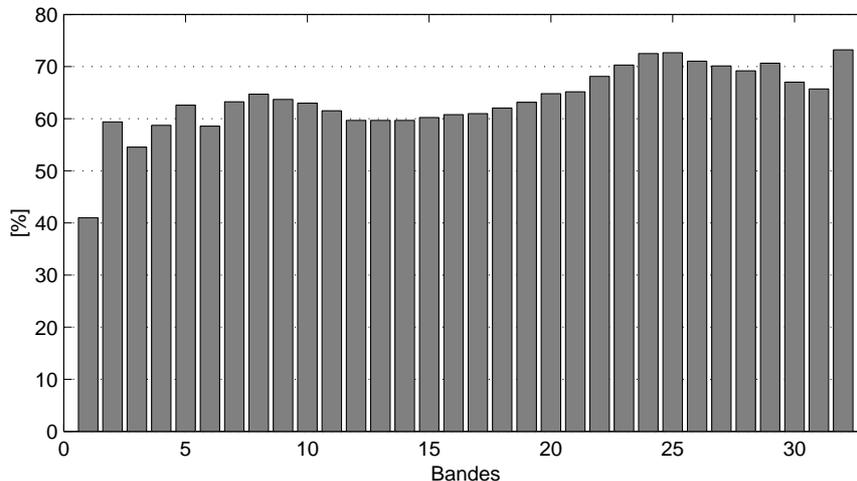


Fig. 5.9 : Probabilités d'utilisation d'une seule bande lors de la sommation de 2 signaux en considérant le phénomène de masquage inter-signaux dans les bandes.

Dans la suite nous explicitons deux méthodes de recombinaison de trames basées sur ce phénomène de masquage en utilisant des critères différents.

5.4.2 Utilisation d'un critère de masquage par coefficient

Dans cette seconde méthode, appelée méthode V2, le test de masquage inter-signaux par bande est basé sur l'énergie des coefficients MDCT. Les différentes étapes de sommation sur les flux codés se déroulent comme suit :

1. Une opération de décodage partiel, y compris la quantification inverse complète, est effectuée sur les différents signaux d'entrée (figure 5.1).
2. Dans la phase de recombinaison, on effectue ce qui suit pour chacune des bandes i , $0 \leq i \leq M - 1$:
 - Recherche parmi les N trames d'entrée de celle qui a l'énergie, $(e_{q,n}(i))^2$, maximale ; n_i étant son numéro :

$$n_i = \arg \max_{n=1..N} \left\{ (e_{q,n}(i))^2 \right\}. \quad (5.6)$$

- Test à partir des coefficients des trames d'entrée dans la bande i pour voir s'ils sont tous masqués par le seuil de masquage $S_{n_i}(i)$ de la trame n_i . Ce seuil étant déterminé lors de la phase du décodage partiel :

- Si $(y_{q,n}(j))^2 < S_{n_i}(i) \quad \forall n \neq n_i$ et $\forall j \in \text{bande}(i)$ alors :

$$y_T(j) = y_{q,n_i}(j) \quad \text{pour tout } j \in \text{bande}(i), \quad (5.7)$$

$$\text{et} \quad e_T(i) = e_{q,n_i}(i). \quad (5.8)$$

Le mot de code de la trame n_i dans la bande i est sauvegardé ainsi que sa version décodée donnant le vecteur des signes, l'index du leader et le rang.

- Sinon, les coefficients de la bande, $y_T(j)$, et le coefficient d'enveloppe spectrale, $e_T(i)$, sont calculés respectivement par les équations (5.1) et (5.4).
3. Le codage partiel se base sur le même schéma de la figure 5.2 sauf pour la quantification. Ainsi, pour une bande i où un seul signal est actif, on effectue les étapes de trans-quantification complémentaire à celle effectuées en quantification inverse lors du décodage partiel. Ceci en tenant compte de l'allocation de bits, $b_{i,T}$, du signal somme et de celle initial de la trame n_i ainsi que des informations du mot de code correspondant à cette dernière. Si la condition permettant d'avoir le même mot de code que celui d'entrée n'est pas vérifiée, l'algorithme décrit dans le paragraphe 5.3.2 est donc repris à partir de l'étape ii.c de la phase 2.

Pour les bandes du signal somme obtenues par trans-quantification un problème dû à la normalisation peut se révéler. En effet, le vecteur quantifié, $Y_{q,n_i}(i)$, a été initialement normalisé par $e_{q,n_i}(i)$. Lors du codage partiel, c'est la version requantifiée, $e_{q,T}(i)$, de cette dernière quantité qui est transmise dans la trame et qui sera utilisée comme facteur d'échelle pour la dénormalisation lors du décodage. La valeur de $e_T(i) = e_{q,n_i}(i)$ et de sa version requantifiée, $e_{q,T}(i)$, peuvent être différentes à cause de l'utilisation du codage différentiel pour l'enveloppe spectrale. En effet, les coefficients d'enveloppe spectrale des bandes voisines issues d'une vraie sommation ont changé. Ce problème ajoute donc une erreur supplémentaire au vecteur des coefficients MDCT de la bande obtenus après décodage final.

5.4.3 Utilisation d'un critère de masquage basé sur l'énergie

Le critère de test de masquage inter-signaux utilisé dans la méthode précédente nécessite un décodage complet des coefficients MDCT. L'algorithme de trans-quantification n'est donc pas exploité de manière optimale pour réduire au maximum la complexité. Pour réaliser une trans-quantification complète sur les bandes où un seul signal intervient, la méthode V3 utilise un critère de masquage basé sur les énergies des bandes. Ainsi, pour chacune des bandes i , le test suivant est effectué après recherche du signal d'entrée, n_i , de forte énergie :

$$\text{Tester si } (e_{q,n}(i))^2 < S_{n_i}(i) \quad \forall n \neq n_i. \quad (5.9)$$

Pour les bandes vérifiant cette hypothèse une opération de trans-quantification sur leur mot de code est effectuée comme elle a été décrite dans le paragraphe 5.3.2.

5.5. Evaluation des performances des différentes méthodes de recombinaison proposées

5.5.1 Gain en complexité et en retard

L'implantation des trois méthodes de recombinaison dans le domaine codé a montré de bonnes performances en terme de réduction de la complexité. Les gains en complexité sont donnés dans le tableau 5.5. Ils représentent des valeurs moyennes obtenues sur la base de données UIT en utilisant un débit binaire de 24 kbit/s. Comme prévu, c'est la méthode V3 qui fournit le meilleur gain en complexité. En terme du délai algorithmique, la réduction est de 20 ms. Elle représente une diminution de 50 % du retard total introduit par cette opération de sommation de trames initialement codées.

Méthode V1	Méthode V2	Méthode V3
35,5 %	44,2 %	51,2 %

Tableau 5.4 : Gains en complexité des différentes méthodes de recombinaison dans le domaine codé par rapport à la méthode de sommation classique.

5.5.2 Tests subjectifs

Des tests subjectifs ont été effectués pour valider la méthode de recombinaison des signaux audio codés par le codeur TDAC de FTR&D au débit binaire de 24 kbit/s. Le but de ces tests est d'une part l'évaluation de la qualité de la méthode de sommation dans le domaine codé par rapport à la méthode classique et d'autre part la comparaison de la qualité obtenue par les deux versions de cette méthode de recombinaison. Trois tests de type triple stimuli A-B-C ont été alors organisés. Ces tests se sont déroulés suivant la même méthodologie décrite dans le paragraphe 4.7.1. Ils utilisent aussi la même échelle de notation. Seize sujets ont participé dans chacun d'eux et 8 types de phrases ont été utilisés.

Les deux premiers tests comparent la qualité des signaux audio obtenus par les méthodes de recombinaison V2 et V3, respectivement, à celle du signal résultant d'une sommation classique dans le domaine temporel. La séquence de référence A correspond donc à la somme de deux signaux originaux sans aucun codage. Les séquences B et C sont, dans un ordre aléatoire, soit la phrase obtenue par la méthode classique soit celle obtenue par la méthode de recombinaison. Notons que la méthode V1 étant utilisée comme une étape intermédiaire avant d'arriver à une sommation avancée sur les flux codés TDAC, elle n'a pas fait l'objet d'évaluation.

Les résultats de ces deux tests sont résumés dans le tableau 5.5. Dans ce dernier nous distinguons deux notes pour la méthode de sommation classique dans le domaine temporel puisque l'auditeur fait une comparaison implicite lors de l'écoute des séquences B et C. La note donnée pour chacune de ces deux séquences dépend donc de l'autre. Ces résultats montrent que la méthode V2 est quasiment transparente par rapport à la méthode classique. Par contre pour la méthode V3 il y a une diminution de 0,77 sur le MOS.

Pour confirmer ces résultats nous avons effectué un troisième test qui compare la qualité des signaux audio issus des méthodes V2 et V3. Dans ce cas, la séquence de référence A est la phrase obtenue par la méthode de sommation classique. Les séquences B et C, sont dans un ordre aléatoire,

soit la phrase obtenue par la méthode de recombinaison V2 ou V3. Les résultats illustrés dans le tableau 5.6 montrent bien que la méthode V2 est de meilleure qualité par rapport à V3.

Pour expliquer cette dernière conclusion, il faut revenir au critère de test de masquage utilisé dans la méthode V3. Ce dernier est basé sur la comparaison de la moyenne des énergies des coefficients de la bande au seuil de masquage du signal de forte énergie. Certaines bandes peuvent être donc considérées comme masquées même s'il y a quelques coefficients qui sont au-dessus du seuil de masquage. Ce "vol" de certaines parties du spectre du signal s'entend sur certaines séquences.

	Méthode de sommation			
	Domaine temporel	Méthode V2	Domaine temporel	Méthode V3
Moyenne	4,24	4,15	4,22	3,65
Ecart-type	0,74	0,80	0,80	0,87
Intervalle de confiance à 95 %	0,13	0,14	0,14	0,15

Tableau 5.5 : Résultats des tests pour évaluation des deux méthodes de recombinaison dans le domaine codé par rapport à la méthode de sommation classique.

	Méthode V2	Méthode V3
Moyenne	4,48	4,10
Ecart-type	0,69	0,72
Intervalle de confiance à 95 %	0,12	0,13

Tableau 5.6 : Résultats du test pour la comparaison de la qualité des deux méthodes de recombinaison.

5.6. Conclusion

Dans ce chapitre, nous avons étudié les possibilités de recombinaison des flux audio codés TDAC. La méthode élaborée exploite pleinement les spécificités de ce codeur pour manipuler le flux codé avec un minimum de décodage et ainsi réduire la complexité algorithmique au maximum. Cette méthode utilise aussi le phénomène de masquage sous une nouvelle forme pour l'élimination des composantes inutiles du signal. Elle profite de la nouvelle donnée sur l'interaction entre les signaux et qui est due à leur sommation. La méthode de recombinaison V2 retenue présente de bonne performance en terme de réduction de la complexité et du délai tout en assurant une qualité transparente par rapport à la méthode de sommation temporelle.

Chapitre 6 Architectures et applications pour le traitement audio dans le domaine compressé

6.1. Introduction

Dans les chapitres précédents, nous avons développé différentes solutions pour le traitement du signal audio dans le domaine compressé. Ces solutions qui concernent essentiellement les codeurs MPEG-1 Layer I et II et le codeur TDAC de FTR&D, portent sur la sommation, la conversion du débit et le filtrage pour la spatialisation sonore de flux partiellement décodés. Dans ce chapitre nous nous intéressons à quelques applications pratiques de telles méthodes. Ces applications mettront en valeur l'intérêt du traitement du signal audio dans le domaine compressé pour la transmission sur le réseau.

Dans le paragraphe 1.4, nous avons déjà évoqué quelques éléments sur les problèmes et contraintes imposés par les réseaux de communication multimédia et en particulier pour la transmission audio temps réel. Nous avons aussi montré l'apport du traitement du signal audio pour faire face à certains de ces problèmes. Le présent chapitre, dont le but va dans le même sens, se focalise essentiellement sur le contexte de communication multipoint. L'application principale exposée concerne le traitement dans un pont audio de téléconférence utilisant le codeur TDAC de FTR&D. Dans cet exemple, l'interaction entre architecture réseau et méthodes de traitement du signal audio dans le domaine compressé sera alors illustrée.

Ce chapitre débute par un bref descriptif de l'environnement actuel de communication audiovisuelle et particulièrement celui décrit par la norme H.323 pour les réseaux à commutation de paquets. Une méthode de récupération des trames effacées dans le domaine fréquentiel est ensuite exposée. Cette méthode permet de faire face au phénomène des pertes de paquets dont souffre ce type de réseaux. Ensuite la description complète du pont audio de téléconférence à traitement dans le domaine compressé, qui a été développé, sera donnée. Finalement, d'autres exemples de configurations de transmission audio mettant en œuvre le traitement dans le domaine compressé seront présentés.

6.2. Communication multimédia de groupe et téléconférence

6.2.1 Normes et environnement

Plusieurs normes dans la série UIT-T H.3xx relatives aux systèmes de communication audiovisuelle et multimédia ont été mises au point pour répondre aux différentes propriétés et contraintes imposées par la transmission sur chaque type de réseau [Schaphorst 99]. Ces différentes normes donnent les spécifications sur les codages audio et vidéo, le multiplexage des différents flux

(audio, vidéo et données) au niveau des terminaux ainsi que sur le contrôle système et la signalisation. Cette série a commencé avec la norme H.320 spécifique à la communication sur le réseau RNIS bande étroite (N-ISDN) [H.320]. Elle inclut aussi les recommandations suivantes [Okubo 97] :

- H.321 pour l'utilisation des terminaux H.320 dans l'environnement du RNIS large bande (ATM). Pour tirer profit pleinement de la disponibilité de la ressource bande passante dans ce type de réseaux, la recommandation H.310 a été développée spécifiquement. Elle offre par exemple la possibilité d'utiliser des codeurs audio haute qualité type MPEG-1.
- H.322 pour l'utilisation des terminaux H.320 sur des réseaux locaux (LAN) à qualité de service garantie type Iso-Ethernet.
- H.324 pour opérer à faible débit sur le réseau téléphonique commuté et aussi radiomobile (GSTN).

