REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Ecole Nationale Polytechnique

D. E. R. Génie Electrique & Informatique

Département d'Electronique

Option : Télécommunications

Thème

Conception et réalisation d'un codeur/décodeur de la parole à large bande (50 - 7000 Hz) et à faible débit (13 kbits/s)

Etudié par : **Mr. M. OULD-CHEIKH** — **Ingénieur d'Etat en Electronique**

Soutenance publiquement le : 23 / 06 / 1999 — devant le jury composé de :

Président :

Mr. D. BERKANI	**Professeur**	**ENP**

Rapporteurs :

Dr. M. GUERTI	**Maître de conférences**	**ENP**
Dr. M. HALIMI	**Chargé de recherche**	**CDTA**

Examinateurs :

Mr. A. GUESSOUM	**Maître de conférences**	**Univ. Blida**
Mr. A. BELOUCHERANI	**Docteur - Enseignant**	**EMP**
Mr B. BOUDRAA	**Chargé de recherches**	**USTHB**

Invité d'honneur :

Mr. A. OULD-ALI	**Docteur – Enseignant**	**EMP**

Ecole Nationale Polytechnique, BP 182 16200, El Harrach , Alger

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Ecole Nationale Polytechnique
D. E. R. Génie Electrique & Informatique
Département d'Electronique
Option : Télécommunications

Thème

Conception et réalisation d'un codeur/décodeur de la parole à large bande (50 - 7000 Hz) et à faible débit (13 kbits/s)

Etudié par : **Mr. M. OULD-CHEIKH** — **Ingénieur d'Etat en Electronique**

Soutenance publiquement le : 23 / 06 / 1999 — devant le jury composé de :

Président :		
Mr. D. BERKANI	**Professeur**	**ENP**
Rapporteurs :		
Dr. M. GUERTI	**Maître de conférences**	**ENP**
Dr. M. HALIMI	**Chargé de recherche**	**CDTA**
Examinateurs :		
Mr. A. GUESSOUM	**Maître de conférences**	**Univ. Blida**
Mr. A. BELOUCHERANI	**Docteur - Enseignant**	**ENP**
Mr B. BOUDRAA	**Chargé de recherches**	**USTHB**
Invité d'honneur :		
Mr. A. OULD-ALI	**Docteur – Enseignant**	**EMP**

Ecole Nationale Polytechnique, BP 182 16200, El Harrach , Alger

Résumé

Un codeur de parole CELP (Code Excited Linear Prediction) large bande a été développé pour un débit inférieur à 13 kbits/s. La complexité de recherche du meilleur code vecteur a été réduite grâce à la transformation des opérations de filtrage en produits matriciels (soustraction de la 'Réponse à entrée nulle'). De plus, nous avons implémenté la technique dite 'Backward Filtering' qui permet de diminuer la complexité de recherche du vecteur d'excitation d'un facteur d'ordre 'p' dans le cas des dictionnaires pauvres en échantillons (Sparse codebook). Nous avons un filtre prédicteur d'ordre élevé (pitch) afin de restaurer la périodicité du signal vocal. Ensuite, nous avons utilisé une quantification scalaire adaptative pour coder les paramètres LSF (Line Spectral Frequencies).

Mots clés : *Codage de la parole, Algorithme CELP, Prédicteur pitch, Backward filtering.*

Abstract

A 13 kb/s wideband CELP speech encoder was developed. This is an area of increasing growth and interest due to some emerging applications like :multimedia devices, videoconferencing, ISDN applications, etc; these scenarios require high-quality speech without the constraint of the limited telephonic bandwidth. Thus, the bandwidth considered in those applications goes from very low frequencies (around 50 Hz) up to 7000 Hz. The sampling frequency typically used is 16 kHz, although higher sampling frequencies are under consideration for some applications. The research goal consists of reducing the bit rate while maintaining the subjective quality. One way to approach the problem is to extend the telephonic bandwidth schemes to this scenario, tuning them to handle chiefly speech, but also music. The CELP algorithm is used for achieving a toll quality of speech at 13 kb/s. We have introduced a pitch predictor to restore the periodicity of speech signal. In order to reduce the computational complexity, we used an algebraic codebook and the Backward Filtering technique.

Keywords : *Speech coding, CELP algorithm, Pitch predictor, Backward filtering.*

Abstract

A 16 kb/s wideband CELP speech encoder was developped. The CELP encoder is used as a narrowband encoder at low and medium bit rates in many applications, e.g. cellular mobil. This is an area of increasing growth and interest due to some emerging applications like :multimedia devices, videoconferencing, ISDN applications, etc; these scenarios require high-quality speech without the constraint of the limited telephonic bandwidth. Thus, the bandwidth considered in those applications goes from very low frequencies (around 50 Hz) up to 7000 Hz. The sampling frequency typically used is 16 kHz, although higher sampling frequencies are under consideration for some applications. The low frequencies added increase the sensation of voice naturalness whereas the extra high frequency range increases the voice intelligibility and speaker recognition capability. Although the aim of these techniques is to code wideband speech, it is assumed that the proposed algorithms must also show a nice performance when coding music signals up to 7000 Hz.

The reference encoder for this bandwidth is the ITU-T standard G.722, a subband-ADPCM encoder with bit rates varying between 64 Kbps, 54 Kbps, and 48 Kbps. The research goal consists of reducing the bit rate while maintaining the subjective quality. One way to approach the problem is to extend the telephonic bandwidth schemes to this scenario, tuning them to handle chiefly speech, but also music.

The CELP algorithm is used for achieving a toll quality of speech at 16 Kbps. We have introduced a pitch predictor to restore the periodicity of speech signal. In order to reduce the computational complexity, we used an algebraic codebook and the Backward Filtering technique.

Objective and subjective measures show that the encoder achieves good synthetic speech quality.

A 16 kb/s wideband CELP encoder is used for multimedia devices and may be used as compression program for speech signals on storage supports.

Résumé

Un codeur de parole CELP large bande à un débit inférieur à 16 kbits/s a été réalisé. Le codeur CELP est utilisé comme un codeur à bande étroite à moyenne et faible débit dans plusieurs applications par exemple les télécommunications mobiles.

C'est un domaine de développement croissant et intéressant dû à plusieurs applications émergentes comme : le multimédia, la vidéoconférence, les applications ISDN (Integrated Services Digital Network), ... etc ; Ces applications demandent une parole de haute qualité sans la contrainte de la largeur de bande téléphonique limitée. Ainsi, la largeur de bande considérée dans plusieurs applications s'étend des fréquences très basses (autour de 50 Hz) jusqu'à 7000 Hz. La fréquence d'échantillonnage utilisée est de 16 kHz, bien que des fréquences d'échantillonnage plus élevées soient sous études pour quelques applications. L'ajout des fréquences basses augmente la sensation du confort d'écoute alors que les fréquences hautes supplémentaires augmentent l'intelligibilité de la voix et la capacité de reconnaissance du locuteur. Bien que le but de ces applications est de coder la parole large bande, il est admis que les algorithmes proposés donnent une bonne performance pour le codage des signaux de musique s'étendant jusqu'à 7000 Hz.

Le codeur de référence pour cette largeur de bande est le standard G.722 d'ITU-T, un codeur ADPCM en sous-bande avec des débits variant entre 64 kb/s, 54 kb/s, et 48 kb/s. Le but essentiel consiste à réduire le débit binaire tout en maintenant une bonne qualité subjective. L'une des approches du problème est l'extension de la largeur de bande téléphonique.

L'algorithme CELP a été utilisé pour atteindre une parole de bonne qualité à 16 kb/s. Un prédicteur pitch a été introduit afin de reproduire la périodicité du signal de parole. Afin de réduire la complexité de recherche, nous avons utilisé un dictionnaire algébrique (ternaire) et la méthode du "Backward Filtering".

Les mesures objectives et subjectives montrent que la qualité de la parole synthétique est de bonne qualité.

Le codeur CELP large bande à 16 kbits/s peut être utilisé pour les applications multimedia ou encore comme programme de compression des signaux parole dans les supports de stockage.

REMERCIEMENTS

Je tiens à exprimer ma gratitude au Dr M. HALIMI , Chargé de Recherche au CDTA, et au Dr M. GUERTI , Maître de Conférences à l'ENP. Durant la période du magister j'ai pu profiter de leurs larges visions du traitement du signal ainsi que de la rigueur scientifique dont ils font preuve. Ils ont su guider ce travail avec enthousiasme tout en me laissant une grande liberté. Je les en remercie.

Je tiens à exprimer tous mes remerciements à Monsieur BERKANI D d'avoir accepté de présider le jury de cette thèse.
Messieurs GUESSOUM A, chargé de recherches à l'université de BLIDA, BELOUCHERANI A, Docteur-Enseignant à l'ENP et BOUDRAA B, chargé de recherches à l'USTHB trouvent ici ma gratitude pour avoir bien voulu porter un jugement sur mon travail.

Je tiens à exprimer ma profonde gratitude à Monsieur DAMOU, le directeur de la recherche et de la formation post-graduée à l'EMP, pour ces encouragements, son intérêt et son aide pour que ce travail aboutisse. Je remercie aussi Monsieur GOUIGAH le chef de l'UER Eléctronique à l'EMP pour son soutien et ses précieux conseils. Je remercie aussi Monsieur KELLALI S, le chef de laboratoire des systèmes de communication à l'EMP pour avoir mit à ma disposition les moyens nécessaires pour l'accomplissement de mon travail , je lui en suis très reconnaissant.

Qu'il me soit permis de témoigner ici toute ma reconnaissance et ma profonde gratitude à tous ceux qui ont contribué de près ou de loin à la concrétisation de ce modeste travail.

Finalement, je remercie toute ma famille pour son soutien morale et pour m'avoir toujours préparer les meilleures conditions de vie.

Sommaire

Liste des figures

Introduction Générale

INTRODUCTION GENERALE

Les avantages du codage d'un signal numériquement sont bien connus et sont largement discutés dans la littérature. Brièvement, la représentation numérique offre une grande robustesse, une régénération efficace du signal, un cryptage aisé, une possibilité de combinaison des fonctions de transmission et de réception, et l'avantage d'un format uniforme pour différents types de signaux. Les câbles (différents moyens de« transporteurs ») devraient avoir une bande passante suffisante afin de permettre la transmission des signaux numériques.

Le codage de la parole est essentiel dans les efforts pour obtenir un usage plus efficace des réseaux de télécommunication numériques, en particulier les réseaux cellulaires, et pour réduire la mémoire nécessaire dans les systèmes de stockage de la parole.

Durant ces dernières décennies, il y a eu un grand progrès dans le développement des algorithmes de codage de la parole à faible débit. Les codeurs de parole de bonne qualité sont maintenant disponibles à des débits de 4.8 kb/s (FS1016). Les efforts des chercheurs ont été concentrés sur les signaux de parole à bande étroite où la bande de transmission est limitée à 300–3400Hz, comme dans le cas des systèmes de téléphone analogique. Cette limitation de la largeur de bande dégrade la qualité du signal. Pour plusieurs applications futures, un élargissement de la largeur de bande est nécessaire dans le but d'atteindre une bonne qualité de communication d'un point à un autre.

L'augmentation de la bande de 50–7000 Hz améliore de façon significative la qualité de la parole. Ceci est en partie du à l'amélioration du confort d'écoute de la parole de l'orateur (augmentation de la bande de 300 à 50 Hz), et partiellement due à une augmentation du naturel fourni par l'accroissement de la haute fréquence (de 3400 à 7000 Hz). La haute qualité de la parole à large bande est généralement désirée dans le cas de l'audioconférence.

Le débit standardisé pour le codage à haute qualité de la parole à 7 kHz est le 64 kbits/s [Recommandation G.722], typiquement pour une application audioconférence en utilisant l'ISDN (Integrated Services Digital Network).

Les progrès réalisés dans les algorithmes de codage de la parole et l'augmentation des capacités de calcul des processeurs permettent de réduire le nombre de bits nécessaires à l'obtention d'une bonne qualité du signal de parole synthétique.

De ce fait, des algorithmes récents ont fournit une bande passante de 7 kHz à un débit de 32 kb/s permettant la stéréo téléconférence. Le débit désiré pour le codage de la parole large bande est de 16 kb/s (ou moins) [Fuldseth, 91][Jayant, 91][Jayant, 90][Kabal, 91].

Les faibles débits pour la parole large bande sont utilisés dans les téléconférences à haute qualité avec l'audio et la vidéo combinés. En général, pour des faibles valeurs de débits comme le 64 kbits/s, la pratique courante est de limiter la parole au téléphone traditionnel de largeur de bande de 4 kHz et d'utiliser le 8 kb/s ou, au plus, le codeur à 16 kb/s pour le codage du canal. Avec les progrès dans la compression de la parole à large bande, le codage à 16 kb/s de l'audio 7kHz est supposé être une composante importante dans la téléconférence, spécialement à de grands débits de l'ISDN (128 et 384 kb/s).

Dans ce mémoire, le but de notre travail est la réalisation d'un codeur / décodeur large bande dont les caractéristiques sont les suivantes :

- débit binaire inférieur à 16 kb/s ;
- largeur de bande : 50 Hz – 7000 Hz.

Une classe importante de codeurs qui utilisent la prédiction linéaire est le codeur analyse par synthèse (LPAS) ou codeur hybride.

Dans le codage LPAS, le signal de parole d'entrée est analysé pour chaque trame et le signal d'excitation est déterminé pour chaque sous trame parole. Généralement, l'analyse par prédiction linéaire est faite sur des trames de parole variant de 160 à 240 échantillons pour des signaux de parole échantillonnés à 16 kHz et le signal d'excitation est déterminé pour des sous trames allant de 40 à 60 échantillons.

Le signal d'excitation est filtré à travers le filtre de synthèse pour produire le signal synthétique. Le signal original est soustrait du signal reconstruit et l'erreur de codage est minimisée en utilisant un critère de pondération quadratique moyenne. Le signal d'excitation qui minimise l'énergie de l'erreur est sélectionné et les paramètres correspondants sont transmis au récepteur.

Le récepteur utilise la même structure de synthèse pour reconstruire le signal synthétique. Puisque la procédure de recherche de la meilleure séquence d'excitation au codeur exige le calcul du signal synthétique, ce type de codage est appelé codage analyse par synthèse.

La raison majeure pour le choix du codage LPAS est qu'il est facile à incorporer dans sa structure la notion du masquage spectral (lié à la perception auditive humaine). Ceci est réalisé par l'utilisation d'une pondération perceptuelle sur le signal d'erreur durant la sélection de la meilleure excitation.

Ce mémoire est constitué de cinq chapitres :

Le **premier** comporte des généralités sur le modèle de production de la parole humaine ainsi que le système auditif humain et un aperçu sur les différentes méthodes de codage de la parole.

Le **deuxième** chapitre est consacré à l'analyse par prédiction linéaire et l'extraction des paramètres LSF(Line Spectral Frequencies).

Le **troisième** chapitre décrit la mise en œuvre du codeur/décodeur large bande avec la procédure de sélection des paramètres à transmettre (LSF, Pitch et dictionnaire).

Le **quatrième** chapitre comporte un aperçu sur la quantification Scalaire et Vectorielle. Une présentation de l'algorithme de quantification des paramètres LSF sera décrite.

Le **cinquième** chapitre expose les résultats et tests objectifs et subjectifs de la qualité de la parole synthétique, ainsi que les formes d'ondes des différentes phrases.

Chapitre 1

Notions fondamentales sur la parole et son codage

1.1 INTRODUCTION

Ce chapitre se compose de deux parties : les généralités et les principales méthodes de codage. Les premières regroupent les notions de production, d'acoustique et de perception de la parole nécessaires à la bonne compréhension de l'évolution des systèmes de codage de la parole. L'exposé de ces méthodes dans la seconde partie reste classique mais est tout particulièrement orienté vers la compréhension et la description des codeurs hybrides en général et en mettant l'accent sur le codeur CELP (Code Excited Linear Prediction).

1.2 MODELE DE PRODUCTION DE LA PAROLE

La parole est généralement générée par expiration de l'air à travers la glotte et le conduit vocal. Le flux d'air, provenant des poumons, est modulé par les vibrations des cordes vocales et la forme du conduit vocal. (Fig. 1.1)

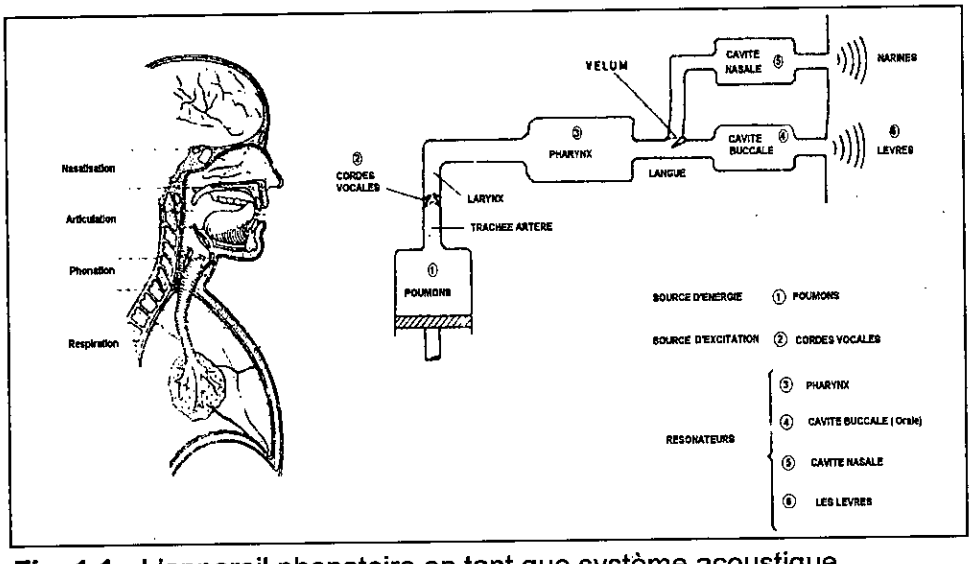


Fig. 1.1 : L'appareil phonatoire en tant que système acoustique.

