UNIVERSITÉ DE SHERBROOKE
Faculté des sciences appliquées
Département de génie électrique et de génie informatique

CODAGE LARGE BANDE DE LA PAROLE PAR ENCAPSULATION DU CODEUR ITU G.729 (CS-ACELP)

Mémoire de maîtrise es sciences appliquées
Spécialité: génie informatique

Romain TRILLING

Sherbrooke (Québec), Canada Août 1998

Résumé

Les technologies modernes en codage numérique de la parole ont atteint un niveau de qualité qui permet de répondre à un grand nombre de besoins. Les communications téléphoniques en bande réduite peuvent ainsi être traitées sans difficulté avec un débit de 8 Kb/s (G.729). Le codage audio de haute qualité (CD) peut être compressé à des débits aussi faibles que 64 Kb/s. À un niveau intermédiaire, le codage large bande est satisfaisant pour un débit de 16 à 24 Kb/s.

Le développement des applications multimédia sur l'internet ainsi que les systèmes de conférence téléphonique feraient bon usage d'un système adaptatif permettant de régler le niveau de qualité du codage selon le débit disponible. Cette étude propose une solution destinée à répondre à ce besoin. Le projet qui va être décrit présente un système de codage encastré permettant d'offrir deux niveaux de qualité bande étroite / bande réduite pour les transmissions de parole. On utilise pour cela un codeur déjà normalisé, soit la norme G.729, que l'on cherche à encapsuler en un codeur large bande. Le débit du codeur de moins bonne qualité est celui du G.729. Pour la qualité supérieure on reprend le débit de départ que l'on complète à 16 Kb/s à l'aide d'un second canal à 8 Kb/s.

Remerciements

Je tiens avant tout à exprimer mes sincères remerciements envers mon directeur de recherche Jean-Pierre Adoul pour son encadrement, son enseignement et pour m'avoir permis de compléter le programme de maîtrise à l'université de Sherbrooke.

Je souhaite également remercier Roch Lefebvre et l'ensemble du groupe de codage de l'université de Sherbrooke pour les explications et conseils qui m'ont été donnés tout au cours de cette étude.

TABLE DES MATIÈRES

TABLE DES FIGURES

LISTE DES TABLEAUX

Introduction

Ce document propose une solution visant à construire un codeur large bande par encapsulation du codeur UIT G.729. L'intérêt d'un tel dispositif est de pouvoir proposer deux niveaux de qualité à partir d'un même codeur. Le principe de fonctionnement est décrit par la figure 0.1.

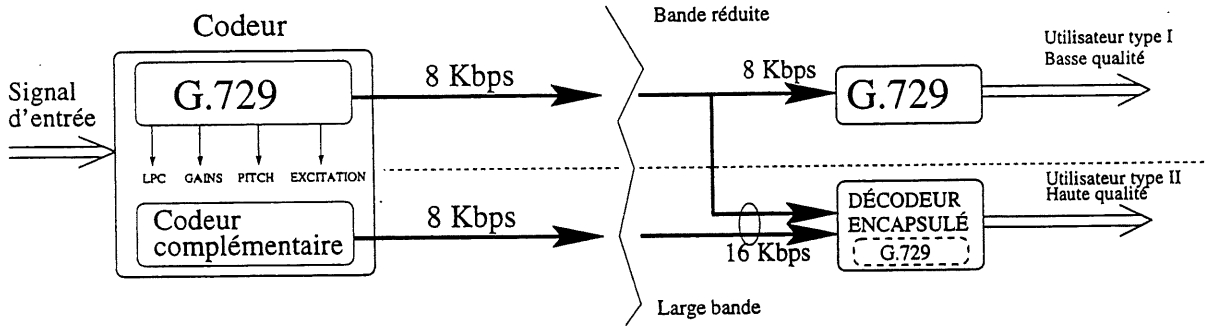

Figure 0.1 - *Dispositif d'encapsulation du G.729*

Le signal sonore est traité par un bi-codeur qui comprend un G.729A et un codeur additionnel qui a pour tâche de compléter le G.729A. Ce système produit deux trains de bits pouvant être transmis sur deux canaux à 8 Kb/s. En bout de ligne, l'usager (utilisation de type I) reçoit toujours les données du premier canal à 8 Kb/s (G.729A) lui permettant la reconstruction d'un signal de synthèse dans la bande téléphonique (200 Hz - 3400 Hz). Le débit additionnel (second canal à 8 Kb/s) n'est reçue que par les usagers qui ont accès à un

niveau de connexion privilégié (utilisation de type II). Pour ces derniers, il est possible de synthétiser un signal large bande (80 Hz - 5500 Hz) de meilleure qualité.

A titre d'exemple, voici deux contextes d'application du codage étudié:

* *Broadcasting à deux niveaux* :

On considère un serveur qui a pour tâche d'émettre un signal sonore sur deux canaux à 8 Kb/s (décris par la figure 0.1). Cette information est distribuée sur l'ensemble du réseau. Les usagers peuvent alors selon leur niveau d'accès recevoir sur un seul canal ou sur les deux canaux. Un tel type d'application étant uni-directionelle (le serveur ne reçoit pas de réponse des usagers), on peut être tolérant au niveau de l'encodeur en ce qui concerne le délai et la complexité dans la mesure où seul le serveur est concerné.

* DSVD à débits variables: cas d'une communication de type « Point à Point ».

Considérons une communication de type DSVD (*Digital Simultanenous Voice and Data*) construite à partir d'un codeur G.729A. Un tel système pourrait vérifier la norme UIT V.70 pour laquelle le G.729A est adapté (la norme V.70 pour le DSVD se base sur des codeurs de parole dont le débit est multiple de 8 Kb/s).

Si la communication est faite à partir du réseau téléphonique commuté à l'aide de modems V.34b, on dispose d'un débit utile de l'ordre de 33 Kb/s. Ce débit est à partager entre les communications vocales et numériques. Pour le moment, le débit utilisé pour la transmission de la parole est d'au plus 16 Kb/s (G.729 en *Full Duplex*). Si aucune autre transmission de données n'est à considérer, il reste un débit d'au moins 16 Kb/s non utilisé. Un système comprenant un codeur de parole avec plusieurs modes permettrait d'offrir une meilleur utilisation du débit disponible. L'intérêt de bâtir un tel dispositif à partir d'un système encapsulé serait de pouvoir rester compatible avec du matériel qui n'utiliserait que le G.729A.

Les principaux scénarios à envisager sont exposés dans le tableau 0.1 (les débits, approximatifs, sont donnés à titre indicatif):

Scénarios	Débits (33 Kb/s au total)	
	Parole	Données
Transferts de données uniquement	0 Kb/s	33 Kb/s
Téléphonie et données	16 Kb/s (deux sens en bande réduite)	17 Kb/s
	24 Kb/s (un sens en bande réduite + un sens en large bande)	9 Kb/s
Téléphonie seulement	32 Kb/s (deux sens large bande)	1 Kb/s

TABLEAU 0.1 - *Divers modes de répartition du débit d'un modem (33 Kb/s) entre les transferts de données et la téléphonie numérique*

La suite de cet exposé présente tout d'abord quelques généralités sur les signaux large bande et les signaux en bande réduite. Après un résumé sur les principaux aspects de la norme G.729, suivra une description de la solution proposée.

Chapitre 1

Bande téléphonique et large bande

Ce chapitre présente quelques rappels et résultats concernant le codage large bande et en bande étroite. Après une petite étude sur les signaux (parole et musique) dans les deux bandes de fréquences, un état de l'art en codage dans les deux domaines sera donné.

1.1 Introduction: bande d'audition

L'oreille ne peut percevoir que certains sons. La figure 1.1 donne une représentation du domaine audible pour un être humain. On remarque tout d'abord que le niveau de perception dépend grandement de la plage de fréquences considérée ainsi que du niveau sonore. On définit alors deux courbes dans le plan fréquence/intensité : un seuil d'audibilité et un seuil de confort. La zone ainsi définie est le domaine dans lequel les sons peuvent être perçus. Tout signal en dehors de cette plage est inaudible, gênant ou même dangereux.

La bande d'audition est composée des fréquences de 20 Hz à 20 000 Hz. En pratique une telle largeur de bande n'est conservée que pour un codage de très haute fidélité (qualité CD).

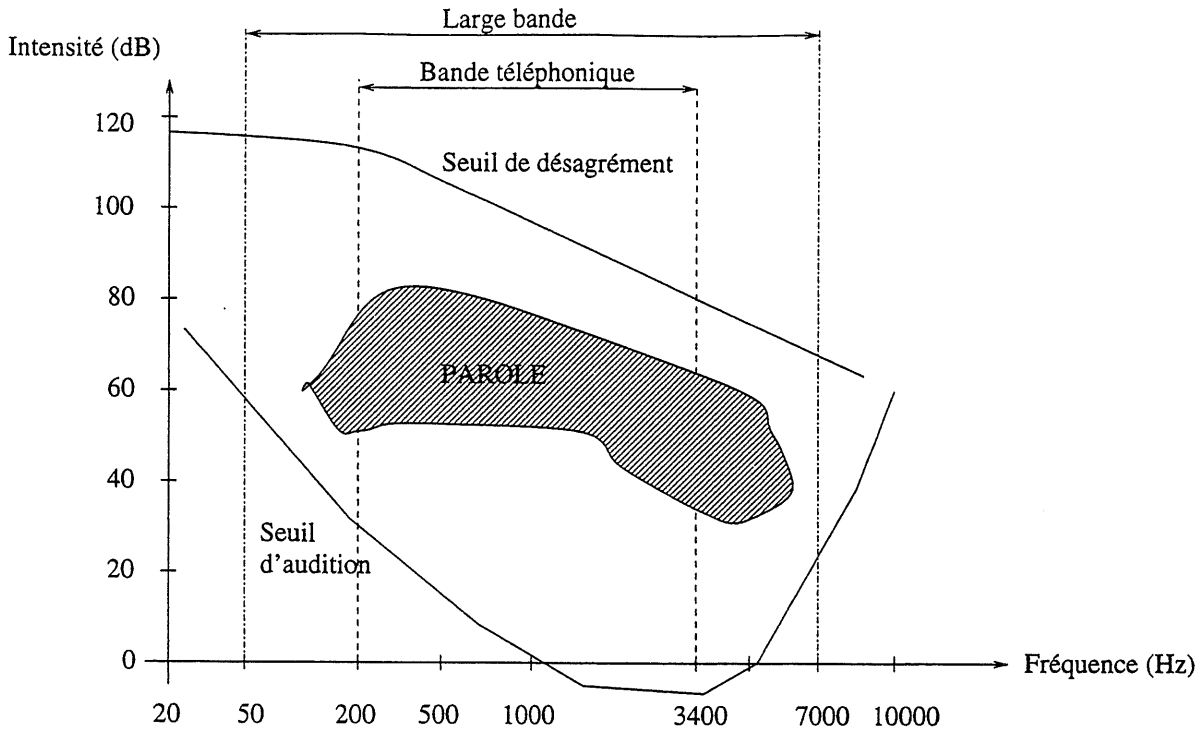

Figure 1.1 - *Perception auditive*

Selon la nature du signal à coder (parole ou musique) on filtre le signal en sélectionnant soit la bande téléphonique, suffisante pour la parole, soit une bande plus large pour traiter des sons plus complexes.

1.2 Codage dans la bande téléphonique

La bande téléphonique correspond à la plage de fréquences 200 Hz - 3400 Hz. En pratique le filtrage IRS pour sélectionner cette bande déborde légèrement des deux côtés. On peut ainsi obtenir un spectre élargi à 140 Hz - 3500 Hz.

Le codage en bande réduite concerne essentiellement les signaux de parole. La bande étroite est juste suffisante pour conserver l'intelligibilté du langage ainsi que les paramètres propres au locuteur (voix, émotion ...). La plage de 200 Hz - 3400 Hz permet de positionner correctement les premiers formants (trois ou quatre) et de contenir toutes les premières harmoniques du pitch. Une transmission de parole en bande réduite est en mesure de communiquer fidèlement les phonèmes voisés et de donner une impression de bruit pour les phonèmes non voisés. Des considérations plus précises concernant la nature des phonèmes seront apportées dans l'étude du large bande.

Les signaux de musique peuvent aussi être quantifiés en bande étroite mais l'intérêt d'une telle manipulation est nettement moins grand que pour la parole. En effet, dans le cas de la musique la perte ou la trop grande atténuation des basses fréquences dégrade beaucoup la perception que l'on a d'un passage musical. Par ailleurs, la suppression des composantes hautes fréquences annule tout simplement une partie de l'information en privant l'auditeur de certaines notes aïgues ou bruit de confort hautes fréquences.

Les codeurs de parole en bande réduite:

Les codeurs actuellement utilisés en téléphonie numérique sont des codeurs de types CELP (*Code Excited Linear Prediction*). Différentes variantes de cette technologie sont en mesure de fournir des résultats comparables en terme de qualité et de débit. Le tableau 1.1 donne quelques exemples de codeurs standardisés au cours des cinq à dix dernières années.

Normes	G.723.1	G.729	GSM		
			Full Rate	*Half Rate*	*EFR*
Date	1995	1995	1987	1994	1996
Technologie	MP-MLQ/ACELP	CS-ACELP	RPE-LTP	VSELP	ACELP
Débit (Kb/s)	5.3 / 6.3	8	13	5.6	12.2
Délai	30 ms + 7.5 ms	10 ms + 5 ms	20 ms	20 ms + 5 ms	20 ms

TABLEAU 1.1 - *Normes actuelles en codage de parole bande réduite*

MP-MLQ : *Multip-Pulse Maximum Likehook Quantization*
GSM : *Global System for Mobile communications*
RPE : *Regular Pulse Excitation with Long-Term Prediction*

La plupart des codeurs comprennent trois couches: prédiction linéaire, prédiction à long terme et codage du code d'excitation. Parmi les technologies CELP les plus connues ont trouve les codeurs de CELP (dictionnaires stochastiques à grande complexité), ACELP (*Algebraic Code Excited Linear Prediction*) et VSELP (*Vector Sum Excited Linear Prediction*). Ces codeurs permettent des transmissions à des débit pouvant descendre à des valeurs aussi faibles que 5 Kb/s tout en maintenant une qualité d'assez grande fidélité..

1.3 Codage large bande

Le codage large bande considère théoriquement aussi bien les signaux de parole que les signaux de musique. Les techniques de codage peuvent alors différer selon le type d'application du codeur.

1.3.1 Codage de la parole large bande

L'agrandissement de la bande passante pour la parole n'apporte pas grand chose du point de vu de l'information. Contrairement à la musique, où la bande élargie peut comporter des phénomènes supplémentaires (notes aïgues), l'information de parole (l'intelligibilité) est intégralement contenue dans la bande téléphonique (sauf peut être pour quelques langues très exotiques). On peut néansmoins espérer deux améliorations :

- Pour les phonèmes voisés, l'addition de la bande de fréquences 50 Hz - 140 Hz donne une meilleur représentation des premières harmoniques. On remarque surtout cela pour un locuteur masculin pour lequel la fréquence de pitch est assez faible. D'une manière plus générale les basses fréquences procurent une sensation de confort et un sentiment de parler « face à face ».

- L'apport des hautes fréquences, supérieures à 3400 Hz, n'a de l'importance que pour les phonèmes plus complexes tels que les fricatives non voisées (ex: « S », « CH », « F »), les fricatives voisées (ex: « Z ») ou encore les plosives (ex: « T », « D »).

La figure 1.2 représente un sonogramme pour une phrase prononcée par un locuteur masculin. Cet extrait a été sélectionné parce qu'il est riche en phonèmes non-voisés. Ce graphique est découpé en deux morceaux au niveau de l'axe $f = 3400Hz$ afin de mettre en évidence la plage hautes fréquences qui est exclue lorsque que l'on ne considère que la bande téléphonique. On remarque que la zone supérieure du sonogramme ne présente de l'énergie que pour les phonèmes non-voisés. Pour les sons voisés l'énergie en hautes fréquences est négligeable en comparaison à celle contenue dans les formants. Les figures 1.3 et 1.4 sont deux exemples de phonèmes respectivement voisé et non-voisé. On remarque bien que la distribution de l'énergie pour les deux situations est très différente.

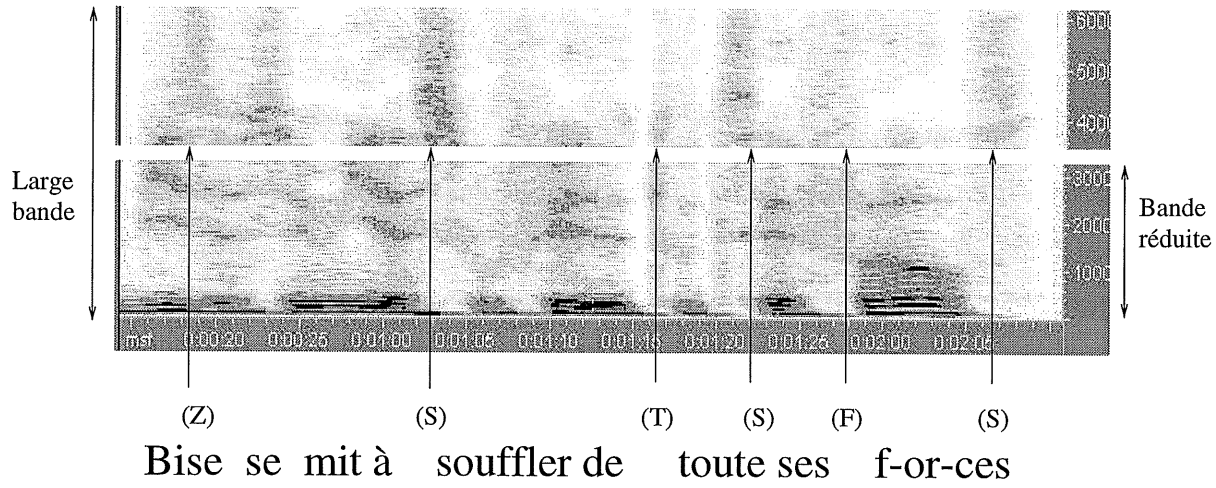

Figure 1.2 - *Sonogramme de "bise se mit à souffler de toutes ses forces"*

Un codeur large bande optimisé pour la parole pourrait prendre soin de bien représenter les formants ainsi que la structure harmonique en limitant le codage des hautes fréquences lorsque le son est voisé. En revanche, un débit plus conséquent pourrait être attribué à la partie supérieure du spectre lorsque le phonème est non-voisé ou composé.

1.3.2 Codage de la musique en large bande

La musique n'est intéressante à coder que lorsque l'on dispose d'une largeur de bande suffisante. Le codage large bande permet d'offrir une telle qualité. Contrairement à la parole, il n'existe pas réellement de modèle permettant de représenter le signal. En revanche les sons en musique sont beaucoup plus stationnaire que les phonèmes en parole. Pour cette raison, on est porté à utiliser des trames d'analyse plus grandes qu'en parole. En pratique on peut travailler avec des blocs d'au moins 20 ms.

Comme il vient d'être précisé, il est difficile de prévoir à l'avance l'allure de l'enveloppe

spectrale. On peut cependant admettre que la musique est une combinaison de bruit et de "tons". Un ton pur est une concentration d'énergie sur une raie spectrale donnée, avec un plancher de bruit faible. A titre d'exemple, une note de musique isolée conduit à un spectre comprenant seulement un ton pur, localisé à la fréquence de la note.