Ces recommandations sont propres à des communications utilisant des mécanismes de transport à qualité de service garantie. Dans ces systèmes le mécanisme de transport fournit une bande passante, un taux d'erreur, un délai et une gigue stables et bien connus au préalable. La norme H.323 [H.323, Thom 96] a été élaborée pour des systèmes de communication multimédia sur des réseaux à commutation de paquets n'offrant pas nécessairement de telles garanties. Certains volets de la norme visent à tenir compte des conditions de cet environnement pour limiter leurs effets négatifs sur la qualité de la communication. Dans sa philosophie de conception, elle a pour objectif d'offrir une interopérabilité entre différents types de terminaux définis dans les recommandations précédentes, en particulier H.320, ainsi qu'une certaine scalabilité : des réseaux à faible débit aux réseaux large bandes (LAN à QoS non garantie, N-ISDN, ATM, GSTN...). Elle offre aussi de nouvelles configurations de conférence multipoint.

Pour fournir ces fonctionnalités, la norme H.323 décrit d'autres entités constituant le système de communication que les terminaux. Il s'agit des passerelles (GW : Gateway), des portiers (GK : Gatekeeper), des contrôleurs multipoints (MC : Multipoint Controller), des processeurs multipoints (MP : Multipoint Processor) en plus des ponts de conférence ou unités de commande multipoint (MCU : Multipoint Control Unit). L'utilisation de passerelles permet l'interopérabilité entre différents types de terminaux. Elles assurent entre autres le transcodage audio (et aussi vidéo) dans le cas de terminaux n'utilisant pas les mêmes codeurs. Les portiers fournissent des fonctions d'administration du réseau telle que le contrôle des admissions dans une zone, la conversion des adresses, la régulation et le contrôle du trafic. Pour cette dernière fonctionnalité, le portier limite et avise les modifications de la largeur de bande passante utilisée par chaque appel à l'aide de mécanismes de signalisation. Elle vise à contrôler l'impact des communications H.323 sur la disponibilité de la bande passante dans une zone et donc à maintenir une certaine qualité de service.

En plus du rôle des portiers, la norme H.323 fournit d'autres outils pour le contrôle et l'amélioration de la qualité de service. Ceci se fait à l'aide de l'utilisation de protocoles applicatifs type RTCP et RSVP. Ainsi par exemple, le protocole RTCP utilise un mécanisme permettant d'avoir au niveau de l'émetteur des informations sur les paramètres QoS ce qui permet d'adapter l'émission (réduction de débit ou autre).

H.323 est une norme ouverte au niveau de l'implantation de ces différentes entités (GW, GK, MCU, MP et MC). Cette ouverture laisse des possibilités d'amélioration pour assurer encore mieux les objectifs caractérisant sa philosophie de conception. Elle donne ainsi l'opportunité de doter ces

composantes d'autres fonctionnalités ou de les implanter de manière plus efficace. L'utilisation de méthodes du traitement du signal avancé dans les ponts de conférence va donc dans ce sens.

6.2.2 Conférence multipoint

Lorsqu'une communication doit s'établir entre plus de deux terminaux on parle de conférence multipoint. Dans ce cas plusieurs types de configurations sont envisagés selon la topologie réseau utilisée pour connecter les différents terminaux. Ainsi on distingue les conférences multipoint centralisée et décentralisée (figure 6.1). La première configuration se base sur une topologie de réseau en étoile. Elle utilise un pont de conférence (ou MCU) pour gérer les communications.

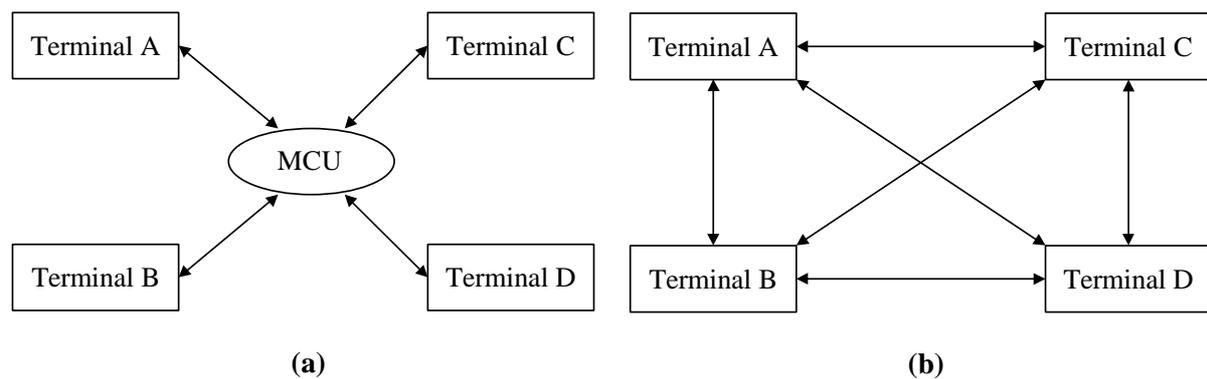


Fig. 6.1 : Configurations de conférence multipoint (a) centralisée avec un réseau en étoile (b) décentralisée avec un réseau de type maillé.

Un pont de conférence comporte entre autres des processeurs audio et vidéo, définis dans la norme H.231 [H.231], et dont le rôle consiste à combiner les flux audiovisuels entrant pour construire ceux de sortie. Pour l'audio, cela revient à réaliser une opération de mixage ou de commutation. Pour la vidéo, il assure l'aiguillage des signaux vidéo en mode affichage commuté et le multiplexage spatial des images en une seule (format écran partagé) en mode présence continue.

Dans la norme H.320, seule la configuration centralisée est possible pour toute communication multipoint. Le pont de conférence est donc un organe nécessaire dans le réseau et séparé des terminaux. La nouvelle norme H.323 subdivise le MCU en deux parties : le contrôleur multipoint (MC) et le processeur multipoint (MP). Elle permet en plus des deux configurations de conférence précédentes l'utilisation de solutions hybrides. Le processeur multipoint assure spécifiquement dans ce cas les fonctionnalités de traitement des flux audiovisuels décrites ci-dessus. Il n'est pas nécessairement situé avec le contrôleur multipoint sur une même machine dédiée. Il peut être localisé par exemple sur un des terminaux.

Dans une configuration de conférence multipoint décentralisée à N terminaux, chaque terminal est connecté aux $N - 1$ autres terminaux. Il doit posséder un MP capable de décoder les $N - 1$ flux audio reçus avant de sortir un flux mixé sur un même haut-parleur. Cette configuration a donc l'inconvénient de nécessiter une capacité de calcul importante au niveau des terminaux et d'utiliser aussi un grand nombre de connexions bidirectionnelles (tableau 6.1) qui augmente rapidement avec le nombre de participants. Malgré l'utilisation du mécanisme multicast pour la transmission sur IP, réduisant le nombre de connexions, la capacité de décodage des terminaux reste toujours un problème. En plus, en multicast on peut être limité par le réseau d'accès.

La configuration de conférence multipoint centralisée pallie cet inconvénient en effectuant les traitements dans un seul processeur multipoint qui joue aussi le rôle de concentrateur de flux. Par contre, cette architecture souffre de certains défauts. Ainsi, la mise en cascade de deux opérations de codage/décodage dans la chaîne de communication entre terminaux induit un retard important surtout lors de l'utilisation d'algorithmes de compression fréquentiels caractérisés par un grand délai algorithmique. La complexité peut toujours être un paramètre pertinent dans cette configuration du fait du nombre d'accès élevé que peut offrir un pont audio. Il faut noter aussi que dans certaines architectures le MP peut être localisé sur un des terminaux jouant aussi le rôle du pont.

Configuration	Nombre de décodeurs par terminal	Nombre de connexions	Nombre d'opérations de codage/décodage
Décentralisée	$N - 1$	$N(N - 1)/2$	1
Centralisée	1	N	2

Tableau 6.1 : Comparaison des configurations centralisées et décentralisées pour un système de téléconférence à N terminaux.

6.3. Algorithme de correction des trames TDAC effacées dans le domaine fréquentiel

La transmission de l'audio temps réel sur les réseaux à commutation de paquets (cas des réseaux IP) est confrontée au problème de pertes de paquets. Lors de la restitution, ce problème donne lieu à un signal sonore haché et donc une dégradation perceptible de la qualité. Les craquements dans le signal reconstruit rendent la compréhension de la parole difficile et l'écoute de la musique moins naturelle. Même à de faibles taux tel que 2 % ces craquements sont perceptibles, à 15% l'intelligibilité est toujours possible mais suscite une grande concentration, et à partir de 20-25% la compréhension devient impossible.

Dans les applications mettant en œuvre un traitement dans le domaine compressé, comme c'est le cas du pont audio utilisant le codeur TDAC, seule l'information sur le signal dans ce domaine est accessible. L'application de la méthode de recouvrement de trames effacées de ce codeur, opérant dans le domaine temporel (voir paragraphe 2.8.7), s'avère donc impossible. Nous développons donc une nouvelle méthode⁸ qui agit directement sur le spectre MDCT. Comme elle est décrite dans ce paragraphe, elle ne concerne que le cas où la correction s'effectue sur un signal individuel. Son extension au cas où plusieurs signaux sont mis en œuvre, pour tirer profit de leur interaction, sera donnée dans le paragraphe suivant.

6.3.1 Description générale

L'algorithme utilisé pour la récupération de pertes de paquets est basé sur une méthode de régénération de la trame TDAC. L'ensemble des coefficients MDCT, constituant l'information

⁸ Méthode développée en collaboration avec Cédric Boisseau pendant son stage à FTR&D [Boisseau 99].

principale de chacune des trames codées, est obtenu par une méthode dite d'extrapolation sinusoidale à partir de ceux correspondant aux deux trames précédentes mises en mémoire. Une mise à jour de cette mémoire à deux trames est effectuée régulièrement en utilisant soit celle correctement reçue soit celle qui vient d'être régénérée.

L'extrapolation sinusoidale n'est utilisée que si la dernière trame en mémoire est de nature voisée ou tonale et l'enveloppe spectrale est calculée directement une fois le spectre MDCT déterminé. Dans le cas contraire, la trame effacée est obtenue par simple répétition de cette dernière. Dans les deux cas de figure la nouvelle trame aura la même nature, en terme de voisement et tonalité, que cette dernière trame en mémoire.

6.3.2 Méthode d'extrapolation sinusoidale

6.3.2.1 Principe de base

Rappelons d'abord l'expression des coefficients, $y(m, k)$, de la trame m obtenus par application de la transformée MDCT au bloc du signal, $x(m, l)$, $0 \leq l \leq 2M - 1$, de taille $2M$:

$$y(m, k) = \sum_{l=0}^{2M-1} h(l)x(m, l) \cos \left[\frac{\pi}{M} \left(k + \frac{1}{2} \right) \left(l + \frac{M+1}{2} \right) \right], \quad 0 \leq k \leq M - 1, \quad (6.1)$$

où $h(l)$, $0 \leq l \leq 2M - 1$, est la fenêtre de pondération sinusoidale. Cette transformation peut être exprimée aussi comme suit :

$$y(m, k) = \operatorname{Re} \left\{ \sum_{l=0}^{2M-1} h(l)x(m, l) e^{-j \frac{\pi}{M} (k + \frac{1}{2}) (l + \frac{M+1}{2})} \right\}, \quad 0 \leq k \leq M - 1. \quad (6.2)$$

où $\operatorname{Re}\{C\}$ désigne la partie réelle du nombre complexe C . Puis :

$$y(m, k) = \operatorname{Re} \left\{ e^{-j \frac{\pi}{2M} (k + \frac{1}{2}) (M+1)} \sum_{l=0}^{2M-1} h(l)x(m, l) e^{-j \frac{\pi}{M} (k + \frac{1}{2}) l} \right\}, \quad 0 \leq k \leq M - 1, \quad (6.3)$$

Considérons la transformée de Fourier discrète du signal pondéré, $h(l)x(m, l)$, évaluée en $(k + \frac{1}{2})$, et donnée par :

$$Y(m, k) = \sum_{l=0}^{2M-1} h(l)x(m, l) e^{-j \frac{2\pi (k + \frac{1}{2}) l}{2M}}, \quad 0 \leq k \leq 2M - 1. \quad (6.4)$$

Ces coefficients de DFT peuvent être aussi exprimés en coordonnées polaires sous la forme suivante :

$$Y(m, k) = |Y(m, k)| e^{j\Phi(m, k)}, \quad 0 \leq k \leq 2M - 1. \quad (6.5)$$

A partir des relations (6.3) à (6.5), on peut en déduire l'expression suivante des coefficients MDCT :

$$y(m, k) = |Y(m, k)| \cos \left[\Phi(Y(m, k)) - \frac{\pi}{2M} \left(k + \frac{1}{2} \right) (M+1) \right], \quad 0 \leq k \leq M - 1. \quad (6.6)$$

Considérons le cas particulier d'un signal $x(n)$ sinusoidal de fréquence f_0 telle que :

$$\frac{(k_0 - 0,5)}{2M} \leq \frac{f_0}{f_e} < \frac{(k_0 + 0,5)}{2M}, \quad (6.7)$$

avec f_e la fréquence d'échantillonnage et k_0 un entier compris entre 0 et $M - 1$. Dans ce cas l'énergie du signal est concentrée autour du coefficient $Y(m, k_0)$. Le module $|Y(m, k_0)|$ est constant pour toute trame m et la phase $\Phi(Y(m, k))$ est une fonction linéaire du numéro de bloc m [Mahieux 89]. D'après l'équation (6.6), le coefficient $y(m, k_0)$ évolue donc comme une sinusoïde en fonction de l'indice m . L'algorithme d'extrapolation sinusoïdale repose sur cette propriété.