La parole peut être classée en deux catégories :

voisée qui est caractérisée par sa quasi-périodicité et en général par des segments de son de haute énergie telles que les voyelles (Fig. 1.2).

non voisée qui décrit généralement des segments de faible énergie telles que les consonnes (Fig.1.3).

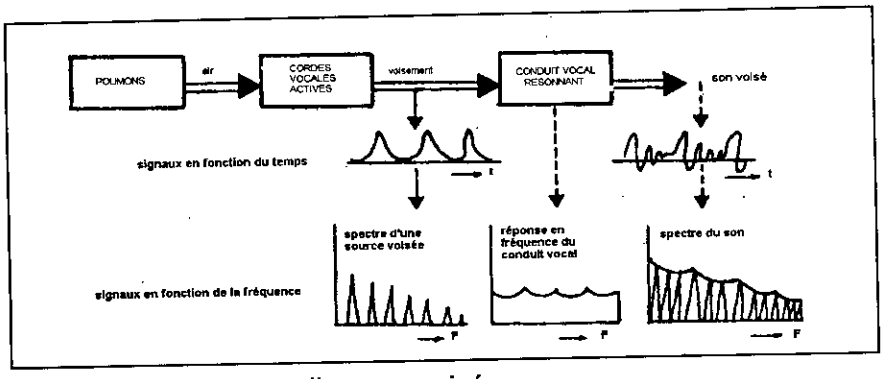


Fig. 1.2 : Production d'un son voisé

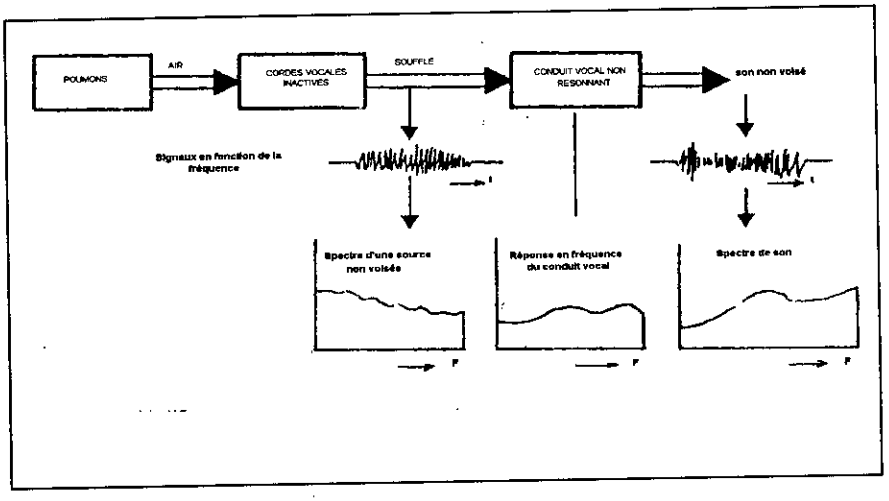


Fig. 1.3 : Production d'un son non voisé

La parole voisée est produite quand la circulation de l'air à partir des poumons est interrompue par une ouverture et une fermeture périodique des cordes vocales, générant ainsi une excitation glottale périodique pour le conduit vocal. Le flux auquel les cordes vocales se ferment et s'ouvrent est appelé *fréquence fondamentale* (dénotée F_0) ou *pitch*. Sa valeur varie avec la dimension des cordes vocales. Les valeurs moyennes typiques sont de 150 Hz et 450 Hz, respectivement, pour l'homme et la femme [fant, 60] [Loo, 1996]. Puisque F_0 et la forme du conduit vocal changent dans le temps, la parole voisée n'est pas vraiment périodique mais peut être caractérisée comme quasi-périodique sur des petits intervalles de temps. La parole non voisée est produite lorsque les cordes vocales ne vibrent pas. Le conduit vocal est ainsi excité par un bruit turbulent généré quand la circulation d'air à partir des poumons passe à travers une constriction étroite dans le conduit vocal.

La production de la parole peut être vue comme une opération de filtrage dans laquelle une source sonore excite un filtre (conduit vocal) [Fant, 60]. La source de son représente un bruit généré par une constriction du conduit vocal durant les sons non voisés ou des impulsions glottales durant les sons voisés, ou une combinaison des deux. Le spectre de la source sonore pour les sons voisés contient des harmoniques espacées de F_0 avec une énergie plus concentrée aux fréquences basses alors que pour les sons non voisés le spectre est approximativement plat et sans structure harmonique.

Le conduit vocal modifie la distribution d'énergie dans le spectre de la source sonore et introduit des résonances (Formants) et des anti-résonances (Antiformants).

Vu que le conduit vocal se comporte comme un filtre variant dans le temps, les résonances et les anti-résonances sont dues, respectivement, aux pôles et aux zéros de la réponse fréquentielle du conduit vocal. La Fig.1.4 représente les formes d'ondes temporelles des segments de parole voisée et non voisée avec leur spectre. On n'omettra pas de souligner le caractère presque aléatoire (bruit) de la forme d'onde non voisée Ffig.1.4 a), comparé au caractère périodique de la forme d'onde voisée (Fig.1.4 c) où les sections de la forme sont répétées approximativement tous les 60 échantillons. La périodicité dans la forme d'onde voisée apparaît aussi dans son spectre (Fig. 1.4 d) en définissant des pics (harmoniques), espacés à la fréquence fondamentale (dans ce cas, 200 Hz). Le spectre du signal non voisé n'a pas une pareille structure et est très aléatoire.

Les codeurs tendent à réduire le débit binaire, tout en préservant la qualité de la parole. Sont essentiellement pris en compte les redondances dans le signal de parole et les limitations perceptuelles de l'oreille humaine [O'Shaughnessy, 1987][Deller, 1993].

A ce niveau, plusieurs observations peuvent être faites :

1- le signal de parole est localement stationnaire ;

2- les périodes du pitch successives sont généralement similaires ;

3- l'énergie du signal de parole est plus concentrée aux fréquences basses pour les sons voisés

Elles sont attribuées aux limitations mécaniques des organes du système phonatoire, c'est à dire le conduit vocal et les cordes vocales. L'oreille humaine est insensible à la phase, plus sensible aux fréquences basses qu'aux fréquences hautes et accorde une plus grande importance aux pôles spectraux qu'aux zéros.

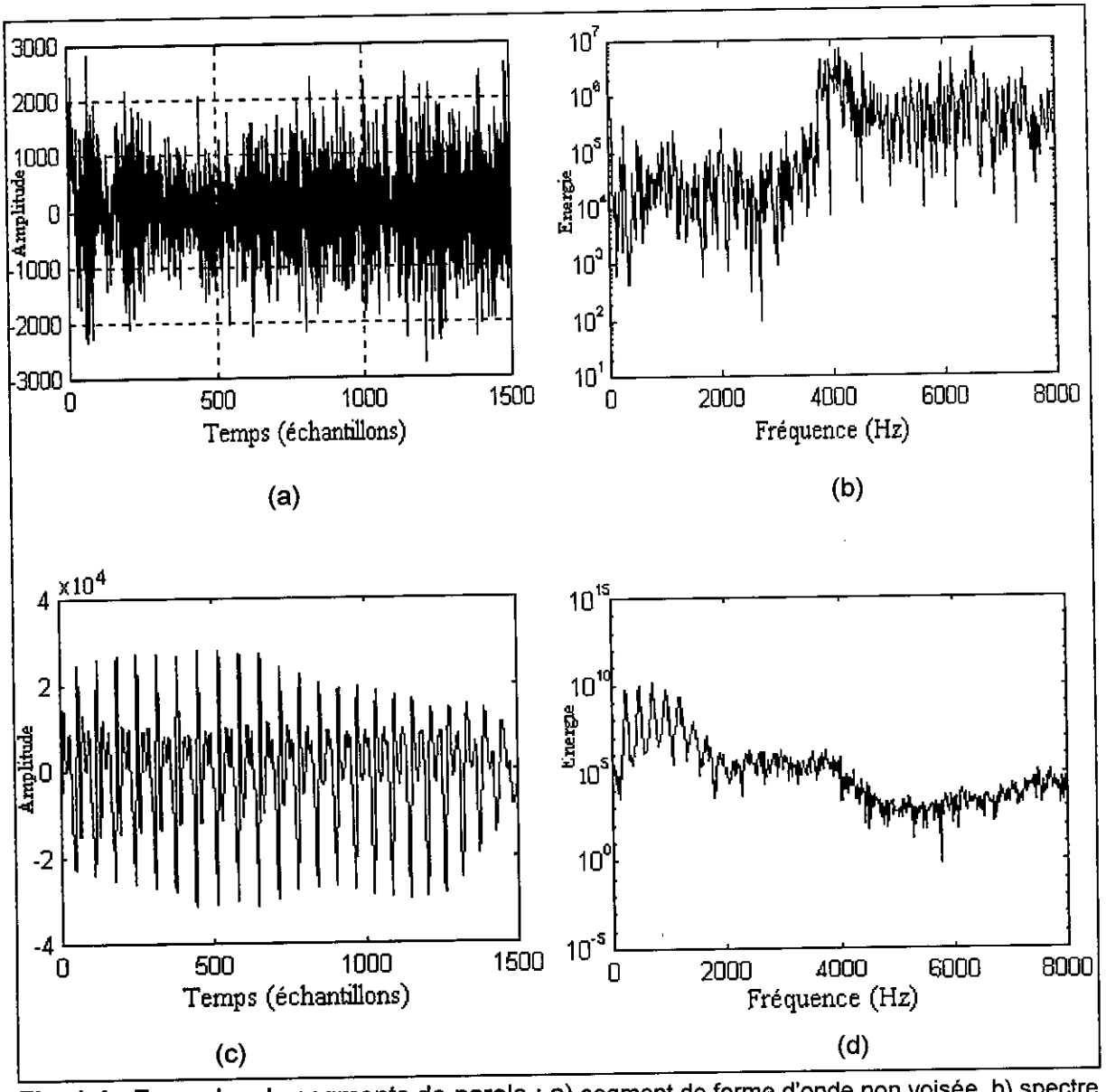


Fig. 1.4 : Exemples de segments de parole : a) segment de forme d'onde non voisée, b) spectre du segment non voisé, c) segment de forme d'onde voisée, d) spectre de forme d'onde voisée

Le signal de parole est échantillonné à 16000 échantillons/ seconde et chaque segment est de longueur de 95 ms.

La redondance dans le signal de parole conduit à la conclusion que les échantillons de parole sont corrélés. L'enveloppe spectrale correspond aux corrélations court terme et la structure harmonique correspond aux corrélations long terme (Cf. chapitre 2).

1.3 PRINCIPALES METHODES DE CODAGE DE LA PAROLE

Dans cette partie, nous présenterons les différentes méthodes de codage de la parole. Nous nous concentrons sur le codage d'analyse par la synthèse et en particulier le codage CELP.

1.3.1 Généralités sur les systèmes de codage

Un signal de parole numérique offre de nombreux avantages tels que l'immunité au bruit, la facilité de stockage, la commodité d'emploi que ce soit pour le multiplexage, le cryptage ou la synthèse.

Nous allons définir les différents éléments constituant un système de compression de parole [Makhoul, 85] [Kailath, 85]

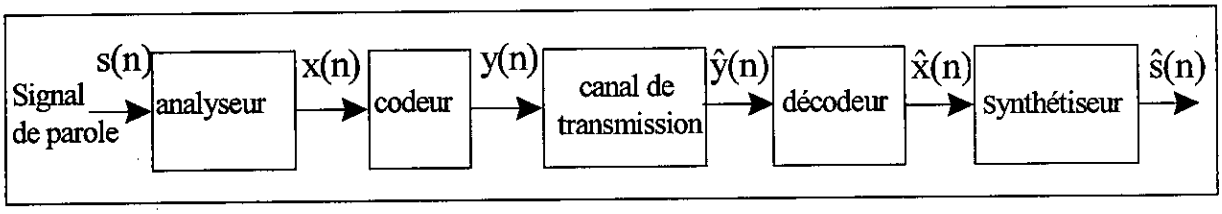


Fig. 1.5 : Système de transmission de la parole

- le premier élément analyse un signal de parole qui aura été filtré et échantillonné au préalable. La sortie de l'analyseur est un vecteur x d'éléments non quantifiés ;
- le codeur quantifie puis code pour la transmission le vecteur **x** ;
- le canal de transmission transmet le vecteur y des éléments codés ;
- le décodeur décode le vecteur reçu $\hat{\mathbf{y}}$ (ou $\hat{\mathbf{y}}=\mathbf{y}$ en l'absence de bruit de transmission) et on extrait un ensemble de paramètres $\hat{\mathbf{x}}(\mathbf{n})$;
- le synthétiseur utilise ce vecteur de paramètres $\hat{\mathbf{x}}(\mathbf{n})$ pour reconstruire le signal de parole.

L'objectif d'un système de compression est de **réduire le débit** binaire exprimé en **bits par seconde (bps)**, pour transmettre l'informations y(n) et tout en gardant une qualité satisfaisante de parole synthétique.

L'analyse du signal détermine l'efficacité du système de compression de parole. Pour les codeurs les plus élémentaires, elle est inexistante. La technique d'analyse à réaliser sur un signal est fixé par la méthode de synthèse. On trouve en général deux éléments pour la

synthèse : la fonction d'excitation et une fonction de transfert. Le synthétiseur détermine le nombre de paramètres nécessaires à la synthèse.Une réduction supplémentaire de débit pourra se faire par un meilleur codage des paramètres [Mauc, 94].

Les deux fonctions principales du codeur sont la quantification et le codage.

La **Quantification Scalaire (QS)** attribue à une valeur de paramètre un nombre choisi dans un ensemble fini et connu de nombres.

La **Quantification Vectorielle (QV)** attribue à un groupe de valeurs un vecteur choisi dans un ensemble fixé de vecteurs appelé **dictionnaire**.

En parole, la *quantification vectorielle* est présentée comme une méthode de suppression des redondances c'est-à-dire des liens qui existent entre les différents paramètres du vecteur [J.Makhoul, 85]. Elle utilise quatre propriétés interdépendantes des paramètres vectoriels :

- la dépendance linéaire (corrélation) ;
- la dépendance non-linéaire ;
- la forme de la fonction de densité de probabilité ;
- la dimension des vecteurs.

Le **codage** traduit les nombres choisis en séquences de nombres binaires qui seront transmises au décodeur. Selon les besoins, chacune de ces opérations peut être améliorée.

La réalisation d'un système de codage efficace dépendra des paramètres suivants :

- le débit de transmission ;
- la qualité de parole synthétique ;
- le coût du système.

1.3.2 Suppression de la redondance dans la parole

La suppression partielle des redondances permet une représentation plus efficace des données. La compression des données peut se faire sans perte d'informations (exemple le codage de Huffmann) ou avec perte d'informations en exploitant dans ce cas la tolérance de l'organe récepteur (exemple l'oreille). Le signal de parole a des caractéristiques particulières. La compression du signal consistera à réduire les redondances qui sont essentiellement [Calliope, 89]:

- le manque de platitude du spectre court terme ;
- la quasi périodicité des signaux voisés ;

- la limitation des formes et des vitesses de mouvements possibles du conduit vocal ;
- les distributions de probabilités non-uniformes des valeurs de paramètres de transmission

Le manque de platitude du spectre court terme est lié au fait que les échantillons de parole adjacents sont corrélés entre eux. On peut décorréler ces échantillons par un filtrage spectral adapté. La quasi périodicité des signaux de parole voisée peut être supprimée en utilisant un prédicteur long terme. La lenteur du conduit vocal permet d'envoyer les paramètres des filtres toutes les 10-30 ms. La dernière des redondances citées peut être exploitée par un codage approprié.

1.3.3 Classification des codeurs

Le classement des codeurs de parole peut se faire selon différentes approches : le débit obtenu, le type de codage...

1.3.3.1 codeurs par formes d'ondes :

Dans cette catégorie, on distingue les codeurs temporels et fréquentiels. Ces derniers n'utilisent aucune connaissance a priori sur la façon dont le signal est généré. Le codeur temporel fait correspondre à l'amplitude du signal analogique une suite d'éléments discrets. Le signal reconstruit est sans doute le plus proche du signal original. Ces codeurs sont conçus pour être indépendants du signal codé et peuvent coder n'importe quel son. Le débit de codage est généralement élevé. En utilisant les propriétés de corrélation du signal, il est possible de diminuer ce débit jusqu'à une certaine limite. En dessous de 16 kbit/s la qualité se dégrade et la réduction de débit en bande étroite (2.4 - 4.8 kbit/s) est peu envisageable.

1.3.3.2 vocodeurs (ou Voicecoder) :

Utilisent une méthode dite par analyse et synthèse, où l'on essaie d'extraire du signal de parole un ensemble de paramètres liés à un modèle simplifié. Ces paramètres sont l'enveloppe du spectre court terme et les informations sur le signal d'excitation (pitch, amplitude). On suppose donc qu'on a des connaissances a priori sur le signal de parole. Ces codeurs sont sensibles aux bruits de transmission et la qualité de la parole est limitée. Le débit de transmission est généralement faible (exemple codeur LPC-10 à 2.4 kbit/s).