Lorsque le signal devient trop complexe, on l'assimile à un bruit coloré. Les figures 1.5 et 1.6 représentent deux exemples de passages de musique. La figure 1.5 correspond à une situation ou le spectre devient rapidement plat lorsque l'on monte en fréquences. On peut ici séparer le signal en deux morceaux: de l'information en basses fréquences et du bruit de confort en hautes fréquences. La figure 1.6 quand à elle, met en évidence un passage de musique beaucoup plus tonal.

1.3.3 Les codeurs large bande

Idéalement un codeur large bande doit pouvoir traiter sans préférences aussi bien les sources de parole que celles de musique. Deux approches sont alors possibles: soit on améliore un codeur de parole (type ACELP par exemple) pour qu'il traite au mieux la musique, soit on part d'un codeur mieux conçu pour la musique (codage par transformée type TCX par exemple) que l'on adapte afin de mieux coder la parole. La seconde stratégie semble plus prometteuse dans la mesure où un codeur ACELP est conçu presque exclusivement pour la parole (pour la musique, le débit consacré au pitch est parfois du gaspillage) tandis qu'un codeur par transformée assure toujours une contribution minimale quelle que soit la source considérée.

Il n'existe pas encore beaucoup de normes en large bande pour le moment. La référence à considérer est encore la norme UIT G.722. Ce codeur est un codeur de forme d'onde temporelle de type ADPCM. Il utilise un débit de 48 Kb/s à 64 Kb/s. Une seconde norme devrait bientôt pouvoir remplacer ce dernier. Les nouveaux débits à considérer seront probablement

16 Kb/s à 32 Kb/s. La technique de codage utilisée est cette fois un codage par transformée MLT avec un codage entropique sur les indices de quantification en bout de ligne (un peu comme MPEG1 layer 3). Le tableau 1.2 donne une brève description des deux codeurs large bande qui viennent d'être évoqués.

Codeur	G.722	Nouveau large bande G.7XX
Année	1988	1998
Débit	3 modes: 48 Kb/s, 56 Kb/s et 64 Kb/s	3 modes: 16 Kb/s, 24 Kb/s et 32 Kb/s
Délai	0.125 ms (+ 1.5 ms lookahead)	20 ms (+ 20 ms lookahead)
Modèle	* Codage en deux sous-bandes (QMF) * Codage ADPCM d'ordre 4 dans chacune des bandes * Hautes fréquences quantifiées à 2 bits par échantillon * Basses fréquences quantifées à 3 bits, 4 bits ou 5 bits par échantillon selon le débit choisi, les trois modes étant encapsulés	* Codage par transformée MTL * Quantification scalaire des raies de la transformée * Attribution du budget par catégorisation selon les bandes de fréquence. * Codage entropique (huffman) sur les indices de quantification

TABLEAU 1.2 - *Codeurs large bande*

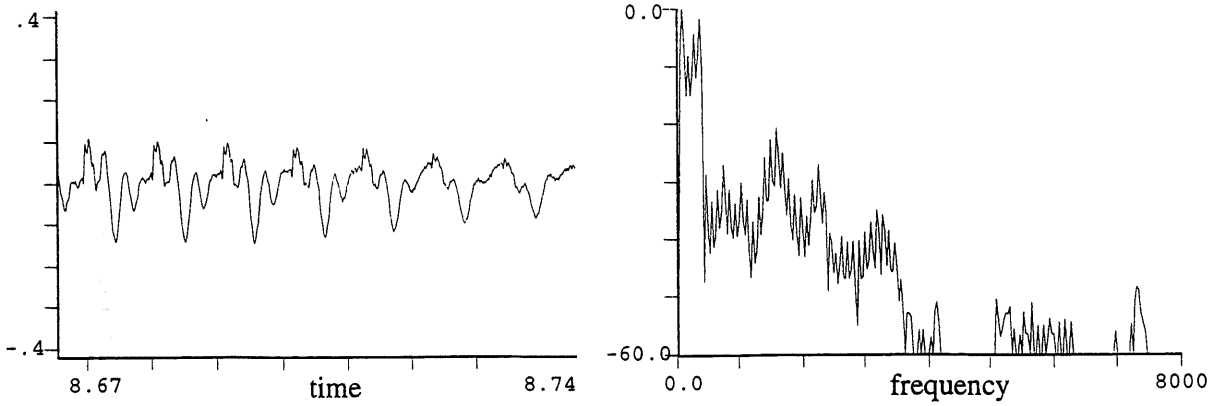

Figure 1.3 - *phonème voisé*

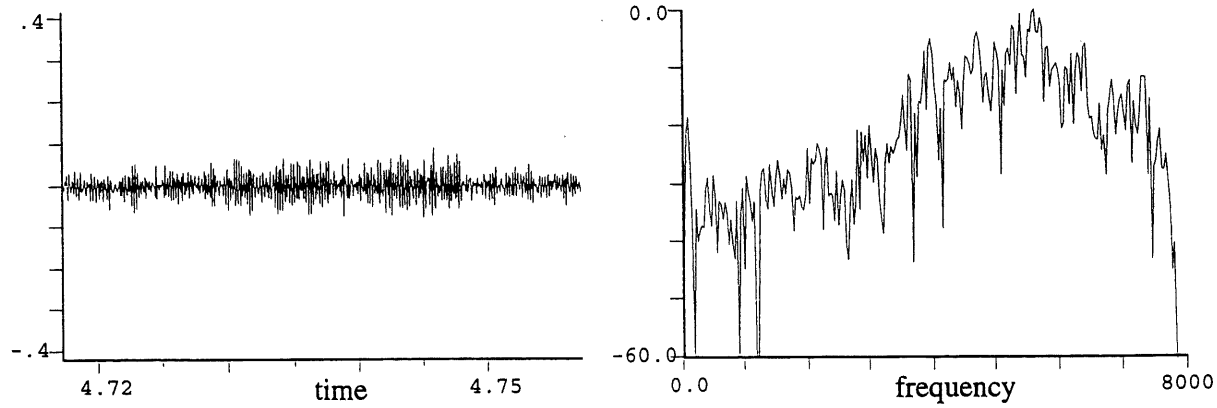

Figure 1.4 - *phonème non voisé*

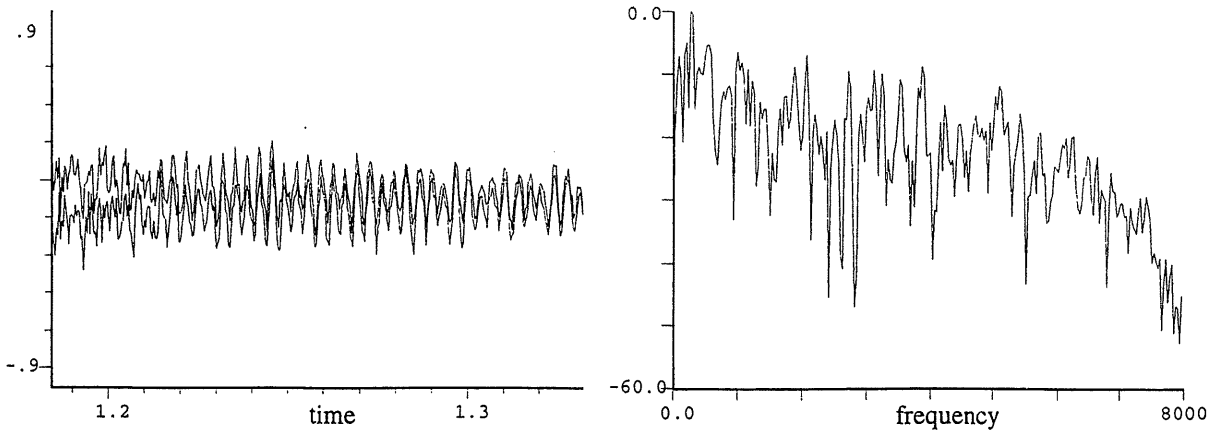

Figure 1.5 - *Musique: bruit en hautes fréquences*

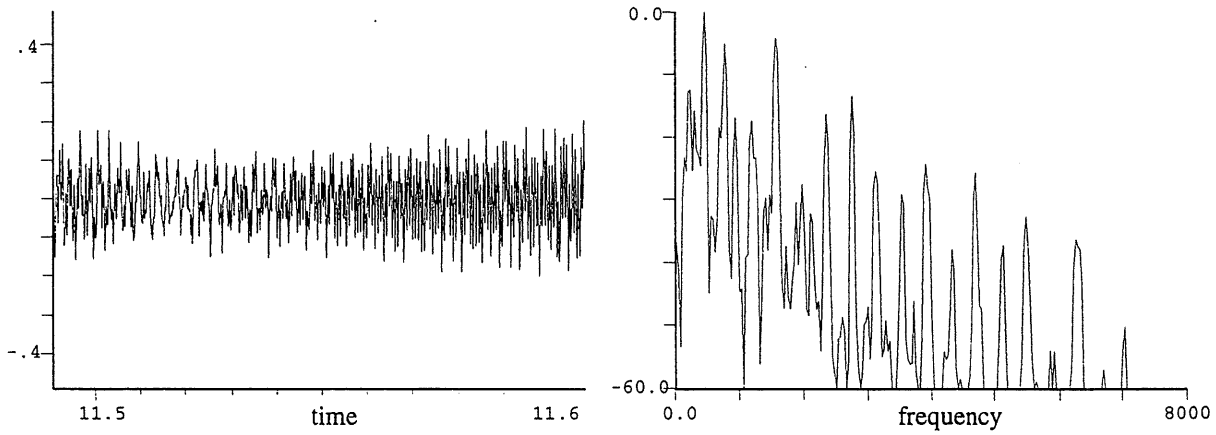

Figure 1.6 - *Musique: tons en hautes fréquences*

Chapitre 2

Présentation de la norme G.729

2.1 La norme G.729 et ses extensions

2.1.1 Le G.729 (CS-ACELP)

Le G.729 ou CS-ACELP (*Conjugate Structure and Algebraic-Code Excited Linear Prediction*) est une norme de codage numérique de la parole qui a été approuvé à l'UIT en novembre 1995. Cette norme permet de coder la parole avec un débit de 8 Kb/s en conservant une qualité de grande fidélité (comparable à celle du G.726 ADPCM 32 Kb/s [4]).

Ce codeur représente un bon compromis en termes de délai (retard de codage), débit, qualité et robustesse (résistance à d'éventuelles dégradations de transmission). La qualité du codeur est optimale pour une bande de fréquences de 100 Hz - 3500 Hz.

Les applications principales du G.729 sont dans la téléphonie numérique, la visio-conférence et les applications sur réseau qui font usage d'un codeur de parole (ex: terminaux à commandes vocales).

2.1.2 Les annexes du G.729

Outre la norme originale, un certain nombre d'annexes sont apparues. Leur intérêt est de pouvoir apporter une amélioration sur un aspect particulier du codeur (débit ou complexité).

Annexe A

Cette extension comprend une assez grande réduction en complexité, par rapport à la norme originale, pour une légère dégradation du signal de synthèse. La charge CPU est diminuée de 50% pour une qualité qui reste très proche du *toll-quality.*

Le G.729A est compatible bit à bit avec le G.729. Cette annexe à été pensée à l'origine pour répondre aux exigences des transmissions de type DSVD (*Digital Simultaneous Voice and Data*) où on envisage de transmettre simultanément de la parole et des données numériques. Contrairement à la norme originale, l'annexe A permet un encodage de la parole en temps réel (même en version de type flottant) sur la quasi-totalité des processeurs pentium. A titre de référence, le codage temps réel sur un P100 occupe un peu plus de 50% du processeur.

L'annexe A du G.729 est la version du codeur qui a été utilisée dans ce projet.

Annexe B

L'extension B vise à réduire le débit moyen. On utilise des techniques VAD (*Voice Activity Detector*) et CNG (*Comfort Noise Generator*). Le principe est de détecter au mieux si un interlocuteur parle ou non. On choisit soit d'encoder le signal d'entrée, soit d'envoyer un bruit de confort. Sachant que pour une communication téléphonique un interlocuteur parle environ 50% du temps, on espère réduire le débit moyen à une valeur proche de 6.5 Kbit/s. Les références [11] et [7] procurent davantage de détails concernant cette annexe.

Autres annexes: C, D et E

L'annexe C comprend les versions flottant des codeurs G.729 et G.729A
Les annexes D et E sont des versions du codeur à débit inférieur (6.4 Kb/s pour l'annexe D) ou supérieur (11.8 Kb/s pour l'annexe E)

Résumé des caractéristiques du codeur

Les spécifications générales du codeur sont données par le tableau 2.1. Les données sont fournies pour la norme originale (G.729) ainsi que pour l'annexe A (G.729A). Les objectifs visés pour le système encapsulé que l'on étudie sont également énoncés.

	G.729	G.729A		encapsulation	G.729A encapsulé
Débit	8 Kb/s	8 Kb/s	+	8 Kb/s	16 Kb/s
Délai	15 ms	15 ms	+	10 ms max	25 ms max
Complexité	20 - 25 MIPS	10 MIPS	+	5 MIPS max	15 MIPS max

TABLEAU 2.1 - *Spécifications du G.729, de l'annexe A et du codeur encapsulé*

2.2 L'algorithme de codage: une vue globale

Le G.729 fait partie de la famille des codeurs ACELP (*Algebraic Code-Excited Linear-Prediction*). Une vision encore assez générale de l'algorithme de codage est donnée par le diagramme de la figure 2.1. L'algorithme peut être décomposé en trois parties: la prédiction linéaire, l'analyse du pitch et la recherche du code d'excitation. Le codeur travail sur des trames de 10 ms que l'on découpe ensuite en deux sous-trames de 5 ms. Un lookahead de 5 ms est utilisé lors de l'analyse LPC.

2.2.1 Le codeur

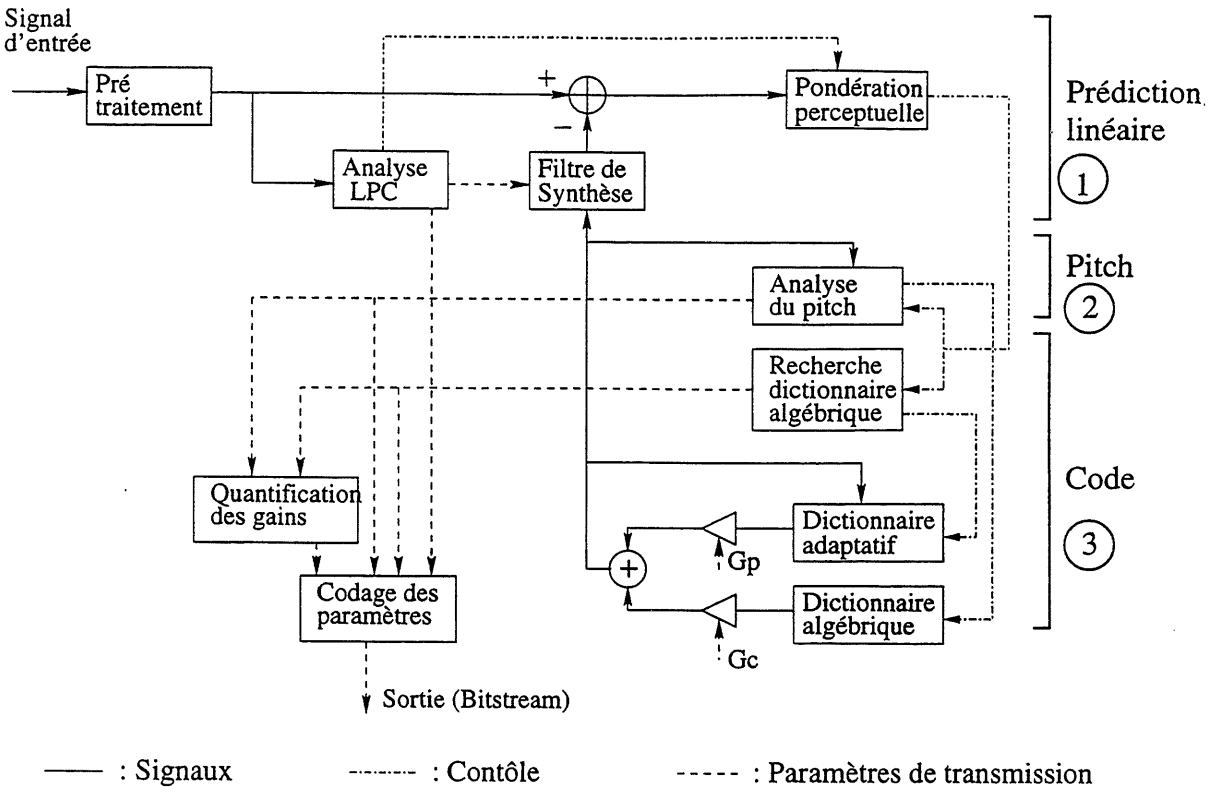

Figure 2.1 - *Diagramme du codeur G.729*

La prédiction linéaire

Le calcul des coefficients de prédiction est effectué toutes les 10 ms. Un filtre LPC d'ordre 10 est déterminé par l'algorithme de Levinson-Durbin. Pour le calcul des coefficients de corrélation on utilise une fenêtre d'analyse (figure 2.2) de 30 ms comprenant 15 ms du signal passé et 5 ms de lookahead. Les paramètres du filtre sont quantifiés sous forme de coefficients LSF (*Line Spectrum Frequencies* [2]). Ces mêmes paires de LSF sont utilisées pour interpoler le filtre de prédiction: on obtient ainsi une mise à jours du filtre sur chacune des sous-trames de 5ms.

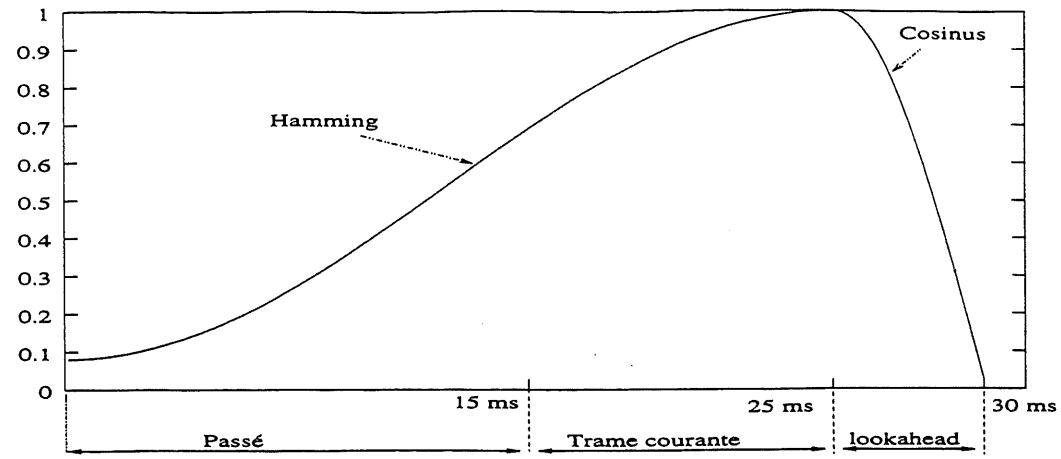

Figure 2.2 - *Fenêtrage pour l'analyse LPC*

L'analyse de pitch

La recherche du pitch est faite de façon très minutieuse, en deux étapes:

- Détermination du pitch en boucle ouverte.