6.3.2.2 Extrapolation sinusoïdale

La propriété qui vient d'être établie sur la forme sinusoïdale des coefficients de transformée MDCT d'un signal sinusoïdal permet d'utiliser une formule de prédiction sur ces coefficients. En effet, toute suite de nombres sinusoïdale $u_n = A \sin(2\pi fn)$, où A et f sont des constantes réelles, vérifie la relation de récurrence suivante :

$$u_{n+2} = 2 \cos(2\pi f) u_{n+1} - u_n, \quad n \in \mathbb{N}. \quad (6.8)$$

Le coefficient numéro k , $0 \leq k \leq M - 1$, de la trame à prédire s'obtient alors à partir des coefficients de même numéro des deux trames précédentes selon l'expression suivante :

$$y_p(m, k) = a(m, k) y(m-1, k) - y(m-2, k), \quad (6.9)$$

où $a(m, k) = 2 \cos[2\pi f(m, k)]$ est le coefficient d'extrapolation, $f(m, k)$ étant la fréquence instantanée de la sinusoïde représentant l'évolution du coefficient $y(m, k)$ en fonction de l'indice de trame m . Cette fréquence peut varier d'une trame à l'autre, l'adaptation de la formule (6.9) revient donc à redéterminer ce dernier paramètre. Comme il sera détaillé ci-après, cette adaptation dépend du coefficient de la trame courante. En cas de perte de paquet, c'est cette trame qu'on cherche à déterminer. Le coefficient $a(m, k)$ n'est donc pas accessible. Il sera remplacé par $a(m-1, k)$ avec l'hypothèse que la fréquence instantanée n'a pas trop changé. La formule de prédiction (6.9) est alors remplacée par :

$$y_p(m, k) = a(m-1, k) y(m-1, k) - y(m-2, k). \quad (6.10)$$

L'extrapolation sinusoïdale n'a un sens que pour les signaux de parole et de musique de nature voisée ou tonale puisqu'ils contiennent des composantes fréquentielles pures. Mais même dans ce cas l'hypothèse d'une évolution sinusoïdale des coefficients MDCT n'est pas tout à fait vraie. En effet, dans le spectre MDCT il y a un effet de modulation et d'étalement (contribution de toutes les composantes à chacun des coefficients MDCT). Ce phénomène d'étalement n'est pas très gênant si la résolution fréquentielle est suffisamment grande (25 Hz dans le cas du codeur TDAC). Pour appliquer cette méthode d'extrapolation sinusoïdale, il est possible de se restreindre aux zones entourant ces composantes. On peut utiliser alors une méthode de détection de pics pour les localiser (par exemple celle du modèle psychoacoustique 1 du codeur MPEG-1) mais celle-ci prend beaucoup de complexité. On choisit plutôt de limiter la bande de régénération des coefficients à [0-4kHz], au lieu de [0-7kHz].

6.3.2.3 Adaptation du coefficient d'extrapolation

La mise à jour des fréquences $f(m,k)$, $0 \leq k \leq M-1$, se fait de manière continue lors du traitement de chaque nouvelle trame. Ainsi, si la trame correspond à un paquet reçu, ces fréquences sont déterminées directement. Si la trame correspond à un paquet perdu, cette détermination est faite après obtention des coefficients de transformée récupérés. Dans les deux cas, le calcul repose sur deux méthodes heuristiques qui fonctionnent en parallèle :

- La première se base sur le changement de signe du coefficient $y(m,k)$ évoluant en fonction de m : une demi-période est détectée après deux changements de signes successifs.
- La seconde méthode utilise la détection de changement de pente du même signal pour le calcul des demi-périodes.

La mise à jour de la fréquence est faite dès que l'une de ces deux méthodes a obtenu une nouvelle valeur.

6.4. Pont audio à traitement dans le domaine compressé

Dans ce paragraphe nous présentons un pont audio ou plus précisément un processeur multipoint audio à traitement dans le domaine compressé et à fonctionnalités enrichies [Benjelloun 99, Benjelloun 01]. Ce pont audio se base sur le codeur TDAC de FTR&D. Il a été conçu dans le but de fonctionner dans l'environnement du système de téléconférence *PVHD* de FTR&D. Ce système se base sur les spécifications de la norme H.323. Le processeur multipoint audio développé a été donc intégré⁹ dans une architecture comportant un MCU H.323 [Lelong 99].

Rappelons que des ponts audio à traitement dans le domaine compressé ont déjà été développés par le passé pour d'autres codeurs. C'est le cas par exemple des systèmes utilisant le codeur large bande UIT-T G.722. En effet, ce type de codage procède par subdivision du spectre en deux sous-bandes puis un codage ADPCM est effectué sur le signal de chaque bande. Pour résoudre le problème du retard et de la complexité introduits par les bancs de filtres miroirs en quadrature, au niveau du pont audio, la norme spécifie dans son appendice I une méthode de recombinaison directe sur les signaux sous-bandes. Elle consiste simplement à réaliser un décodage ADPCM des deux échantillons sous-bandes des trames d'entrée, à faire la sommation sur l'ensemble des voies d'entrée considérées, et enfin à effectuer un codage ADPCM avant de construire la trame de sortie [G.722, Taka 88].

6.4.1 Fonctionnalités et avantages

Le contexte considéré est celui d'une téléconférence multipoint à N terminaux ($N \geq 2$) connectés à un pont audio. Ce dernier reçoit N flux codés d'entrée à partir desquels il construit les flux de sortie et les redirige vers les terminaux correspondants. La transmission se fait par paquets. Chaque paquet transporte deux trames concaténées en plus de l'entête dû à la pile de protocoles RTP/UDP/IP. Il correspond donc à des durées de 40 ms du signal audio. Les fonctionnalités assurées par ce pont sont les suivantes :

⁹ Cette intégration a été réalisée par Franck Lelong du laboratoire DSE/SGV de FTR&D.

- Somme, dans le domaine compressé, des $N - 1$ flux audio codés TDAC correspondant à chacune des voies de sortie sur les terminaux. L'algorithme utilisé correspond à la méthode de somme V2 exposée dans le chapitre 5. Notons que le débit spécifié sur une voie de sortie peut être différent de ceux des voies d'entrée.
- Prise en compte et gestion des discontinuités des flux audio reçus dus, d'une part, au phénomène des pertes de paquets sur le réseau, et d'autre part, à l'émission discontinue dans le cas de l'utilisation d'un mécanisme de VAD/DTX dans les codeurs des terminaux.
- Correction des trames effacées dues aux pertes des paquets audio en utilisant la méthode décrite dans le paragraphe 6.3.

Le pont donne aussi la possibilité de faire des connexions/déconnexions dynamiques en cours de communication. Une conférence peut donc commencer en mode point à point entre deux terminaux et évoluer vers le mode multipoint au fur et à mesure que d'autres terminaux la rejoignent. Le schéma fonctionnel complet du pont audio sera donné plus loin dans la figure 6.4.

En utilisant le traitement dans le domaine compressé, ce pont audio fournit l'avantage de réduire le délai global de la communication entre terminaux et donc d'améliorer sa qualité en termes d'interactivité. Il a aussi la caractéristique d'avoir une complexité réduite. Cette complexité est d'autant plus diminuée par l'utilisation du mécanisme de gestion profitant des trames inactives (perdus ou de silence).

La mise en œuvre d'un traitement de recouvrement des trames effacées dans le pont audio cherche à atténuer la dégradation de qualité due aux pertes de paquets. Elle sert comme mécanisme de renforcement pour les corrections effectuées dans les terminaux. En effet, dans cette architecture de communication multipoint centralisée, la correction des pertes paquets dans le terminal est insuffisante. Elle ne concerne que les paquets perdus sur le segment multipoint-terminal relatifs au flux audio mixé dans le pont et descendant vers le terminal récepteur. Certains paquets perdus sur le segment terminal-multipoint et qui concernent les flux audio individuels dirigés vers le pont audio ne sont pas récupérés (voir l'exemple illustratif de la figure 6.2). Malgré la correction dans le terminal, des artefacts peuvent toujours s'entendre surtout si le niveau du signal individuel qui a subi une perte est assez élevé par rapport aux autres signaux dans le flux mixé. D'autre part, le traitement de correction dans le pont permet aussi de diminuer la probabilité au niveau d'un terminal de voir un paquet perdu. Elle correspond alors à la seule probabilité de perte sur le segment multipoint-terminal récepteur. Une telle diminution permet à la méthode de correction du codeur TDAC dans le terminal de mieux fonctionner. En effet, cette dernière est basée sur la restitution du signal perdu par une prédiction linéaire utilisant un filtre LPC dont les coefficients sont fournis à partir d'une analyse sur la trame précédente (voir paragraphe 2.8.7). Toute augmentation du taux de trames effacées dégrade l'efficacité de cette méthode.

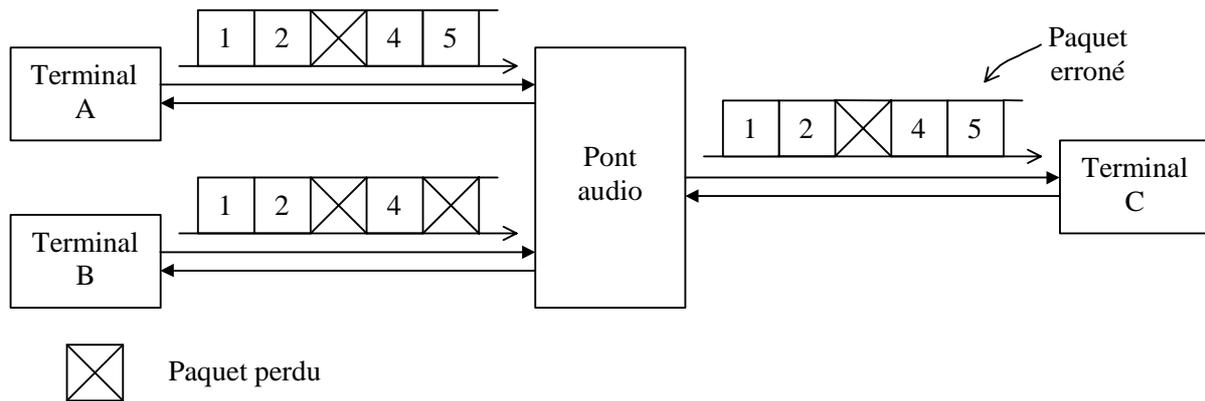


Fig. 6.2 : Illustration de l'utilité de mise en œuvre d'une correction de perte de paquets dans le pont audio.

6.4.2 Gestion des flux discontinus

La détection d'activité vocale (DAV) est une fonction optionnelle du codeur TDAC. Elle permet de réduire le débit binaire, et donc la surcharge du réseau, en éliminant l'émission des périodes de silence du signal. Quand cette fonction est activée, le pont doit en tenir compte. Un mécanisme de gestion est alors utilisé. Il repose sur le principe de ne réaliser une opération de décodage partiel sur une trame d'une voie d'entrée que si cela est nécessaire. Seules les voies actives sont prises en compte dans l'opération de sommation.

Cette stratégie de réduction du nombre de décodeurs/codeurs pour diminuer la complexité a été déjà introduite dans certains ponts audio indépendamment du traitement dans le domaine compressé. Elle se base sur un certain critère pour n'utiliser qu'une partie des signaux reçus dans le mixage. Ainsi, dans [Aoyagi 96] le mixage est toujours fait sur les N' signaux ayant les gains les plus forts, où N' est prédéfini et fixe, le gain étant lu directement à partir des mots de code d'entrée. Dans [Nahumi 95], la sommation n'est réalisée que sur les voies actives, le nombre d'entrées prises en compte est donc variable. Quant au pont audio de PictureTel, il réalise toujours la recombinaison des deux voies les plus puissantes quelque soit le nombre d'entrées [PictureTel].

L'information sur la coupure d'émission est fournie par l'activation d'un bit de la dernière trame de la période de hangover (trame SID). Cette information pour les différentes voies d'entrée au niveau du pont permet de gérer le processus de sommation/commutation au niveau du pont. Les différents cas de figure possibles sont les suivants :

- *Une voie active parmi N* : les trames de cette voie sont commutées vers les autres sorties. Aucune opération de décodage/codage (partiels) n'est effectuée sauf si une conversion de débit est nécessaire.
- *Deux voies actives parmi N* : une commutation croisée des trames de ces deux voies vers leurs sorties correspondantes est effectuée. Les autres sorties reçoivent leur somme.
- Dans tous les autres cas restants, la sommation est réalisée sur l'ensemble des voies actives correspondants à chacune des sorties.

La mise en œuvre de ce mécanisme de gestion apporte une réduction importante de la complexité par rapport à un pont classique à sommation dans le domaine temporel et aussi par rapport à celui à

sommation dans le domaine compressé (voir figure 6.3). Ce résultat est dû au fait que dans la majorité du temps, un seul participant à la conférence parle. Le pont audio effectue donc surtout de la commutation.