1.3.3.3 codeurs hybrides :

Ces codeurs font intervenir les techniques d'analyse par synthèse et les techniques de codage par formes d'ondes. Au prix d'une complexité parfois élevée, ils permettent d'obtenir une bonne qualité de signal à des débits intermédiaires (fig 1.6).

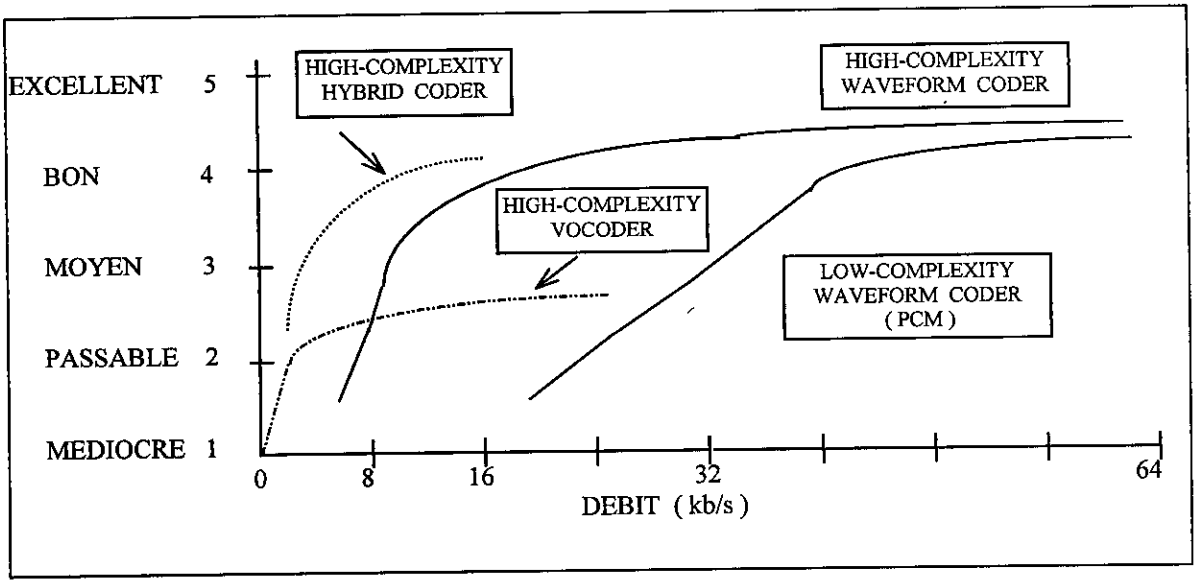


Fig. 1.6 : Relation entre le débit de codage et la qualité de parole

1.3.4 Aperçu sur les méthodes de codage de la parole

Pratiquement, les premiers systèmes de codage utilisaient la Modulation par Impulsions Codées ou codage MIC dit uniforme pour transformer un signal analogique en un signal de parole numérique.

Cette transformation se fait en deux étapes : la conversion du Continu au Discret (C/D) ou échantillonnage qui transforme une forme d'onde continue dans le temps en une forme d'onde définie à des instants discrets et la quantification qui transforme l'amplitude du signal à un instant discret en un nombre.

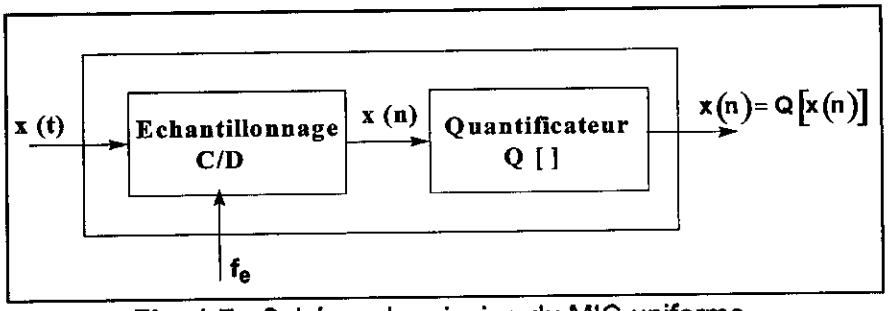


Fig. 1.7 : Schéma de principe du MIC uniforme

Typiquement, les fichiers parole sont uniformément quantifiés en utilisant 8 à 16 bits par échantillon. Cependant, la quantification non uniforme peut être utilisée pour coder le signal de parole avec moins de bits que ceux utilisés dans la quantification linéaire (ou

uniforme) . Les systèmes de télécommunication utilisent pour la compression soit la loi μ (Japonais et Nord Américains) soit la loi **A** (Européens) [Jayant,84].

Le codage MIC a donné des résultats assez satisfaisant mais un débit assez élevé puisqu'il ne prend pas en compte l'existence d'une forte corrélation entre les échantillons les plus proches.

Le codage différentiel MICD (**M**odulation à **I**mpulsions **C**odées **D**ifférentielle) réduit ce débit en quantifiant la différence d_n entre le signal de parole original s_n et une prédiction $\hat{s}_n$ de sa valeur à partir d'une combinaison linéaire des P échantillons passés. La prédiction $\hat{s}_n$ du signal s_n peut utiliser ou bien un prédicteur dont les coefficients sont fixes ou bien un prédicteur dont les coefficients sont actualisés au cours du temps (Fig. 1.8)

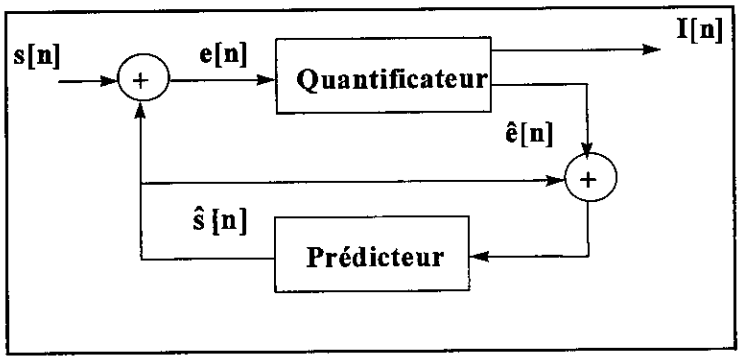


Fig. 1.8 : Codeur MICD

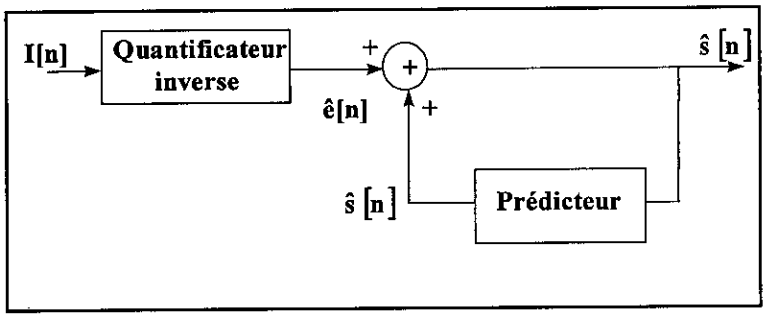


Fig. 1.9 : Décodeur MICD

Le codeur MIC **D**ifférentiel ou MICD utilise un prédicteur dont les coefficients sont fixes. La complexité de ces algorithmes est *basse à moyenne* et le débit obtenu est de l'ordre de 3 à 4 bits par échantillon. [Jayant, 84]. Dans sa forme la plus simple, le prédicteur A(z) est donné par :

$$A(z) = 1 - z^{-1} \tag{1.1}$$

On utilise des prédicteurs fixes jusqu'à l'ordre 4.

Plutôt que d'employer un seul prédicteur, et un seul quantificateur pour coder la différence du signal de parole s[n], le type de prédicteur et quantificateur pourrait être remplacé par une fonction issue des caractéristiques statistiques locales de s(n) .
Le Codage Différentiel Adaptatif ou MICDA utilise une/ou les deux méthodes d'adaptation : au niveau de la quantification et de la prédiction.

Dans la quantification adaptative, la sortie du quantificateur et les niveaux de décision sont échelonnés par correspondance à la puissance du signal d'entrée. Dans la prédiction adaptative, les coefficients du prédicteur sont dynamiquement compensés et sont basés sur les statistiques court terme des échantillons passés. La recommandation G.721 d'ITU-T (International Telecommunication Union Telephon) est un standard international qui utilise le MICDA (ou ADPCM), dans lequel le quantificateur et le prédicteur sont adaptatifs, pour coder le signal de parole à 32 kb/s.

Un cas particulier du Codage Différentiel est le Codage Prédictif Adaptatif ou APC (Adaptive Predictive Coding) [Atal, Schroeder, 70], qui permet d'améliorer les performances du codeur en utilisant un prédicteur adaptatif qui modélise le spectre court terme et un prédicteur long terme également adaptatif qui modélise les périodicités de la forme d'onde.

La nature quasi-périodique du signal original se retrouve dans le signal résiduel obtenu après prédiction linéaire. La périodicité du signal résiduel peut être supprimée en utilisant un second prédicteur : Le prédicteur long-terme.

Atal et Schroeder, proposaient un prédicteur dont la forme limitée à l'ordre 3 serait [Atal, Shroeder, 79] :

$$P(z) = 1+\beta_1 z^{-(D+1)} +\beta_2 z^{-D} +\beta_3 z^{-(D-1)} \qquad (1.2)$$

où D : retard correspondant à la période pitch, et
$\{\beta_j\}$ est l'ensemble des coefficients du prédicteur pitch.

Ce prédicteur pitch a deux effets majeurs : Il réduit l'amplitude des impulsions et améliore le rapport signal à bruit. [Atal, Schroeder, 70,79].

Le codage APC a été employé pour reproduire une bonne qualité de parole pour les communications à 9.6 kb/s et une parole presque d'excellente qualité (near to toll quality) à 16 kb/s [Deller, 93].

Les systèmes LPC (**L**inear **P**rediction **C**oding) exploitent les redondances de la parole humaine en modélisant le signal de parole avec un système de filtre linéaire à des débits variant entre 16 kb/s et 32 kb/s. Pour un codage à des débits allant de 4 kb/s à16 kb/s, le codage **A**nalyse par **S**ynthèse ou hybride basé sur la **P**rédiction **L**inéaire (LPAS) peut être utilisé pour augmenter l'efficacité d'une quantification du signal de parole [Kleijn, 94][Kroon, 95]. Le signal de parole est en premier filtré à travers un filtre d'analyse trame par trame, produisant un signal résiduel. Le résiduel est quantifié sur une base d'une sous-trame par sous-trame, et le résiduel quantifié devient le signal d'excitation pour le filtre de synthèse. Dans chaque sous-trame, le meilleur signal d'excitation est choisi à partir d'un ensemble fini des signaux d'excitation en utilisant un critère de distorsion minimum qui compare la sous-trame du signal original avec le signal reproduit basé sur chaque signal d'excitation.

Dans le codage LPAS, le décodeur est intégré dans le codeur. Pour un signal d'entrée donné, un filtre de synthèse et un modèle d'excitation donné, les paramètres (excitation, prédicteurs long-terme et court terme) sont calculés et transmis. Plusieurs méthodes sont utilisées pour représenter le signal d'excitation. Dans le codage LPAS avec excitation Multi-Impulsionnelle [Atal, 84], l'excitation est une séquence d'impulsions localisées à des instants quelconques. Le codage LPAS employant un dictionnaire vectoriel pour coder le signal d'excitation est dit Code Excited Linear Prediction (CELP). La corrélation court terme (ou enveloppe spectrale) dans le signal parole est modélisée par le filtre de synthèse $1/A(z)$. Le filtre $1/P(z)$ modélise la corrélation long terme (ou structure fine) dans le signal parole.

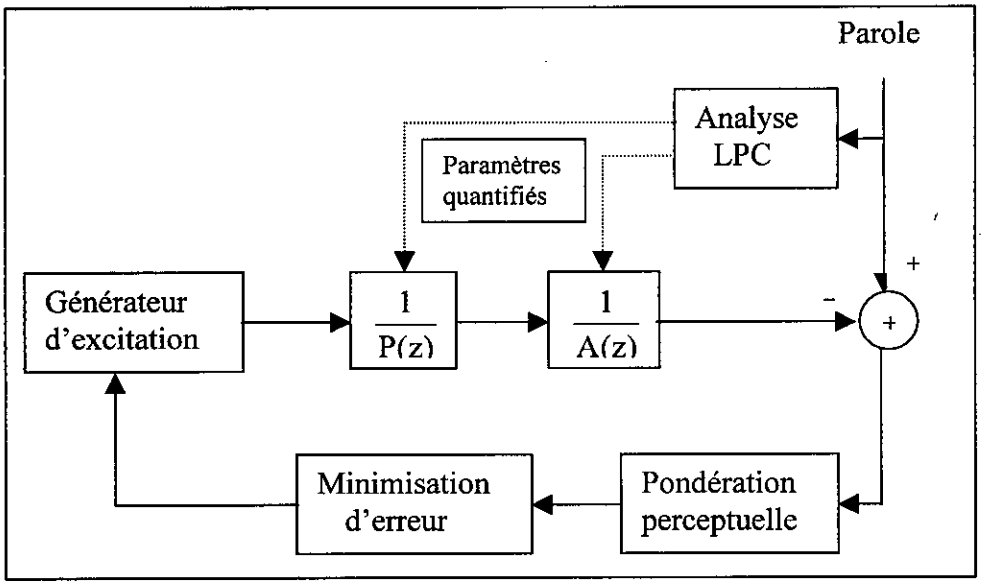


Fig.1.10 : Codeur Linéaire Prédictif basé sur l'Analyse par la Synthèse.

1.3.5 Mise en forme du spectre de bruit

Le bruit de quantification a généralement un spectre plat. On sait que lorsqu'un signal perturbateur est masqué par un signal utile de plus grande amplitude, il est moins audible. On va donc essayer de masquer le bruit par le signal en réduisant la densité spectrale du bruit dans les bandes de fréquences où l'énergie du signal est faible et en l'augmentant dans les zones formantiques (fig. 1.11).

Atal et Schroeder ont proposé pour leur codeur prédictif adaptatif une pondération du type :

$$W(z)=\frac{A(z)}{A(z/\gamma)} \tag{1.3}$$

Pour $\gamma=1$, W(z) = 1. Il n'y a pas de pondération et comme le spectre moyen décroît en fonction de la fréquence, le bruit sera surtout audible aux fréquences élevées.

Pour $\gamma=0$, W(z) = A(z). Il sera perçu sous forme de bruit basse fréquence essentiellement au niveau du premier formant. Une bonne répartition fréquentielle est obtenue pour des valeurs de γ comprises entre 0.7 et 0.9.

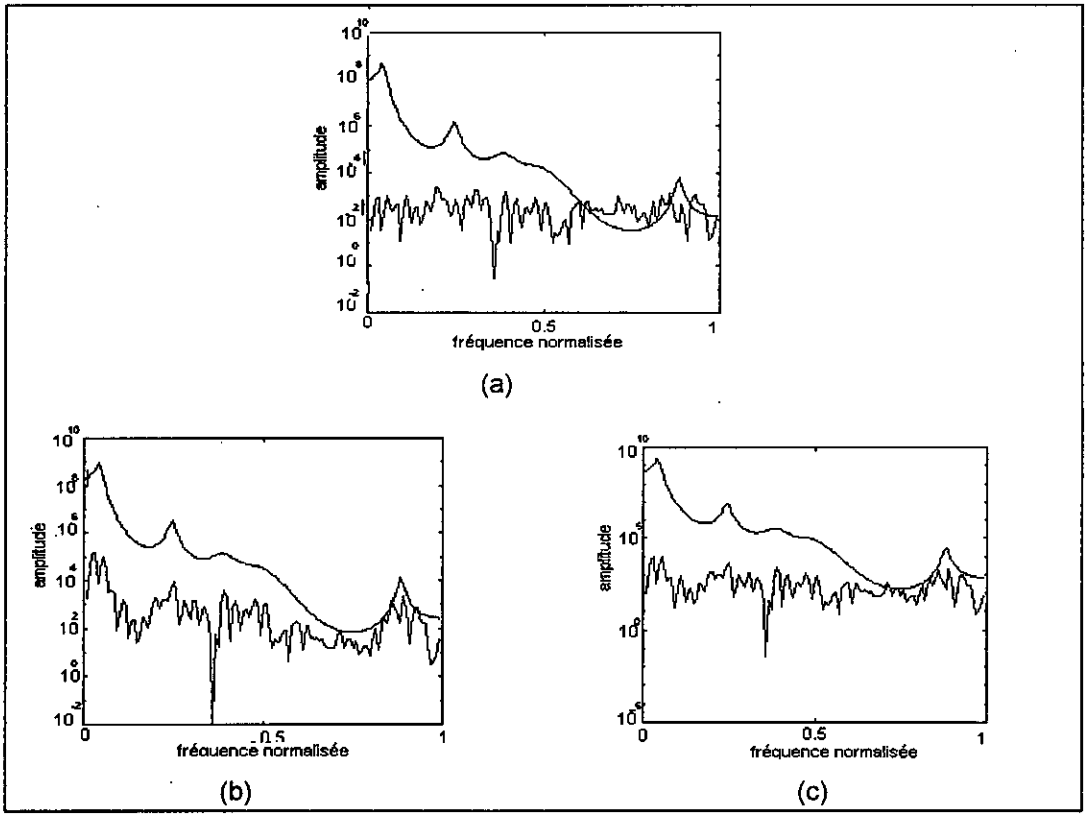


Fig. 1.11 : Effet du filtre perceptuel sur le signal de parole

(a) Spectre du modèle Auto-Regressif (p =16) & spectre d'erreur, (b) et (c) Spectre du modèle AR & filtre perceptuel pour $\gamma=0.7$ et $\gamma=0.9$

1.4 CONCLUSION

Le but du codage est de réduire le nombre d'informations à envoyer chaque seconde tout en gardant une qualité adéquate aux besoins. L'exposé des techniques de codage montre que la connaissance du mode de production et d'audition de la parole chez l'homme permet d'améliorer la qualité du codage. Ainsi, le couple «source - conduit vocal» se modélise en codage par le couple « signal d'excitation - filtre LPC»; le filtre de mise en forme du bruit a été définit.