  Cette recherche vise à donner une première approximation du pitch. Elle se fait en boucle ouverte sur la trame courante, pondérée par un filtre perceptuel W(z). La valeur cherchée est le décalage qui maximise la corrélation entre le signal et ses versions décalées. On essaie par ailleurs de favoriser les valeurs faibles du délai afin d'éviter de choisir un multiple du pitch.

- Détermination du pitch en boucle fermée

  On établit pour chaque sous-trame (5ms) une recherche autour de l'évaluation en boucle ouverte. On détermine la nouvelle valeur du délai en minimisant l'erreur quadratique dans un domaine perceptuel entre la cible et le signal reconstruit à partir du passé. Pour ce faire, on cherche à maximiser la cross-corrélation entre la cible et la synthèse.

Le délai de pitch est déterminé à une fraction $\frac{1}{3}$ près. La partie fractionnaire est calculée par une recherche d'extrema sur une version interpolée des cross-corrélations.

Recherche du code d'excitation

Le G.729 (CS-ACELP) comporte deux dictionnaires. Chacun de ces dictionnaires a une taille de 40 échantillons soit 5 ms.

* Un dictionnaire adaptatif:

Le mot choisit est une portion du passé du signal de synthèse, le décalage étant égal au délai de pitch. Ce dictionnaire représente la contribution du pitch

* Un dictionnaire fixe et algébrique: codage de l'excitation à bas débit.

On utilise un dictionnaire ISPP (*Interleaved Single-Pulse Permutations*) à 4 pulses. Chaque sous-trame de 40 échantillons est décomposée en 5 *tracks* (figure 2.3) numérotées de 0 à 4.

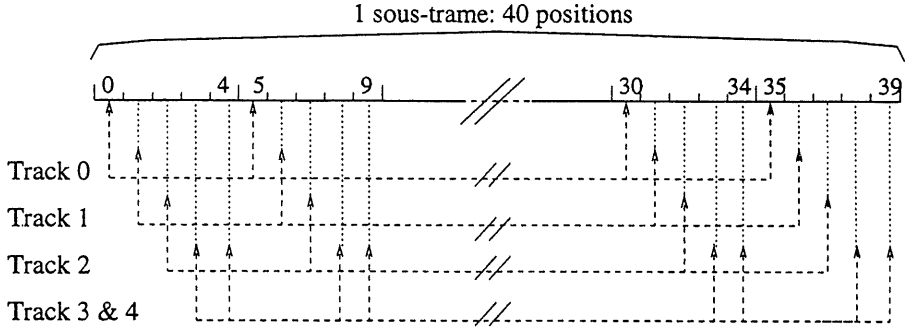

Figure 2.3 - *Positions des pulses pour le dictionnaire fixe*

On positionne un pulse dans chacune des *track* 0 à 2 et un pulse dans une des deux *track* 3 et 4. Les modules des pulses sont indépendants et peuvent prendre comme valeurs +1 ou -1.

L'algorithme permettant de positionner les 4 pulses est celui utilisé dans les codeurs ACELP. On minimise l'erreur quadratique dans un domaine perceptuel entre la cible (résidu dont on a retiré la contribution du pitch) et la synthèse. On positionne les pulses un par un à l'aide d'un système de quatres boucle emboitées. On cherche à maximiser le critère suivant:

$$\frac{C_k^2}{\alpha_k} \tag{2.1}$$

C_k représente la corrélation entre la cible et la synthèse et α_k est l'énergie du mot de code essayé. Pour les calculs de C_k et α_k les signaux sont pris dans le domaine perceptuel (après filtrage par $W(z)$). La structure particulière du dictionnaire permet un calcul rapide du critère lors de chaque passage dans la boucle. Lors de la construction d'un mot de code, après chaque positionnement on met à jours C_k et α_k en ajoutant la contribution apportée par le nouveau pulse.

2.2.2 Les paramètres transmis - train de binaire

Pour une trame de 80 échantillons, on transmet 80 bits. Le train de bits comprend l'encodage des paramètres suivants: coefficients LSF, délai de pitch, code d'excitation ainsi que les gains pour les dictionnaires. Une description du train de bits est donnée par le tableau 2.2:

2.2.3 Le décodeur

Le diagramme de l'algorithme de décodage du train de bits et de génération du signal de synthèse est donné par la figure 2.4.

On retrouve la partie synthèse du codeur. On forme à partir du code d'excitation le signal de synthèse par prédiction linéaire. On effectue ensuite un post filtrage de la sortie. Ce post-

Paramètres	sous-trame 1	sous-trame 2	total trame
LSF	18		18
délai de pitch	8	5	13
délai de pitch : parité	1	0	1
Dictionnaire algébrique: indices	13	13	26
Dictionnaire algébrique: signes	4	4	8
Gains (deux dictionnaires)	7	7	14
total			80

TABLEAU 2.2 - *Composition du train de bits pour une trame*

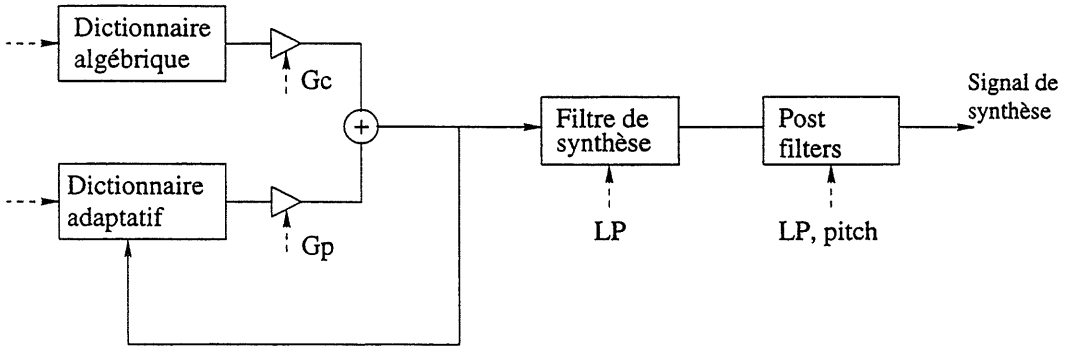

Figure 2.4 - *Diagramme du décodeur*

traitement comprend essentiellement un filtre passe haut, un filtre perceptuel (accentuation des formants) et un post-filtre harmonique.

2.2.4 Le G.729 annexe A: les simplifications apportées

Comme il a été mentionné précédemment, le G.729A est une version de la norme où la complexité a été réduite de moitié pour une légère perte de qualité . Cette dégradation n'est réellement sensible que dans des situations où l'on est en présence de bruit de fond ou

bien lorsque l'on effectue des codages successifs en "tandem". Au niveau de l'algorithme de codage, les principales modifications qui ont été apportées sont les suivante [13]:

- Le filtre perceptuel à été simplifié sous la forme:

$$W(z) = \frac{\hat{A}(z)}{\hat{A}(z/\gamma)} \qquad \gamma = 0.75 \qquad \hat{A}(z) : \textit{filtre LPC quantifié} \tag{2.2}$$

  Cela permet de simplifier les opérations qui combinent une pondération perceptuelle avec un filtrage de synthèse. Ces manipulations se trouvent essentiellement dans les calculs de vecteur cible et de réponse impulsionelle.

- La recherche du pitch en boucle ouverte a été accélérée en effectuant une décimation lors du calcul des corrélations du signal pondéré.

- Le calcul du délai de pitch en boucle fermée a été simplifié en ne cherchant plus qu'à maximiser la corrélation entre la cible et la synthèse. Le critère de sélection n'est donc plus normalisé par un terme d'énergie.

- La recherche dans le dictionnaire algébrique ISPP à été grandement écourtée grâce à un algorithme de parcours en profondeur d'abord (*Depth first algorithm*). Dans cette méthode on partitionne les 4 pulses en deux groupes de 2 pulses. On fixe dans un premier temps les deux premiers pulses, puis on cherche ensuite les positions des deux derniers pulses. Une desciption plus détaillée de cet algorithm est donnée dans [13]. Le gain de complexité obtenu est assez important: au lieu de parcourir autour de 10 % du dictionnaire (G.729), on ne considère plus que 4 % des mots du code.

- Au niveau du post-traitement sùr le signal de synthèse, le post-filtre harmonique n'utilise plus que des délai entiers.

Chapitre 3

Modèle d'encapsulation du G.729A

3.1 Introduction, principe de codage

Ce chapitre présente le modèle de codage proposé pour encapsuler le G.729A en un codeur large bande. Le système étudié est avant tout conçu pour synthétiser de la parole. Les caractéristiques du codeur encapsulé sont:

- Largeur de bande prise en compte: 80 Hz - 5500 Hz
- Débit: 16 Kbits/s soit 8 Kbits/s supplémentaires par rapport au G.729A
- Délai de codage: 23 ms soit 8 ms de plus que le G.729A.
- Complexité modérée

  Les dépenses en calcul sont essentiellement en quantification vectorielle algébrique (quantification sphérique dimension 8 et parcours de dictionnaires ACELP de tailles modestes). On amorti par ailleurs le plus possible les paramètres et les calculs mis à disposition par le codeur G.729A.

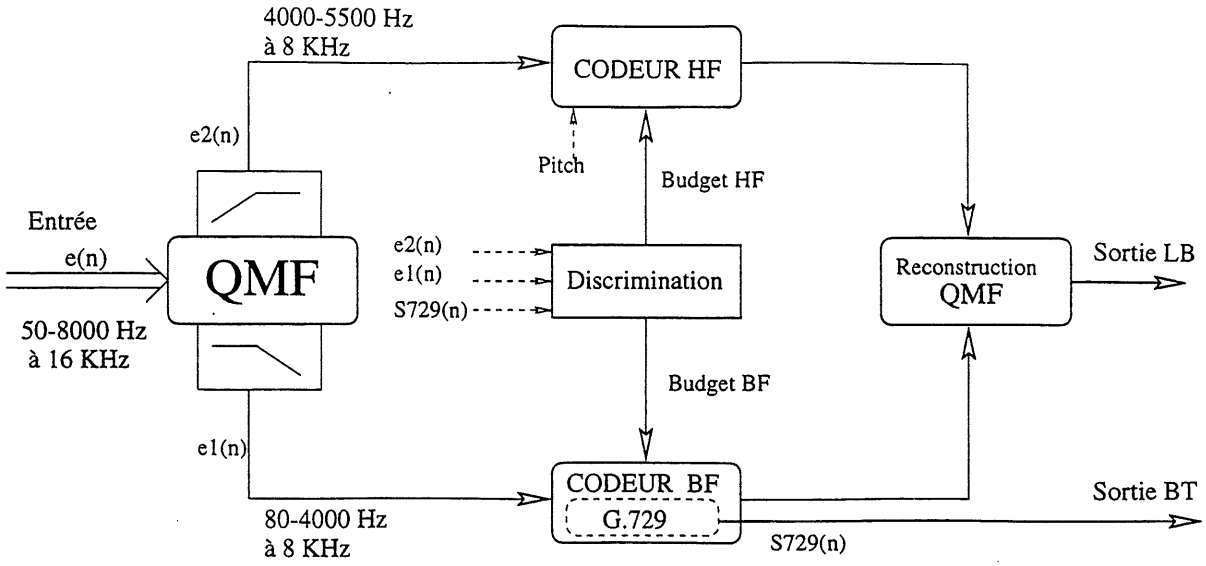

Figure 3.1 - *Modèle d'encapsulation du G.729A*

La figure 3.1 présente le principe de fonctionnement. On effectue avant tout un découpage en deux sous-bandes. La décomposition est faite par une paire de filtres QMF (*Quadrature Mirror Filters*). Le gabarit de ces filtres est donné par le graphique 3.2.

Comme le montre la figure 3.1 le système comprend essentiellement trois traitements:

- Une analyse à partir du signal original et du résultat de codage par le G.729A. Cette étude vise à classifier la nature de la portion de signal contenue dans la trame à coder. Cette discrimination permet d'attribuer le budget en bits pour chacune des sous-bandes. On considère 4 modes, correspondant à quatre distributions possibles du débit.

- Un codeur de signaux hautes fréquences.

- Un codeur additionnel pour la bande basse. Ce codeur reprend et complète les informations contenues dans le train de bits du G.729A et génère un second signal de synthèse

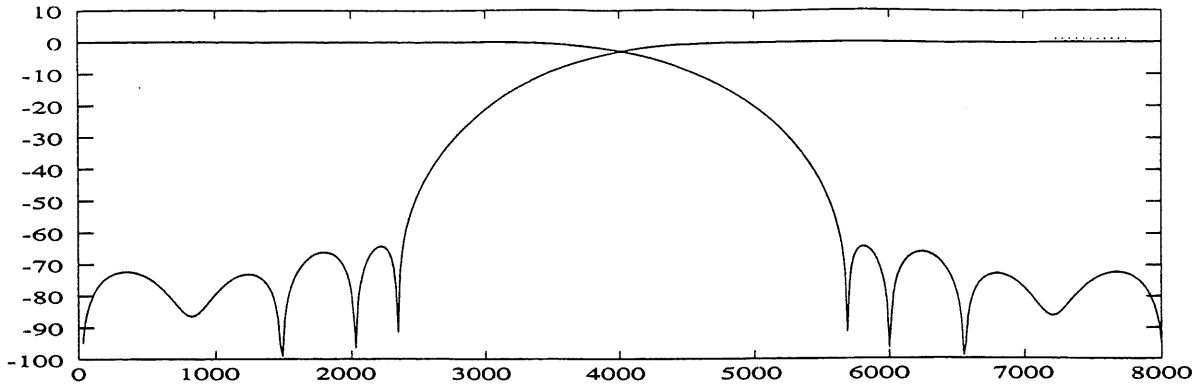

Figure 3.2 - *Gabarit des filtres QMF*

de meilleur qualité.

3.2 Synchronisation et délai de codage

Ce paragraphe apporte quelques précisions concernant la synchronisation des trames de codage des deux sous-bandes. Le système étant composé de deux codeurs, il est nécessaire de pouvoir assurer une bonne synchronisation temporelle entre les parties du signal dans chacune des deux bandes. La figure 3.3 donne une représentation des décalages de trames pour les deux codeurs.

Dans la partie hautes fréquences, on prévoit deux retards, respectivement en début et en fin de traitement. Ces délai proviennent des opérations de filtrage anti-recouvrement lors du changement d'échantillonnage. Le codeur hautes fréquences opère en effet sur un signal échantillonné à 3.2 KHz qu'il faut générer à partir d'une source cadencée à 8 KHz. Les filtres passe-bas utilisés sont de type FIR linéaires de phases. Le retard introduit pour chacun de ces filtres est de taille égale au bloc de *lookahead* pris en considération dans le G.729A, soit 5 ms.

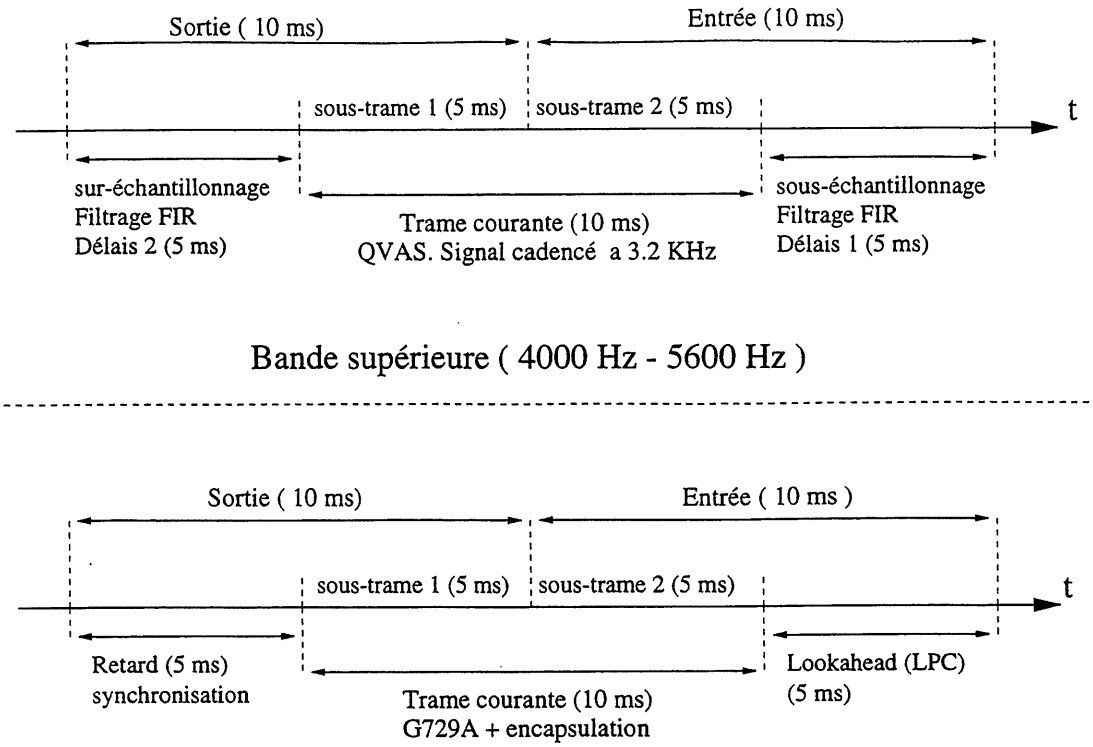

Figure 3.3 - *Synchronisation des bandes et délai de codage*

Dans la bande basse, deux retards sont à prendre en considération: le *lookahead*, utilisé par le G.729A pour l'analyse LPC en début de trame, et un retard de synchronisation en fin de trame. Ces deux retards sont de taille égale à 5 ms. Ces deux délai sont en vis à vis avec les deux délai respectifs de la bande haute. La synchronisation des trames des bandes de fréquences est ainsi assurée.

Le délai de codage global à prendre en compte est de 23 ms (10+5+5+3), soit 10 ms de trame, 5 ms de *lookahead*, 5 ms de synchronisation (attente passive) et 3 ms pour le découpage en sous-bandes (filtres QMF).

3.3 Décision sur le mode de fonctionnement du codeur

On définit quatre modes de fonctionnement pour le codeur large bande. Ces modes seront notés par la suite **mode 0**, **mode 1**, **mode 2** et **mode 3**. Ils correspondent à quatre manières différentes de répartir le débit disponible entre les deux bandes de fréquences. La suite de cette section décrit l'algorithme de décision sur l'attribution des budgets de quantification.

3.3.1 Définition des modes de fonctionnement

On effectue une observation sur le signal d'entrée ainsi que sur le signal de synthèse du G.729A. L'objectif est de classifier la portion de signal en entrée parmi quatres catégories de sons. Les modes à considérer sont définis comme suit:

- **Mode 0**: La plus grande partie du débit est investie dans la bande basse (80-4000Hz). On sélectionne ce mode dans la situation, fréquente en musique et plus rare pour la parole, où la contribution apportée part le G.729A est jugée insuffisante.

- **Mode 1**: La répartition du débit est légèrement pondérée pour les basses fréquences. On choisit ce partage du débit pour des phonèmes voisés, pour lesquels l'information pertinente est essentiellement les premières partie du spectre.

- **Mode 2**: Ce mode correspond à une répartition équitable du débit entre les deux bandes. On pense choisir cette distribution du budget pour des phonèmes dont l'énergie est répartie de manière presque uniforme sur la totalité de la bande. Les phonèmes concernés sont souvent des fricatives voisées ([z],[v]) ou des plosives ([p],[d]). Le mode 2, tout comme le mode 1, sert également d'état de transition entre les modes extrêmes (0 et 3).