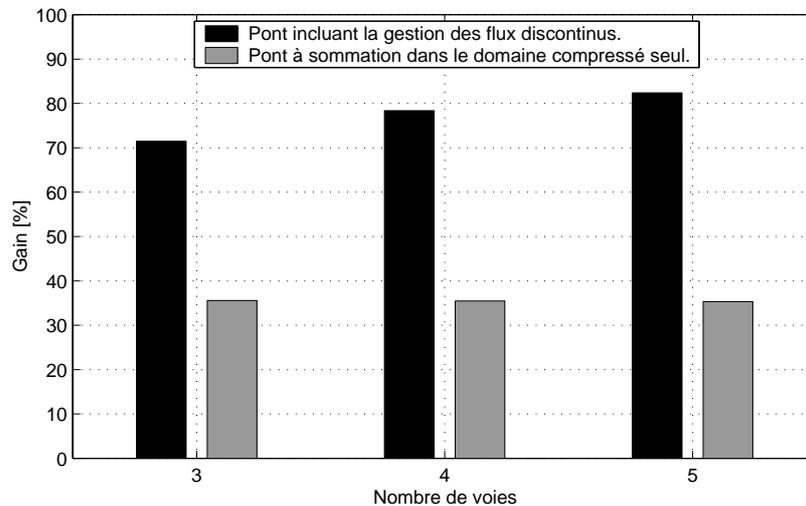


Fig. 6.3 : Comparaison des gains en complexité du pont à sommation dans le domaine compressé par rapport à celui réalisant la sommation en temporel avec et sans utilisation du mécanisme de gestion des flux discontinus.

Si on choisit de ne pas activer le dispositif de correction d'erreurs dans le pont, les trames des paquets perdus seront assimilées à des trames inactives. La même stratégie de gestion précédente est alors utilisée. Dans le cas contraire, la distinction entre ces deux types de flux discontinus est assurée par le dernier paquet de la période de parole contenant la trame SID. Pour le cas particulier où ce paquet est perdu, les trames inactives seront considérées comme effacées. Ce cas ne pose pas de problème car la méthode de correction d'erreurs utilise en plus un critère de masquage inter-signaux des différentes voies pour qu'elle soit déclenchée. La dernière trame en mémoire de la voie inactive étant faible, le critère a une faible probabilité d'être satisfait si les autres voies sont actives.

6.4.3 Correction des pertes de paquets dans le pont

6.4.3.1 Principe

La méthode d'extrapolation sinusoïdale développée dans le paragraphe 6.3 pour réaliser un recouvrement des trames effacées dans le domaine de la transformée MDCT peut s'appliquer directement à chacun des signaux audio entrant dans le pont subissant une perte. Par ailleurs, vu que ces différents signaux seront soumis à une opération ultérieure de sommation, il est intéressant de tenir compte de leur interaction. En effet, la dégradation de la qualité due à une perte de paquet sur un des signaux peut ne pas être aussi intense que dans le cas d'un signal individuel. Le phénomène de masquage engendré par un niveau élevé des signaux bien reçus peut masquer les craquements dus aux coupures sur les autres. Il est alors possible de profiter de ce phénomène pour limiter l'utilisation du procédé de correction d'erreurs aux seuls cas où l'absence d'un segment d'un signal, dû à une coupure, sera vraiment perceptible. Le déclenchement de la correction d'erreur sera donc conditionné par un

critère testé au préalable. En adoptant cette stratégie la méthode de correction sera plus efficace en termes de limitation de la complexité et du risque de mauvaise correction.

L'utilisation de la méthode de récupération de trames par la méthode du paragraphe 6.3 nécessite une mise en mémoire des trames antérieures pour chacun des signaux d'entrée. En cas de perte d'un paquet, la seule possibilité est que le critère de décision soit appliqué sur les trames en mémoire.

6.4.3.2 Critère de déclenchement de correction d'erreurs

Le décodage partiel des flux audio TDAC reçus fournit les coefficients du spectre MDCT ainsi que les seuils de masquage par bandes calculés pour la détermination de l'allocation des bits. Le critère de masquage peut être donc spécifié par bande. Ainsi la correction d'erreur va s'appliquer à un ensemble restreint de bandes de la trame du signal individuel qui ne sont pas masquées par les bandes leur correspondant des autres signaux.

Considérons l'exemple de la sommation des $N-1$ signaux des entrées, E_n , $1 \leq n \leq N-1$, pour fournir le signal sur la sortie S_N . Si tous les paquets correspondants à ces différents signaux sont perdus alors la correction est faite directement sur l'ensemble des bandes des trames constituant chacun d'eux. Sinon, nous utilisons un critère de masquage similaire à celui décrit dans le paragraphe 5.4.2. Ainsi, la décision sur le déclenchement du mécanisme de correction d'erreur pour la bande i des trames constituant le paquet perdu courant du signal n_0 est réalisé suivant la procédure suivante :

- Recherche, parmi les trames en mémoire des signaux $n \neq n_0$ et pour lesquels le paquet courant est bien reçu, celle qui a l'énergie $(e_{q,n}^M(i))^2$ dans la bande i maximale. L'indice du signal correspondant à cette trame est noté n_i .
- Test si les coefficients de la trame en mémoire du signal n_0 dans la bande i sont tous masqués par le seuil de masquage $S_{n_i}^M(i)$ de la trame n_i dans la même bande :

$$(y_{q,n_0}^M(j))^2 < S_{n_i}^M(i), \forall j \in \text{bande}(i). \quad (6.11)$$

Dans les notations ci-dessus, l'exposant M indique qu'une telle quantité correspond à la trame précédente mise en mémoire.

Si la condition (6.11) n'est pas réalisée, alors la correction d'erreur est déclenchée pour cette bande suivant la méthode décrite dans le paragraphe 6.3.

6.4.3.3 Mécanisme global de récupération des paquets perdus dans le pont audio

Le schéma fonctionnel du pont audio à traitement dans le domaine compressé et intégrant la correction des pertes de paquets est donné dans la figure 6.4. Ce dispositif intègre aussi un procédé de factorisation du traitement de récupération de trames puisque chaque signal d'entrée contribue aux signaux mixés sur les $(N-1)$ autres voies de sortie. Pour illustration, considérons le cas où un paquet est perdu sur une entrée E_{n_0} contribuant aux signaux des voies de sorties, $\{S_n\}_{n \neq n_0}$. Le critère de déclenchement de correction d'erreur est d'abord testé pour l'ensemble des signaux des entrées $\{E_n\}_{n \neq n_0, n \neq n_1}$ utilisés pour la détermination de la première sortie n_1 faisant intervenir E_{n_0} et où les paquets sont bien reçus. Une correction d'erreur est réalisée pour toutes les bandes des trames du paquet de cette voie d'entrée si la décision est positive. Dès que l'ensemble des paramètres

correspondant aux bandes des trames de ce paquet sur cette voie est déterminé une fois, il sera utilisé dans la détermination de toutes les autres sorties faisant intervenir E_{n_0} et ne sera plus recalculé.

Dans le cas où les terminaux utilisent la transmission discontinue, les signaux des voies d'entrée pris en compte dans le mécanisme de déclenchement de correction d'erreur seront restreints aux seules voies où la trame est active. Dans ce cas aussi, la correction d'erreur ne sera pas déclenchée si seule la voie, où le paquet est perdu, est active parmi toutes les voies d'entrée contribuant à une sortie donnée. Dans ce cas la correction est laissée au terminal de réception.

Une fois les paramètres de la trame effacée déterminés, ils sont soumis au processus de recombinaison. Ils sont utilisés de la même manière que pour une trame d'un paquet correctement reçu.

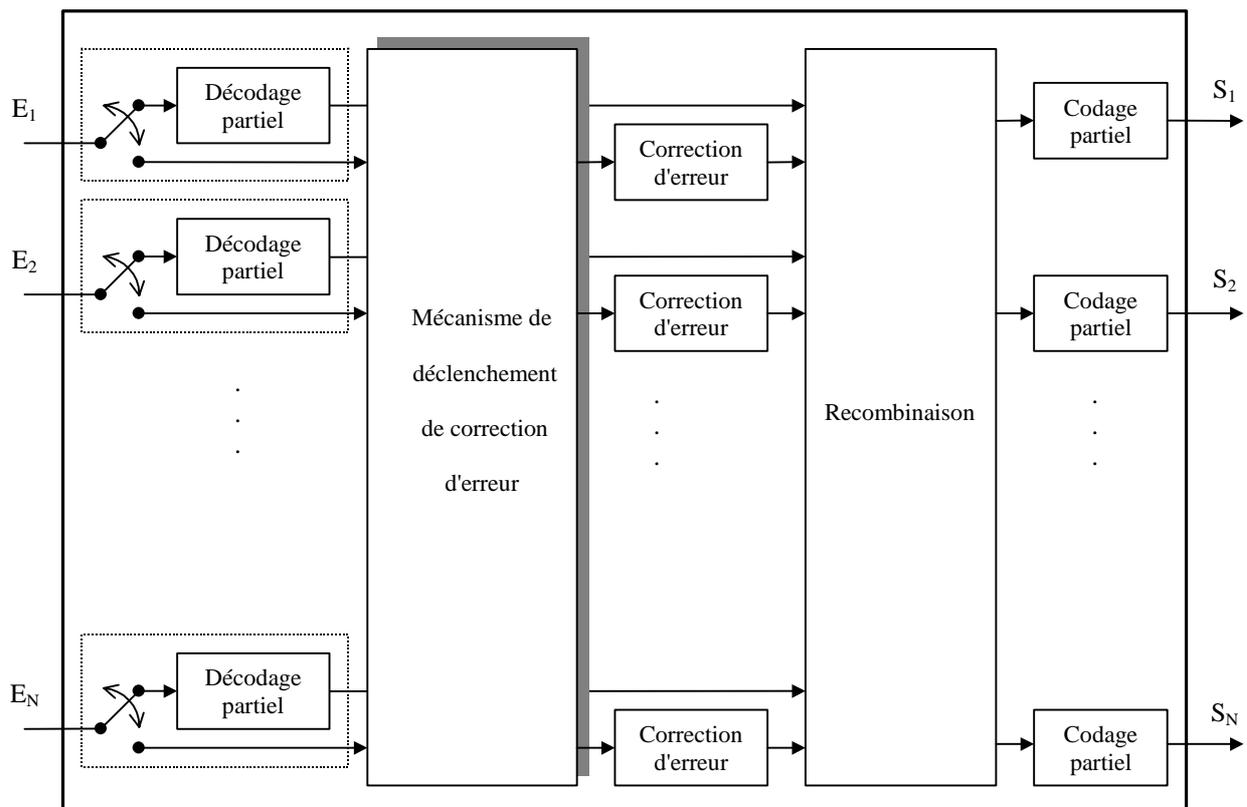


Fig. 6.4 : Schéma fonctionnel du pont audio de téléconférence opérant dans le domaine compressé et incluant la correction des pertes paquets.

6.5. Autres exemples d'application

6.5.1 Terminal de téléconférence multipoint décentralisée

En architecture de téléconférence multipoint décentralisée chaque terminal effectue la sommation des $N-1$ signaux reçus à partir des autres terminaux pour les reproduire sur le même haut-parleur. Dans le cas classique, les différents flux audio sont complètement décodés avant d'être mixés dans le domaine temporel. Pour les codeurs en sous-bandes en général, il y a possibilité de remplacer les

$N - 1$ opérations de synthèse par banc de filtres par une seule en réalisant une sommation directement sur les paramètres fréquentiels obtenus après décodage partiel. Ces paramètres sont les échantillons des sous-bandes pour le codeur MPEG-1 et les coefficients MDCT pour le codeur TDAC. La complexité algorithmique est ainsi diminuée. Dans ce cas, elle augmente moins rapidement en fonction du nombre, N , de terminaux connectés. Ceci est d'autant plus vrai si la synthèse par banc de filtres détient un grand pourcentage du temps de calcul de l'opération de décodage, ce qui est le cas par exemple des décodeurs MPEG-1 Layer I et II.

Comme pour le pont en architecture centralisée, il est possible aussi de mettre en œuvre pour chacun des signaux d'entrées la même méthode de correction de trames effacées basée sur l'extrapolation sinusoïdale. En plus, elle peut être consolidée par le mécanisme de déclenchement conditionnel de correction profitant de l'interaction entre les différents signaux. La figure 6.5 donne une illustration de l'ensemble du procédé dans le terminal incluant toutes ces fonctionnalités.

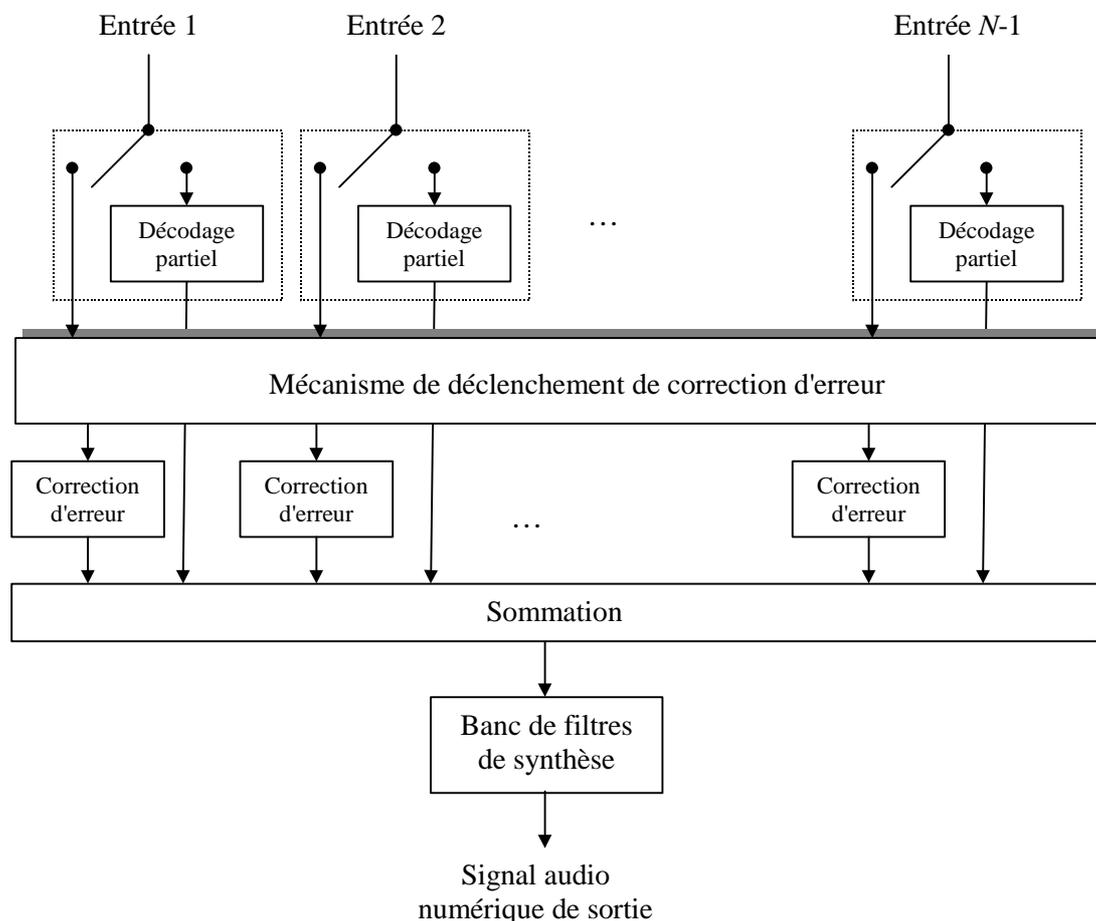


Fig. 6.5 : Terminal de téléconférence en configuration multipoint décentralisée opérant dans le domaine compressé.