Chapitre 2
Codage de la parole par prédiction linéaire

2.1 INTRODUCTION

Dans ce chapitre, nous nous concentrons sur le Codage Linéaire Prédictif, lequel est communément utilisé dans les algorithmes du codage de la parole à bas débit. Plusieurs représentations des coefficients prédicteurs qui donnent un codage spectral efficace sont introduites.

2.2 PREDICTION LINEAIRE

La prédiction linéaire est l'un des plus importants outils dans l'analyse d'un signal de parole. Sa simplicité relative de calcul et sa capacité à fournir une estimation exacte des paramètres du signal, rend cette méthode prédominante dans le codage à bas débit du signal de parole. Elle peut être définie comme suit : un échantillon de parole peut être approximé comme une combinaison linéaire des échantillons passés. Ainsi, en minimisant l'erreur quadratique moyenne entre les échantillons actualisés et ceux prédits linéairement sur un intervalle fini, un ensemble de coefficient prédicteur est déterminé.

La prédiction linéaire est ainsi utilisée pour enlever les redondances du signal de parole, ou modéliser le conduit vocal. La suppression des redondances est réalisée avec un filtre Prédicteur Linéaire (LP) (ou *filtre d'analyse LP*).

Le filtre d'analyse LP enlève la structure formantique du signal de parole. Le filtre d'analyse inverse (ou filtre de synthèse) modélise le conduit vocal et sa fonction de transfert décrit l'enveloppe spectrale du signal de parole. Un autre affinement peut être obtenu en considérant les corrélations long terme du signal voisé en utilisant la prédiction long terme. Dans ce cas le filtre de prédiction linéaire peut être utilisé pour enlever les redondances des échantillons trop espacés. Ce filtre exploite la périodicité du signal. L'inverse de ce filtre s'appelle prédicteur long terme. Ce dernier modélise l'effet de la glotte et sa fonction de transfert décrit la structure harmonique du signal de parole.

2.2.1 Prédiction court terme

Le modèle source - filtre nous permet d'utiliser la prédiction linéaire pour enlever les redondances court terme du signal de parole. Dans une trame de N échantillons, le signal de parole s[n] peut être considéré comme la sortie d'un certain système avec une entrée inconnue d'excitation u[n] [Makhoul, 75] :

$$s[n] = \sum_{k=1}^{p} a_k s[n-k] + G \sum_{l=0}^{q} b_l u[n-l] \qquad (2.1)$$

$b_0 = 0$

$\{a_k\},\{b_l\}$ et G sont les paramètres du système et

p, q sont les ordres de prédiction.

Dans l'équation (2.1), le signal de parole est prédit comme une combinaison linéaire des sorties passées et des excitations courantes et passées.

La transformée en z de ce système, est ainsi donnée par :

$$H(z)=\frac{S(z)}{U(z)}=G\frac{1+\sum_{l=1}^{q} b_l z^{-l}}{1-\sum_{k=1}^{p} a_k z^{-k}} \qquad (2.2)$$

où S(z) et U(z) sont respectivement les transformées en z de s[n] et u[n].

H(z) est le modèle général pôle / zéro et s'appelle Modèle Auto-Régressif à Moyenne Ajustée (ARMA). Les racines polynomiales du numérateur et dénominateur correspondent, respectivement, aux zéros et pôles du système.

Ce modèle peut se réduire à deux cas :

- modèle tout zéro Moyenne Ajustée (MA) : $a_k = 0$ pour k = 1, . . . ,p.
- modèle tout pôle Auto Régressif (AR) : $b_l = 0$ pour l = 1, . . . ,q.

Le modèle tout pôle est préféré pour plusieurs applications parce qu'il est très efficace en calcul et convient au modèle du tube acoustique pour la production de la parole [Deller,93]. Bien qu'il fournisse une très bonne représentation des effets du conduit vocal pour les voyelles (résonances), c'est seulement une approximation pour les classes de phonèmes comme les nasales et les fricatives qui sont bien modélisées par les zéros de la fonction de transfert du conduit vocal. Néanmoins, comme déjà affirmé, l'oreille humaine est plus sensible aux pôles qu'aux zéros ce qui rend la simplification acceptable. De plus, il a été montré que l'effet d'un zéro dans la fonction de transfert peut être obtenu en incluant plus de pôles [Atal, 71].

En se basant sur le modèle tout pôle, l'échantillon de parole courant est prédit par une combinaison linéaire de p échantillons passés ; c'est à dire :

$$s[n]=\sum_{k=1}^{p} a_k\, s[n-k] \qquad (2.3)$$

et la sortie r [n], appelée erreur de prédiction ou signal résiduel, est donnée par

$$r[n]=s[n]-\sum_{k=1}^{p} a_k\, s[n-k] \qquad (2.4)$$

En prenant la transformée en z des deux membres de l'équation 2.4.

$$R(z)=A(z)S(z) \tag{2.5}$$

où R(z) est la transformée en z du signal résiduel, et

$$A(z)=1-\sum_{k=1}^{p} a_k z^{-k} \tag{2.6}$$

Le filtre A(z) s'appelle filtre d'analyse. Le filtre de synthèse tout pôle H(z),

$$H(z)=\frac{1}{A(z)} \tag{2.7}$$

modélise l'enveloppe spectrale de puissance court terme du signal de parole (fig. 2.1).

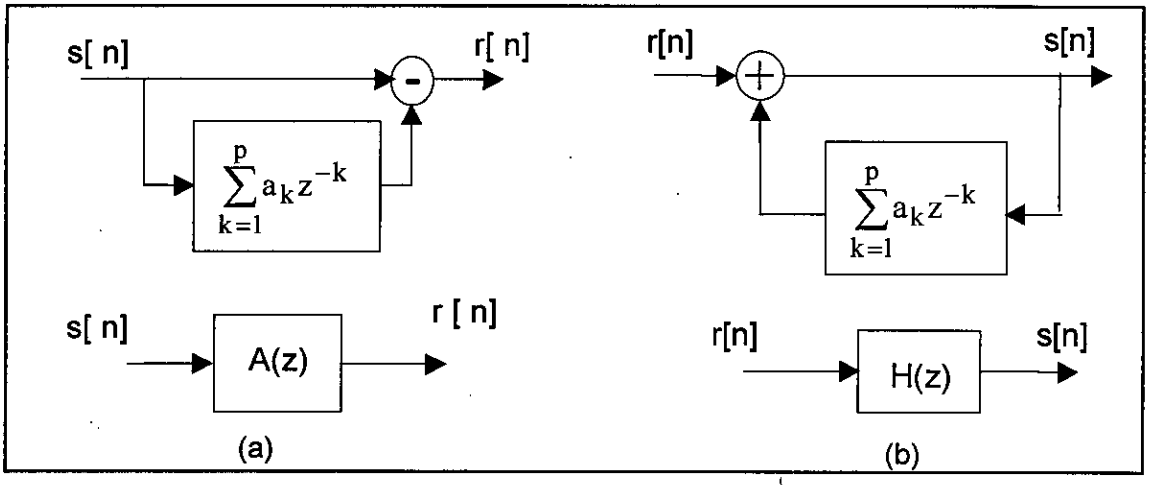


Fig. 2.1 : Diagrammes bloc du formant

(a) Etage d'analyse du formant(Prédiction court terme)

(b) Etage de synthèse du formant.

Le choix de l'ordre p du modèle est un compromis entre la précision spectrale, le calcul temps/mémoire et le débit de transmission. En général, une paire de pôles est allouée pour chaque formant présent dans le spectre de parole, plus des pôles additionnels pour approximer des zéros possibles. Pour un signal échantillonné à 16 kHz, p varie typiquement de 14 à 20.

La prédiction linéaire peut être classée comme :

- une Adaptation Progressive Avant *"forward adaptative"* dans ce cas la prédiction est basée sur les échantillons passés et les coefficients prédicteurs vont être transmis au récepteur comme "side information".

- Une Adaptation Rétrograde où la prédiction est basée sur les échantillons reconstruits passés $\hat{s}[n]$. Nous n'avons pas besoin de transmettre le "side information" au récepteur.

Les coefficients du filtre $\{a_k\}$ (aussi appelés coefficients LPC) sont estimés à chaque trame des échantillons du signal de parole. On utilise pour cela soit la méthode des moindres carrés soit celle en Treillis [Makhoul, 75] [Rabiner, 78] [O'shaughnessy, 87].

Cette dernière méthode est très complexe en calcul . Dans la méthode des moindres carrés, le signal de parole ou le signal erreur est pondéré par une fenêtre de Hamming et l'ensemble des coefficients $\{a_k\}$ est choisi de façon à minimiser l'énergie du signal d'erreur.

Selon la pondération on aboutit aux méthodes de covariance et d'autocorrélation :

- dans la première, la pondération se fait sur le signal erreur.
- dans la seconde, c'est le signal de parole qui est fenêtré. Cette méthode garantit que le filtre d'analyse LPC résultant A(z) est à phase minimum, ce qui signifie que le filtre de synthèse tout pôle H(z) est toujours stable. Cette propriété fait que la méthode d'autocorrélation est la technique la plus populaire pour l'estimation des coefficients du filtre.

2.2.2 Prédiction long terme

La parole voisée montre une forte corrélation long terme et est maintenue dans le signal résiduel LP. Ces redondances peuvent être nouvellement exploités de nouveau par l'utilisation d'un *prédicteur pitch*. Dans ce contexte, un filtre long terme peut être employé :

$$P(z) = \beta z^{-D} \tag{2.8}$$

β et D sont, respectivement, le coefficient prédicteur et la période du pitch estimé en échantillons.

Le signal erreur est exprimé par :

$$e(n) = r(n) - \beta\, r(n-D), \tag{2.9}$$

Dans le domaine temporel, le Filtre d'Analyse Pitch (FAP) soustrait de l'échantillon courant de parole (pondéré par β) correspondant à un retard égal à la période estimée à partir de l'échantillon courant. Dans le domaine fréquentiel, le FAP enlève la structure harmonique du signal d'entrée (dans notre cas le résiduel). L'analyse pitch n'aura pas un effet utile au niveau du signal non voisé puisque son excitation est aléatoire (pas de structure harmonique). Le coefficient prédicteur se relate au degré de périodicité de la forme d'onde et prend les valeurs $0 \le \beta < 1$.

Ainsi β est proche de 0 pour une structure non périodique (et dans ce cas la valeur de D est sans signification) et est pratiquement égale à l'unité pour l'état stable de la parole voisée.

Au décodeur, le filtre de synthèse pitch est donné par

$$P_s(z) = \frac{1}{P(z)} = \frac{1}{1-\beta z^{-D}} \tag{2.10}$$

et est utilisé pour introduire une structure harmonique du signal de parole synthétisé.

2.2.3 Estimation des paramètres prédicteurs

Une formulation générale pour la détermination des coefficients prédicteurs pour le prédicteur court terme et le prédicteur long terme dans la forme transversale est présenté dans la figure 2.2 [Ramachandran, 84].

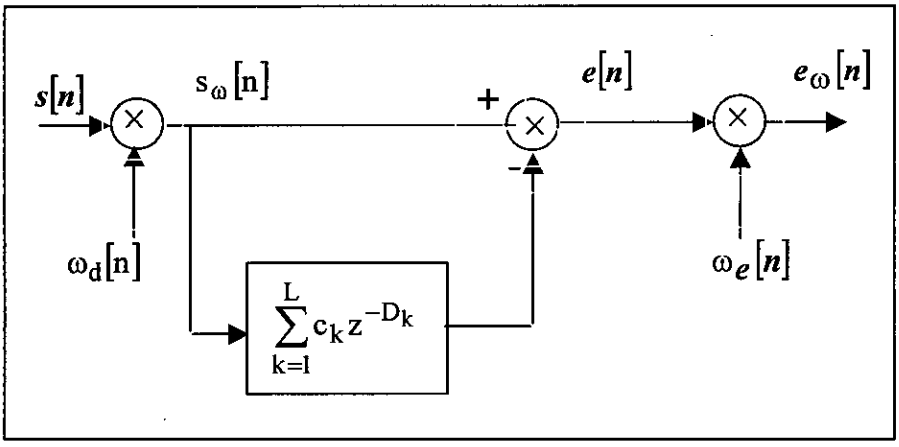


Fig. 2.2 : Modèle d'analyse pour les prédicteurs transversaux

En se basant sur le modèle montré en Fig. 2.2, le signal erreur fenêtré $e_\omega[n]$ est donné par :

$$\begin{aligned} e_\omega[n] &= \omega_e[n] e[n] \\ &= \omega_e[n] s_\omega[n] - \omega_e[n] \sum_{k=}^{L} c_k s_\omega[n - D_k] \end{aligned} \tag{2.11}$$

Où $s[n]$ est le signal d'entrée et $\omega_d[n]$, $\omega_e[n]$ sont les fenêtres de pondération utilisées pour les données et l'erreur. Les valeurs de D_k sont arbitraires mais des entiers distincts correspondant aux retards du signal d'entrée pondéré $s_\omega[n]$.

L'énergie de l'erreur, ou l'Erreur Quadratique Moyenne (EQM), est donnée par :

$$\varepsilon = \sum_{n=-\infty}^{\infty} e_{\omega}^{2}[n]. \tag{2.12}$$

Les coefficients c_k sont calculés par minimisation de ε. Ceci est accompli en prenant la dérivée partielle de l'équation (2.13) par rapport à chacun des coefficients c_k, pour $k = 1,\ldots, L$ et en posant chacune des L équations résultantes à zéro. Ceci conduit à un système linéaire d'équations qui peuvent être écrites sous forme matricielle ($\Phi \mathbf{c} = \mathbf{a}$) :

$$\begin{bmatrix} \phi(D_1,D_1) & \phi(D_1,D_2) & \cdots & \phi(D_1,D_L) \\ \phi(D_2,D_1) & \phi(D_2,D_2) & \cdots & \phi(D_2,D_L) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(D_L,D_1) & \phi(D_L,D_2) & \cdots & \phi(D_L,D_L) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_L \end{bmatrix} = \begin{bmatrix} \phi(0,D_1) \\ \phi(0,D_2) \\ \vdots \\ \phi(0,D_L) \end{bmatrix} \tag{2.13}$$

où

$$\Phi(i,j) = \sum_{n=-\infty}^{\infty} \omega_e^2[n]\, s_{\omega}[n-i]\, s_{\omega}[n-j] \tag{2.14}$$

La matrice Φ est toujours définie symétrique et positive. Elle est aussi une matrice de Toeplitz si les retards d'inter-coefficients sont égaux. Selon que Φ est de Toeplitz ou non, la résolution du système d'équations peut se faire par la récursion de Levinson ou la décomposition de Cholesky.

Dans le cas d'un prédicteur court terme $D_k = k$ pour k = 1, . . . ,p et pour un prédicteur long terme d'ordre N_p, $D_k = D + k$ pour k = 0, . . . , N_p-1.

Quand $w_e[n] = 1 \quad \forall n$, la formulation ci-dessus aboutit à la méthode d'autocorrélation. La méthode de covariance est utilisée si $w_d[n] = 1 \quad \forall n$.

Pour la prédiction à court terme, on peut montrer que la méthode d'autocorrélation donne un filtre d'analyse à phase minimale alors que le filtre de synthèse résultant de la méthode de covariance pourrait être instable. Dans le cas de la prédiction du pitch, la méthode d'autocorrélation (excepté le prédicteur pitch du premier ordre) et la méthode de covariance pourraient aboutir à un filtre de synthèse pitch instable. Par contre une légère instabilité est souvent utile pour modéliser les amplitudes croissantes dans le signal

d'excitation. Comme la méthode de covariance donne des gains de prédiction élevés, elle est généralement préférée pour la prédiction pitch. Chaque fois que le filtre de synthèse pitch est instable, un schéma efficace de stabilisation peut être employé pour limiter l'instabilité à des limites désirables [Ramachandran, 87].

Souvent un filtre prédicteur pitch à seul retard est utilisé et D est déterminé séparément du coefficient prédicteur en utilisant la méthode de covariance. Cette procédure évite une recherche exhaustive pour le D optimal dans le prédicteur pitch à plusieurs retards [Ramachandran, 87]. En utilisant la méthode de covariance, l'erreur quadratique moyenne dans l'équation. (2.12) peut être réécrite sous forme matricielle par :

$$\varepsilon=\phi(0,0)-2c^{T}a+c^{T}\Phi c \tag{2.15}$$

Les coefficients optimaux sont donnés par :

$$\mathbf{c}=\Phi^{-1}\mathbf{a}$$

l'erreur quadratique moyenne résultante est :

$$\varepsilon=\phi(0,0)-\mathbf{c}^{T}\mathbf{a}. \tag{2.16}$$

Pour le prédicteur pitch d'ordre N_p=1 et l'équation (2.15) est réduite à :

$$\phi(D,D)\,c_1=\phi(0,D) \tag{2.17}$$

et le coefficient optimal β_{opt} est donné par

$$\beta_{opt}=c_1=\frac{\phi(0,D)}{\phi(D,D)} \tag{2.18}$$

Par conséquent, l'erreur quadratique moyenne résultante se réduit à

$$\varepsilon=\phi(0,0))-\frac{\phi^{2}(0,D)}{\phi(D,D)} \tag{2.19}$$

L'Erreur Quadratique Moyenne résultante dans l'Eq. (2.20) est minimisée en maximisant :

$$\frac{\phi^{2}(0,D)}{\phi(D,D)} \tag{2.20}$$

Cette fonction est calculée pour toutes les valeurs possibles de D, et son maximum indique le meilleur choix pour la période pitch D. Le champ des valeurs sur lequel la période pitch est cherchée est typiquement compris entre 20 et 147 échantillons pour une fréquence d'échantillonnage de 8 kHz et entre 40 et 295 échantillons pour une fréquence

d'échantillonnage de 16 kHz ce qui couvre la plupart des valeurs pitch rencontrées dans la parole humaine (54.2 Hz – 400 Hz). D'autres méthodes pratiques pour le choix de D peuvent être trouvées dans [Ramachandran, 89] [Atal, 79].