- **Mode 3**: Le débit est essentiellement investi dans le codage de la bande supérieure. Ce mode est dédié à certaines fricatives non-voisées (ex: [s],[ch]) pour lesquelles l'énergie est davantage concentrée dans les hautes fréquences.

3.3.2 Discrimination selon la nature des phonèmes

Méthode de discrimination

La sélection du mode de fonctionnement peut être vue comme une discrimination musique (mode 0)/ parole voisée (modes 1 et 2) / parole non-voisée (mode 4). L'analyse vise à obtenir, si possible à moindre coût, des informations pertinentes sur la distribution spectrale du signal d'une part et sur l'efficacité du codage G.729A d'autre part. La stratégie choisie consiste à dégager un certain nombre de paramètres de discrimination. Ces variables permettent ensuite de représenter le signal à l'aide de points dans un espace de dimension égale au nombre de paramètres. Il est alors possible de classifier un son en formant des groupements de points obtenus par partitionnement de l'espace. Pour être en mesure d'observer sans trop de difficultés des zones remarquables de l'espace, on limitera sa dimension à 3. Le nombre de critères de discrimination est ainsi borné à 3.

Critères de discrimination

Les critères de discrimination sélectionnés doivent procurer des informations sur le voisement, la performance du G.729A pour la trame à coder ainsi que la répartition d'énergie entre les deux sous-bandes. Il existe un grand nombre de paramètres permettant de mettre en valeur ces caractéristiques. On opte cependant pour des variables qui ne requièrent que peu de calculs supplémentaires par rapport au travail déjà effectué par le G.729A. On retient

ainsi trois paramètres intéressants d'un point de vue pertinence/coût en complexité.

- Le **Gain de pitch**: on prend la valeur obtenue comme moyenne des gains de pitch calculés par le G.729A sur chacune des deux sous-trames. Ce paramètre rend très bien compte de la nature voisée ou non voisée du signal. On précise que les valeurs du gain de pitch sont prises au codeur sous formes non quantifiées et non tronquées (ces valeurs peuvent en effet être limitées à 1.2 pour des raisons de stabilité).

- Le **RSB** (Rapport Signal à Bruit): on détermine la valeur logarithmique du rapport signal à bruit entre le signal d'entrée en bande basse (signal e_1 de la figure 3.1), et le signal de synthèse S_{729} obtenu par travail du G.729 sur e_1.

$$RSB = 10 * \log\left(\frac{E[e_1^2]}{E[S_{729} - e_1]^2}\right)$$

- Le **rapport des énergies** ΔE entre les deux sous-bandes. On calcule le quotient des énergies du signal dans les deux sous-bandes. La valeur est prise sous forme logarithmique. Ce paramètre permet de rendre compte de la répartition de l'information entres les deux bandes de fréquences considérées, représentées respectivement par les signaux e_1 et e_2.

$$\Delta E = 10 * \log\left(\frac{E[e_2^2]}{E[e_1^2]}\right)$$

Pour la plupart des signaux de parole ou de musique, l'énergie est surtout concentrée en basses fréquences et la valeur de ΔE est relativement faible($< -20dB$). Cependant, dans des cas particuliers, comme celui des fricatives non voisées, la répartition de l'énergie est déplacée vers la zone hautes fréquences.

Observation des signaux dans l'espace des paramètres

L'espace des paramètres a été défini dans le paragraphe précédent.

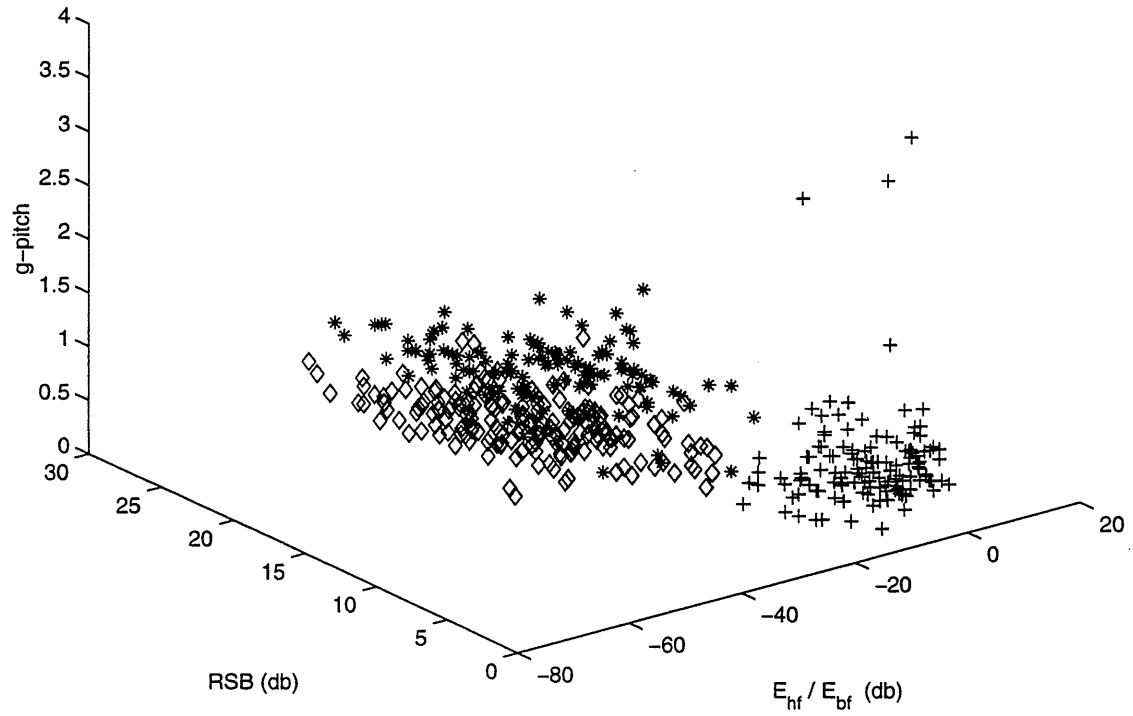

Figure 3.4 - *Espace des paramètres (RSB, g-pitch, ΔE). Points de l'espace pour de la musique ($\Diamond$), des sons voisés (*) et des sons non-voisés (+)*

On observe sur la figure 3.4 que les points appartenant respectivement à des trames de phonèmes voisés, non-voisés ou à des sons de musique, occupent des positions particulières dans l'espace. On remarque également que la projection sur le plan (RSB, ΔE) à elle seule semble déjà très sélective (figure 3.5). Ceci tend à prouver que le paramètre g-pitch, responsable du voisement, n'est pas indispensable. En réalité cette variable est redondante par rapport à la combinaison des deux autres critères. On constate qu'un résultat quantitatif sur le travail effectué par le G.729A, associé à une connaissance de la répartition de l'énergie, procure une information pertinente sur le voisement. En conséquence, on choisit de ne retenir

pour la discrimination que le plan (RSB, ΔE).

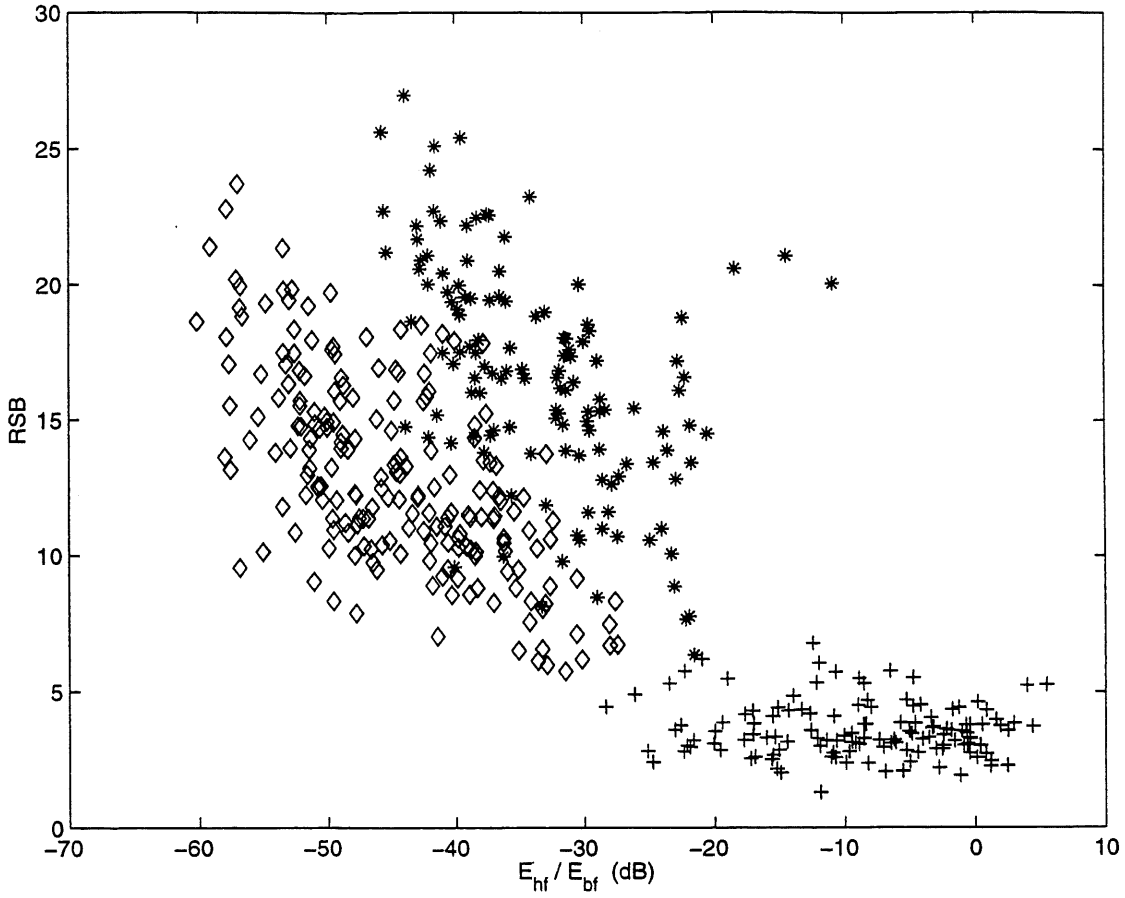

Figure 3.5 - *Plan des paramètres (RSB, ΔE). Points du plan pour de la musique ($\Diamond$), des sons voisés (*) et des sons-non voisés (+)*

Choix des modes de fonctionnement: découpage du plan (RSB, ΔE) en une partition de quatre ensembles

Comme le montre la figure 3.5, à chaque classe de signaux (musique, voisé et non-voisé), on peut associer une partie du plan. Ces parties semblent quasiment disjointes, sauf au niveau

de la frontière voisé/musique. En effet, une trame de musique peut avoir des propriétés semblables à celles d'une trame voisée, pour peu que la voix d'un chanteur soit présente voir dominante.

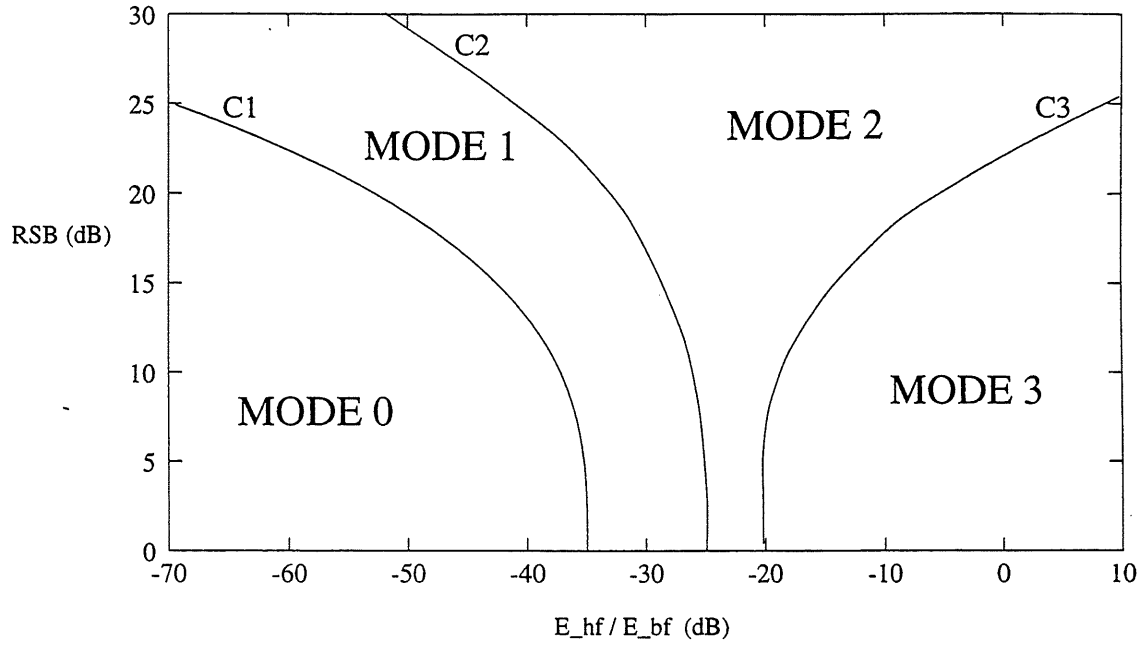

Figure 3.6 - *Découpage du plan (RSB, ΔE) selon 4 modes de fonctionnement*

On définit les modes de fonctionnement en partitionnant le plan en quatre morceaux. Ces parties sont séparées à l'aide de trois courbes frontière C_1, C_2 et C_3 (figure 3.6). Plusieurs familles de courbes sont en mesure de pouvoir découper le plan selon les modes de fonctionnement. Les équations qui on été choisies sont les fonctions puissances ainsi définies (y représente la valeur du RSB en dB et $C_i(y)$ celle du rapport E_{hf}/E_{bf} en dB):

$$C_1(y) = -35(\frac{y}{25})^3 - 35\,, \quad C_2(y) = -45(\frac{y}{35})^3 - 25 \quad C_3(y) = 30(\frac{y}{25})^3 - 20$$

Transitions d'états

Le passage d'un mode de fonctionnement à un autre est contrôlé à l'aide d'une fonction de transitions. Cette fonction attribue un mode effectif selon l'état précédant et le mode demandé par l'algorithme de discrimination (position du signal dans le plan (RSB, ΔE). Les transitions inter-modes sont décrites par le diagramme d'états finis représenté sur la figure 3.7.

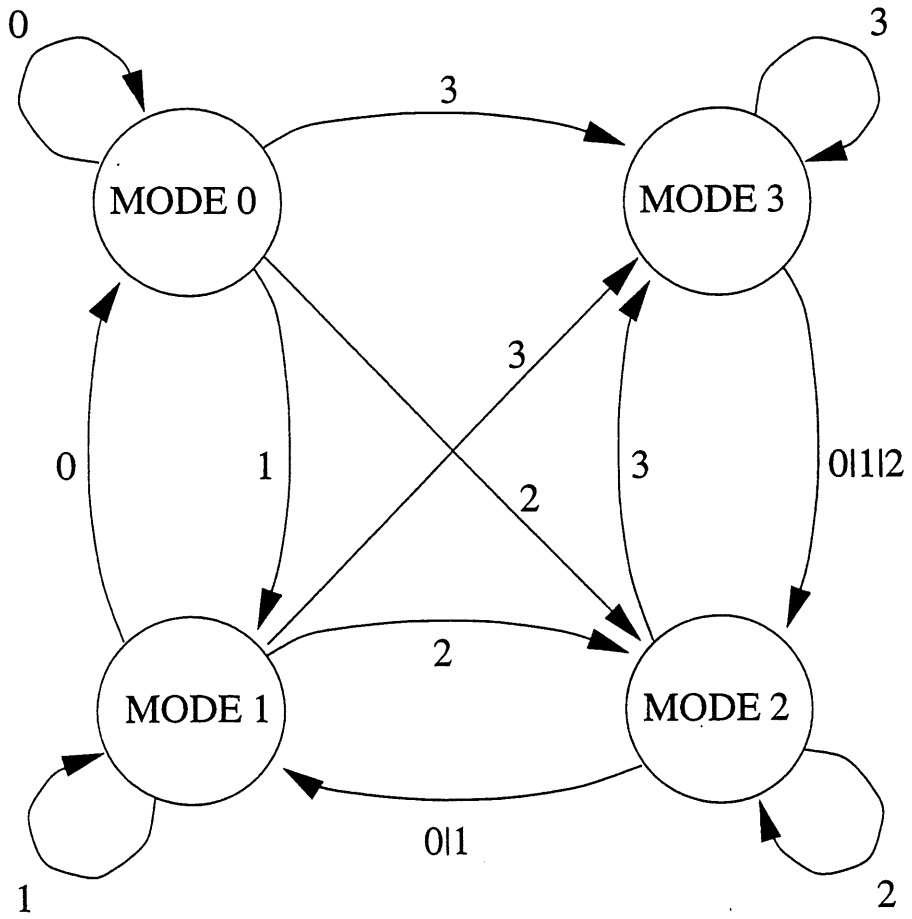

Figure 3.7 - *Diagramme de transitions inter-modes*

Le contrôle des transitions du débit attribué à chacune des bandes est important pour la

bande supérieure. La bande basse dispose toujours d'un débit minimal provenant du G.729. En revanche, la plage hautes fréquences, peut passer d'un état de bonne quantification (mode 3) à un mode de représentation plus sommaire (mode 0). Les écarts de qualité peuvent être sensibles si on passe trop rapidement d'un mode à l'autre. La fonction de transition décrite par le diagramme de la figure 3.7 vise à imposer une décroissance modérée du débit de la bande hautes fréquences. Ainsi, si l'algorithme de discrimination choisit le mode 0 sachant que l'état précédent est le mode 3, on force un passage par les états intermédiaires (mode 1 et mode 2). Par contre, aucune limitation n'est imposée lorsque l'on cherche à augmenter le débit dans la bande supérieure.

3.4 Codage de la bande basse (encapsulation)

Le codeur basses fréquences opère dans la plage 80 Hz - 4000 Hz pour des signaux échantillonnés à 8 Khz. Cette partie du codeur global est précisément le secteur où il y a encapsulation du G.729A. Le codeur est en fait une extension du G.729A qui produit deux signaux de synthèse en sortie: la synthèse provenant du G.729A uniquement et un second signal de synthèse, plus proche du signal original, pouvant être combiné avec le signal hautes fréquences pour construire un signal large bande de bonne qualité.

3.4.1 Codeur ACELP à deux étages

Principe de fonctionnement

Le travail dans la bande inférieure du spectre consiste à transformer le G.729A en un codeur ACELP à deux étages. Le premier étage génère le signal de synthèse du G.729A et le second étage code le signal d'erreur. Le résultat des deux étages combinés permet de

synthétiser le signal de meilleur qualité. La figure 3.8 décrit le fonctionnement des deux étages et met en évidence le rôle joué par le G.729A dans ce système.