6.5.2 Conversion du débit dans les serveurs audio et les passerelles

Dans un service d'audio à la demande, les clients peuvent se connecter sur différents types de réseaux. La largeur de bande passante disponible pour chacun peut varier selon le type de connexion. Elle peut être inférieure à 56 kbit/s pour un accès modem et de 10 Mbit/s pour un accès via un réseau Ethernet. Chaque client doit recevoir la meilleure qualité possible correspondant à sa connexion. Une solution consiste à stocker dans le serveur les signaux audio à différents débits. Cette duplication a l'inconvénient de nécessiter une grande capacité de stockage (il suffit de se rappeler qu'une minute d'un signal audio compressé à 96 kbit/s nécessite une taille mémoire de 5,76 Mbits). Une solution plus adaptée consiste à stocker les signaux avec un taux de compression faible assurant la qualité transparente. Lors de chaque demande une conversion de débit peut être effectuée pour adapter le débit du flux audio à la bande passante du réseau d'accès. Vu le nombre de demandes élevées que peut recevoir simultanément un serveur audio, l'utilisation d'une méthode de conversion du débit dans le domaine compressé avec une faible complexité peut être d'une grande utilité.

Dans une communication multicast ou multi-unicast entre terminaux situés sur des réseaux hétérogènes le même problème se manifeste. Une première possibilité consiste à négocier le débit lors de la phase d'établissement de la communication. Dans ce cas c'est le terminal ayant les plus faibles ressources en termes de bande passante qui impose son débit pénalisant ainsi les autres. L'utilisation d'une passerelle remédie à ce problème en offrant la possibilité de transcodage et de conversion de débit entre différents types de réseaux. Un traitement dans le domaine compressé permet à une telle solution d'être utilisée sans avoir à subir ses inconvénients à savoir la complexité dans ce nœud du réseau et aussi le retard qui a un effet négatif pour l'interactivité de la communication. Si l'environnement réseau présente des problèmes de pertes de paquets, il est possible d'introduire aussi à ce niveau une méthode de correction dans le domaine compressé pour consolider celle effectuée au niveau des terminaux.

6.5.3 Contexte du système MPEG-4

MPEG-4 est une norme de codage multimédia par objets. Elle va au-delà du codage de médias individuels. Ainsi, en plus d'une large panoplie de codeurs audio et vidéo, elle fournit de nouveaux outils tels que celui permettant la description d'une scène audiovisuelle. Dans un contexte d'un serveur multimédia, le terminal MPEG-4 client reçoit un flux binaire multiplexé correspondant à chacun des objets média primitifs codés en plus d'instructions sur leur composition pour construire la scène audiovisuelle. La norme MPEG-4 Système fournit un format spécial, appelé *BIFS* (BInary Format for Scene description), pour transmettre ces instructions.

La partie du BIFS qui contrôle la composition d'une scène sonore est dite *AudioBIFS* [Scheirer 99, Väänänen 00, Scheirer 00]. Son rôle est de décrire la composition spatio-temporelle des objets audio. Un objet audio étant un flux binaire élémentaire obtenu par un des codeurs MPEG-4 Audio. Pour construire la scène sonore et assurer un certain rendu, ces différents flux décodés peuvent subir un post-traitement de type composition (mixage, retards pour synchronisation), spatialisation sonore, ajout d'effet de salle (réverbération, effets spéciaux...) ou un filtrage quelconque par un filtre téléchargé. AudioBIFS représente les manipulations à effectuer par un graphe. A l'entrée du graphe il y a les signaux audio décodés, et chacun de ses nœuds représente un type d'opération de traitement du

signal à réaliser. A sa sortie sont obtenus les différents signaux sonores à restituer ou à associer à d'autres objets média.

Parmi les nœuds définis dans la version 1 de MPEG-4 on trouve *AudioMix* qui permet de mixer M voies d'entrée pour obtenir N voies de sortie en spécifiant une matrice de mixage. *AudioDelay* permet de synchroniser les flux sonores par ajout d'un certain retard. Quant au nœud *AudioFx* il offre la possibilité d'effectuer un traitement du signal quelconque sur un ensemble de voies sonores d'entrée. Les algorithmes utilisés sont mis à jour dynamiquement et sont transmis dans le graphe de la scène. Ils sont décrits sous forme de routines écrites dans le langage spécifique SAOL (Structured Audio Score Language) qui possède des fonctions prédéfinies telles que filtres en peigne et passe-tout, lignes à retards fractionnels, filtres FIR et IIR.

Dans les outils de compression audio fournis par la norme MPEG-4 on trouve des codeurs par transformée utilisés surtout pour la transmission audio haute qualité (monophonique et multivoies) à des débits entre 16 et 64 kbit/s [Brandenburg 00]. C'est le cas des codeurs AAC et TwinVQ basés sur la transformée MDCT.

Dans ce contexte du système MPEG-4, il est possible de tirer profit des avantages fournis par le traitement dans le domaine compressé et précisément le mixage et le filtrage.

6.6. Conclusion

Plusieurs applications pour le traitement dans le domaine compressé ont été présentées dans ce chapitre. Ils concernent aussi bien des mises en œuvre dans les nœuds intermédiaires du réseau que dans les terminaux. Ces types de traitements visent soit à remédier à des problèmes réseau (adaptation à la bande passante, pertes de paquets) soit à améliorer la qualité globale de communication. Par ses avantages liés principalement à une faible complexité et délai, le traitement dans le domaine compressé permet d'atteindre ces objectifs sans affecter les autres paramètres de qualité. L'exemple du pont audio illustre un exemple complet de mise en œuvre de tels traitements en environnement réseau.

Chapitre 7 Conclusion générale

7.1. Bilan

Cette thèse a trait au traitement du signal audio dans le domaine codé pour des codeurs de type fréquentiel perceptuel. Elle vise comme domaine d'application préférentiel les systèmes de communication de groupe. Les études menées dans le cadre de ce travail se focalisent sur les exemples des codeurs MPEG-1 Layer I-II et TDAC de FTR&D. Elles s'intéressent particulièrement aux traitements de type filtrage, mixage, et conversion de débit.

La première facette de cette thèse a un aspect de recherche algorithmique. Elle a eu pour objet l'élaboration d'algorithmes permettant la transposition des traitements cités ci-dessus dans le domaine codé. La démarche suivie dans la conception de tels algorithmes a eu toujours pour objectif la réduction au maximum de la complexité et du délai algorithmiques avec la contrainte de maintenir une qualité transparente par rapport au traitement dans le domaine temporel. Les principales contributions de cette thèse sur cet axe de recherche sont les suivantes :

- Méthode générique pour le filtrage dans le domaine des sous-bandes. Elle permet de réaliser la conversion de tout filtre rationnel (FIR ou IIR), défini initialement par sa fonction de transfert, en une matrice de filtres de sous bandes. La propriété de reconstruction parfaite est la seule condition à satisfaire par le banc de filtres. Cette méthode a été appliquée aux bancs de filtres TDAC et pseudo-QMF de MPEG-1 et utilisée pour réaliser la synthèse binaurale par des filtres HRTF dans le domaine des sous-bandes.
- Algorithme de sommation de flux audio codés MPEG-1 qui résout principalement le problème d'allocation de bits pour les trames du signal somme. Il propose une nouvelle méthode d'estimation des seuils de masquage et par suite des rapports signal à masque nécessaire pour la procédure d'allocation de bits. Cet algorithme a été dérivé dans le cas d'un seul signal pour convertir le débit d'un flux codé.
- Algorithme de sommation de flux audio codés TDAC de FTR&D qui profite pleinement des caractéristiques de ce codeur. Il exploite la structure particulière des dictionnaires de la quantification vectorielle utilisés pour minimiser au maximum la phase de décodage partiel et ainsi réaliser un traitement avancé sur le flux codé. Ceci est facilité par l'utilisation du phénomène de masquage dans un mélange de signaux.
- Méthode de correction de trames effacées dans le domaine de la transformée MDCT basée sur l'extrapolation de ses coefficients.

Ces études algorithmiques ont montré la difficulté de tirer des règles génériques pouvant s'appliquer à toute la classe de codeurs fréquentiels perceptuel. En effet, si la transformation temps/fréquence est bien décrite mathématiquement et permet ainsi une "manipulation" plus ou moins facile, les blocs

postérieurs de quantification et codage relèvent plus dans leur conception du domaine l'ingénierie et présentent des différences caractéristiques d'un codeur à l'autre.

Le second volet de cette thèse concerne l'étude d'architectures algorithmiques tirant profit des avantages des traitements audio dans le domaine compressé pour des applications de transmission du son sur le réseau. La contribution essentielle de la thèse sur cet axe est concrétisée par l'implantation et la mise en œuvre d'un pont audio de téléconférence multipoint en architecture centralisée pour un environnement H.323. Ce pont audio se base sur le codeur TDAC de FTR&D et joue le rôle de concentrateur de flux par sommation des flux dans le domaine codé. Il inclut la récupération des paquets perdus par correction des trames dans le domaine de la transformée MDCT et aussi la gestion des flux discontinus.

7.2. Perspectives et travaux futurs

Plusieurs voies de recherche peuvent être suggérées comme continuité des travaux entamés dans cette thèse. Elles concernent aussi bien les développements algorithmiques que les aspects architecture et implantation.

Ainsi, pour le filtrage dans le domaine des sous-bandes la méthode d'implantation de la matrice générée des filtres de sous-bandes reste générale et ne tient compte d'aucune propriété spécifique au banc de filtres utilisé. L'examen de l'écriture de cette solution dans le cas particulier d'un banc de filtres MDCT et pour le filtrage FIR peut s'avérer intéressant pour le développement d'un algorithme de filtrage rapide en sous-bandes. En effet, d'une part, une relation directe existe entre une transformée MDCT et une DCT type IV [Malvar 92], d'autre part, des algorithmes efficaces ont déjà été proposés pour le filtrage dans le domaine DCT [Kresh 99]... De même, il serait intéressant de chercher des structures d'implantations plus efficaces lors de l'utilisation de filtres IIR possédant des caractéristiques particulières tels que les filtres passe-tout.

Plusieurs études ont déjà été faites sur le filtrage adaptatif en sous-bandes et ont démontré son efficacité en terme de vitesse de convergence et de réduction de la complexité [Gilloire 92]. Etendre la solution de filtrage en sous-bandes proposée dans cette thèse au cas adaptatif et faire la relation avec les travaux déjà existants serait aussi un axe de recherche intéressant.

L'étude de l'interaction entre codage fréquentiel et d'autres traitements de type annulation d'écho et débruitage est un sujet algorithmique à développer. Notons que des travaux ont déjà été effectués dans ce sens [Carnero 97].

Au niveau de l'architecture de traitement du son pour la transmission sur le réseau, d'autres schémas incluant plus de fonctionnalités peuvent être imaginés. Ceci peut concerner particulièrement l'exemple de l'architecture de téléconférence centralisée où l'on peut centraliser d'autres traitements de type annulation d'écho acoustique, débruitage et aussi spatialisation sonore. L'implantation de cette dernière peut s'étendre à des méthodes plus complètes de spatialisation, au-delà de la synthèse binaurale, incluant la réverbération et l'effet de salle tel que celle utilisée dans le système *SPAT* [Jot 97].

Finalement, la nouvelle norme MPEG-4 présente des fonctionnalités plus complètes pour les traitements audio et suggère une large panoplie de codeurs notamment l'AAC et la Twin-VQ. L'étude des possibilités de traitement dans le domaine codé de ces nouveaux codeurs serait très intéressante.

Annexes

Annexe A Décomposition polyphasée des filtres récursifs

A.1. Principe

Considérons un filtre rationnel à réponse impulsionnelle infinie (IIR) donné par sa fonction de transfert $S(z)$. Cette fonction de transfert peut s'exprimer comme suit [Oppenheim 99]:

$$S(z) = \prod_{k=1}^{N_s} \frac{a_{0k} + a_{1k}z^{-1} + a_{2k}z^{-2}}{1 + b_{1k}z^{-1} + b_{2k}z^{-2}} \cdot \prod_{k=1}^{N_f} \frac{c_{0k} + c_{1k}z^{-1}}{1 + d_{1k}z^{-1}}, \quad (\text{A.1})$$

avec les a_{lk} , b_{lk} , c_{lk} et d_{lk} sont tous des coefficients réels. Cette expression correspond à l'implémentation du filtre en une structure en cascade de cellules du premier et second ordre.