2.3. REPRESENTATION SPECTRALE DES PARAMETRES PREDICTEURS

Les coefficients de Prédiction Linéaire ne sont pas toujours codés directement mais sont transformés en un ensemble de paramètres qui ont des propriétés désirables. Plusieurs représentations des coefficients ont été proposées. Les plus populaires actuellement sont les Paires de Fréquences Spectrales (LSF) [Soong et Juang, 84]. D'autres représentations incluent les coefficients de réflexion, logarithmes des rapports des aires des sections, les coefficients cepstraux, la réponse impulsionnelle du filtre LP (Rabiner et Schafer, 78)...etc.

2.3.1. Coefficients de réflexion

Une procédure d'amorçage peut être utilisée pour trouver les coefficients LP à partir des coefficients de réflexion $\{k_m\}$. Initialement on calcule l'énergie moyenne dans la trame de parole telle que :

$$E_0 = R(0) \tag{2.21}$$

On résout alors récursivement les équations suivantes pour chaque itération de m, avec m = 1, 2, . . . ,p.

$$k_m = \frac{1}{E_{m-1}}\left[R(m) - \sum_{k=1}^{m-1} \alpha_{m-1}(k) R(m-k)\right] \tag{2.22}$$

avec

$$\alpha_k(m) = \alpha_k(m-1) - k_m \alpha_{m-k}(m-1), \quad 1 \le k \le m\text{-}1 \tag{2.23}$$

et

$$E_m = (1 - k_m^2) E_{m-1} \tag{2.24}$$

Les $\alpha_k(m)$ représentent les coefficients de prédiction du prédicteur linéaire d'ordre m :

$$a_k = \alpha_k(m) \qquad 1 \le k \le m \tag{2.25}$$

Ainsi, les coefficients de prédiction résultants du prédicteur linéaire d'ordre p sont quand m = p.

Une propriété importante des coefficients de réflexion est que $|k_m| < 1$ ce qui implique la stabilité du filtre. Quand on utilise la méthode de covariance pour trouver les coefficients de prédiction , les convertir en coefficients de réflexion peut être utile dans la détermination de la stabilité du filtre.

On calcule récursivement pour m = p, p-1, . . . ,2, initialement avec $\alpha_p(k) = a_k$.

$$\alpha_{m-1}(i) = \frac{\alpha_m(i) k_m \alpha_m(m-i)}{1-k_m^2}, \qquad 1 \le i \le m\text{-}1 \tag{2.26}$$

$$k_{m-1} = \alpha_{m-1}(m-1) \tag{2.27}$$

Si $|k_m| \ge 1$, alors soit on réduit artificiellement l'amplitude inférieure à l'unité. Le spectre du signal de parole est modifié, mais les sorties instables sont éliminées.

Quand on désire quantifier les coefficients de réflexion, une prudence est demandée afin de ne pas avoir des valeurs égales à 1 ou (-1). La transformation non linéaire des coefficients de réflexion en coefficients Log ARea nous permet d'éviter ce problème. Les coefficients LAR sont calculés de la manière suivante :

$$g_m = \log\left(\frac{1+k_m}{1-k_m}\right), \qquad 1 \le m \le p \tag{2.28}$$

En les convertissant en coefficients de réflexion, on obtient :

$$k_m = \frac{e^{g_m} - 1}{e^{g_m} + 1}, \qquad 1 \le m \le p \tag{2.29}$$

2.3.2 Coefficients cepstraux

Le cepstre d'un signal de parole est la Transformée de Fourier inverse du spectre de puissance logarithmique :

$$\log\left[\frac{1}{\left|A\left(e^{jw}\right)\right|^2}\right]=\sum_{n=-\infty}^{\infty}c_n e^{-jw} \tag{2.30}$$

où $c_n = c_{-n}$, et $c_0 = 0$, sont les coefficients cepstraux.

Un nombre infini de coefficients cepstraux peuvent être calculés à partir des coefficients de prédiction [Atal, 79] :

$$c_n = a_n + \sum_{k=1}^{n-1}\frac{k}{n}a_{n-k}\,c_k \tag{2.31}$$

Pour un prédicteur linéaire d'ordre p, $a_n = 0$ pour n > p.

En outre, un filtre à phase minimale implique que $c_n = 0$ pour $n \leq 0$.

2.3.3 Paires de fréquences spectrales (LSF)

Les LSF ont été introduits pour la première fois par Soong comme une alternative de la représentation paramétrique des coefficients de la prédiction linéaire [Soong, 84].

Le polynôme à phase minimum A(z) d'ordre p peut être décomposé en deux polynômes P(z) et Q(z) d'ordre (p+1) où :

$$A(z)=\frac{1}{2}\left[P(z)+Q(z)\right] \tag{2.32}$$

Les polynômes sont calculés comme suit :

$$P(z)=A(z)+z^{-(p+1)}A(z^{-1}) \tag{2.33}$$

$$Q(z)=A(z)-z^{-(p+1)}A(z^{-1}) \tag{2.34}$$

où le coefficient de réflexion k_{p+1} prend la valeur (+1) pour P(z) et (-1) pour Q(z).

Les zéros de P(z) et Q(z) qui se trouvent sur le cercle unité sont entrelacés. Les 'p' LSF correspondent aux positions angulaires 'ω' des 'p' zéros localisés sur le cercle unité entre 0 et π radians.

On obtient deux zéros particuliers à $\omega = 0$ et $\omega = \pi$ qui peuvent être ignorés lors de la quantification (car ils sont connus).

Ainsi, les p LSF $\{\omega_i\}$ ont une propriété d'ordonnancement ascendante qui assure la stabilité du filtre de synthèse LPC:

$$0 < \omega_1 < \omega_2 < \ldots < \omega_p < \pi \qquad \text{[radians / s]} \tag{2.35}$$

Les LSF correspondent explicitement au spectre du filtre LP. Ils se groupent autour des pics spectraux (fig. 2.3). De plus, la sensibilité spectrale de chaque LSF est localisée. Chaque changement dans un LSF donné génère une altération dans la forme du spectre seulement dans le voisinage proche de LSF.

La fig. 2.4 illustre deux exemples du spectre LP dans lequel chaque spectre a un seul LSF changé.

Les LSF peuvent être calculés de diverses manières. Soong et Juang [Soong, 84] calculèrent les LSF en appliquant la Transformation en Cosinus Discrète (TCD) aux coefficients des polynômes :

$$G(z) = \begin{cases} \dfrac{P(z)}{1+z^{-1}}, & p \text{ pair} \\ P(z), & p \text{ impair} \end{cases} \tag{2.36}$$

et

$$H(z) = \begin{cases} \dfrac{Q(z)}{1-z^{-1}}, & p \text{ pair} \\ \dfrac{Q(z)}{1-z^{-2}}, & p \text{ impair} \end{cases} \tag{2.37}$$

Kabal et Ramachandran [Kabal, 86] utilisèrent une expansion du polynôme de Chebyshev d'ordre m en x :

$$T_m(x) = \cos(m\omega) \tag{2.38}$$

Où $x = \cos(\omega)$ figure sur le demi cercle supérieur dans le plan z sur l'intervalle réel [-1, +1].

Les polynômes $G'(\omega)$ et $H'(\omega)$ peuvent être exprimés tels que :

$$G'(x)=2\sum_{i=0}^{l} g_i T_{l-i}(x) \tag{2.39}$$

$$H'(x)=2\sum_{i=0}^{l} h_i T_{m-i}(x) \tag{2.40}$$

avec l = m = p / 2 quand p est pair,
et l = (p+1) / 2 et m = (p-1) / 2 quand p est impair.

Les racines du polynôme sont déterminées itérativement en cherchant le changement de signe dans l'intervalle [-1, +1].
Les LSF correspondent aux racines polynomiales en utilisant la Transformation $\omega = \cos^{-1}(x)$.

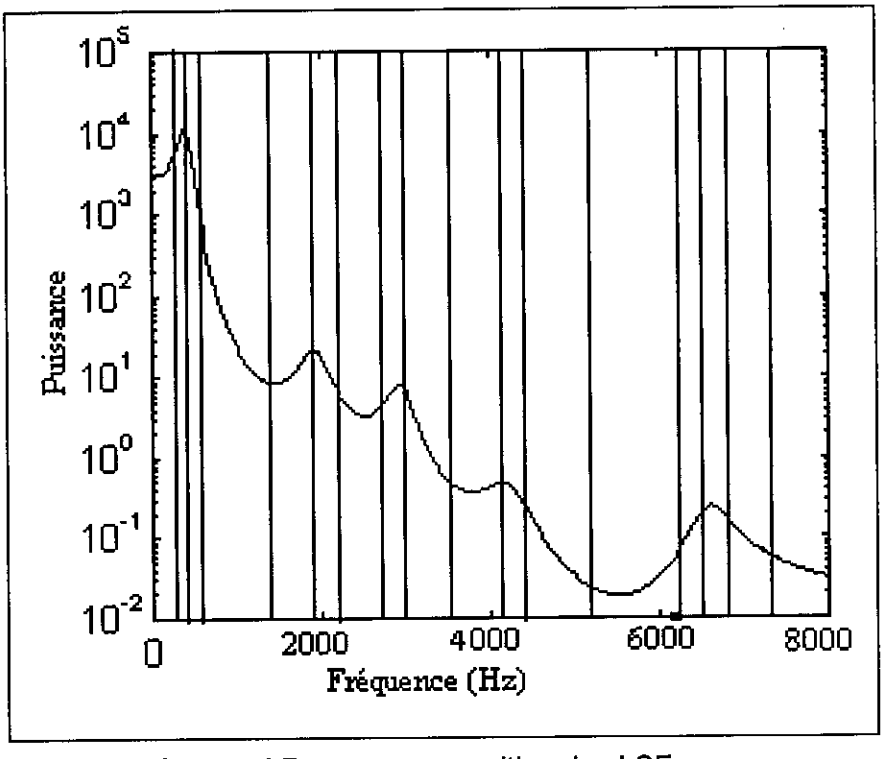


Fig. 2.3 : Spectre LP avec superposition des LSF

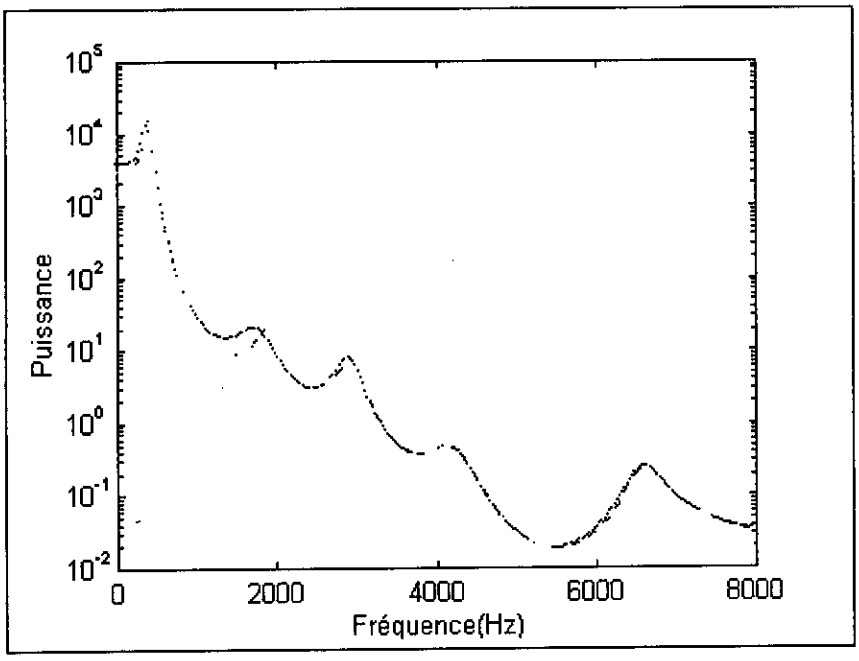


Fig.2.4 : Effet de changement d'une valeur LSF sur le spectre LP

(----) LSF original, (.....) 5 ème LSF changé, (__) 14 ème LSF changé

2.3.4 Interpolation des LSF

Puisque les LSF correspondent aux fréquences de résonances du conduit vocal, ils ne changent pas radicalement. Nous pouvons donc les interpoler sans causer beaucoup de distorsion. C'est pourquoi on utilise un ensemble de LSF par 240 échantillons (trame) au lieu de 60 échantillons (sous trame). Dans le traitement de la sous trame, nous avons besoin de LSF pour calculer le signal résiduel ; ces LSF sont les résultats de l'interpolation. Ceci est illustré dans la figure 2.5.

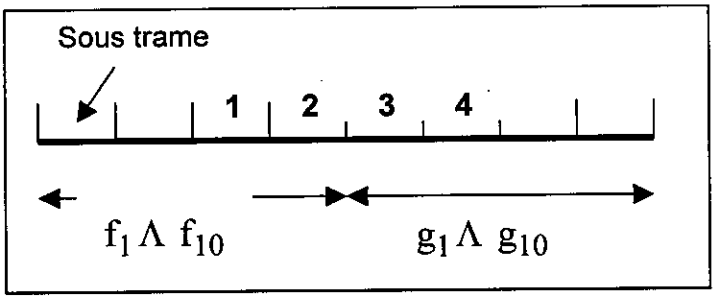


Fig. 2.5 : Interpolation des LSF

Soient $f_1 \Lambda f_{10}$ les LSF calculés à partir de l'analyse LPC des 4 sous trames, et $g_1 \Lambda g_{10}$ à partir des 4 sous trames suivantes.

Les LSF pour la première sous trame **1** sont interpolés par $(7/8)f+(1/8)g$, $(5/8)f+(3/8)g$ pour la sous trame **2**, $(3/8)f+(5/8)g$ pour la sous trame **3**, $(1/8)f+(7/8)g$ pour la sous trame **4**.

Ce schéma d'interpolation cause un retard d'encodage de deux sous trames (7.5 ms)

2.4 MESURES DE DISTORSION OBJECTIVE

L'appareil auditif humain est l'ultime évaluateur de la qualité d'un codeur de la parole et de sa performance (préservation de l'intelligibilité). Les mesures objectives peuvent donner un estimateur immédiat et fiable de la qualité perceptuelle d'un algorithme de codage.

2.4.1 Mesures dans le domaine temporel

La mesure objective de qualité la plus couramment utilisée, pour les codeurs qui essaient de préserver la forme du signal, reste le Rapport Signal à Bruit (RSB) et le RSBsegmental

2.4.1.1 Rapport signal à bruit

Le RSB mesure la longueur relative de la puissance du signal sur la puissance de bruit. La mesure RSB, en décibels (dB), est définie de la manière suivante :

$$RSB=10\log_{10}\frac{\sum_{n=-\infty}^{\infty}s^2[n]}{\sum_{n=-\infty}^{\infty}(s[n]-\hat{s}[n])^2}\ dB, \qquad (2.41)$$

où $\hat{s}[n]$ est la version codée de l'échantillon s[n] du signal de parole original. Cependant, la mesure RSB n'est pas un très bon estimateur de la qualité de la parole [Deller, 93]. La mesure RSB pondère de la même manière toutes les erreurs dans le signal, négligeant le fait que l'énergie du signal de parole est variable dans le temps.

Le signal de parole étant par nature non - stationnaire, certains segments du signal peuvent avoir une énergie plus ou moins grande. En supposant que l'énergie de l'erreur soit à peu près constante, le RSB pourra être soit très important soit très faible. Pour avoir une idée de la qualité de la parole synthétique, on utilise plutôt le RSB segmental (RSBseg).

2.4.1.2 Rapport signal à bruit segmental

Le RSBseg est la moyenne géométrique des mesures RSB calculées sur différentes trames. La mesure RSBseg, exprimée en dB, sur M segments de parole est définie comme suit :

$$RSBseg=\frac{1}{M}\sum_{m=0}^{M-1}10\log_{10}\left[\frac{\sum_{n=1}^{N}s^2[n+Nm]}{\sum_{n=1}^{N}(s[n+Nm]-\hat{s}[n+Nm])^2}\right]dB \qquad (2.42)$$

où chaque segment "m" est de longueur N. Pour un signal de parole avec une fréquence d'échantillonnage de 16 kHz les valeurs de N varient typiquement entre 160 et 240 échantillons (10 à 15 ms). Cette mesure présente l'avantage de tenir compte de l'évolution du RSB au cours du temps et, en particulier, de bien prendre en compte les segments de faible énergie. On essaie en outre de limiter les trop grands écarts ; si le signal pour un RSB(m) est supérieur à 40 dB, on le remplace par 40 dB et de même, dans les zones de silence, le RSB peut atteindre des valeurs très négatives : dans ce cas, on peut ou bien retirer du calcul ces zones ou bien fixer un seuil inférieur à T tel que $0 \leq T \leq -20$ dB.