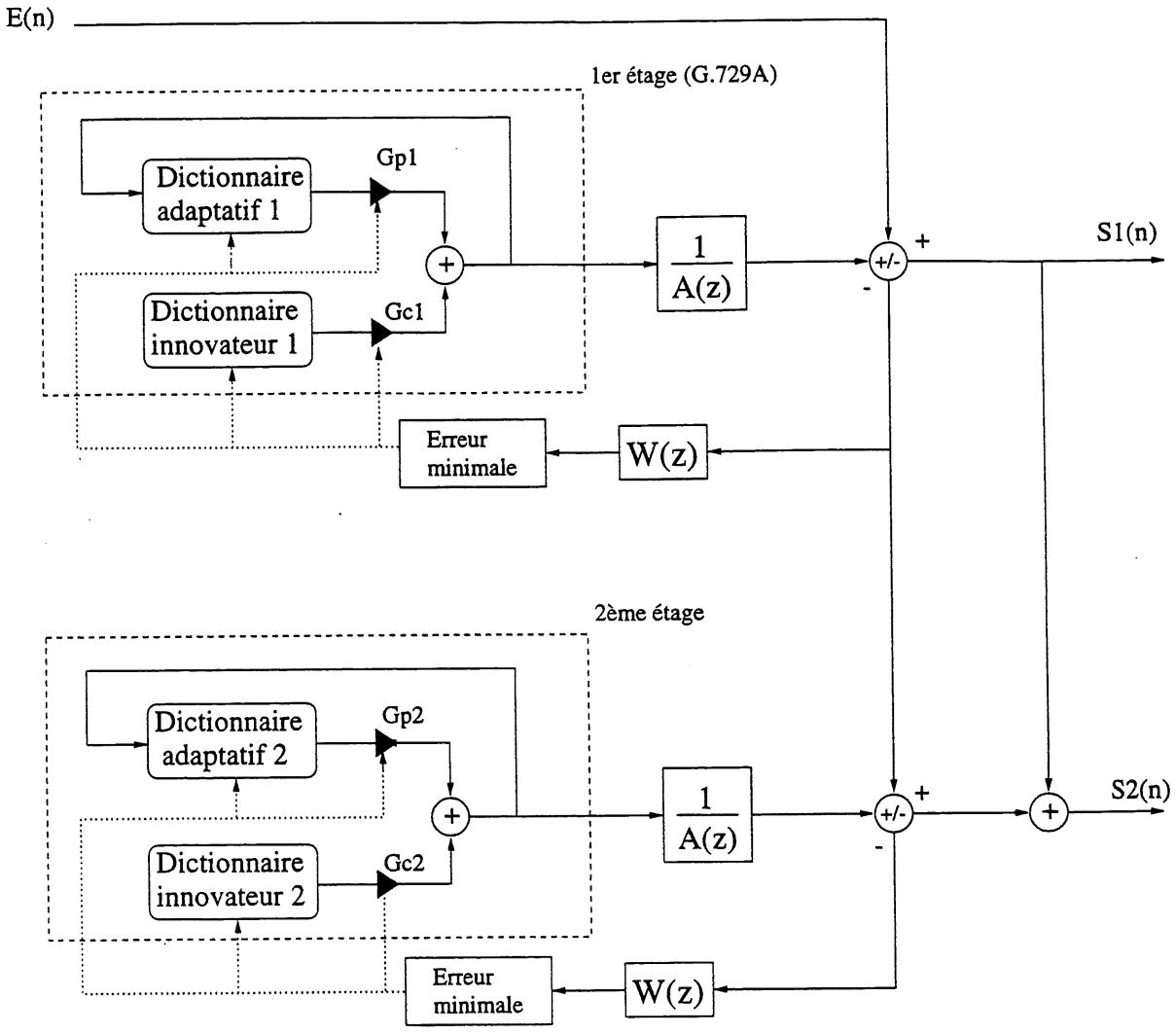

Figure 3.8 - *Codeurs ACELP à deux étages incluant le G.729A*

La plupart des paramètres du signal comme le délai de pitch, les filtres LPC (noté $1/A(z)$) ainsi que le filtre de pondération perceptuel (noté W(z)) sont communs aux deux étages et proviennent du train de bits du G.729A. Par ailleurs l'analyse du signal dans le second étage

se fait toujours par sous-trames de 5 ms.

Algorithme de codage pour le second étage

Le traitement effectué dans le second étage consiste à calculer les mots de codes d'une nouvelle paire de dictionnaires (adaptatif, innovateur). Le signal cible est l'erreur de quantification entre l'entrée du codeur et la synthèse du G.729A.

Le principe de fonctionnement est similaire à celui du premier étage (G.729A). Le dictionnaire adaptatif est déterminé comme un morceau du passé de la synthèse du signal d'erreur (synthèse du second étage). On utilise pour obtenir ce mot de code le délai de pitch en valeur fractionnaire (1/3). Le second dictionnaire innovateur, tout comme le premier, est de type ISPP (*Interleaved Single-pulse Permutation*). Le nombre d'impulsions à positionner dépend du mode de fonctionnement du codeur global. Le calcul de la cible, ainsi que l'algorithme de recherche dans le dictionnaire innovateur sont similaires à ceux décris dans la norme G.729A ([5]). Enfin, les gains sont calculés de manière à minimiser l'erreur quadratique dans le domaine perceptuel (filtrage par W(z)) entre l'original et la combinaison linéaire des deux mots de codes, désormais fixes. Pour plus de détails concernant ces algorithmes sur la technologie ACELP, le lecteur est invité à consulter les références [5], [13] et [8].

Quelques mots sur la complexité

La complexité de calcul pour le second étage est peu élevée. On bénéficie de calculs déjà effectués lors de la construction du premier étage. On utilise entre autre le calcul des coefficients des filtres A(z) et W(z), ainsi que de certaines variables utiles pour la recherche dans le dictionnaire ISPP par la méthode du *Backward filtering* ([5]). Les calculs dont on tire le plus profit sont ceux de la réponse impulsionelle h du filtre W(z)/A(z) ainsi que de la

matrice des corrélations de h: $\Phi = H^tH$ (H est la matrice de Toepliz triangulaire inférieure de diagonale h(0) et de diagonales inférieures h(1),...,h(39))

Données transmise pour le second étage (encapsulation)

Outre les paramètres fournis par le train de bits du G.729A, les données qu'il reste à transmettre pour le second étage sont pour chaque sous-trame: les gains des dictionnaires, les indices des mots de code pour le dictionnaire innovateur ainsi que les signes des impulsions sélectionnées.

3.4.2 Utilisation du débit selon le mode de fonctionnement

Comme il a été précisé dans la section précédente, le codeur global peut fonctionner sous 4 modes différents (notés mode 0 à mode 3) selon la décision qui a été faite lors de la discrimination sur le signal à coder. À chacun de ces modes correspond un budget de quantification pour le second étage du codeur ACELP.

La variation du débit influe sur le nombre d'impulsions du dictionnaire ISPP et la quantification des gains. On choisit de positionner 1, 2 ou 5 pulses et de quantifier les gain avec 0, 3 ou 4 bits par valeur. Lorsque le budget pour le dictionnaire innovateur est limité, on favorise pour les positions des impulsions, les indices de la *track* (3/4) qui a été laissée de côté lors du calcul du code innovateur du G.729A. On rappelle à cet effet que chaque sous-trame est décomposée en 5 *track* (numérotées de 0 à 4) de 8 positions chacune. Le dictionnaire innovateur du G.729A positionne quatre pulses par sous-trame dont une impulsion dans chacune des trois premières *track* et une impulsion dans une des deux *track* 3 ou 4, laissant ainsi une *track* de côté. On numérote 3/4 cette *track*.

Les dépenses du budget de quantification pour chacun des modes sont:

- **Mode 0** : 6 impulsions dans chaque sous-trame. On positionne un pulse dans chacune des *track* et on ajoute une impulsion supplémentaire pour la *track* 3/4. Les gains sont tous quantifiés avec 3 bits par valeur.

- **Mode 1** : 3 impulsions par sous-trame. On positionne un pulse dans la *track* 3/4. Les deux autres impulsions sont plaçées dans les quatres autres *track* prises deux par deux. Les gains sont tous quantifiés avec 2 bits par valeur.

- **Mode 2** : 2 impulsions par sous-trame. On postionne un pulse dans une des *track* 0 ou 1. Le second pulse est positionné dans une des *track* 2 et 3/4. Les gains sont tous quantifiés avec 2 bits par valeur.

- **Mode 3** : 1 pulse par sous-trame, pouvant prendre 8 positions repérées par la *track* 3/4. Le gain de pitch est mis à zéro (le mode est non-voisé et le dictionnaire adaptatif n'est pas pris en compte). Le gain du code d'excitation est quantifié avec 2 bits.

Quantification des gains

Le deux gains attribués respectivement au code d'excitation et au dictionnaire adaptatif sont quantifiés scalairement. Le gain du code d'excitation représente pour beaucoup l'énergie de la trame. Cette valeur varie souvent lentement, notamment dans les parties voisées, plus stationnaires.

Le gain de pitch peut prendre des valeurs comprises entre 0 et 1.2. La figure 3.9 donne la distribution de la valeur de ce gain. On constate dans un premier temps que les points extrêmes 0 et 1.2 présentent une concentration de probabilité beaucoup plus importante (surtout pour la valeur nulle). Sur le reste de l'intervalle, la répartition est quasiment uniforme.

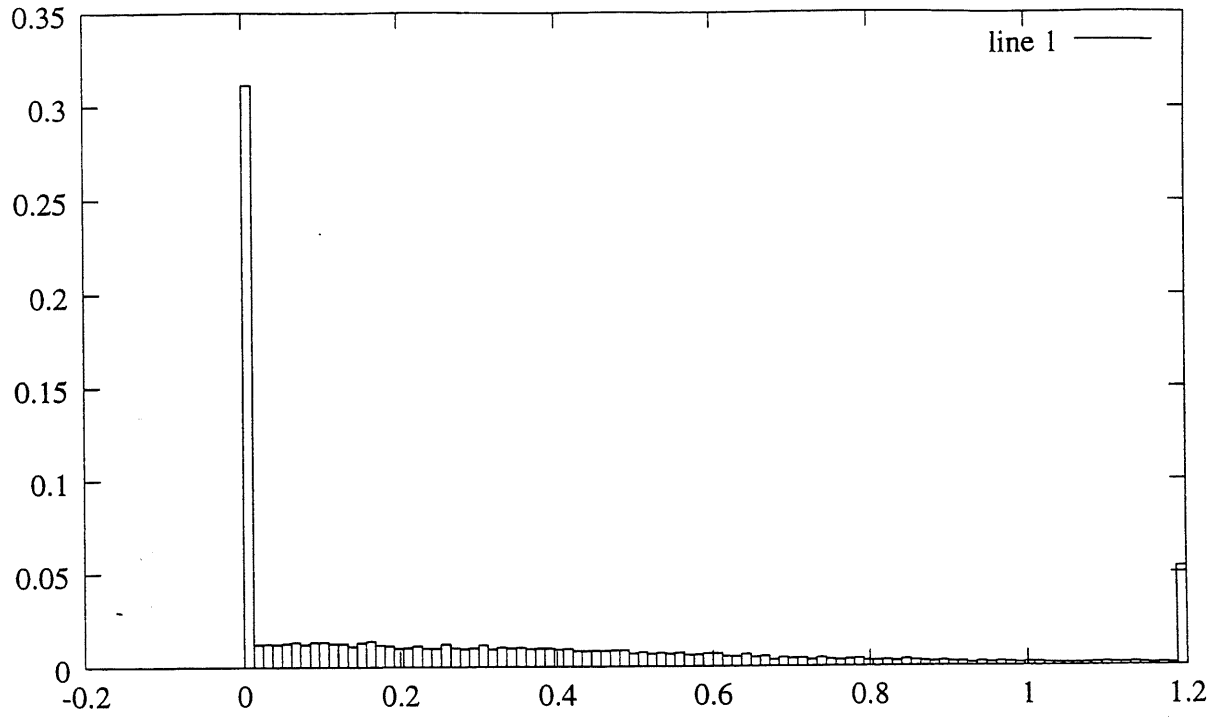

Figure 3.9 - *Distribution du gain de pitch pour le second étage*

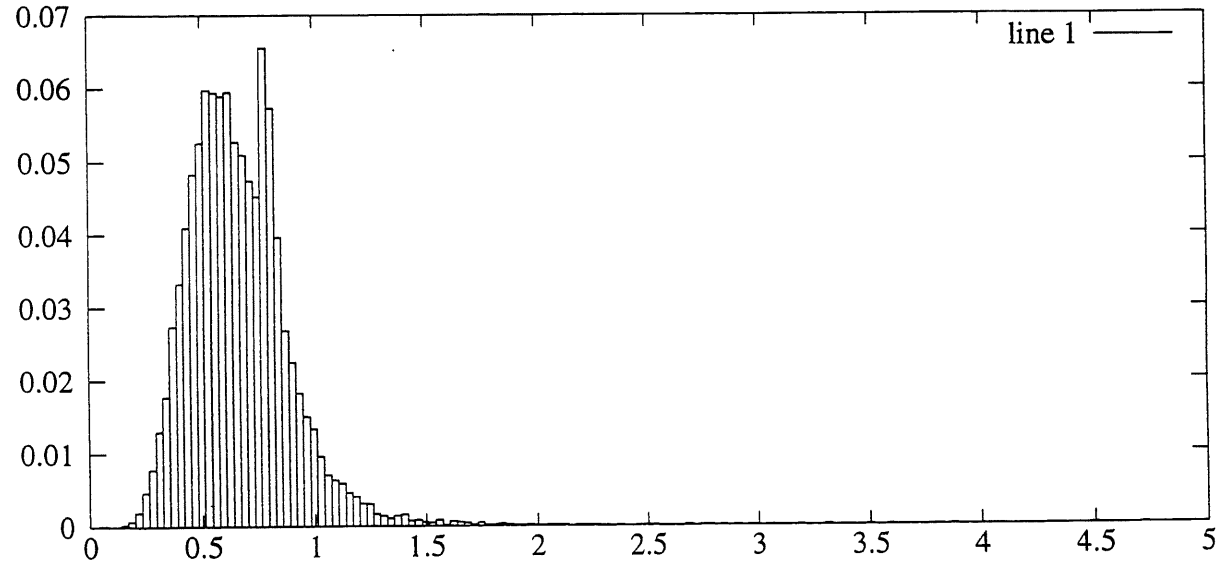

Figure 3.10 - *Distribution du rapport des gains innovateurs (G.729A/second étage)*

Le gain attribué au code innovateur du second étage est très prédictible. Sa valeur est fortement liée aux valeurs précédentes de ce paramètre. Cependant la corrélation la plus intéressante à utiliser est encore celle qui lie les gains respectifs des codes innovateurs des deux étages (gain innovateur du G.729A et gain innovateur du second étage). Le graphique 3.10 donne la distribution du rapport de ces gains. Cette distribution est centrée autour d'une valeur moyenne de 0.6. On constate par ailleurs la présence d'un second pic proche de la valeur unitaire. Cette valeur correspond au cas où les deux étages apportent des contributions équivalentes en terme d'énergie (cas des trames nulles ou très mal représentées par le G.729A seul). Le gain de prédiction obtenu pour la quantifiction du gain innovateur est de l'ordre de 9 à 10 dB.

Les paramètres que l'on quantifie sont donc le gain de pitch et la prédiction du gain du code innovateur à partir du premier étage. Les dictionnaires sont construits par l'algorythme de la K-moyenne à partir d'une banque d'apprentissage comprenant des extraits de fichiers de parole et de musique. La banque mère est divisée en quatre sous-banques correspondant aux quatres modes de fonctionnement. Pour chaque mode, les dictionnaires sont calculés à l'aide de la banque d'apprentissage adaptée.

3.5 Codage de la bande supérieure

3.5.1 Principe de fonctionnement du codeur

La figure 3.11 donne le modèle de codage. Pour rester synchrone avec les trames du G.729, le traitement se fait par des blocs de 80 échantillons. On effectue en premier lieu un sous-échantillonnage afin de ne garder qu'un nombre minimal d'échantillons. La bande de fréquences conservée est 4000-5600 Hz . Le signal est désormais cadencé à une fréquence de 3.2 KHz et chaque trame contient deux sous-trames de 16 échantillons.

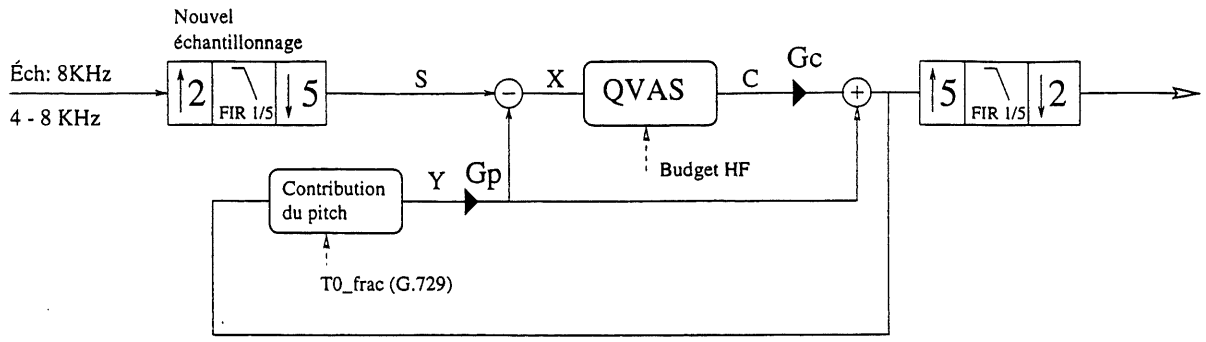

Figure 3.11 - *Codeur hautes fréquences*

On retire ensuite les corrélations à long terme entre les échantillons. Ce traitement vise surtout à retirer d'éventuelles impulsions de pitch. Cette opération est également motivée par l'observation d'harmonies en hautes fréquences pour certaines fricatives ([z],[v]...). Le dictionnaire adaptatif est calculé à l'aide du signal passé et du délai de pitch obtenu gratuitement à partir du train de bits du G.729A. On quantifie enfin le signal résiduel.

Pour chaque sous-trame de 16 échantillons on quantifie la direction (QVAS) (figure 3.12 de chacun des deux vecteurs de dimension 8 formant la sous-trame. On dispose pour cela de quatres quantificateurs différents selon le débit disponible pour la représentation des hautes fréquences. Ce processus fournit en sortie un mot de code de dimension 16 (deux mots de code de dimension 8) noté **C** qui représente le signal résiduel.