Propriété : Soit $S_c(z)$ le filtre équivalent à la mise en cascade des deux filtres scalaires $S_1(z)$ et $S_2(z)$ ($S_c(z) = S_1(z)S_2(z)$). Soient $\mathbf{S}_c(z)$, $\mathbf{S}_1(z)$ et $\mathbf{S}_2(z)$ les matrices de transfert correspondants respectivement aux versions blocs de ces filtres. Alors, on a :

$$\mathbf{S}_c(z) = \mathbf{S}_1(z)\mathbf{S}_2(z). \quad (\text{A.2})$$

En effet, la mise en cascade des représentations bloc de ces deux filtres fera intervenir une succession d'un bloc de conversion parallèle-série et d'un bloc de conversion série-parallèle, comme dans la figure 3.8a, qui est équivalente à l'identité. Ceci permet de déduire directement la relation (A.2).

Comme corollaire à cette propriété, on peut déduire les formules des composantes polyphasées du filtre produit $S_c(z)$:

$$\begin{cases} S_{c,k}(z) = \sum_{l=0}^k S_{1,l}(z)S_{2,(k-l)}(z) + \sum_{l=k+1}^{M-1} S_{1,l}(z)S_{2,(M+k-l)}(z)z^{-1}, & 0 \leq k \leq M-2, \\ S_{c,(M-1)}(z) = \sum_{l=0}^{M-1} S_{1,l}(z)S_{2,(M-1-l)}(z). \end{cases} \quad (\text{A.3})$$

Ces équations s'obtiennent par simple écriture de l'expression des éléments de la première ligne de la matrice produit $\mathbf{S}_c(z)$, et qui sont les composantes polyphasées du filtre $S_c(z)$, à partir de (A.2).

A partir de la propriété ci-dessus des matrices de transfert et de la décomposition en une structure cascade de l'équation (A.1), on peut conclure qu'il suffit de trouver les matrices de transfert des cellules du premier et second ordre pour déterminer celle d'un filtre récursif quelconque. Dans le

paragraphe A.3, nous donnerons les formules des décompositions en filtres polyphasés des cellules du premier et second ordre.

A.2. Forme générale des composantes polyphasées d'un filtre IIR

Soit $S(z)$ un filtre IIR donné par ses pôles, p_k , et ses zéros, z_k :

$$S(z) = \frac{\prod_{k=1}^N (1 - z_k z^{-1})}{\prod_{k=1}^N (1 - p_k z^{-1})}. \quad (\text{A.4})$$

Nous donnerons dans la suite le principe de calcul ainsi que la forme de ses composantes polyphasées, ce développement est tiré de [Bellanger 90]. En effet, en utilisant l'identité :

$$1 - pz^{-1} \equiv \frac{1 - p^M z^{-M}}{\sum_{k=0}^{M-1} p^k z^{-k}}, \quad p \in \mathbb{C}. \quad (\text{A.5})$$

On peut écrire :

$$S(z) = \frac{\prod_{k=1}^N (1 - z_k z^{-1}) (1 + p_k z^{-1} + \dots + p_k^{M-1} z^{-(M-1)})}{\prod_{k=1}^N (1 - p_k^M z^{-M})}. \quad (\text{A.6})$$

Cette relation peut s'écrire aussi sous la forme suivante :

$$S(z) = \frac{\sum_{i=0}^{NM} a_i z^{-i}}{1 + \sum_{j=0}^N b_j z^{-Mj}}. \quad (\text{A.7})$$

Ce qui permet de déduire donc la forme décomposée de $S(z)$:

$$S(z) = \sum_{k=0}^{M-1} z^{-k} S_k(z^M), \quad (\text{A.8})$$

avec

$$S_k(z) = \frac{\sum_{i=0}^N a_{Mi+k} z^{-i}}{1 + \sum_{j=0}^N b_j z^{-j}}, \quad 0 \leq k \leq M-1, \quad (\text{A.9})$$

soit

$$S_k(z) = \frac{N_k(z)}{D(z)}, \quad 0 \leq k \leq M-1. \quad (\text{A.10})$$

Les filtres polyphasés, $S_k(z)$, $0 \leq k \leq M-1$, ont donc tous le même dénominateur $D(z)$. Ils ont pour pôles les p_i^M , $1 \leq i \leq N$. Si le filtre $S(z)$ est stable alors $|p_i| < 1$, $\forall i$, donc $|p_i^M| < 1$ et par conséquent

ses composantes polyphasés seront aussi stables (et donc tous les éléments de la matrice de transfert lui correspondant).

A.3. Composantes polyphasées de cellules du premier et second ordre

L'utilisation du principe de calcul du paragraphe précédent, basé sur l'identité (A.5), permet de déterminer les composantes polyphasés des cellules du premier et second ordre purement récurrentes. Le tableau A.1 résume les résultats du calcul pour les différents cas possibles selon le type des pôles.

Pour une cellule du second ordre général de la forme :

$$S(z) = \frac{a_0 + a_1 z^{-1} + a_2 z^{-2}}{1 + b_1 z^{-1} + b_2 z^{-2}}, \quad (\text{A.11})$$

avec $a_i, b_i \in \mathbb{R}$ et $b_2 \neq 0$. Le cas d'une cellule du premier ordre revient à prendre $b_2 = 0$ et $b_1 \neq 0$, et le même raisonnement reste valable.

On décompose d'abord la partie purement récurrente :

$$\frac{1}{1 + b_1 z^{-1} + b_2 z^{-2}} = \sum_{k=0}^{M-1} z^{-k} \frac{N'_k(z^M)}{D(z^M)}. \quad (\text{A.12})$$

Les composantes polyphasés de $S(z)$ auront le même dénominateur $D(z)$. Les numérateurs $N_k(z)$ sont données par les formules suivantes :

$$\begin{cases} N_0(z) = a_0 N'_0(z) + a_1 z^{-1} N'_{M-1}(z) + a_2 z^{-1} N'_{M-2}(z), \\ N_1(z) = a_0 N'_1(z) + a_1 N'_0(z) + a_2 z^{-1} N'_{M-1}(z), \\ N_k(z) = a_2 N'_{k-2}(z) + a_1 N'_{k-1}(z) + a_0 N'_k(z), \quad \text{pour } 2 \leq k \leq M-1, \end{cases} \quad (\text{A.13})$$

Ces relations s'obtiennent aisément en substituant (A.12) dans (A.11) et en identifiant avec la forme polyphasé de $S(z)$.

	Cellule du second ordre			Cellule du premier ordre
	Pôles complexes conjugués	Pôles réels distincts	Pôle réel double	
$S(z)$	$\frac{1}{(1-pz^{-1})(1-p^*z^{-1})}$ $p = re^{j\theta}$	$\frac{1}{(1-az^{-1})(1-bz^{-1})}$ $a, b \in \mathbb{R}^*$	$\frac{1}{(1-az^{-1})^2}$ $a \in \mathbb{R}^*$	$\frac{1}{1-az^{-1}}$ $a \in \mathbb{R}^*$
$N_k(z),$ $0 \leq k \leq M-2$	$r^k \frac{\sin[(k+1)\theta]}{\sin\theta} + r^{k+M} \frac{\sin[(M-1-k)\theta]}{\sin\theta} z^{-1}$	$\frac{a^{k+1} - b^{k+1}}{a-b} + \frac{a^M b^{k+1} - a^{k+1} b^M}{a-b} z^{-1}$	$(k+1)a^k + (M-1-k)a^{k+M} z^{-1}$	a^k
$N_{M-1}(z)$	$r^{M-1} \frac{\sin(M\theta)}{\sin\theta}$	$\frac{a^M - b^M}{a-b}$	Ma^{M-1}	a^{M-1}
$D(z)$	$1 - 2r^M \cos(M\theta) z^{-1} + r^{2M} z^{-2}$	$1 - (a^M + b^M) z^{-1} + a^M b^M z^{-2}$	$1 - 2a^M z^{-1} + a^{2M} z^{-2}$	$1 - a^M z^{-1}$

Tableau A.1 : Formules des composantes polyphasées des cellules du 1^{er} et 2^d ordre purement récurives.

Annexe B Algorithme de Schalkwijk

Considérons un dictionnaire, \mathcal{C} , à union de codes à permutation dans \mathbb{R}^N . Soit $\mathbf{y} = [y_1 \ y_2 \ \dots \ y_N]^T$ un vecteur appartenant à la classe j de ce dictionnaire. Le problème inhérent à ce type de dictionnaire consiste, d'une part, à trouver le rang (ou index lexicographique), r , du vecteur \mathbf{y} parmi les $\kappa(j)$ éléments de sa classe. Et inversement, de retrouver \mathbf{y} sachant que son rang dans la classe j est r . Dans cette annexe nous décrivons les algorithmes de Schalkwijk direct et inverse [Schalkwijk 72] qui permettent d'assurer ces tâches.

B.1. Détermination du rang d'un vecteur dans sa classe

Soit q le nombre de valeurs distinctes que prennent les composantes du vecteur \mathbf{y} . On dénote $\mu_0, \mu_1, \dots, \mu_{q-1}$ ces valeurs avec $\mu_0 > \mu_1 > \dots > \mu_{q-1}$. On peut remarquer au passage que si on note w^d le nombre de répétition de la valeur μ_d , $0 \leq d \leq q-1$, alors le vecteur $\tilde{\mathbf{y}}_j$, leader de \mathbf{y} , s'écrit comme suit :

$$\tilde{\mathbf{y}}_j = \left[\begin{array}{cccc} \leftarrow w^0 & \rightarrow & \leftarrow w^1 & \rightarrow \\ \mu_0 & \dots & \mu_1 & \dots & \mu_{q-1} & \dots & \mu_{q-1} \end{array} \right]^T. \quad (\text{B.1})$$

On associe à $\mathbf{y} = [y_1 \ y_2 \ \dots \ y_N]^T$, le vecteur $\mathbf{D} = [d_1 \ d_2 \ \dots \ d_N]^T$ tel que $d_k = d$ si et seulement si $y_k = \mu_d$, $0 \leq d \leq q-1$: d_k est l'indice de la valeur, μ_d , que prend la $k^{\text{ème}}$ composante, y_k .

On appelle w_k^d le nombre de répétitions de la valeur μ_d , dans le vecteur \mathbf{y} , qui restent dans les positions k à N .

Avec ces notations, la formule de Schalkwijk qui donne le rang, r , du vecteur \mathbf{y} dans sa classe, s'écrit comme suit :

$$r = \sum_{k=1}^N \sum_{d=0}^{d_k-1} \frac{(N-k)!}{(w_k^d - 1)! \prod_{\substack{i=0 \\ i \neq d}}^{q-1} (w_k^i)!}, \quad (\text{B.2})$$

avec les conventions : $0! = 1$, $(-1)! = \infty$ et $\sum_{d=0}^{-1} = 0$.

Cette formule s'écrit aussi sous la forme suivante :

$$r = \sum_{k=1}^N \frac{(N-k)!}{\prod_{i=0}^{q-1} (w_k^i)!} \left(\sum_{d=0}^{d_k-1} w_k^d \right). \quad (\text{B.3})$$

B.2. Algorithme de Schalkwijk inverse

Connaissant la classe j du vecteur \mathbf{y} ainsi que son rang, r , dans cette classe on cherche à déterminer le vecteur \mathbf{y} . L'information sur l'index j de la classe permet de déterminer le leader $\tilde{\mathbf{y}}_j$. On connaît donc : le nombre q de composantes distinctes, leurs valeurs $(\mu_0 > \mu_1 > \dots > \mu_{q-1})$ et le nombre w^i de répétition de chaque valeur μ_i dans $\tilde{\mathbf{y}}_j$ et par conséquent dans \mathbf{y} . Pour retrouver le vecteur \mathbf{y} , on détermine successivement $d_1(y_1)$ puis $d_2(y_2)$ jusqu'à $d_N(y_N)$. d_k est tel que :

$$\frac{(N-k)!}{\prod_{i=0}^{q-1} (w_k^i)!} \sum_{d=0}^{d_k-1} w_k^d \leq \alpha_k(r) < \frac{(N-k)!}{\prod_{i=0}^{q-1} (w_k^i)!} \sum_{d=0}^{d_k} w_k^d, \quad (\text{B.4})$$

où :

$$\alpha_k(r) = r - \sum_{l=1}^{k-1} \frac{(N-l)!}{\prod_{i=0}^{q-1} (w_l^i)!} \sum_{d=0}^{d_l-1} w_l^d.$$

La référence [Lamblin 88] illustre en détail un exemple de détermination du rang, r , ainsi que le passage de r au vecteur \mathbf{y} .

Annexe C **Base de données parole UIT**

La base de données de signaux parole de l'UIT a été utilisée dans les statistiques et les tests concernant la méthode de sommation pour le codeur TDAC de FTR&D. Cette base de données contient des signaux de parole échantillonnés à $f_e = 16$ kHz. Elle regroupe des voix d'hommes et de femmes d'une longueur totale de 3 minutes 36 secondes. Le tableau suivant donne des statistiques sur cette base de données en nombre de trames de 20 ms.

	Trames voisée	Trames non voisée	Nombre total de trames
Voix d'hommes	2563	2249	4812
Voix de femmes	2605	2207	4812

Tableau C.1 : Statistique sur la base de donnée parole de l'UIT.