2.4.2 Mesures dans le domaine spectrale

La mesure de distorsion $\mathbf{d}(\mathbf{x},\hat{\mathbf{x}})$ entre deux vecteurs de parole $\mathbf{x}$ et $\hat{\mathbf{x}}$ satisfait à deux conditions [Gray, Buzo, 80]

$$\begin{aligned} d(x,x) &= 0 \\ d(x,\hat{x}) &\geq 0 \end{aligned} \qquad (2.43)$$

Une mesure très rigoureuse est la mesure de distance, ou métrique, qui demande en plus à satisfaire deux autres conditions [Gray, Markel, 76] :

$$\begin{aligned} d(x,\hat{x}) &= d(\hat{x},x) \\ d(x,\hat{x}) &\leq d(x,y)+d(y,\hat{x}) \end{aligned} \qquad (2.44)$$

En général, toute mesure de performance est la moyenne long terme d'une mesure de distorsion ou distance.

$$D = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}d(X_i,\hat{X}_i) \qquad (2.45)$$

Une mesure de distorsion dans le cas de la parole devrait avoir une signification dans le domaine fréquentiel. La mesure est généralement faite en utilisant des trames de parole entre 10 et 30 ms de longueur. Les différences entre l'enveloppe spectrale originale et codée qui peuvent conduire perceptuellement à des sons différents qui sont dues à [Rabiner et Juang, 93] :

- Les résonances ou formants de l'enveloppe spectrale originale et codée se produisent à des fréquences considérablement différentes.
- Les largeurs de bande du formant de l'enveloppe spectrale originale et codée diffèrent considérablement.

Plusieurs mesures de distorsion spectrale ont été proposés dans la littérature. On citera par exemple la mesure de distorsion log spectrale, la mesure d'Itakura-Saito, la distance cepstrale et la mesure de distance Euclidienne pondérée.

2.4.2.1 Mesure de distorsion Log Spectrale

La mesure de distance log spectrale dans la norme $\mathbf{L_p}$ est définie par :

$$d_{SD}^{p} = \frac{2}{2\pi}\int_{-\pi}^{\pi}\left|10\log_{10}S(\omega) - 10\log_{10}\hat{S}(\omega)\right|^{p} d\omega \qquad (2.46)$$

où le spectre d'amplitude fréquentiel $S(\omega)$ est donnée par :

$$S(\omega) = \frac{G}{\left|A\left(e^{j\omega}\right)\right|^2} \qquad (2.47)$$

$$= \frac{G}{\left[1 - \sum_{n=1}^{p} a_n e^{jn\omega}\right]^2} \qquad (2.48)$$

G est le facteur gain du filtre LP, et $\{a_n\}$ sont les coefficients de Prédiction Linéaire.

Quand p=2 (norme L_2), la mesure de distorsion log spectrale se réduit à une distance quadratique moyenne (rms). Elle est définie en décibels par

$$d_{SD} = \sqrt{\frac{1}{\omega_u - \omega_l}\int_{\omega_l}^{\omega_u}\left[10\log_{10}\frac{S(\omega)}{\hat{S}(\omega)}\right]^2 d\omega} \ \text{dB} \qquad (2.49)$$

où ω_l et ω_u définissent, respectivement, les limites fréquentielles inférieures et supérieures de l'intégration. Idéalement, ω_l est égale à zéro et ω_u correspond à la demie fréquence d'échantillonnage.

En pratique, la distance log spectrale rms est calculée par discrétisation sur une largeur de bande limitée. Pour un signal de parole échantillonné à 8 kHz filtré à travers un filtre passe-bas de 3 kHz, la distorsion log spectrale (SD) rms est calculée comme une sommation, avec une résolution d'approximativement 31.25 Hz par échantillon, sur 96 points éspacés uniformément de 0 Hz à 3 kHz.

Ceci peut être exprimé par :

$$SD = \sqrt{\frac{1}{n_1 - n_0}\sum_{n=n_0}^{n_1-1}\left[10\log_{10}\frac{S\left(e^{j2\pi n/N}\right)}{\hat{S}\left(e^{j2\pi n/N}\right)}\right]^2}\ dB \qquad (2.50)$$

où pour N = 256, n_0 et n_1 correspondent , respectivement, à 1 et 255.

La distance log spectrale rms fournie le meilleure point de référence pour la comparaison [Gray et Markel, 76]. [Paliwal et Atal, 93] ont suggéré que la qualité de codage transparente est atteinte quand les résultats de quantification sont approximativement de 1 dB dans le cas de la distance log spectrale rms

2.4.2.2 Mesure de distorsion d'Itakura-Saito

Aussi connue comme la mesure de distance du rapport de vraisemblance, la distorsion d'Itakura-Saito (d_{IS}) mesure le rapport d'énergie entre le signal résiduel qui résulte quand on utilise le filtre LP quantifié et le signal résiduel qui résulte quand on utilise le filtre LP non quantifié. La mesure d'Itakura-Saito est définie par :

$$d_{SD}^{p} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[e^{V(\omega)} - V(\omega) - 1\right]d\omega \qquad (2.51)$$

où la différence log spectrale $V(\omega)$ entre deux spectres est définie comme

$$V(\omega) = \log S(\omega) - \log\hat{S}(\omega) \qquad (2.52)$$

En évaluant les intégrales, cette mesure peut être exprimée comme le polynôme

$$d_{IS} = \left(\frac{G}{\hat{G}}\right)^2\frac{\hat{a}^T R\hat{a}}{a^T Ra} - 2\log\left(\frac{G}{\hat{G}}\right) - 1 \qquad (2.53)$$

où $\hat{\mathbf{a}} = \left[1, \hat{a}_1, \hat{a}_2, ..., \hat{a}_p\right]^T$, $\mathbf{a} = \left[1, a_1, a_2, ..., a_p\right]^T$, et **R** est la matrice d'autocorrélation.

Quand les gains sont égaux, alors la mesure d'Itakura-Saito se simplifie à :

$$d_{IS} = \frac{\hat{a}^T R \hat{a}}{a^T R a} - 1 \tag{2.54}$$

Cependant, la mesure d'Itakura-Saito n'est pas symétrique. Pour la symétrie, la mesure d'Itakura modifiée peut être utilisée :

$$d_{IS} = \frac{1}{2}\left[\frac{\hat{a}^T R \hat{a}}{a^T R a} - \frac{a^T \hat{R} a}{\hat{a}^T \hat{R} \hat{a}} - 2\right] \tag{2.55}$$

2.4.2.3 Distance Cepstrale

La mesure de distorsion log spectrale souffre de l'inconvénient des calculs du logarithme et de la transformée de Fourier demandés pour chaque point de la sommation. La distance cepstrale (d_{CD}) est une approximation efficace de calcul de la mesure de distance log spectrale en mesurant la différence totale entre le cepstre original et le cepstre codé du signal de parole. Le cepstre d'un signal de parole est la transformée de Fourier du logarithme du spectre de parole :

$$\log S(\omega) = \sum_{n=-\infty}^{\infty} c_n e^{-jn\omega} \tag{2.56}$$

où $\{c_n | c_n = c_{-n}, c_0 = 0\}$ sont étiquetés comme les coefficients cepstraux

En utilisant l'équation de Parseval, la distance cepstrale est liée directement à la distance log spectrale rms (norme L_2) :

$$d_{CD}^2 = \sum_{n=-\infty}^{\infty} (c_n - \hat{c}_n)^2 \tag{2.57}$$

$$= 2\sum_{n=1}^{\infty} (c_n - \hat{c}_n)^2 \tag{2.58}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|\log S(\omega) - \log \hat{S}(\omega)\right|^2 d\omega \tag{2.59}$$

Bien que la sommation est infinie, la sommation est généralement tronquée en un nombre fini N_c. Généralement, le nombre de coefficients cepstraux est pris égal à 3 fois l'ordre p du filtre d'analyse

$$d_{CD} = 10\log_{10}\sqrt{2\sum_{n=1}^{N_c}(c_n - \hat{c}_n)^2} \tag{2.60}$$

2.4.2.4 Mesure de Distance LSF Euclidienne Pondérée

Les LSF (Line Spectral Frequencies) ont une relation directe avec la forme de l'enveloppe spectrale. Une mesure de distance de l'erreur quadratique devrait être utilisée pour comparer les vecteurs LSF original et codé. Soient deux vecteurs LSF **x** et $\hat{\mathbf{x}}$ de dimensions m ; la mesure de distance Euclidienne est :

$$d(x,\hat{x}) = (x-\hat{x})^T(x-\hat{x}) = \| x-\hat{x} \|^2 \tag{2.61}$$

Pour obtenir un estimateur de la qualité perceptuelle de l'enveloppe spectrale, une mesure de distance LSF Euclidienne pondérée est utilisée :

$$d_\omega(x,\hat{x}) = (x-\hat{x})^T W (x-\hat{x}) \tag{2.62}$$

où **W** est une matrice de pondération m × m définie symétrique et positive qui devrait être dépendante de **x**. Si **W** est une matrice diagonale avec les éléments $\omega_{ii} > 0$, la distance peut être exprimée par :

$$d(x,\hat{x}) = \sum_{i=1}^{m}\omega_{ii}(x_i - \hat{x}_i)^2 \tag{2.63}$$

Quand on ne désire pas de pondération, la matrice de pondération est prise égale à la matrice identité **W = I**.

Paliwal et Atal [Palival, 93] proposaient comme matrice de pondération le produit entre une matrice de pondération fixée et une matrice de pondération adaptative :

$$\mathbf{W} = \mathbf{W_f}\,\mathbf{W_a}$$

La matrice de pondération adaptative $\mathbf{W_a}$ varie d'une trame à une autre, par accentuation des pics spectraux dans les régions de formants sur les autres régions qui sont présents dans le spectre LPC de la trame courante.

Les éléments de la diagonale ω_i dans $\mathbf{W_a}$ sont chacun assignés au $i^{ème}$ LSF par la composante ω_i :

$$\omega_i = [S(\omega_i)]^r \tag{2.64}$$

où $S(\omega_i)$ est l'amplitude du spectre de puissance à la fréquence ω_i et r une constante arbitraire. Paliwal et Atal ont choisi r = 0.30.

Un schéma de la pondération fixée peut être déterminé en tenant compte de l'incapacité de l'oreille humaine à discerner les différences aux fréquences hautes. Pour un vecteur LSF d'ordre 10, Paliwal et Atal utilisaient les poids suivants :

$$c_i = \begin{cases} 1.0, & \text{pour } 1 \le i \le 8, \\ 0.8, & \text{pour } i = 9, \\ 0.4, & \text{pour } i = 10. \end{cases} \qquad (2.65)$$

2.5 MESURES SUBJECTIVES DE LA QUALITE DE LA PAROLE

Les essais d'écoute sont nécessaires car la qualité d'un système de codage de la parole ne vaut que par le jugement humain. De plus, le RSB n'est pas nécessairement corrélé avec la qualité d'écoute.

Les méthodes les plus utilisées sont le

1. *Diagnostic Rhytme Test* (DRT) qui mesure l'intelligibilité sur un grand nombre de mots ;
2. *Diagnostic Acceptability Mesure* (DAM) qui mesure le naturel perçu de la parole ;
3. *Mean opinion score* (MOS) ou l'auditeur évalue un codeur sur une échelle absolue allant de 1 à 5 (Cf. fig.1.6).

2.6 ENVIRONNEMENT D'EVALUATION DE LA PERFORMANCE

L'évaluation de performance du codage des paramètres spectraux est basée sur un ensemble de vecteurs d'apprentissage et un ensemble de vecteurs test .

Une base de données d'approximativement de 1minute de parole à large bande échantillonnée à 16 kHz, est utilisée pour construire une séquence d'entraînement. Une autre base additionnelle de 30 secondes est utilisée pour la séquence de test. L'analyse LP d'ordre 16 est exécutée en utilisant la méthode d'autocorrélation. La corrélation entre les trames adjacentes est kept à un minimum avec un recouvrement en utilisant une fenêtre de Hamming de 20 ms. Ce sont 7200 vecteurs LSF pour l'entraînement et 3600 vecteurs LSF pour le test. Les vecteurs LSF seront convertis en d'autres représentations paramétriques comme les coefficients LPC.

Les tests d'écoute sont faits pour un groupe sélectionné de codage spectral. L'environnement de simulation utilisé pour l'évaluation subjective du codec spectral est illustré en figure 2.6.

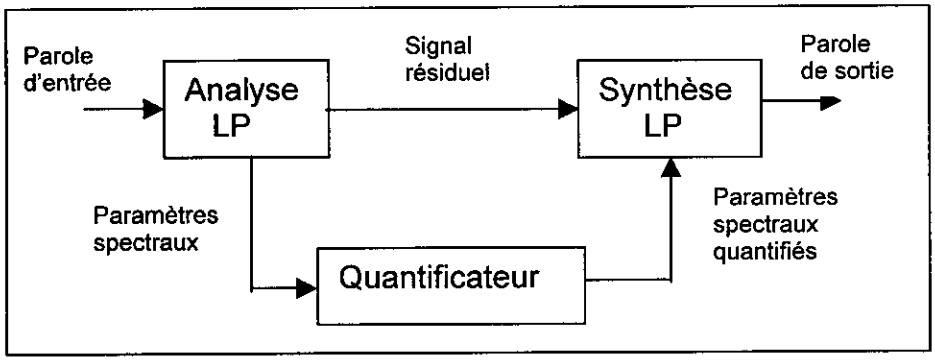


Fig. 2.6 : Environnement de simulation pour l'évaluation du codec de la parole

A titre d'illustration, on donne l'histogramme de chaque vecteur LSF

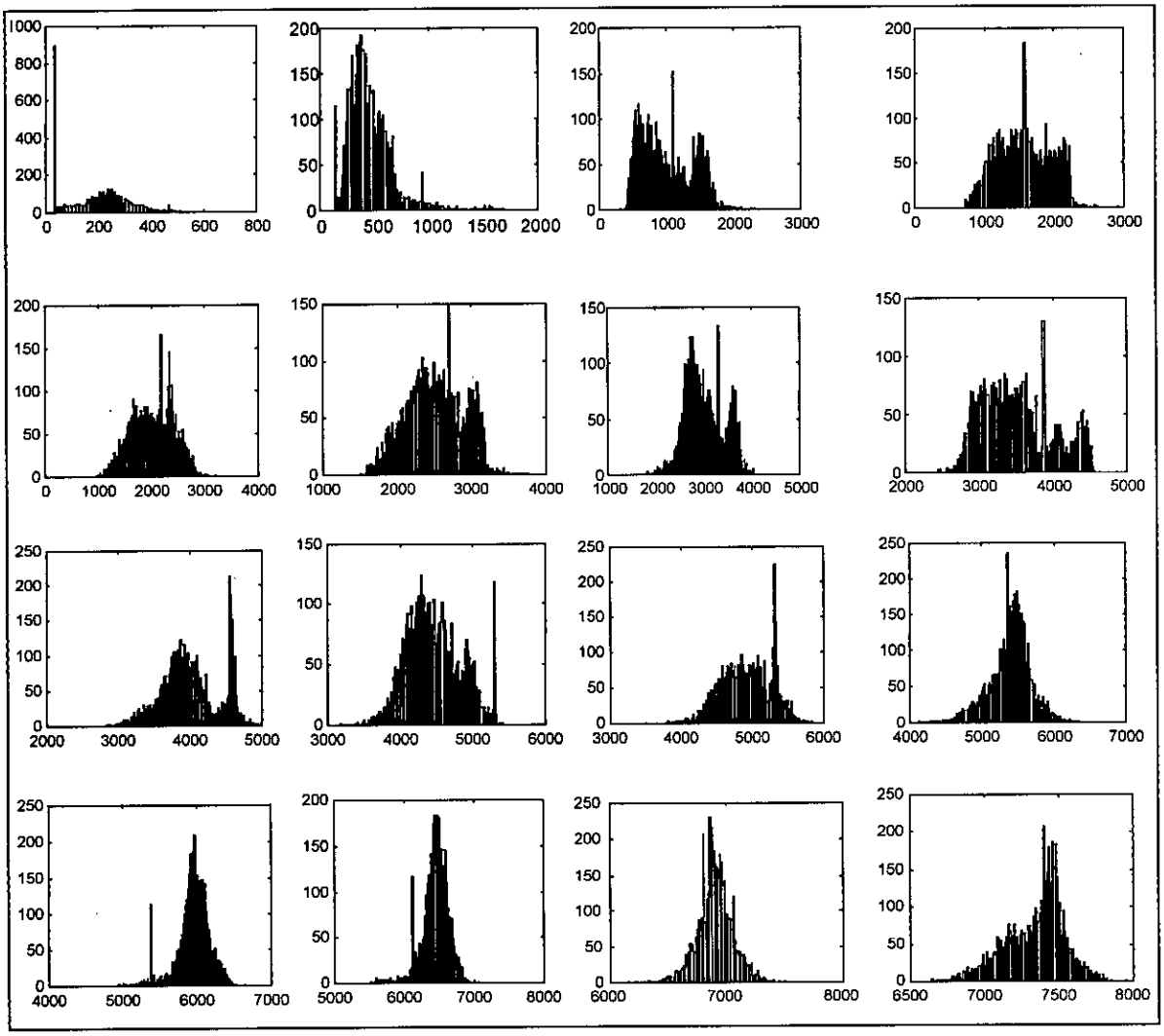


Fig. 2.7 : Histogramme de chaque paramètre LSF

2.7 CONCLUSION

Les coefficients du filtre LP sont déterminés à partir du signal de parole en utilisant les techniques de prédiction linéaire. Le taux d'actualisation des coefficients LP est relaté aux caractéristiques du conduit vocal. Variant de 30 à 100 périodes par seconde (chaque 30 à 10 ms).