Une fois le dictionnaire adaptatif **Y** et le mot code **C** déterminés, on calcule les gains respectifs *Gp* et *Gc* des deux dictionnaires. Ces gains sont calculés en minimisant l'erreur quadratique **e** entre le signal reconstruit et le signal original **S**. On annule pour cela les deux dérivées partielles de la fonction d'erreur. Pour des raisons de stabilité, le gain de pitch est limité à 1.2.

$$min\|e^2(Gc, Gp)\| = \|S - Gc.C - Gp.Y\|^2$$

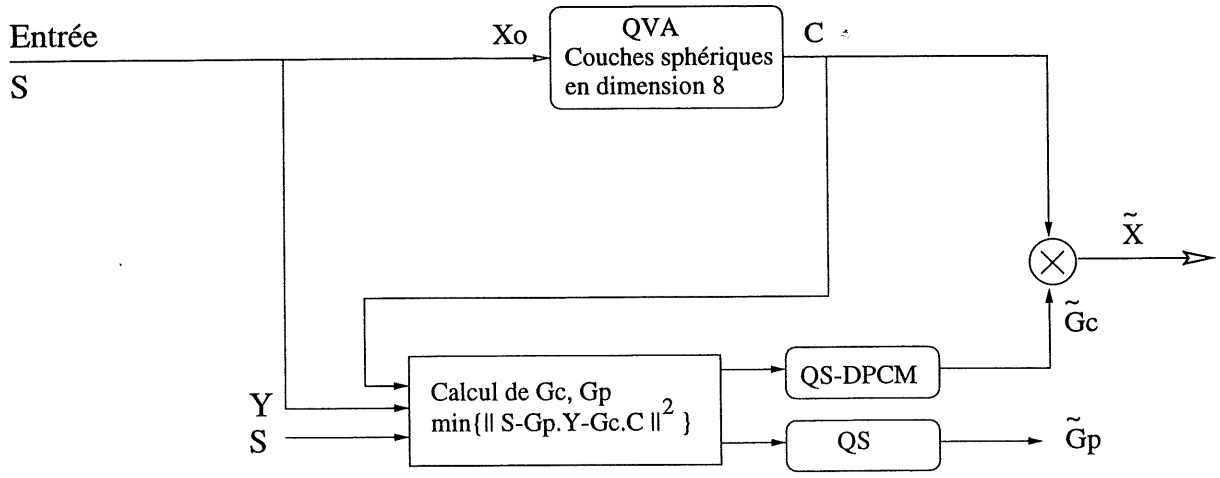

Figure 3.12 - *Quantification sphérique (QVAS)*

Cas particulier: Sons non-voisés

Le dictionnaire adaptatif n'est pas toujours pris en compte. En effet, parmi les quatres modes de fonctionnement du système d'encapsulation du G.729, seuls les modes 1 et 2 considèrent des sons voisés. Lorsque le signal est non-voisé (mode 3) ou si l'énergie de la bande supérieure du spectre est faible (mode 0), le gain de pitch est automatiquement mis à zéro et les gains Gp et Gc deviennent:

$$Gp = 0 \qquad Gc = \frac{< S, C >}{< C, C >}$$

Les gains Gp et Gc sont ensuite quantifiés scalairement. Comme Gc varie peu d'une sous-trame à l'autre, sa valeur est très prédictible. On utilise pour sa quantification un codage différentiel DPCM. La prédiction est statique, d'ordre 2. Les valeurs des coeefficients de prédiction peuvent varier d'un mode à l'autre (surtout pour les modes extrêmes). Les quantificateurs sont composés de dictionnaires stochastiques à une dimension. La construction de ces dictionnaires est similaire à celle utilisée pour la quantification des gains du codeur

basses-fréquences.

Les sections suivantes présentent de manière plus détaillée la quantification du signal résiduel **X**.

3.5.2 Quantification d'une source gaussienne

Cette partie présente quelques rappels concernant le codage de source pour un modèle gaussien.

Modèle du signal résiduel

Le résidu de prédiction du codeur hautes fréquences est assimilable à un bruit blanc gaussien. Les échantillons composant ce signal sont supposés décorrélés et suivent une loi normale centrée dont la variance est donnée par l'énergie de la sous-trame (gain). Ce modèle est relativement fiable, il reste cependant des cas où l'on est en présence d'un échantillon particulièrement grand. Une telle impulsion représente un événement de pitch qui n'a pu être suffisamment atténué par la prédiction à long terme.

Vecteurs gaussiens

Une sous-trame du signal résiduel est composée de 16 échantillons que l'on assimile à 16 variables aléatoires X_i indépendantes. Après avoir normalisé les 16 composantes, les variables ont toutes pour densité de probabilité la loi normale centrée réduite $\mathcal{N}(0,1)$.

Le quantificateur sphérique (QVAS) décode des vecteurs de dimension 8. Chacun de ces vecteurs est gaussien et se note $V_8 = [X_1, ..., X_8]$. Les composantes X_i étant indépendantes, la distribution de V_8 s'exprime comme produit des distributions marginales des variables X_i.

On s'intéresse davantage à la variable aléatoire $R_8 = \|V_8\|$ qui représente le rayon du vecteur gaussien. Cette variable présente une distribution particulière. L'orientation du vecteur est elle isotrope.

On expose à présent quelques résultats de probabilités dont l'utilité est ici de localiser des vecteurs gaussiens dans l'espace à n dimensions. On introduira d'abord les lois de Chi-deux et de Rayleigh. On conclura ensuite sur une description de la distribution spatiale des vecteurs gaussiens de variance unitaire et de composantes indépendantes.

Distributions de Chi-deux et distributions de Rayleigh

La distribution de Chi-deux à n degrés de liberté a pour densité de probabilité:

$$\chi_n^2(0) : \qquad f_{\chi^2}(x) = \frac{1}{\Gamma(n/2)2^{n/2}} x^{n/2-1} e^{-x/2} \tag{3.1}$$

$$E[\chi_n^2(0)] = n \qquad Var(\chi_n^2(0)) = 2n \tag{3.2}$$

Cette loi est un cas particulier de la loi gamma: $\chi_n^2(0) \sim \Gamma(n/2, 1/2)$

La distribution de Rayleigh a pour densité de probabilité:

$$\mathcal{R}_n : \qquad f_{\mathcal{R}_n}(x) = \frac{1}{\Gamma(n/2)2^{n/2-1}} x^{n-1} e^{-x^2/2} \tag{3.3}$$

$$E[\mathcal{R}_n] = \sqrt{2}\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \qquad Var(\mathcal{R}_n) = n - E[\mathcal{R}_n] \tag{3.4}$$

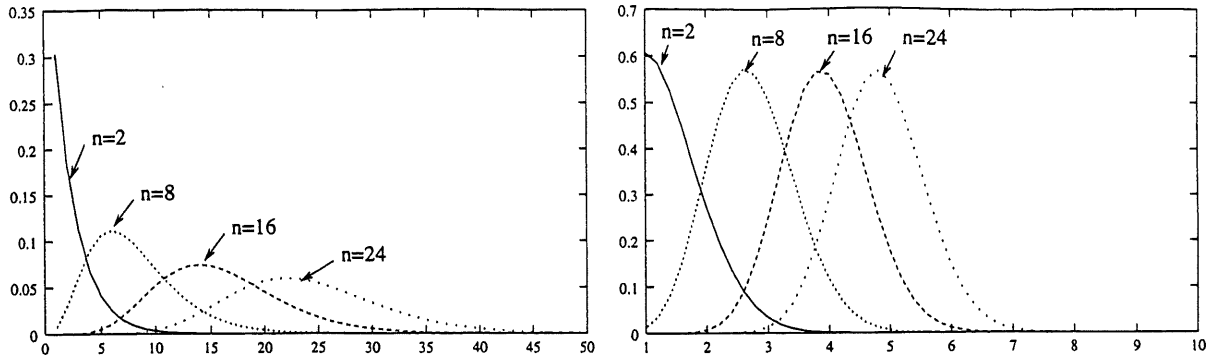

Figure 3.13 - *Distributions de Chi-deux (à gauche) et de Rayleigh (à droite)*

Ces expressions, bien qu'un peu complexes peuvent être approximées pour les grandes valeurs de n:

$$E[\mathcal{R}_n] \sim \sqrt{n} \qquad et \qquad \frac{\sigma(\mathcal{R}_n)}{E[\mathcal{R}_n]} \sim \frac{1}{\sqrt{2n-1}} \tag{3.5}$$

Ces équivalences ne sont à conserver que pour des vecteurs gaussiens de dimensions suffisamment élevées. En particulier, les approximations sont mauvaises pour $n = 2$ et encore approximatives en dimension 4. On peut garder ces simplifications à partir de $n = 8$.

Quantification d'un vecteur gaussien

Il a été vu au paragraphe précédent que le rayon d'un vecteur gaussien a une valeur dont la densité de probabilité présente une concentration proche de la valeur moyenne $E(\mathcal{R}_n)$. La dispersion des valeurs du rayon autour de la moyenne est régit par le rapport $\frac{\sigma(\mathcal{R}_n)}{E[\mathcal{R}_n]}$ que l'on peut approximer par $\frac{1}{\sqrt{2n-1}}$. On considère alors que les vecteurs gaussiens sont localisables près d'une couche sphérique de rayon $\sqrt{n}$. Par ailleurs, en augmentant la dimension de l'espace, on diminue la dispersion des vecteurs.

D'après ce qui précède, le dictionnaire d'un quantificateur adapté à une source gaussienne

devrait être composé de vecteurs régulièrement répartis (direction isotrope) sur une sphère de l'espace. Si en outre on a la possibilité de choisir la dimension des vecteurs, il préférable de prendre la valeur plus grande possible.

3.5.3 Composition des dictionnaires utilisés

Les sphères du réseau de Gosset

Les dictionnaires utilisés dans le quantificateur sphérique (QVAS) utilisent des points des premières sphères du réseau tourné de Gosset :

$$RE_8 = 2D_8 \cup 2D_8 + [11111111]$$

Les points du réseau RE_8 sont répartis régulièrement sur les sphères de rayons $2\sqrt{2n}$. De plus, sur chacune de ces sphères les points sont regroupés en classes de points repérées par leurs *leader* respectifs. Un *leader* est un point du réseau dont les composantes sont ordonnées de manière décroissante. Il définit une classe composée de tous les points obtenus par permutations des composantes . Ces *leader* sont eux aussi regroupés par des *Leader* absolus. Ces vecteurs ont toutes leurs coordonnées positives et ordonnées de manière décroissante. Ils permettent de regrouper tous les *leader* qui ne diffèrent que par les signes des composantes. On précise que les *leader* absolus, contrairement aux *leader* qu'ils englobent, n'appartiennent pas nécessairement au réseau RE_8.

Dans le cas d'un quantificateur vectoriel, la décomposition en classes associées à des *Leader* permet un repérage facile et rapide du point d'une sphère du réseau approximant au mieux (plus proche voisin) le vecteur réel introduit en entrée du quantificateur. Les références [18]] et [[22] apportent plus de détails sur ces considérations.

Structure des dictionnaires choisis

Pour être en mesure de répondre aux exigences des différents modes du système d'encapsulation, on construit quatres quantificateurs notés Q0, Q1, Q2 et Q3. On choisit de construire les quantificateurs à partir des sphères **S1**, **S2** et **S5**.

Dans la mesure où ces sphères ne contiennent pas toujours un nombre de points égal à une puissance de 2, on retire ou on ajoute quelques points à ces sphères afin d'obtenir des dictionnaires indexables par des nombres entiers d'unités binaires. Par ailleurs, lorsqu'un quantificateur ne prend pas en compte la totalité des *Leader* d'une sphère, on conserve en priorité les *Leader* impairs. Ce choix est fait pour favoriser les vecteurs de composantes non nulles qui peuvent être indésirables pour la quantification d'échantillons réels. Le tableau 3.1 donne un inventaire des *Leader* des sphères **S1** et **S2** et **S5** de RE_8.

Les quantificateurs retenus sont les suivants (un résumer de leurs caractéristiques est donné par le tableau 3.2):

- **Mode 0** :Le dictionnaire Q0 a pour taille 16 (4 bits/vecteur) et est constitué des 16 mots du code de Hamming étendu (8,4,4) en notation +/- 1.
- **Mode 1**: Le dictionnaire Q1, de taille 256 (8 bits/vecteur), inclut tous les points le la première sphère. Q1 est ensuite complété par 16 points supplémentaires. Ces vecteurs sont les points des classes de *Leader* absolu [$2\sqrt{2}$ 0 0 0 0 0 0 0] .
- **Mode 2**: Le quantificateur Q2 est composé des points de la seconde sphère. Comme le nombre de vecteurs contenu dépasse légèrement la valeur $2^{11} = 2048$, on retire 112 points. Les vecteurs non retenus sont repérés par les *Leader* 3.1 ([3 1 1 1 1 1 1 -1]) et 3.8 ([1 -1 -1 -1 -1 -1 -1 -3]) (notation du tableau 3.1).

SPHÈRE 1 (S1)									
No	Leaders								Cardinal
1	2	2	0	0	0	0	0	0	112
1.1	2	2	0	0	0	0	0	0	28
1.2	2	0	0	0	0	0	0	-2	56
1.3	0	0	0	0	0	0	-2	-2	28
2	1	1	1	1	1	1	1	1	128
2.1	1	1	1	1	1	1	1	1	1
2.2	1	1	1	1	1	1	-1	-1	28
2.3	1	1	1	1	-1	-1	-1	-1	70
2.4	1	1	-1	-1	-1	-1	-1	-1	28
2.5	-1	-1	-1	-1	-1	-1	-1	-1	1
Total									240

SPHÈRE 2 (S2)									
No	Leaders								Cardinal
1	4	0	0	0	0	0	0	0	16
1.1	4	0	0	0	0	0	0	0	8
1.2	0	0	0	0	0	0	0	-4	8
2	2	2	2	2	0	0	0	0	1120
2.1	2	2	2	2	0	0	0	0	70
2.2	2	2	2	0	0	0	0	-2	280
2.3	2	2	0	0	0	0	-2	-2	420
2.4	2	0	0	0	0	-2	-2	-2	280
2.5	0	0	0	0	-2	-2	-2	-2	70
3	3	1	1	1	1	1	1	1	1024
3.1	3	1	1	1	1	1	1	-1	56
3.2	3	1	1	1	1	-1	-1	-1	280
3.3	3	1	1	-1	-1	-1	-1	-1	168
3.4	3	-1	-1	-1	-1	-1	-1	-1	8
3.5	1	1	1	1	1	1	1	-3	8
3.6	1	1	1	1	1	-1	-1	-3	168
3.7	1	1	1	-1	-1	-1	-1	-3	280
3.8	1	-1	-1	-1	-1	-1	-1	-3	56
Total									2160

SPHÈRE 5 (S5)									
No	Leaders								Cardinal
1	4	4	2	2	0	0	0	0	6720
1.1	4	4	2	2	0	0	0	0	420
1.2	4	4	2	0	0	0	0	-2	840
1.3	4	4	0	0	0	0	-2	-2	420
1.4	4	2	2	0	0	0	0	-4	840
1.5	4	2	0	0	0	0	-2	-4	1680
1.6	4	0	0	0	0	-2	-2	-4	840
1.7	2	2	0	0	0	0	-4	-4	420
1.8	2	0	0	0	0	-2	-4	-4	840
1.9	0	0	0	0	-2	-2	-4	-4	420

SPHÈRE 5 (S5) (Suite)									
No	Leaders								Cardinal
2	6	2	0	0	0	0	0	0	224
2.1	6	2	0	0	0	0	0	0	56
2.2	6	0	0	0	0	0	0	-2	56
2.3	2	0	0	0	0	0	0	-6	56
2.4	0	0	0	0	0	0	-2	-6	56
3	4	2	2	2	2	2	2	0	7168
3.1	4	2	2	2	2	2	2	0	56
3.2	4	2	2	2	2	2	0	-2	336
3.3	4	2	2	2	2	0	-2	-2	840
3.4	4	2	2	2	0	-2	-2	-2	1120
3.5	4	2	2	0	-2	-2	-2	-2	840
3.6	4	2	0	-2	-2	-2	-2	-2	336
3.7	4	0	-2	-2	-2	-2	-2	-4	56
3.8	2	2	2	2	2	2	0	-4	56
3.9	2	2	2	2	2	0	-2	-4	336
3.10	2	2	2	2	0	-2	-2	-4	840
3.11	2	2	2	0	-2	-2	-2	-4	1120
3.12	2	2	0	-2	-2	-2	-2	-4	840
3.13	2	0	-2	-2	-2	-2	-2	-4	336
3.14	0	-2	-2	-2	-2	-2	-2	-4	56
4	5	3	1	1	1	1	1	1	7168
4.1	5	3	1	1	1	1	1	-1	336
4.2	5	3	1	1	1	-1	-1	-1	1120
4.3	5	3	1	-1	-1	-1	-1	-1	336
4.4	5	1	1	1	1	1	1	-3	56
4.5	5	1	1	1	1	-1	-1	-3	840
4.6	5	1	1	-1	-1	-1	-1	-3	840
4.7	5	-1	-1	-1	-1	-1	-1	-3	56
4.8	3	1	1	1	1	1	1	-5	56
4.9	3	1	1	1	1	-1	-1	-5	840
4.10	3	1	1	-1	-1	-1	-1	-5	840
4.11	3	-1	-1	-1	-1	-1	-1	-5	56
4.12	1	1	1	1	1	-1	-3	-5	336
4.13	1	1	1	-1	-1	-1	-3	-5	1120
4.14	1	-1	-1	-1	-1	-1	-3	-5	336
5	3	3	3	3	1	1	1	1	8960
5.1	3	3	3	3	1	1	1	1	70
5.2	3	3	3	3	1	1	-1	-1	420
5.3	3	3	3	3	-1	-1	-1	-1	70
5.4	3	3	3	1	1	1	-1	-3	1120
5.5	3	3	3	1	-1	-1	-1	-3	1120
5.6	3	3	1	1	1	1	-3	-3	420
5.7	3	3	1	1	-1	-1	-3	-3	2520
5.8	3	3	-1	-1	-1	-1	-3	-3	420
5.9	3	1	1	1	-1	-3	-3	-3	1120
5.10	3	1	-1	-1	-1	-3	-3	-3	1120
5.11	1	1	1	1	-3	-3	-3	-3	70
5.12	1	1	-1	-1	-3	-3	-3	-3	420
5.13	-1	-1	-1	-1	-3	-3	-3	-3	70
Total									30240

TABLEAU 3.1 - *Leader des sphères 1, 2 et 5 du réseau* RE_8

- **Mode 3**: Le quantificateur Q3 correspond au débit le plus élevé, avec un dictionnaire de taille 2^{14}. Les éléments considérés sont des points de la cinquième sphère. Pour ne conserver que le nombre de points souhaité, on ne retient pas la sphère au complet. Seuls les vecteurs des *Leader* impairs (*Leader* 4 et 5) et certains vecteurs pairs (*Leader* 3) sont conservés.

Mode	Dictionnaire (Sphères/*Leader*)	Taille	Débit (bit/dim)
0	Q0=(8,4,4) $\subset$ S1	$16 = 2^4$	0.5
1	Q1={ S1, *Leader* [$2\sqrt{2}$ 0 0 0 0 0 0 0] }	$256 = 2^8$	1
2	Q2= S2-{*Leader* nos 3.1 et 3.8}	$2048 = 2^{11}$	1.375
3	Q3= *Leader* 3, 4 et 5 de la sphère S5	$16352 \approx 2^{14}$	1.75

TABLEAU 3.2 - *Dictionnaires sphériques utilisés selon le mode de fonctionnement*

On note que le dictionnaire du quantificateur Q3 n'a pas une taille exactement égale à 2^{14}. Q3 comprend 32 indices non utilisés. Ces positions vides ne représentent que 0.1% de la taille et sont donc négligeables. On pourrait cependant introduire 32 vecteurs supplémentaires pour des configurations particulières (ex: impulsion de pitch).

Performances des quantificateurs Q_i

Le tableau 3.3 montre les performances des quantificateurs utilisés. Ces mesures ont été faites à partir d'une séquence de $16\,10^5$ nombres aléatoires distribués selon la loi normale $\mathcal{N}(0,1)$. Pour chacun des quantificateurs, le tableau 3.3 donne les valeurs suivantes:

- RSB_{SLB}: Limite de théorique de Shannon (*Shannon Lower Bound*) donnant le RSB maximal pour le débit considéré.

- RSB_0: RSB obtenu en quantifiant la séquence par des vecteurs de dimension 8. Chaque vecteur est normalisé par un gain que l'on suppose transmit "gratuitement" sans distorsion.

- RSB_1: RSB obtenu en quantifiant la séquence par des vecteurs de dimension 16 (une sous-trame). On quantifie en décomposant chaque vecteur en deux sous-vecteurs de dimension 8 quantifiés par Q_i. Les deux sous-vecteurs ont la même norme. On calcule le gain global optimal, qui est transmis "gratuitement", sans distorsion.