Bibliographie

- [Adoul 86] J.-P. Adoul, "La quantification vectorielle des signaux : approche algébrique," *Annales des Télécommunications*, tome 41, n° 3-4, Mars-Avril 1986.
- [Allamanche 00] E. Allamanche and J. Herre, "Secure Delivery of Compressed Audio by Compatible Bitstream Scrambling," 108th AES Convention 2000 February 19-22, Paris, France.
- [Aoyagi 96] Y. Aoyagi, "Multipoint Communication Apparatus with Reduced Number of Audio Decoders," United States Patent, US5727075, March 1996.
- [Assunção 98] P. A. A. Assunção and M. Ghanbari, "A Frequency-Domain Video Transcoder for Dynamic Bit-Rate reduction of MPEG-2 Bit Stream," *IEEE Trans. On Circuits and Systems for Video Technology*, Vol. 8, No. 8, December 1998.
- [Bargar 99] R. Bargar and al., "Networking Audio and Music Using Internet2 and Next-Generation Internet Capabilities," AES White Paper – 1001, *J. Audio Eng. Soc.* Vol. 47, No. 4, April 1999.
- [Beerends 92] J. G. Beerends, J. A. Stemerdink, "A Perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation" *J. Audio Eng. Soc.* Vol. 40, No. 12, 1992 December.
- [Belanger 90] M. Bellanger, "Traitement numérique du signal : théorie et pratique," 4^{ème} édition, Masson, 1990.
- [Benjelloun 99] A. Benjelloun Touimi, Y. Mahieux, C. Lamblin, "Procédé de traitement d'au moins un flux binaire audio codé organisé sous la forme de trames," Brevet Français n° 99 15574, 8 décembre 1999.
- [Benjelloun 00a] A. Benjelloun Touimi, Y. Mahieux, C. A. Lanciani, "A Summation Algorithm for MPEG-1 Coded Audio Signals: a First Step Towards Audio Processing in the Compressed Domain," *Annales des Télécommunications*, tome 55, n° 3-4, Mars-Avril 2000.
- [Benjelloun 00b] A. Benjelloun Touimi, "A Generic Framework for Filtering in Subband Domain," *IEEE 9th Workshop on Digital Signal Processing*, Hunt, Texas, USA, October 2000.
- [Benjelloun 01] A. Benjelloun Touimi, C. Boisseau, C. Lamblin, F. Bouteille, D. Deleam "Méthode et dispositif de traitement d'une pluralité de flux binaires audio" Brevet Français n° 01 01662, 2 février 2001.

- [Benyassine 97] A. Benyassine, E. Shlot, H.-Y. Su, D. Massaloux, C. Lamblin and J.-P. Petit, "ITU-T Recommendation G.729 Annexe B: A Silence Compression Scheme for Use with G.729 Optimized for V.70 Digital Simultaneous Voice and Data Applications," *IEEE Communications Magazine*, September 1997.
- [Berger 72] T. Berger and F. Jelinek, "Permutation Codes for Sources," *IEEE Trans. On Information theory*, Vol. IT-18, No. 1, January 1972.
- [Boisseau 99] C. Boisseau, "Traitements dans les ponts de téléconférence pour réduction de complexité et robustesse aux pertes de paquets," Stage Ingénieur/DEA INPG/ENSERG, Septembre 1999.
- [Bosi 97] M. Bosi and al., "ISO/IEC MPEG-2 Advanced Audio Coding," *J. Audio Eng. Soc.*, Vol. 45, No. 10, 1997 October, pp. 789-814.
- [Bosi 99] M. Bosi, "Filter Banks in Perceptual Audio Coding," AES 17th Conference on high-quality Audio Coding, September 2-5, 1999, Florence, Italy.
- [Botte 88] M. C. Botte, G. Canevet, L. Demany and C. Sorin, "Psychoacoustique et Perception Auditive," INSERM/SFA/CNET, 1988.
- [Brandenburg 94] K. Brandenburg and G. Stoll, "ISO-MPEG-1 Audio: A Generic Standard for Coding of High-Quality Digital Audio," *J. Audio Eng. Soc.*, Vol. 42, No. 10, 1994 October, pp. 780-792
- [Brandenburg 97] K. Brandenburg, M. Bosi, "Overview of MPEG Audio : Current and Future Standards for Low-Bit-rate Audio Coding," *J. Audio Eng. Soc.*, Vol. 45, No. 1/2, 1997 January/February, pp. 4-21.
- [Brandenburg 98] K. Brandenburg, "Perceptual Coding of High Quality Digital Audio," In *Applications of Digital Signal Processing to Audio and Acoustics*. Edited by M. Kahrs and K. Brandenburg, Kluwer Academic Publishers, 1998.
- [Brandenburg 00] K. Brandenburg, O. Kunz and A. Sugiyama, "MPEG-4 Natural Audio Coding," *Signal Processing: Image Communications* 15 (4-5) (2000) pp. 423-444.
- [Broadhead 95] M. A. Broadhead and C. B. Owen, "Direct Manipulation of MPEG Compressed Digital Audio," *ACM Multimedia Conference*, San Francisco, 1995.
- [Carnero 97] B. Carnero, "Perceptual Coding and Enhancement of Wide-Band Speech," Thèse de doctorat de l'Ecole Polytechnique Fédérale de Lausanne, n° 1675, 1997.
- [Chan 96] D. Chan, J. Yang, C. Fang, "Fast Implementation of MPEG Audio Coder Using Recursive Formula with Fast Discrete Cosine transforms," *IEEE Trans. on Speech and Audio Processing*, Vol. 4, No. 2, March 1996, pp. 144-148.
- [Chang 95] S.-F. Chang and D. G. Messerschmitt, "Manipulation and Composition of MC-DCT Compressed Video," *IEEE Journal on Selected Areas in Communications*, Vol. 13, No. 1, January 1995.

- [Chen 97] T. Chen, A. Katsaggelos and S. Y. Kung, Editors, "The Past, Present, and Future of Multimedia Signal Processing," IEEE Signal Processing Magazine, Vol. 14, No. 4, pp. 28-51, July 1997.
- [Chen 76] W.-H. Chen and S. G. Fralick, "Image Enhancement Using Cosine Transform Filtering," Image Sci. Math. Symp., Monterey, CA. November 1976.
- [Chitprasert 90] B. Chitprasert and K. R. Rao, "Discrete Cosine Transform Filtering," Signal Processing 19 (1990) pp. 233-245.
- [Colomes 94] C. Colomes, M. Lever and Y. F. Dehery, "A Perceptuel Measurement system (POM) for the Quality Assesment of Perceptuel Coders," AES 96th Convention, Preprint 3801, 1996 February, Amsterdam.
- [Combescure 99] P. Combescure and al. "A 16, 24, 32kbits/s Wideband Speech Codec Based on ATCELP," IEEE Int. Conf. On Acoust., Speech, Signal Proc., March 15-19, 1999, Arizona , USA.
- [Crochiere 81] R. E. Crochiere and L. Rabiner, "Interpolation and Decimation of Digital Signals – A Tutorial Review," Proceeding of the IEEE, Vol. 69, pp. 300-331, Mars 1981.
- [Dawood 99] A. M. Dawood and M. Ghanbari, "Scene Content Classification from MPEG Coded Bit Streams," IEEE 3rd Workshop on Multimedia Signal Processing, pp. 253-258, September 13-15, 1999, Copenhagen, Denmark.
- [Dia 93] H. Dia, "Codage par transformée de la parole à bande élargie (0 à 7kHz)," Thèse de doctorat de l'Institut National Polytechnique de Grenoble, Octobre 1993.
- [DT&FT 98] Deutsche Telekom and France Telecom / CNET, "High-Level Description for ITU-T Wideband (7kHz) ATCELP Speech Coding Algorithm of Deutsche Telekom, Aachen University of Technology (RWTH) and France Telecom (CNET)," Delayed contribution COM16-129 / 130, ITU-T Study Group 16, Q.20, Geneva, January 1998.
- [Duhamel 91] P. Duhamel, Y. Mahieux and J.-P. Petit, "A Fast Algorithm For The Implementation of Filter Banks Based on Time Domain Aliasing Cancellation," Int. Conf. on Acoust., Speech, Signal Proc., pp. 2209-2212, 1991.
- [Fernando 99] W. A. C Fernando and C. N. Cangarajah, "Sudden Scene Change Detection in MPEG-2 Sequences," IEEE 3rd Workshop on Multimedia Signal Processing, pp. 259-264, September 13-15, 1999, Copenhagen, Denmark.
- [Gardner 98] W. G. Gardner, "Reverberation Algorithms," In Applications of Digital Signal Processing to Audio and Acoustics. Edited by M. Kahrs and K. Brandenburg, Kluwer Academic Publishers, 1998.
- [Gänsler 98] T. Gänsler and P. Eneroth, "Influence of Audio Coding on Stereophonic Acoustic Echo Cancellation," Int. Conf. On Acoust., Speech, Signal Proc., 1998.

- [Gersho 92] A. Gersho and R. M. Gray, "Vector Quantization and Signal Compression," Kluwer Academic Publishers, 1992.
- [Gilloire 92] A. Gilloire, M. Vetterli, "Adaptative Filtering in Subbands with Critical Sampling: Analysis, Experiments, and Application to Acoustic Echo Cancellation," IEEE Trans. On Signal Processing, Vol. 40, No. 8, August 1992.
- [Grill 99] B. Grill, "The MPEG-4 General Audio Coder," AES 17th Conference High Quality Audio Coding, September 2-5 1999, Florence, Italy.
- [Guimares 98] M. Perreau Guimares, "Optimisation des ressources binaires et modélisation psychoacoustique pour le codage audio," Thèse de doctorat de l'Université Paris V, 1998.
- [G.722] Recommandation UIT-T G.722, "Codage audiofréquence à 7 kHz à un débit inférieur ou égal à 64 kbit/s," Livre bleu, Novembre 1988.
- [G.722.1] ITU-T Recommendation G.722.1, "Coding at 24 and 32 kbit/s for Hands-free Operation in Systems with Low Frame Loss," September 1999.
- [Hans 97] M. C. Hans, V. Bhaskaran, "A Compliant MPEG-1 Layer II Audio Decoder with 16-B Arithmetic Operations," IEEE Signal Processing Letters, Vol. 4, No. 5, May 1997.
- [Hans 98] M. C. Hans, "Optimisation of Digital Audio for Internet Transmission," Ph.D. thesis, Georgia Institute of Technologie, May 1998.
- [Hardman 00] V. Hardman and O. Hodson, "Internet/Mbone Audio," in Handbook of Emerging Communications Technologies: The Next Decade, Editor R. Osso, CRC Press, Springer, 2000.
- [Humes 89] L. E. Humes, W. Jesteadt, "Models of the Additivity of Masking," J. Acoust. Soc. of Am., 85 (3), March 1989.
- [Humes 92] L. E. Humes, L. W. Lee, "Two Experiments on The Spectral Boundary Conditions for Non Linear Additivity of Simultaneous Masking," J. Acoust. Soc. of Am., 92 (5), November 1992.
- [H.231] Recommandation UIT-T H.231, "Unités de commande multipoint pour les systèmes audiovisuels utilisant des canaux numériques fonctionnant à des débits inférieurs ou égaux à 1920 kbits/s," Mars 1996.
- [H.310] Recommandation UIT-T H.310, "Systèmes et terminaux de communication audiovisuels à large bande," Septembre 1998.
- [H.320] Recommandation UIT-T H.320, "Systèmes et équipements terminaux visiophoniques à bande étroite," Mars 1996.
- [H.323] Recommandation UIT-T H.323, "Systèmes de communication multimédia en mode paquet," Septembre 1999.

- [IS 11172-3] IS 11172-3, "Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbit/s - Part 3 : Audio," ISO/IEC, 1993.
- [Jayant 84] N. S Jayant, P. Noll, "Digital Coding of Waveforms : Principles and Applications to Speech and Video," Prentice Hall, 1984.
- [Jbira 99] A. Jbira, "Codage audio hiérarchique à faible retard," Thèse de doctorat de l'ENST de Paris, 1999.
- [Jot 95] J.-M. Jot, V. Larcher and O. Warusfel, "Digital Signal Processing Issues in the Context of Binaural and Transaural Stereophony," AES 98th Convention, 1995 February 25-28 Paris.
- [Jot 97] J.-M. Jot, "Real-Time Spatial Processing of Sounds for Music, Multimedia and Interactive Human-Computer Interfaces". ACM multimedia systems journal (special issue on audio and multimedia) 1997.
- [Kostas 98] T. J. Kostas, M.S. Borella, I. Sidhu, G. M. Schuster, J. Grabiec and J. Mahler, "Real-Time Voice Over Packet-Switched Networks," IEEE Network Magazine, January/February 1998.
- [Koilpillai 92] R. D. Koilpillai, P. P. Vaidyanathan, "Cosine-Modulated FIR Filter Banks Satisfying Perfect Reconstruction," IEEE Trans. On Signal Processing, Vol. 40, No. 4, April 1992.
- [Konstantinides 94] K. Konstantinides, "Fast Subband Filtering in MPEG Audio Coding," IEEE Signal Processing Letters, Vol. 1, No. 2, February 1994.
- [Kim 95] D. Kim and B. G. Lee, "Transform Domain IIR Filtering," IEEE Trans. On Signal Processing, Vol. 43, No. 10, October 1995.
- [Kresch 99] R. Kresch and N. Merhav, "Fast DCT Domain Filtering Using the DCT and the DST," IEEE Trans. On Image Processing, Vol. 8, No. 6, June 1999.
- [Kumar 96] M. Kumar, M. Zubair, "A High Performance Software Implementation of MPEG Audio," IEEE Int. Conf. on Acoust., Speech, Signal Proc., pp. 1049-1052, 1996.
- [Lamblin 88] C. Lamblin, J.-P. Adoul, "Algorithme de quantification vectorielle sphérique à partir du réseau de Gosset d'ordre 8," Annales des Télécommunications, tome 43, n° 3-4, Mars-Avril 1988.
- [Lanciani 99a] C. A. Lanciani and R. W. Schafer, "Subband-Domain Filtering of MPEG Audio Signals," IEEE Int. Conf. on Acoust., Speech, Signal Proc., 1999.
- [Lanciani 99b] C. A. Lanciani, "Compressed-Domain Processing of MPEG Audio Signals," Ph.D. thesis, Georgia Institute of Technology, June 1999.
- [Lee 92] J. B. Lee and B. G. Lee, "Tranform Domain Filtering Based on Pipelining Structure," IEEE Trans. On Signal Processing, Vol. 40, No. 8, August 1992.