Généralement, les coefficients LP ne sont pas codés directement mais transformés en un ensemble de paramètres qui ont des propriétés de codage désirables. Les plus utilisés sont les LSF (Line Spectral Frequencies) aussi appelés LSP (Line Spectral Pairs).

Chapitre 3
Codeur-Décodeur CELP à large bande

3.1 INTRODUCTION

Le codage de la parole a été un domaine en cours de recherche durant ces dernières décennies. Maintenant le niveau d'activité et d'intérêt dans ce domaine s'est intensivement développé. Le développement important d'algorithmes pour le codage de la parole a récemment émergé et un excellent progrès a été atteint en produisant une parole de bonne qualité à des débits de 4.8Kb/s. La complexité de ces nouveaux algorithmes plus sophistiqués dépassent grandement ceux des anciennes méthodes (comme L'ADPCM).

En raison des différentes normalisations en cours, de nombreux travaux depuis 1985 ont été effectués sur le codeur CELP.

Le codeur CELP est utilisé comme un codeur à bande étroite à moyen et bas débits dans plusieurs applications, par exemple dans les télécommunications mobiles. Ils fournissent une bonne qualité de parole (near to toll quality) avec une largeur de bande de 3.4 kHz à des débits de 4.8 à 8 kb/s. Cependant, il y a plusieurs applications où il n'est pas nécessaire d'opérer à des débits inférieurs à 4.8 kb/s et où une augmentation en qualité et en largeur de bande de la parole devraient être acceptables. Par exemple un vidéophone devrait utiliser un codeur CELP à bande étroite à moins de 16 kb/s pour coder la parole. Si un codeur de la parole à large bande est utilisé, la qualité perceptuelle totale devrait être améliorée. Il a été constaté que l'addition de la bande de 50 à 300 Hz procure une amélioration appréciable du confort d'écoute de même que l'accroissement de la bande de 3.4 à 7 kHz améliore l'intelligibilité.

Cependant, pour une utilisation vidéophone sur les lignes ISDN à 64 ou 128 kbits/s il est inefficace d'utiliser des débits de 48 à 64 kbits/s pour le codage de la parole large bande 7kHz (G.722) Pour plusieurs applications, telle la téléconférence, il est préférable de pouvoir coder la parole à des débits moyens (7.2 à 16 kbits/s) tout en gardant une bonne qualité de la parole.

Le but de ce chapitre est de présenter le codeur/décodeur CELP large bande (50-7000 Hz).

3.2 DEFINITION D'UN CODEUR CELP IDEAL

On peut définir un codeur vectoriel idéal.

Ce codeur serait composé d'un :

- dictionnaire de taille T contenant des formes d'ondes ;
- critère de comparaison idéal.

L'intérêt d'utiliser un dictionnaire est que la séquence d'excitation, connue du codeur et du décodeur, n'est pas transmise. On ne transmet que son index (Fig.3.1).

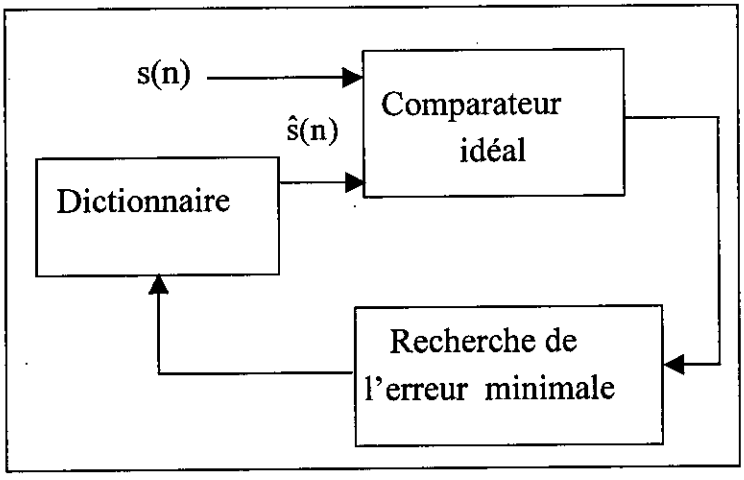


Fig. 3.1 : Codeur vectoriel idéal

Construire un dictionnaire constitué de formes d'ondes nécessiterait d'en stocker suffisamment afin de couvrir toutes les situations possibles. La construction d'un tel dictionnaire serait pratiquement impossible sans compter que le nombre de séquences à tester rendrait l'algorithme de codage complexe.

Aussi, dans le cas d'un codeur hybride, utilise-t-on un modèle H(z). Ce modèle a pour fonction de supprimer les redondances (décorrélation du signal de parole). Le dictionnaire de code contient alors des excitations. Un facteur de gain g_i peut éventuellement exister.

On obtient alors le schéma type d'un codeur hybride vectoriel idéal (Fig.3.2).

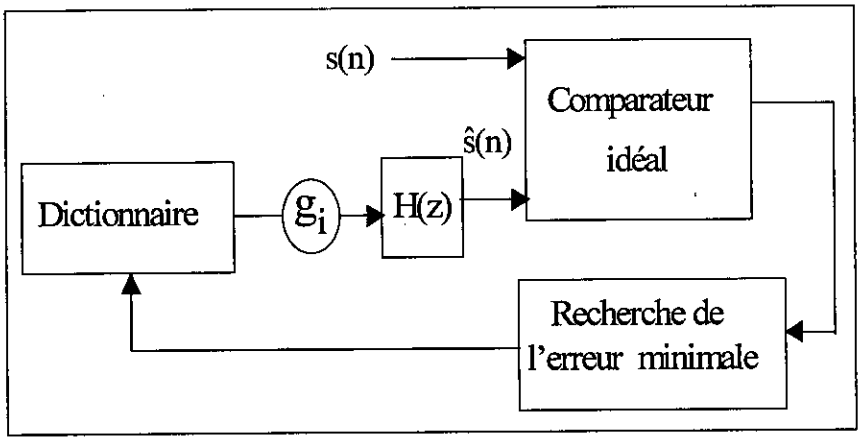


Fig. 3.2 : Codeur hybride idéal

On transmet les paramètres du modèle ainsi que l'index de la séquence d'excitation optimale au décodeur qui produit alors un son synthétique qui a la même qualité que le son original s(n).

Le signal $\hat{s}(n)$ ne ressemble pas nécessairement échantillon par échantillon au signal original. Le comparateur idéal est une distance au sens mathématique. Le signal $\hat{s}(n)$ étant destiné à être écouté, le comparateur idéal doit faire intervenir un critère perceptuel.

En précisant les caractéristiques de chacun des éléments ainsi que les hypothèses réelles, nous définirons l'algorithme itératif standard du codeur CELP

3.2.1 Traitement en trames

Le modèle H(z) utilisé est le filtre linéaire 1/A(z) défini précédemment. Pour des raisons de simplicité de calculs, le critère quadratique est communément utilisé dans les codeurs. Le signal de parole étant, comme nous l'avons vu, localement stationnaire sur des intervalles de temps inférieur à 15 ms. Le signal modélisé est limité en taille. Pour une fréquence d'échantillonnage de 16 kHz, la fenêtre d'analyse comprend typiquement de 160 à 240 échantillons.

Bien évidemment la longueur de la fenêtre de traitement introduit un retard intrinsèque de codage dans le système.

L'intérêt du traitement en trame est le traitement en blocs, aussi appelé vectoriel, du signal de parole. On bénéficie d'emblée de l'avantage de la quantification vectorielle [Gray ,80] par rapport à la quantification scalaire utilisée dans les premières normes de téléphonie numérique d'ITU : MIC G.711 et MICDA G.721, G.722.

3.2.2 Critère de choix

La recherche de la séquence d'excitation optimale nécessite le filtrage de l'ensemble des séquences du dictionnaire.

Soit $\hat{S}_j(z)$ le signal de parole synthétique, réponse du filtre 1/A(z) à la $i^{ème}$ séquence du dictionnaire.

On note:

$$E_w(z) = (S(z) - \hat{S}(z))\, W(z) \tag{3.1}$$

Avec $W(z) = \dfrac{A(z)}{A(z/\gamma)}$

W(z) est le filtre perceptuel [Atal, 79].

$E_w(z)$ est le signal d'erreur entre les signaux original et synthétique filtrés par le filtre perceptuel $W(z)$. On note ε_w la représentation temporelle de $E_w(z)$. Pour des raisons de simplicité, le critère de choix le plus utilisé est le critère quadratique.

La séquence d'excitation c_j du dictionnaire qui est choisie est celle qui minimise l 'erreur quadratique moyenne :

$$Q(j) = \langle \varepsilon_w, \varepsilon_w \rangle \tag{3.2}$$

Le diagramme donnant le principe de la modélisation devient :

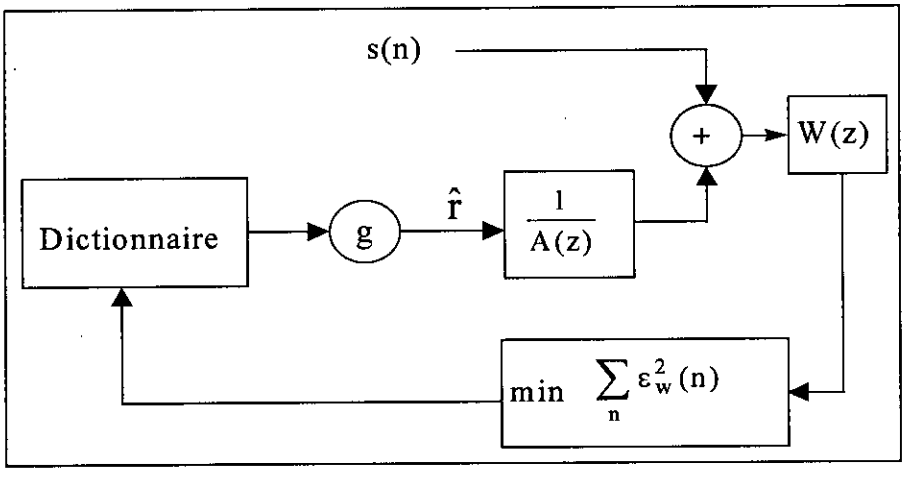


Fig. 3.3 : Schéma d'un codeur hybride

On peut placer en amont le filtre perceptuel. Le signal original **s** filtré par le filtre perceptuel est appelé *signal perceptuel* et est noté **p**. Le signal synthétique perceptuel est noté $\hat{\mathbf{p}}$. Cette modification dans la structure (Fig. 3.4) nous permet de supprimer un filtrage car le signal perceptuel **p** n'est calculé qu'une seule fois et le coût en calcul du filtrage perceptuel dans la boucle de synthèse est nul.

Le diagramme devient :

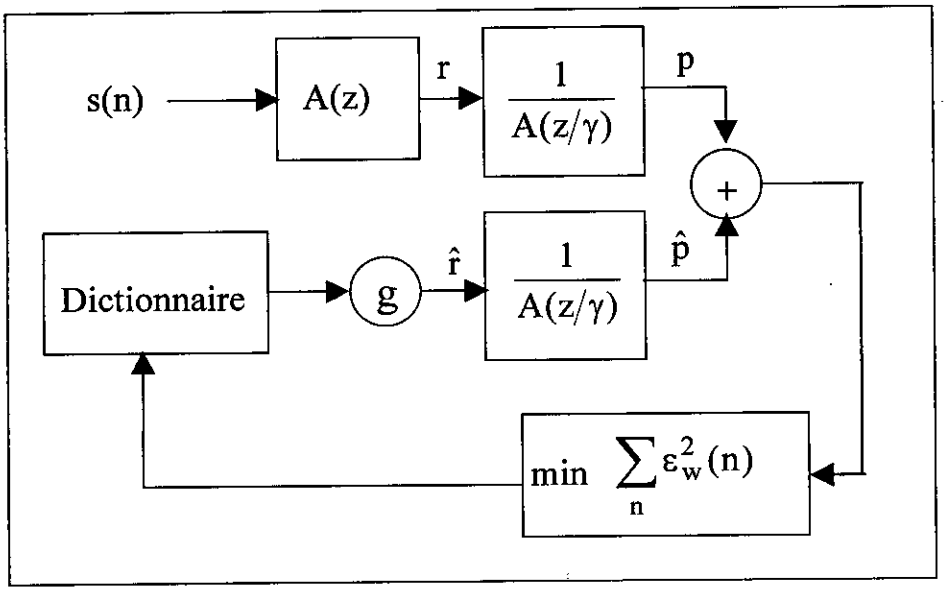


Fig. 3.4 : Schéma du codeur hybride classique

3.3 Générateur de codes

Le générateur de code (dictionnaire) est connu de l'émetteur et du récepteur. Il est composé d'un ou de plusieurs *modules d'excitations*. On cherche dans le générateur de code une séquence d'excitation $\hat{r}_n$ qui minimise au mieux l'erreur quadratique perceptuelle.

Dans sa forme la plus simple, l'excitation est donné par :

$$\hat{r}_n = g_j \, c_n^j \qquad \text{pour } n = 0, \ldots, N-1 \tag{3.3}$$

où g_j est le gain optimal au sens du critère de choix.

$c^j(n)$ est une séquence d'index j choisie dans un dictionnaire.

Plus généralement, en appelant K le nombre de dictionnaires, cette excitation s'écrit

$$\hat{r}_n = \sum_{k=1}^{K} g_{j(k)} \, c_n^{j(k)} \tag{3.4}$$

et dans ce cas $\hat{p} = \sum_{k=1}^{K} g_{j(k)} \, f^{j(k)}$

$f^{j(k)}$ étant le résultat de filtrage du vecteur $c^{j(k)}$ par le filtre perceptuel.

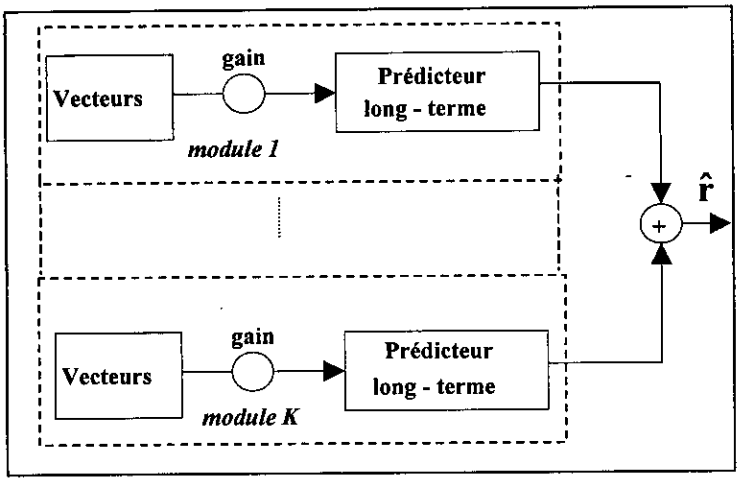


Fig. 3.5 : Générateur de codes

3.4 Contenu d'un module d'excitations

Dans le cas d'un codeur CELP, un module d'excitations est composé d'un :

- dictionnaire de séquences connues de l'émetteur et du récepteur ;
- filtre optionnel dont le rôle est de reproduire la structure périodique du son voisé.

3.4.1 Le dictionnaire

Outre le dictionnaire aléatoire Gaussien introduit par [Schroeder, 85], le contenu du dictionnaire peut être adapté à un algorithme particulier.

On trouve des dictionnaires :

- mono-impulsionnelles, [Shingal, 84] [Atal, 82] ;
- lacunaires [Davidson, 87] [LIN, 86], ;
- d'impulsions régulièrement espacées [Kroon, 86]
- algébriques structurés [Adoul, 87],
- construits avec des séquences multipulses [Woo, 88],
- binaires où les échantillons prennent les valeurs -1 et +1 [Le guyader, 88] [Salami, 89],
- ternaires où les échantillons prennent les valeurs -1, 0 et +1 [Campbell, 91].

3.4.2 Module avec filtre prédicteur long terme

L'équation en z du prédicteur long-terme en synthèse est donnée dans le cas le plus simple par :

$$P(z) = \frac{1}{1-\beta\, z^{-D}} \qquad (3.5)$$

C'est un filtre à réponse impulsionnelle infinie.

Soient h_n , e_n , $\hat{r}_n$,respectivement, la réponse impulsionnelle du filtre, son excitation et la sortie du filtre P(z).

Nous aurons

$$\hat{r}_n = \sum_{i=0}^{N'-1} h_i \, e_{n-i} + \sum_{i=N'}^{\infty} h_i e_{n-i} \qquad \text{pour } n=0...N'-1 \tag{3.6}$$

N' étant la taille de la sous trame

ou encore

$$\hat{r}_n = e_n + \beta \hat{r}_{n-D} \tag{3.7}$$

La gamme de variation du coefficient D s'étend de $D \in [D_{min} = 40, D_{max} = 295]$. Ceci permet de reconstituer des fréquences du fondamental allant de 54.4 Hz à 400 Hz.

Pour n = 0, . . . , N', $\hat{r}_n$ correspond à la sortie du filtre P(z) pour la fenêtre courante. Le terme $\hat{r}(n-D)$ pour n = 0, . . . , N'-1 et pour D = 40, . . . , 295 peut prendre des valeurs allant de $\hat{r}(-295)$ jusqu'à $\hat{r}(N'-1-40)$. Les valeurs comprises entre $\hat{r}(-295)$ et $\hat{r}(-1)$ sont la mémoire des fenêtres d'analyse précédentes.