- On donne en outre les écarts relatifs entre les résultats de quantification et la limite théorique.

Mode	Débit (bits)	RSB_{SLB}(dB)	RSB_0 (dB)	$\frac{\Delta RSB_{(0/SLB)}}{RSB_{SLB}}$	RSB_1 (dB)	$\frac{\Delta RSB_{(1/SLB)}}{RSB_{SLB}}$
Q0	4	3.01	2.46	0.192	2.34	0.228
Q1	8	6.02	5.62	0.066	5.28	0.123
Q2	11	8.28	7.98	0.032	7.33	0.115
Q3	14	10.54	10.30	0.022	9.20	0.127

TABLEAU 3.3 - *Performances des quantificateurs Q0, Q1, Q2 et Q3*

3.6 Composition du train de bits pour chaque mode

Les tableaux 3.4, 3.5, 3.6 et 3.7 donnent pour chaque mode la composition du train de bits pour le canal d'encapsulation (8 Kbits/s supplémentaires par rapport au G.729A). Chaque tableau donne un inventaire des paramètres à transmettre pour chacune des sous-bandes, ainsi que le nombre de bits investis pour leur quantification.

Codeur BF (encapsulation)				Codeur HF				Total
Données	**Sous-trames**		**Trame**	**Paramètres**	**Sous-trames**		**Trame**	
	1	2			1	2		
Pulses	6 IPs	6 IPs		QVAS				
Positions	23	23	46	Q0	4 + 4	4 + 4	16	
et Signes								
Gain code	3	3	6	Gain code	2	2	4	
Gain pitch	3	3	6	Gain pitch	0	0	0	
Total			**58**				**20**	**78**
+ 2 (décision)								**80**

TABLEAU 3.4 - *Mode 0*

Codeur BF (encapsulation)				Codeur HF				Total
Données	**Sous-trames**		**Trame**	**Paramètres**	**Sous-trames**		**Trame**	
	1	2			1	2		
Pulses	3 IPs	3 IPs		QVAS				
Positions	11	11	22	Q1	8 + 8	8 + 8	32	
et Signes	3	3	6					
Gain code	2	2	4	Gain code	3	3	6	
Gain pitch	2	2	4	Gain pitch	2	2	4	
Total			**36**				**42**	**78**
+ 2 (décision)								**80**

TABLEAU 3.5 - *Mode 1*

Codeur BF (encapsulation)				Codeur HF				Total
Données	**Sous-trames**		**Trame**	**Paramètres**	**Sous-trames**		**Trame**	
	1	2			1	2		
Pulses	2 IPs	2 IPs		QVAS				
Positions	6	6	12	Q2	11+11	11+11	44	
Signes	2	2	4					
Gain code	2	2	4	Gain code	3	3	6	
Gain pitch	2	2	4	Gain pitch	2	2	4	
Total			**24**				**54**	**78**
+ 2 (décision)								**80**

TABLEAU 3.6 - *Mode 2*

Codeur BF (encapsulation)				Codeur HF				Total
Données	**Sous-trames**		**Trame**	**Paramètres**	**Sous-trames**		**Trame**	
	1	2			1	2		
Pulses	1 Imp	1 Imp		QVAS				
Positions	3	3	6	Q3	14+14	14+14	56	
Signes	1	1	2					
Gain code	2	2	4	Gain code	5	5	10	
Gain pitch	0	0	0	Gain pitch	0	0	0	
Total			**12**				**66**	**78**
+ 2 (décision)								**80**

TABLEAU 3.7 - *Mode 3*

Chapitre 4

Performances du codeur large bande

On présente dans ce chapitre des commentaires sur les performances du codeur large bande obtenu par encapsulation du G.729A. On expose divers types de résultats. On considère avant tout des mesures quantitatives. On fournit pour se faire, dans chaque cas, les résultats des calculs de rapports signal à bruit entre la sortie du codeur et le signal d'entrée pour différents fichiers. Des observations qualitatives sur les signaux d'erreur sont ensuite décrites. On expose enfin quelques commentaires subjectifs sur la qualité du signal généré par le codeur.

4.1 Banque de fichiers test

L'exposé des résultats a été bâti après avoir effectué les opérations de codage-décodage sur chacun des 14 fichiers de sons. Cette banque de sources comprend avant tout 6 fichiers de parole formés de séquences prononcées dans différentes langues (français, anglais et hongrois), par différents locuteurs féminin et masculin. Outre les fichiers de parole, 8 fichiers complémentaires ont été traités par le codeur. On essaie de quantifier différents types de

musique pour voir si le système initialement conçu pour synthétiser de la parole, peut être jugé fiable pour d'autres sources.

4.2 Analyse quantitative

À partir du codage de chaque fichier de la banque qui vient d'être définie, on calcule quatres valeurs de rapport signal à bruit (RSB):

- RSB du G.729A dans la bande inférieure. Cette valeur donne une mesure objective de la quantification du signal par le G.729A utilisé seul.
- RSB dans la bande basse pour le codeur global. Cette mesure met en évidence l'apport du second étage dans le codage dans la bande inférieure.
- RSB dans la bande hautes fréquences.
- RSB sur la totalité du spectre (large-bande).

Le tableau 4.1 donne les résultats obtenus pour chacun des fichiers. Les conclusions sont variables selon la nature du signal et la contribution apportée par le G.729A. Pour la parole on obtient une amélioration de 1 à 3 dB dans la bande basse pour un RSB global de l'ordre de 16 dB. Les hautes fréquences sont elles quantifiées avec un rapport signal à bruit proche de 7 à 8 dB.

En ce qui concerne les fichiers de musique, le comportement du codeur est très inégal. Dans certains cas, comme pour le fichier de piano, la totalité du débit de codage est investi dans la bande basse (mode 0). Le signal basses fréquences est sensiblement amélioré avec un RSB augmenté de 4.7 dB. En revanche, pour d'autres situations comme celle du fichier de

Fichiers	Bande basses fréquences		Hautes fréquences	-Total-
	Codeur G.729A seul	G.729A encapsulé		
Voix d'homme 1	11.79	14.02	6.43	14.08
Voix de femme 1	14.63	16.24	8.00	16.20
Voix d'homme 2	13.68	16.46	7.71	16.57
Voix de femme 2	15.95	17.89	7.38	17.96
Voix d'homme 3	14.74	17.07	7.28	17.12
Voix de femme 3	12.17	13.18	6.43	13.00
Chanteuse	15.42	16.57	7.29	16.42
Chanteurs	9.64	12.26	6.78	12.47
Musique pop	8.71	10.23	6.14	10.29
Instruments à vent	10.86	13.11	6.54	13.12
Piano	10.15	14.89	3.87	14.94
Orgue	7.75	8.80	10.27	9.00
Musique rock	6.35	7.00	8.90	7.23
Musique slow	6.52	7.77	8.1	7.86

TABLEAU 4.1 - *Rapports signal à bruit pour différents fichiers*

musique rock, la bande supérieure du spectre consomme beaucoup de débit laissant la bande inférieure avec un niveau de qualité médiocre.

4.3 Observations sur le bruit de codage

Les figures 4.1 à 4.5 présentent des courbes permettant d'observer le bruit de codage introduit par le système encapsulé pour différents types de signaux. Pour chaque exemple on donne une représentation du signal temporel ainsi qu'un graphique mettant en évidence les densités spectrales de puissance pour le signal original ainsi que pour le bruit de quantification

(signal d'erreur).

La figure 4.1 correspond au cas d'un phonème voisé. Dans une telle situation, le mode de fonctionnement sélectionné est le mode 3. Le débit d'encapsulation est essentiellement investi dans les hautes fréquences. On constate en effet que dans la bande hautes fréquences, où se situe la plus grande partie l'énergie du signal, la courbe de bruit est bien en dessous du signal.

Dans le cas des figures 4.2 et 4.3, on s'intéresse au comportement du système pour un phonème voisé. Le mode sélectionné dans un tel contexte est le mode 0 ou le mode 1. Les courbes 4.2 donnent les résultats pour une trame voisée prise sur toute la bande. On constante que les hautes fréquences ont été mal quantifiées. Cependant le bruit introduit, bien qu'audible, a la forme spectrale du signal, ce qui le rend moins désagréable. La figure 4.2 présente un cas similaire, où on n'observe le spectre que dans la bande inférieure (fréquences inférieures à 4000 Hz). On met ainsi en relief le travail effectué par le codeur dans la plage spectrale contenant l'essentiel de l'information (bande de base). On observe, sans surprise que le résultat est très proche de ce que l'on peut obtenir avec le G.729A seul. L'amélioration apportée est une petite réduction du plancher de bruit (quelques dB).

Les graphiques des figures 4.4 et 4.5 permettent de rendre compte de l'efficacité du codeur lorsque le signal introduit est de la musique. Le fichier choisi est un morceau de piano. La trame étudiée présente plusieurs tons correspondant à des notes de musique. La bande de fréquences observée est la bande basse (0-4000Hz). Le mode sélectionné par le système dans ce type de scénario est le mode 0. Pour cette raison, on ne s'intéresse pas au bruit de confort introduit dans la bande supérieure. Les courbes 4.4 représentent le signal temporel ainsi que les courbes spectrales relatives au bruit du codeur. Les courbes 4.5 permettent de faire des comparaisons avec le signal de synthèse généré par le G.729A.

L'analyse des graphiques 4.4 et 4.5 permet d'apprécier les améliorations apportées par le

second étage du codeur ACELP dans la bande inférieure. On conclut d'abord que le plancher de bruit est globalement diminué. On constate ensuite que le bruit entre deux tons est souvent fortement atténué. Le bruit dans une telle bande est en principe masqué par la présence des tons. Cependant, le G.729A introduit parfois trop de bruit entre les formants. Ceci est une conséquence de l'ordre trop faible du filtre LPC utilisé (ordre 10). Un tel filtre est insuffisant pour modéliser avec fidélité la totalité d'un spectre de musique.

4.4 Résultats subjectifs

Les commentaires précédant ont permis de mesurer de manière objective la quantification des signaux sonores par le codeur encapsulé large bande. On s'intéresse à présent à des mesures subjectives. Ces résultats sont basés sur des écoutes faites à partir des signaux quantifiés.

Le tableau 4.2 donne les résultats des écoutes. Pour chaque fichier, on affecte une note subjective. On utilise une échelle de cinq notes possibles, définies comme suit:

- **Excellent**: Il est impossible de reconnaître la version quantifiée par rapport à l'original. Malheureusement, aucun fichier n'a pu vérifier ce critère de manière parfaitement fiable.
- **Très bon**: Le fichier quantifié est de très bonne qualité sur toute la bande 80-5500Hz.
- **Bon**: La qualité d'écoute est globalement satisfaisante mais quelques défauts peuvent parfois apparaître.
- **Passable**: Le signal quantifié est bien différentiable du signal original. L'écoute reste cependant agréable.

- **Mauvais:** Le résultat du codage est insuffisant. Des défauts majeurs peuvent apparaître.

Le test d'écoute n'a été effectué que par un seul auditeur. L'objectif de ce test n'est pas de prétendre fournir une mesure subjective fiable de type MOS (*Mean Opinion Score*), mais simplement de donner une première idée générale de la qualité d'écoute des résultats de codage.

Fichiers	Notes	Fichiers	Notes
Voix d'homme 1	Très bon	Piano	Bon
Voix de femme 1	Très bon	Musique pop	Assez bon
Voix d'homme 2	Très bon	Instruments à vent	Passable
Voix de femme 2	Très bon	Orgue	Mauvais
Voix d'homme 3	Très bon	Musique rock	Passable
Voix de femme 3	Bon	Musique slow	Assez bon
Chanteuse	Bon	Chanteurs	Passable

TABLEAU 4.2 - *Tests subjectifs sur différents fichiers*

La qualité d'écoute est en générale bonne ou très bonne pour les fichiers de parole. Pour les autres signaux les conclusions sont parfois moins encourageantes. Le cas de l'orgue, par exemple, est problématique. Les tons présents en hautes fréquences sont mal représentés et le bruit de quantification a généralement un niveau beaucoup trop élevé entre deux pics spectraux consécutifs.