- [Le Guyader 00] A. Le Guyader, P. Phillipe, J. B. Rault, "Synthèse de normes de codage de la parole et du son (UIT-T, ETSI et ISO/MPEG)," *Annales des Télécommunications*, tome 55, n° 9-10, Septembre-Octobre 2000.
- [Lelong 99] F. Lelong, S. Megnin, "Architecture fonctionnelle du mixeur MP," Document de travail CNET DSE/SGV, octobre 1999.
- [LeMair 97] M. H. Willebeek-LeMair and Z.-Y. Shae, "Videoconferencing over Packet-Based Networks," *IEEE J. on Selected Areas in Communications*, Vol. 15, No. 6, August 1997.
- [Levine 96] S. N. Levine, "Effects Processing on Audio Subband Data," *ICMC Proceeding*, Hong Kong, 1996.
- [Levine 98] S. N. Levine, "Audio Representation for Data Compression and Compressed Domain Processing," Ph.D. thesis, Stanford University, 1998.
- [Mahieux 89] Y. Mahieux, A. Charbonnier and J.-P. Petit, "Transform Coding Using Correlation Between Successive Blocks," *IEEE Int. Conf. on Acoust., Speech and Signal Proc.*, Glasgow, 1989.
- [Mahieux 94] Y. Mahieux, J. P. Petit, "High- Quality Audio Transform Coding at 64 kbps," *IEEE Trans. on Comm.*, vol. 42, n° 11, November 1994.
- [Malvar 92] H. S. Malvar, "Signal Processing with Lapped Transforms," Artech House, Inc. 1992.
- [Martucci 94] S. A. Martucci, "Symmetric Convolution and the Discrete Sine and Cosine Transforms," *IEEE Trans. On Signal Processing*, Vol. 42, No. 5, May 1994.
- [Mervah 95a] N. Merhav and V. Bhaskaran, "A Fast Algorithm for DCT Domain Filtering," *HPL Technical Report #HPL-95-56*, Hewlett-Packard Labs., May 1995.
- [Mervah 95b] N. Merhav and R. Kresch, "Approximate Convolution Using DCT Coefficient Multipliers," *HPL Technical Report #HPL-95-141*, Hewlett-Packard Labs., 1995.
- [Moreau 95] N. Moreau, "Techniques de compression des signaux," Masson 1995.
- [Morrison 94] D. G. Morrison, M. E. Nilsson and M. Ghanbari, "Reduction of the Bit-Rate of Compressed Video while in its Coded Form," *Proc. 6th Int. Workshop Packet Video*, Portland, pp. D17.1-D17.4, Sept. 1994.
- [Murano 90] K. Murano, S. Unagamic and F. Amano, "Echo Cancellation and application," *IEEE Communication Magazine*, January 1990, pp. 49-54.
- [Nahumi 95] D. Nahumi, "Conferencing Arrangement for Compressed Information," *United States Patent*, US5390177, February 1995.
- [Nakajima 95] Y. Nakajima, H. Hori and T. Kanoh, "Rate Conversion of MPEG coded Video by Re-Quantization Process," *IEEE Int. Conf. on Acoust., Speech and Signal Proc.*, 1995.

- [Nakajima 98] Y. Nakajima, H. Yangihara, A. Yoneyama, M. Sugano, "MPEG Audio Rate Scaling on Coded Data Domain," IEEE Int. Conf. on Acoust., Speech and Signal Proc., 1998.
- [Nakajima 99] Y. Nakajima, Y. Lu, M. Sugano, A. Yoneyama, H. Yanagihara and A. Kurematsu, "A Fast Audio Classification From MPEG Coded Data," IEEE Int. Conf. On Acoust., Speech, Signal Proc., 1999.
- [Nayebi 90] K. Nayebi, T. P. Barnwell III, and M. J. T. Smith, "The Time Domain Analysis and Design of Exactly Reconstructing FIR Analysis / Synthesis Filter Banks," IEEE Int. Conf. on Acoust., Speech, Signal Proc., April 1990.
- [Neubauer 00] Ch. Neubauer and J. Herre, "Audio Watermarking of MPEG-2 AAC Bit Streams," 108th AES Convention 2000 February 19-22, Paris, France.
- [Noll 93] P. Noll, "Wideband Speech and Audio Coding," IEEE Communication Magazine, November 1993.
- [Noll 95] P. Noll, "Digital Audio Coding for Visual Communications," Proc. of the IEEE, Vol. 83, No. 6, June 1995, pp. 925-942.
- [Noll 97] P. Noll, "MPEG Digital Audio Coding," IEEE Signal Processing Magazine, September 1997, pp. 59-81.
- [Okubo 97] S. Okubo, S. Dunstan, G. Morrison, M. Nilsson, H. Radha, D. L. Skran and G. Thom, "ITU-T Standardization of Audiovisual Communication Systems in ATM and LAN Environments," IEEE J. on Selected Areas in Communications, vol. 15, n° 6, August 1997.
- [Oppenheim 99] A. V. Oppenheim, R. W. Shafer and J. R. Buck, "Discrete-Time Signal Processing," Prentice-Hall, 1999.
- [Painter 00] T. Painter and A. Spanias, "Perceptuel Coding of Digital Audio," Proc. of the IEEE, Vol. 88, No. 4, April 2000.
- [Pan 95] D. Pan, "A Tutorial on MPEG/Audio Compression," IEEE Trans. on Multimedia, Vol. 2, No. 2, Summer 1995, pp. 60-74.
- [Perkins 98] C. Perkins, O. Hodson and V. Hardman, "A Survey of Packet Loss Recovery Techniques for Streaming Audio," IEEE Network, September-October 1998.
- [PictureTel] Site web de PictureTel : <http://www.picturetel.com>
- [PT724] PictureTel Corporation, "A Wideband audio coder for multimedia communications with 16kHz sample rate – a detailed description of the PictureTel candidate algorithm," COM16 -E, ITU-T Group 16, Q.20, Santiago, Chile, May 1999.
- [Princen 86] J. P. Princen, A. B. Bradley, "Analysis/Synthesis filter Bank Design Based on Time Domain Aliasing Cancellation," IEEE Trans. on ASSP, Vol. ASSP-34, No. 5, October 1986.

- [Purnahagen 99] H. Purnahagen, "An Overview of MPEG-4 Audio Version 2," AES 17th Conference High Quality Audio Coding, September 2-5 1999, Florence, Italy.
- [P.800] Recommendation UIT-T P.800. "Méthodes d'évaluation subjective de la qualité de transmission - Annexe D," 1996.
- [Ramstad 82] A. Ramstad, "Sub-band Coder with a simple Adaptive Bit-Allocation Algorithm a Possible Candidate for Digital Mobile Telephony," IEEE Int. Conf. on Acoust., Speech, Signal Proc., pp. 203-207, 1982.
- [Rao 96] K. R. Rao and J. J. Hwang, "Techniques and Standards for Image, Video and Audio Coding," Prentice-Hall, 1996.
- [Rothweiler 83] J. H. Rothweiler, "Polyphase Quadrature Filters - A New Subband Coding Technique," IEEE Int. Conf. On Acoust., Speech, Signal Proc., pp. 1280-1283, Boston 1983.
- [Sánchez 95] V. Sánchez, P. García, A. M. Peinado, J. C. Segura and A. J. Rubio, "Diagonalizing Properties of Discrete Cosine Transforms," IEEE Trans. on Signal Processing, Vol. 43, No. 11, November 1995.
- [Schalkwijk 72] J. P. M. Schalkwijk, "An algorithm for Source Coding," IEEE Trans. On Information Theory, Vol. IT-18, No. 3, May 1972.
- [Schaphorst 99] R. Schaphorst, "Videoconferencing and Videotelephony : Technology and Standards," 2nd edition, Artech House, Norwood, 1999.
- [Scheirer 99] E. D. Scheirer, R. Väänänen and J. Huopaniemi, "AudioBIFS: Describing Audio Scenes with the MPEG-4 Multimedia Standard," IEEE Trans. on Multimedia, Vol. 1, No. 3, September 1999.
- [Scheirer 00] E. D. Scheirer, Y. Lee and J.-W. Yang, "Synthetic and SNHC Audio in MPEG-4," Signal Processing: Image Communications 15 (4-5) (2000) pp. 445-461.
- [Schmitt 98] J.-C. Schmitt, J.-P. Jullien, I. Haigneré and N. Château, "Paramètres techniques pour la qualité audiovisuelle," L'écho des recherches, n° 171, 2^{ème} trimestre 1998, CNET/ENST.
- [Shibata 99] Y. Shibata, Z. Chen and R.-H. Campbell, "A Fast Degradation-Free Algorithm for DCT Block Extraction in the Compressed Domain," IEEE Int. Conf. on Acoust. Speech, Signal Proc. 1999, Phoenix, USA.
- [Shlien 97] S. Shlien, "The Modulated Lapped Transform, Its Time-varying Forms, and Its Applications to Audio Coding Standards," IEEE Trans. On Speech and Audio Processing, Vol. 5, No. 4, July 1997.
- [Taddei 99] H. Taddei, "Codage hiérarchique faible retard 8-14.4-24 kbit/s pour les nouveaux réseaux et services," Thèse de doctorat de l'université de Rennes 1, 1999.

- [Taka 88] M. Taka, S. Shimada and T. Aoyama, "Multimedia Multipoint Teleconference System Using the 7 kHz Audio Coding Standard at 64 kbits/s," *IEEE J. on Selected Areas in Communications*, Vol. 6, No. 2, February 1988.
- [Thom 96] G. A. Thom, "H.323: The Multimedia Communications Standard for Local Area Networks," *IEEE Communications Magazine*, December 1996.
- [Tzanetakis 00] G. Tzanetakis, P. Cook, "Sound Analysis Using MPEG Compressed Audio," *Int. Conf. On Acoust., Speech, Signal Proc.*, 2000.
- [Vaidyanathan 88] P. P. Vaidyanathan and S. K. Mitra, "Polyphase Networks, Block Digital Filtering, LPTV Systems, and Alias-Free QMF Banks: A Unified Approach Based on Pseudocirculants," *IEEE Trans. On ASSP*, Vol. 36, No. 3, March 1988.
- [Vaidyanathan 93] P. P. Vaidyanathan, "Multirate Systems and Filter Banks," Prentice Hall, Englewood Cliffs, NJ, 1993.
- [Väänänen 00] R. Väänänen, "Synthetic Audio Tools in MPEG-4 Standard," 108th AES Convention 2000 February 19-22, Paris, France.
- [Venugopal 99] S. Venugopal, K. R. Ramakrishnan, S. H. Srinivas and N. Balakrishnan, "Audio Scene Analysis and Scene Change Detection in the MPEG Compressed Domain," *IEEE 3rd Workshop on Multimedia Signal Processing*, pp. 191-196, September 13-15, 1999, Copenhagen, Denmark.
- [Wei 97] X. Wei, M. J. Shaw and R. Varley, "Optimum Bit Allocation and Decomposition for High Quality Audio Coding," *IEEE Int. Conf. on Acoust. Speech and Signal Proc.*, pp. 315-318, 1997.
- [Werner 99] O. Werner, "Requantization for Transcoding of MPEG-2 Intraframes," *IEEE Trans. On Image Processing*, Vol. 8, No. 2, February 1999.
- [Yu 99] H. H. Yu, "Fast Image/video Retrieval on Compressed Image and Video Databases," *IEEE 3rd Workshop on Multimedia Signal Processing*, pp. 129-134, September 13-15, 1999, Copenhagen, Denmark.
- [Zhu 98] W.-Zhu., K. H. Yang, and M. J. Beacken, "CIF-to-QCIF Video Bitstream Down-Conversion in the DCT Domain," *Bell Labs Technical Journal*, July-September 1998.
- [Zhu 99] Q.-F. Zhu, L. Kerofsky and M. B. Garrison, "Low-Delay, Low-Complexity Rate Reduction and Continuous Presence for Multipoint Videoconferencing," *IEEE Trans. On Circuits and Systems for Video Technology*, Vol. 9, No. 4, June 1999.
- [Zwicker 99] E. Zwicker, H. Fastl, "Psychoacoustics - Facts and Models," Springer, 1999.

Bibliographie personnelle

- [1] A. Benjelloun Touimi, Y. Mahieux, "Un algorithme de sommation des signaux audio codés MPEG-1 dans le domaine compressé," Journées CORESA, Sophia-Antipolis, 14-15 juin, 1999.
- [2] A. Benjelloun Touimi, Y. Mahieux, C. Lamblin, "Procédé de traitement d'au moins un flux binaire audio codé organisé sous la forme de trames," Brevet Français n° 99 15574, 8 décembre 1999.
- [3] A. Benjelloun Touimi, Y. Mahieux, C. A. Lanciani, "A Summation Algorithm for MPEG-1 Coded Audio Signals: a First Step Towards Audio Processing in the Compressed Domain," Annales des Télécommunications, tome 55, n° 3-4, Mars-Avril 2000.
- [4] A. Benjelloun Touimi, "A Generic Framework for Filtering in Subband Domain," 9th IEEE Workshop on Digital Signal Processing, Hunt, Texas, USA, October 2000.
- [5] A. Benjelloun Touimi, C. Boisseau, C. Lamblin, F. Bouteille, D. Deleam, "Méthode et dispositif de traitement d'une pluralité de flux binaires audio," Brevet Français n° 01 01662, 2 février 2001.