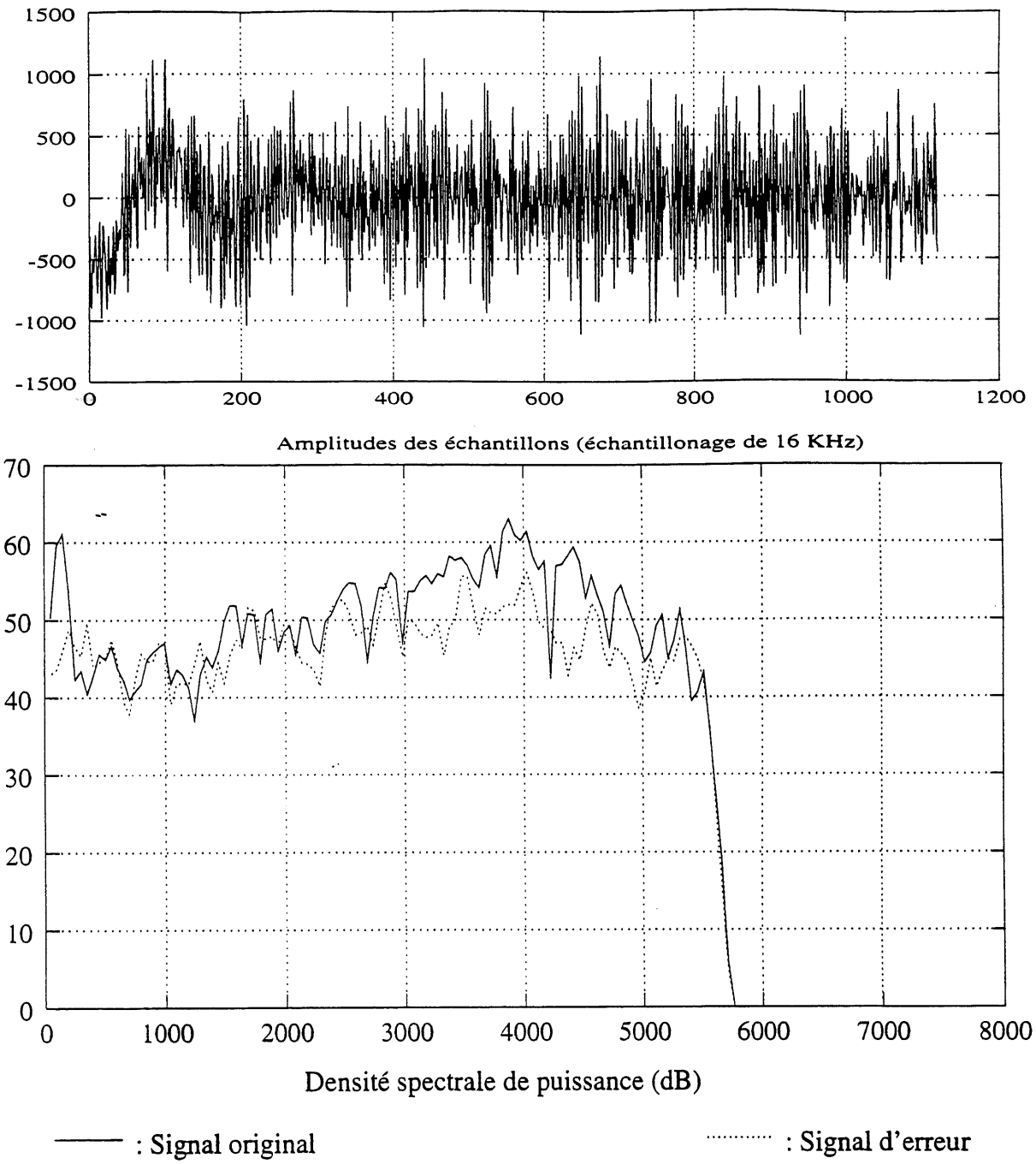

Figure 4.1 - *Cas d'un phonème non-voisé ([S])*

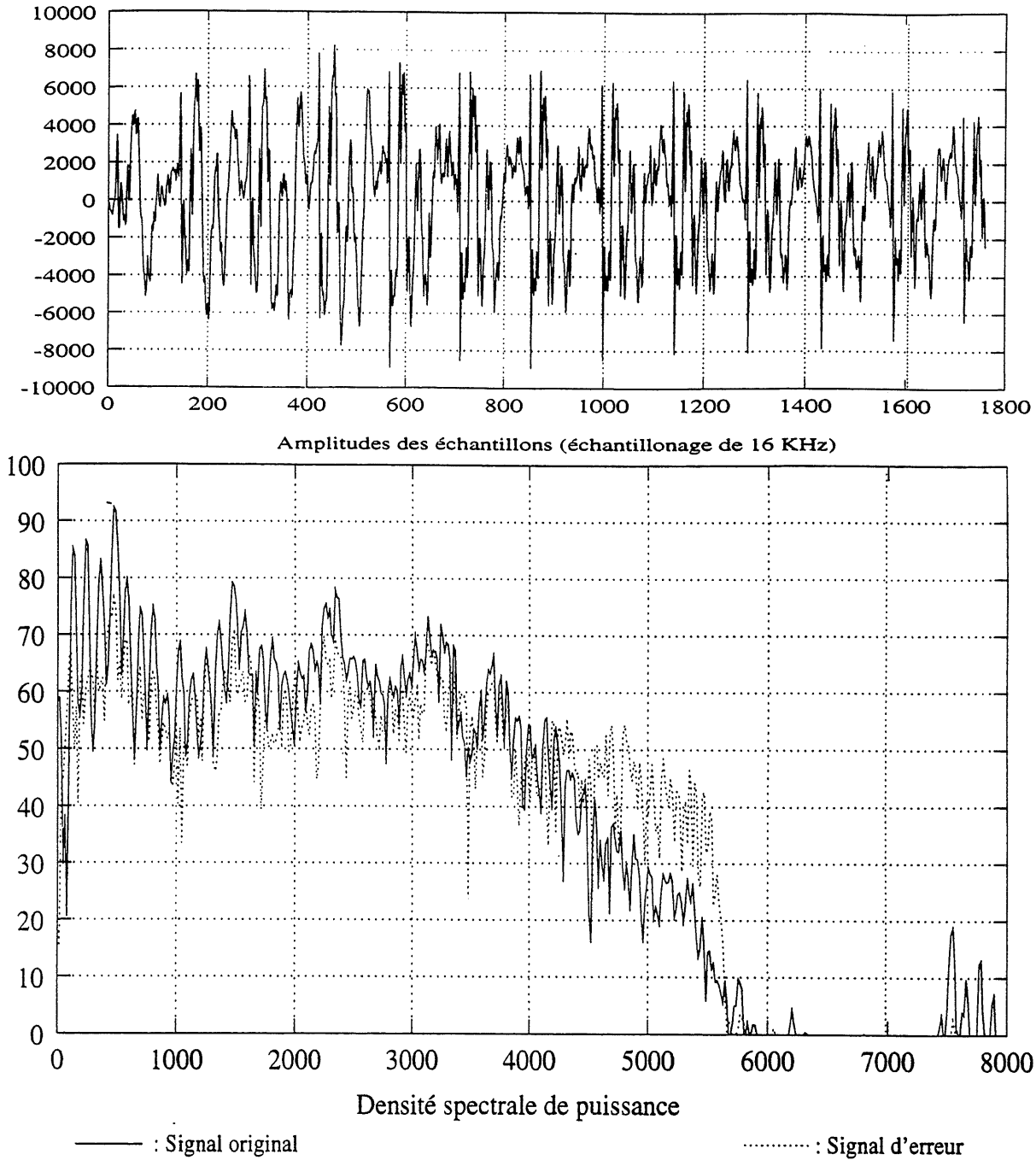

Figure 4.2 - *Cas d'un phonème voisé*

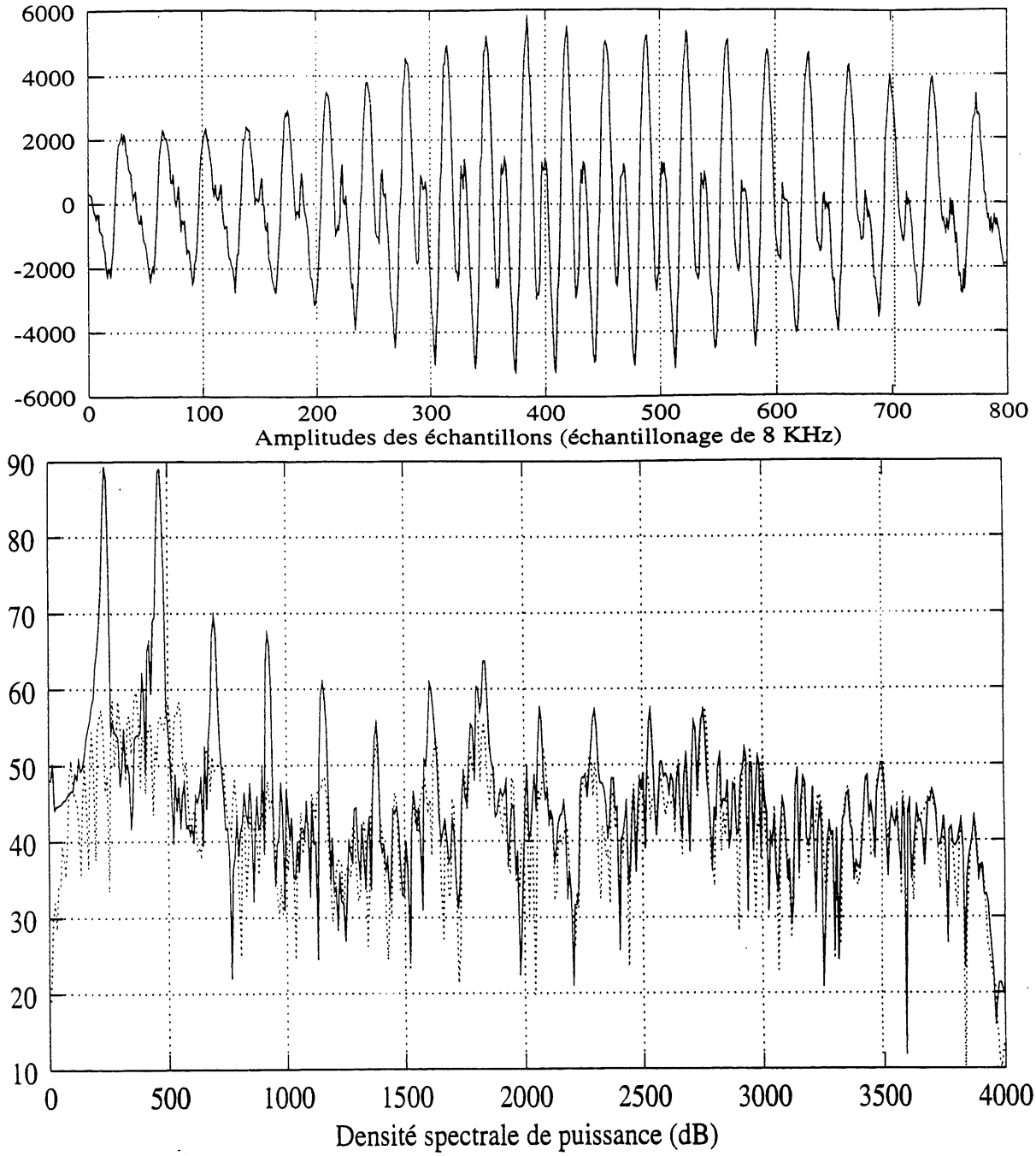

Figure 4.3 - *Observation de la bande inférieure (cas voisé)*

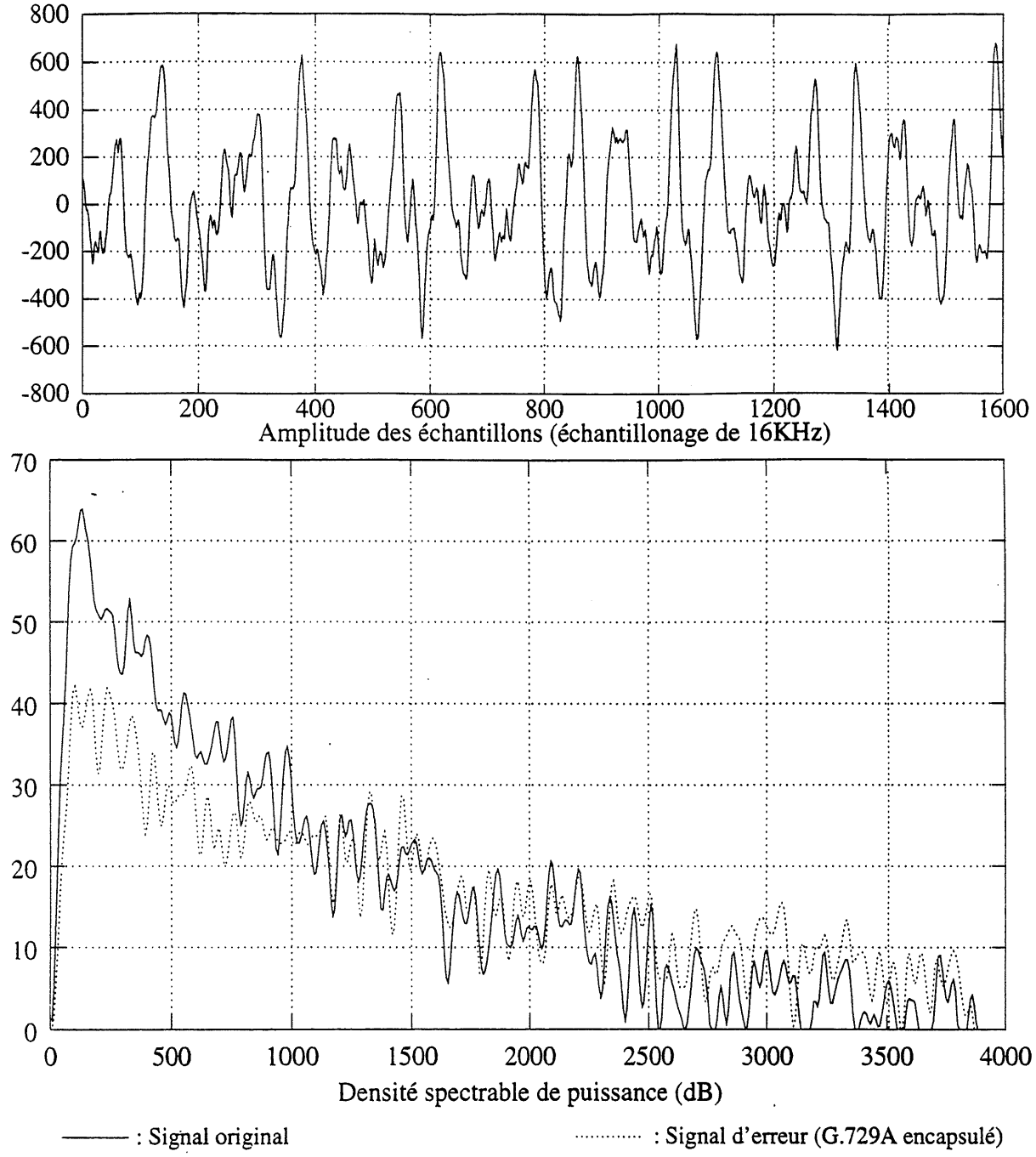

Figure 4.4 - *Résultat du codage pour de la musique (bande inférieure)*

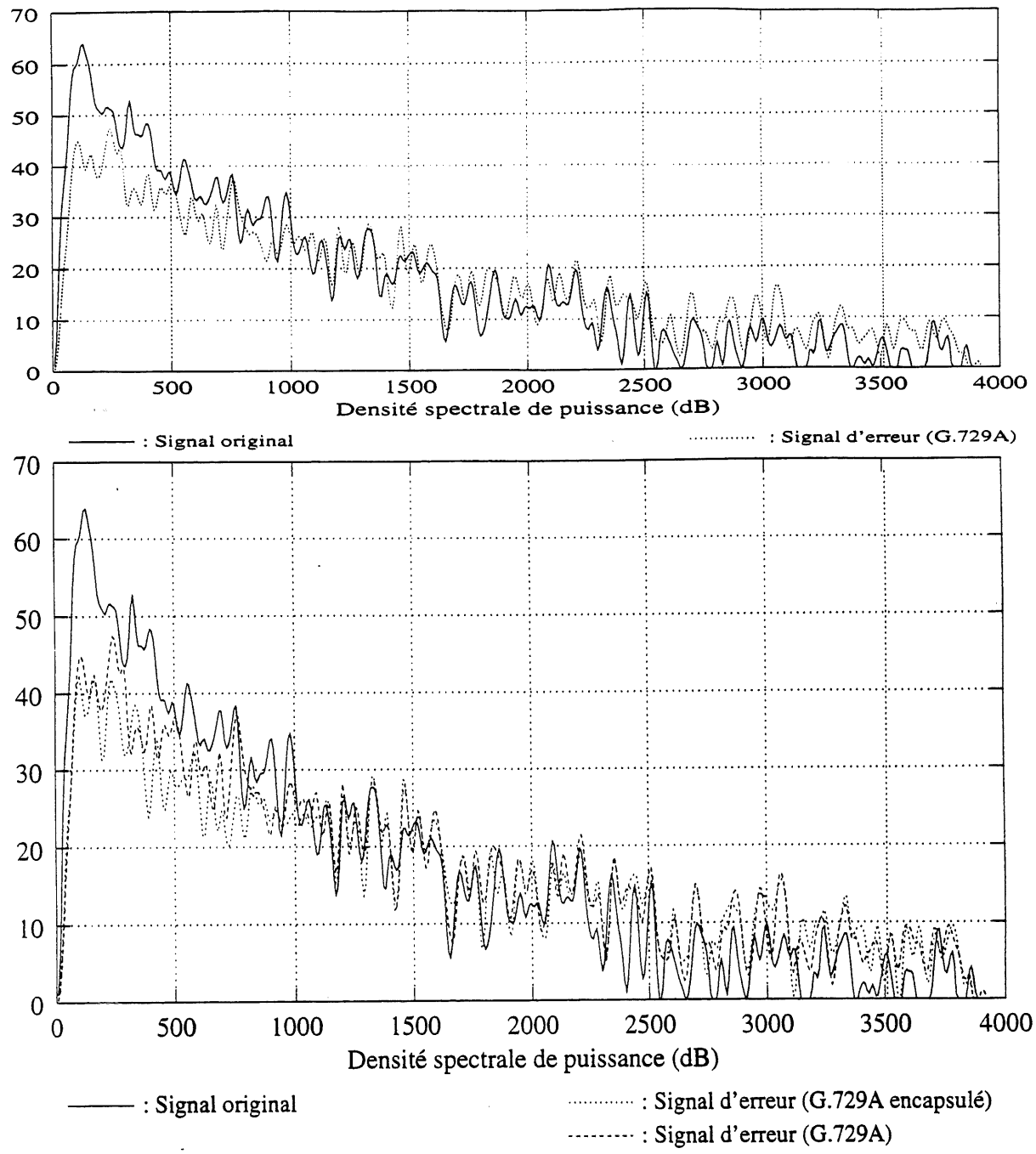

Figure 4.5 - *Codage de la musique: réduction du plancher de bruit*

Chapitre 5

Conclusion

Le principe d'encapsulation qui a été présenté permet de construire un codeur large bande de bonne qualité pour la parole. Les résultats sont moins encourageant lorsque l'on cherche à coder des signaux plus complexes comme la musique. Cet obstacle est sans grosse surprise. Il est encore difficile de construire un codeur large bande fiable avec un débit aussi faible que 16 Kb/s. Par ailleurs, dans le cas d'un codeur encastré, certains axes de traitement sont imposés par le codeur bas débit. Dans la situation qui nous intéresse, le G.729A apporte une contribution insuffisante lorsque l'on cherche à synthétiser certains types de musique. La contribution du codeur ACELP dans ce cas ne justifie pas le débit qu'il consomme.

Le codeur large bande permet cependant d'offrir une qualité de parole de haute qualité. L'extension de la bande de fréquence à la plage 80-5600Hz est très appréciable et le gain de qualité par rapport au G.729A est sensible. On précise de plus que le système d'encapsulation consomme une quantité de calculs modérée. Au niveau du codeur, certains calculs sont obtenus gratuitement grâce au travail déjà effectué par l'encodeur du G.729A.

Une autre approche au codage large bande par encapsulation consiste à quantifier le

signal d'erreur à 16 KHz par un codage par transformée de type TCX (*Transformed Coded eXcitation*). Le travail vise alors à sélectionner les bandes de fréquences à quantifier. Ces dernières correspondent aux parties du spectre qui ont été mal quantifiées ou pas du tout traitées (extension de la bande) par le codeur ACELP. Un tel modèle a été étudié au cours de ce projet. Cette méthode n'a cependant pas été retenue parce que considérée comme pas assez efficace à bas débit. Une autre étude, détaillée dans la référence [20], utilise un modèle basé sur un codage par transformée du signal d'erreur. Malheureusement le modèle proposé limite la bande de codage à 4 KHz ce qui ne permet pas de considérer un tel système comme étant un codeur large bande.

Critères	G.729A	G.723.1
Délais	15 ms	37.5 ms
Débit	8 Kb/s	5.3 Kb/s
Complexité	10 MIPS	20 MIPS
Débit d'encapsulation	**8 Kb/s**	**10.6 Kb/s**

TABLEAU 5.1 - *Comparaison G.723.1/G.729A pour une encapsulation large bande*

Précisions enfin que le modèle proposé dans ce projet pourrait très bien être adapté à d'autres codeurs ACELP. Une attention plus particulière est à porter à la norme G.723.1 à 5.3 Kb/s. Le débit en bande réduite est plus faible que celui consommé par le G.729A. Ce gain en débit se fait au prix d'un plus grand délai de trame (30 ms au lieu de 10 ms) et d'une petite perte de qualité pour le codeur en bande réduite. En revanche, le faible débit utilisé par le G.723.1 laisse une plus grande marge de manoeuvre pour le codeur large bande. Le tableau 5.1 présente les caractéristiques du G.723.1 ainsi que des comparaisons avec le G.729A.

Un codeur large bande obtenu par encapsulation du G.723.1 devrait donner de meilleurs résultats dans le mesure où l'on dispose d'un débit de 10.6 Kb/s pour l'encapsulation. Cependant le codeur en bande réduite serait moins performant.

BIBLIOGRAPHIE

[1] S. SASAKI et S. HAYASHI A. KATAOKA, S. KURIHARA. A 16-kits/s wideband speech codec scalable with G.729. Dans *ESCA. Eurospeech*, pages 1491–1494, 1997.

[2] F. Itakura. Line spectrum representation of linear predictive coefficients of speech signals. Dans *J. Acoust. Soc. Amer.*, volume 57, page S35, avril 1975.

[3] ITU-T, Draft Recommendation G.722. *7 kHz audio-coding within 64 kbit/s*, 1988.

[4] ITU-T, Draft Recommendation G.726. *40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (*ADPCM*)*, Décembre 1990.

[5] ITU-T, Draft Recommendation G.729. *Coding of Speech at 8 Kbits/s Using Conjugate-Structure Algebraic-Code-Excited Linear-Predictive (*CS-ACELP*) Coding.*, Mars 1996.

[6] ITU-T, Draft Recommendation G.729 Annex A. *Reduced Complexity 9 kbits/s* CS-ACELP *Speech Coder.*, Novembre 1996.

[7] ITU-T, Draft Recommendation G.729 Annex B. *A Silence Compression Scheme for* G.729 *Optimized for Terminals Conforming to* ITU-T V.720., Novembre 1996.

[8] C. LAFLAMME et J-P ADOUL R. SALAMI. Enhanced full rate speech codec for IS-136 digital cellular system. Dans *ICASSP*, pages 731–734, 1997.

[9] Claude LAFLAMME Roch LEFEBVRE. Shaping coding noise with frequency-domain companding. Dans *ICASSP*, volume 1, pages 335–339, 1997.

[10] Bishnu S. ATAL et Manfred R. SCHROEDER. Code-excited linear prediction (CELP): high-quality speech at very low bit rates. Dans *ICASSP*, volume 3, pages 937–940, 1985.

[11] Adil BENYASSINE Eyal SHLOMOT Huan-yu SU Dominique MASSALOUX Claude LAMBLIN Jean-Pierre PETIT. ITU-T recommendation G.729 annex B: A silence compression scheme for use with G.729 optimized for V.70 Digital Simultaneous Voice and Data Applications. *IEEE Communications Magazine*, pages 64–72, September 1997.

[12] Toby BERGER. *Rate Distortion Theory.* Prentice-Hall, 1971.

[13] Bruno BESSETTE Claude LAFLAMME Jean-Pierre ADOUL Redwan SALAMI. ITU-T G.729 annex A: Reduced complexity 8 KB/S CS-ACELP. codec for digital simultaneous voice and data. *IEEE Communications Magazine*, pages 56–63, September 1997.

[14] Edward J. DUDEWICZ. *Introduction to Statistics and Probability.* Holt, Rinehart and Winston, 1976.

[15] Allen GERSHO et Robert M. GRAY. *Vector quantization and signal compression.* Kluwer academic publishers, 1992.

[16] Ira A. GERSON et Mark A. JASIUK. Vector sum excited linear prediction (VSELP). Dans *IEEE Workshop on speech coding for telecommunications*, pages 66–68, 1989.

[17] W.B. KLEIJN et K.K. PALIWAL. *Speech coding and synthesis.* Elsevier, 1995.

[18] Claude LAMBLIN. *Quantification vectorielle algébrique sphérique par le réseau de Barnes-Wall: application au codage de parole.* Thèse de doctorat, Université de Sherbrooke, 1988.

[19] Valerio MUZZOLINI. Discrimination automatique voisé / non-voisé / silence / modem. détecteur d'activité de voie téléphonique. Mémoire de maîtrise, Université de Sherbrooke, 1979.

[20] Sean A. RAMPRASHAD. A two stage hybrid embedded speech/audio coding structure. Dans *ICASSP*, pages 337–340, 1998.

[21] Anil UBALE et Allen GERSHO. A multi-band CELP wideband speech coder. Dans *ICASSP*, pages 1367–1370, 1997.

[22] Minjie XIE. *Quantification vectorielle algébrique et codage de la parole en bande élargie.* Thèse de doctorat, Université de Sherbrooke, 1996.

[23] Minjie XIE et Jean-Pierre ADOUL. Embedded algebraic vector quantizers (EAVQ) with application to wideband speech coding. Dans *ICASSP*, pages 240–244, 1996.