Notons **m** le vecteur contenant toutes les valeurs de la mémoire :

$$\mathbf{m} = [\hat{r}(-D_{max}) \; \hat{r}(-D_{max}+1) \Lambda \, \hat{r}(-D_{max}+N'-1) \; \hat{r}(-D_{max}+N') \Lambda \, \hat{r}(-D_{min}) \Lambda \, \hat{r}(-1)] \tag{3.8}$$

Supposons que $D = D_{max}$, alors la sortie du filtre P(z) pour la fenêtre d'analyse courante sera la somme de deux vecteurs de taille N' :

- un vecteur $e^j = g_j [c_0^j, \Lambda \, , c_{N'-1}^j]$ correspondant à une séquence issue du dictionnaire.

- un vecteur $\beta \, m'_{D_{max}} = \beta \, [\hat{r}(-D_{max}) K \; \hat{r}(-D_{max}+N'-1)]$ correspondant à N' valeurs successives de **m** dont le premier élément a pour indice $D = D_{max}$.

Si $D = D_{max} - 1$ alors $\beta \, m'_{D_{max}-1} = \beta \, [\hat{r}(-D_{max}+1) K \; \hat{r}(-D_{max}+N')]$

La mémoire **m** constitue un dictionnaire dont le premier vecteur est $m_{D_{max}}$, le second $m_{D_{max}-1}$, etc. Le vecteur suivant étant obtenu en prenant comme premier élément l'échantillon d'indice augmenté de 1 par rapport à l'indice du premier élément du vecteur précédent.

Vecteur 2

$$m = [\hat{r}(-D_{max})\ \hat{r}(-D_{max}+1)K\ \hat{r}(-D_{max}+N'-1)\ \hat{r}(-D_{max}+N')K\ \hat{r}(-D_{min})K\ \hat{r}(-1)]$$

Vecteur 1

Nous représentons ce dictionnaire sous une forme plus conventionnelle ; chaque vecteur est représenté séparément (Fig 3.6).

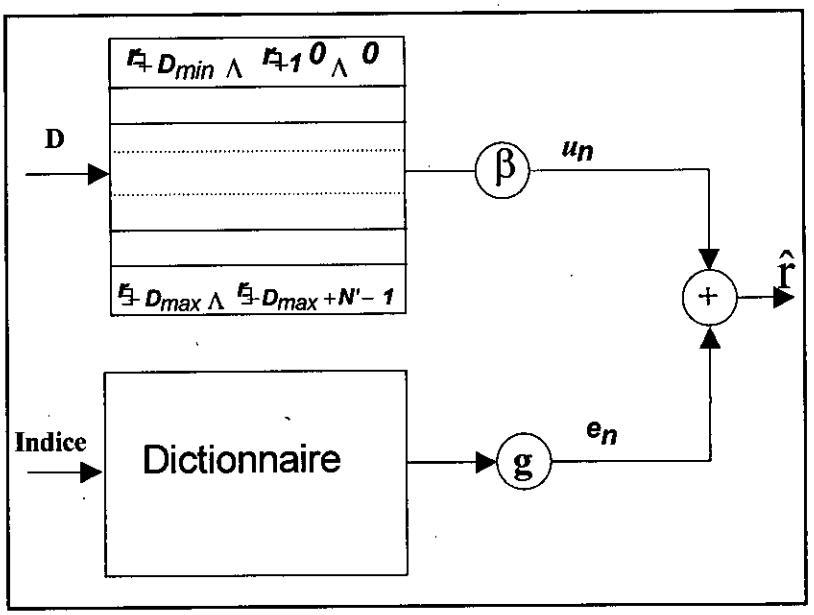


Fig. 3.6 : Modélisation du filtre LTP

- lorsque D≥ N', on peut considérer que l'excitation reconstruite courante $\hat{r}$ est la somme d'un vecteur de taille N' issu de la mémoire des excitations passées $\hat{r}_{-D}$ Λ $\hat{r}_{-D+N'-1}$ avec un vecteur du dictionnaire multiplié par un gain.
- lorsque $N'/2 \le D \le N'$, le vecteur issu de la mémoire des excitations n'est plus complet :

$$\hat{r}_D = [\hat{r}_{-D}\ \ \Lambda\ \ \hat{r}_{-1}\ 0\ \ \Lambda\ \ 0] \qquad (3.9)$$

L'excitation reconstruite peut s'écrire :

$$\begin{aligned}
\hat{r}_0 &= e_0 + \beta\, \hat{r}_{-D} \\
\hat{r}_1 &= e_1 + \beta\, \hat{r}_{-D+1} \\
&\mathrm{M} \\
\hat{r}_{D-1} &= e_{D-1} + \beta\, \hat{r}_{-1} \\
\hat{r}_D &= e_D + \beta\, \hat{r}_0 \\
&\mathrm{M} \\
\hat{r}_{N'-1} &= e_{N'-1} + \beta\, \hat{r}_{N'-1-D}
\end{aligned} \qquad (3.10)$$

L'excitation sera donc la somme de 3 vecteurs :

$$\hat{r}_D = \left[\hat{r}_{-D} \quad \Lambda \quad \hat{r}_{-1}\ 0 \quad \Lambda \quad 0\right] \qquad (3.11)$$

$$e^j = g_j \left[c_0^j \quad \Lambda \quad c_{N'-1}^j\right] \qquad (3.12)$$

$$\begin{aligned}
\hat{r}'_D &= \left[0 \quad \Lambda \quad 0\ \beta\, \hat{r}_0 \quad \Lambda \quad \beta\, \hat{r}_{N'-1-D}\right] \\
&= \left[0 \quad \Lambda \quad 0\ \beta\left(e_0 + \beta \hat{r}_{-D}\right) \quad \Lambda \quad \beta\left(e_{N'-1-D} + \beta\, \hat{r}_{N'-1-2D}\right)\right]
\end{aligned} \qquad (3.13)$$

- lorsque D < N'/2, on retrouve des résultats identiques avec une écriture un peu plus compliquée.

On a :

$$\text{Mémoire} \begin{cases} \hat{r}_0 = e_0 + \beta\, \hat{r}_{-D} \\ \mathrm{M} \quad \mathrm{M} \quad \mathrm{M} \\ \hat{r}_{D-1} = e_{D-1} + \beta\, \hat{r}_{-1} \end{cases} \qquad (3.14)$$

$$\text{Mémoire + excitation courante} \begin{cases} \hat{r}_D = e_D + \beta\, \hat{r}_0 \\ \mathrm{M} \quad \mathrm{M} \quad \mathrm{M} \\ \hat{r}_{2D-1} = e_{2D-1} + \beta\, \hat{r}_{D-1} \end{cases} \qquad (3.15)$$

$$\text{Mémoire + 2 échantillons de l'excitation courante} \begin{cases} \hat{r}_{2D} = e_{2D} + \beta\, \hat{r}_D \\ \mathrm{M} \quad \mathrm{M} \quad \mathrm{M} \\ \hat{r}_{N'-1} = e_{N'-1} + \beta\, \hat{r}_{N'-1-D} \end{cases} \qquad (3.16)$$

On voit que pour connaître $\hat{r}_{2D}$, il nous faut connaître $\hat{r}_D$. On pourrait l'écrire :

$$\begin{aligned}\hat{r}_{2D} &= e_{2D} + \beta\,\hat{r}_D \\ &= e_{2D} + \beta\,(e_D + \beta\hat{r}_0) \\ &= e_{2D} + \beta\,(e_D + \beta(e_0 + \beta\hat{r}_{-D})) \\ &= e_{2D} + \beta e_D + \beta^2 e_0 + \beta^3 \hat{r}_{-D}\end{aligned} \tag{3.17}$$

L'excitation pour un module avec filtre prédicteur long terme (PLT) est codée par 4 paramètres :

$$\underbrace{\text{codage}(D) + \text{codage}(\beta)}_{\text{long-terme}} + \underbrace{\text{codage}(\text{indice}) + \text{codage}(g)}_{\text{dictionnaire}}$$

La séquence d'excitation synthétique se décompose en deux termes :

$$\hat{r}_n = g_j c_n^j + \beta\,\hat{r}_{n-D} \tag{3.18}$$

Le premier terme $g_j c_n^j$ est à déterminer, quant au second, il est a priori connu.

En notant h_n la réponse impulsionnelle du filtre de synthèse, la recherche de l'excitation consiste alors à chercher l'indice D et le gain β, dans la mémoire des excitations passées puis une séquence optimale et un gain dans le dictionnaire.

On obtient alors :

$$\hat{p}_n = \sum_{i=0}^{\infty} h_i \hat{r}_{n-i} \qquad n=0,\ldots,N'-1 \tag{3.19}$$

$$\hat{p}_n = \sum_{i=0}^{n} h_i \hat{r}_{n-i} + \sum_{i=n+1}^{\infty} h_i \hat{r}_{n-i} \tag{3.20}$$

$$\hat{p}_n = g_j \sum_{i=0}^{n} h_i c_{n-i}^j + \beta \sum_{i=0}^{n} h_i \hat{r}_{n-i-D} + \sum_{i=n+1}^{\infty} h_i \hat{r}_{n-i} \tag{3.21}$$

On appelle

$\hat{p}^0$ le vecteur tel que :

$$\hat{p}^0 = \sum_{i=n+1}^{\infty} h_i \hat{r}_{n-i} \tag{3.22}$$

où $\hat{p}^0$ représente donc la mémoire du filtre $1/A(z/\gamma)$.

Soient :

$$\hat{p}^1 = \beta \sum_{i=0}^{n} h_i \hat{r}_{n-i-D} \tag{3.23}$$

$$\hat{p}^2 = g_j \sum_{i=0}^{n} h_i c^j_{n-i} \tag{3.24}$$

Le vecteur $\hat{p}^2$ est a priori connu. Dans le cas où D<N', le calcul du vecteur $\hat{p}^1$ fait intervenir, comme nous l'avons vu auparavant, des échantillons de l'excitation courante.

3.5 Mémoire des filtres

La séquence d'excitation $\hat{r}$ est choisie si le signal $\hat{p}$, réponse du filtre $1/A(z/\gamma)$ à $\hat{r}$, minimise le critère $\langle \boldsymbol{p} - \vec{\boldsymbol{p}}, \boldsymbol{p} - \vec{\boldsymbol{p}} \rangle$.

Le terme $\hat{p}^0_n = \sum_{i=n+1}^{\infty} h_i \hat{r}_{n-i}$ de l'équation (3.20) correspond aux conditions initiales du filtre de synthèse $1/A(z/\gamma)$.

Ces conditions initiales sont dues aux excitations des sous trames précédentes [Hernandez, 86].

Comme la séquence $\hat{p}^0$ est constante et ne dépend pas du mot de code choisi, on peut la soustraire du signal de parole original perceptualisé **p**. Soit **t** la séquence d'excitation telle que :

$$Ht = p - \hat{p}^0 \tag{3.25}$$

La séquence **t** est appelée *séquence cible* [Kleijn, 90]. Cette séquence est l'excitation pour la sous trame qui permet une reconstruction parfaite de **p**. Cette excitation ne dépend que de la sous trame vu que la mémoire a été soustraite.

En notant p^1 tel que : $$p^1 = p - \hat{p}^0 = Ht \tag{3.26}$$

Le critère à minimiser sera donc :

$$\langle p^1 - \vec{p}^1, p^1 - \vec{p}^1 \rangle \tag{3.27}$$

Le filtre de synthèse $1/A(z/\gamma)$ est donc un filtre sans conditions initiales.

3.6 Modélisation optimale de l'erreur au sens des moindres carrés

Dans la forme la plus simple, on peut poser le problème comme suit :
On part d'une observation qui est, dans notre cas, le signal de parole original perceptualisé auquel on a enlevé la contribution des fenêtres précédentes. On cherche alors à construire un estimateur $\hat{p}$ fonction de l'observation et optimisant le critère des moindres carrés.

$\hat{p}$ est une combinaison linéaire de K vecteurs où K est l'ordre de la modélisation

$$\begin{aligned} \hat{p} &= \sum_{k=1}^{K} g_{j(k)}\, f^{j(k)} \\ &= Ag \end{aligned} \tag{3.28}$$

Les colonnes de la matrice A représentent les vecteurs $f^{j(k)}$ que l'on supposera connus.

Pour que $\hat{p}$ soit la meilleure estimation en moyenne quadratique de p, il faut minimiser l'énergie :

$$\begin{aligned} E &= \| p - \hat{p} \|^2 \\ &= \| p - A g \|^2 \end{aligned} \tag{3.29}$$

Les vecteurs colonnes de A définissent un sous espace H.

On cherche alors un vecteur $\hat{p}$ de ce sous espace dont la distance à p est minimale. Le problème consiste alors à trouver g tel que l'énergie soit minimale. Le théorème de projection nous donne la solution suivante :

$$\hat{p} = \mathrm{proj}(p/H) \tag{3.30}$$

La meilleure approximation $\hat{p}$ de p est la projection orthogonale de p dans le sous espace engendré par les vecteurs colonnes de A.

$$\left(f^{j(k)}\right)^T (p - Ag) = 0 \qquad k = 1, K, K \tag{3.31}$$

ce qui s'écrit

$$\begin{aligned} A^T \hat{p} &= A^T p \\ A^T A g &= A^T p \\ g &= \left(A^T A\right)^{-1} A^T p \end{aligned} \tag{3.32}$$

Pour un sous espace H, donné, la solution optimale est :

$$\hat{p} = A\left(A^t A\right)^{-1} A^t p \tag{3.33}$$

Les vecteurs colonnes sont les vecteurs des dictionnaires filtrés.

La minimisation de E consiste à :

- choisir une combinaison $j(1),K\ j(K)$
- calculer le vecteur g optimal associé.

Le vecteur $\hat{p}$, estimation de p est celui qui minimise E. La recherche de la solution optimale impose donc d'essayer chacun des K uplets formés de K séquences issues des K dictionnaires et de calculer le vecteur g associé.

Si les K vecteurs sont issus de K différents dictionnaires de taille $T_1,\Lambda\ ,T_k$. Le nombre de possibilités est :

$$n_1 = T_1 T_2 \Lambda \qquad T_K$$

Si les K vecteurs sont issus du même dictionnaire, le nombre de possibilités est :

$$n_2 = \frac{T!}{K!(T-K)!} = C_T^K$$

On utilise plus généralement un algorithme sous optimal ; les vecteurs filtrés sont cherchés itérativement.

Sachant que $\hat{p}$ est égale à :

$$\hat{p} = \sum_{k=1}^{K} g_{j(k)} f^{j(k)} \tag{3.34}$$

On cherche à minimiser

$$E_n = \left\| p - \sum_{k=1}^{K} g_{j(k)} f^{j(k)} \right\|^2 \qquad \text{pour n = 1, . . . , K} \tag{3.35}$$

- on trouve alors une combinaison j(1), . . . , j(K).
- on résout le système $g = \left(A^t A\right)^{-1} A^t p$ une seule fois avec les séquences j(1), . . . , j(K) trouvées [Moreau, 91].

3.7 Algorithme itératif standard

Le calcul de la solution optimale nécessite, à chaque essai d'une séquence filtrée, d'inverser la matrice A^tA.

L'algorithme standard est sous optimal ; il trouve la solution itérativement :

Pour n = 1, . . . ,K

$$E_n = \left\| p^1 - \sum_{i=1}^{n} g_{j(k)} f^{j(k)} \right\|^2 \tag{3.36}$$

Dans ce cas, la matrice A^tA s'écrit $A^tA = \left\langle f^j, f^j \right\rangle$ et pour la première étape, on doit choisir $j(1)$ qui maximise :

$$\left\langle p, f^j \right\rangle \left\langle f^j, f^j \right\rangle^{-1} \left\langle f^j, p \right\rangle = \frac{\left\langle f^j, p \right\rangle^2}{\left\langle f^j, f^j \right\rangle} \tag{3.37}$$

puis calculer le gain :

$$g_{j(1)} = \frac{\left\langle f^{j(1)}, p \right\rangle}{\left\langle f^{j(1)}, f^{j(1)} \right\rangle} \tag{3.38}$$

A la $k^{ième}$ itération, la contribution des k-1 premiers vecteurs $f^{j(i)}$ est retirée de p :

$$p^k = p - \sum_{i=1}^{k-1} g_{j(i)} f^{j(i)} \tag{3.39}$$

et un nouvel index j(k) et un nouveau gain $g_{j(k)}$ sont calculés vérifiant :

$$\begin{aligned} j(k) &= \arg \max \frac{\left\langle f^j, p^k \right\rangle^2}{\left\langle f^j, f^j \right\rangle} \\ g_{j(k)} &= \frac{\left\langle f^{j(k)}, p \right\rangle}{\left\langle f^{j(k)}, f^{j(k)} \right\rangle} \end{aligned} \tag{3.40}$$

Etant donné que :

$$\left\langle f^j, p^{k+1} \right\rangle = \left\langle f^j, p^k \right\rangle - g_{j(k)} \left\langle f^j, f^{j(k)} \right\rangle$$

$$\left\langle f^j, p^{k+1} \right\rangle = \left\langle f^j, p^k \right\rangle - \frac{\left\langle f^j, f^{j(k)} \right\rangle}{\left\langle f^{j(k)}, f^{j(k)} \right\rangle} \left\langle f^{j(k)}, p^k \right\rangle \quad (3.41)$$

L'intercorrélation $\left\langle f^j, p^{k+1} \right\rangle$ nécessaire à l'étape k+1 peut être calculée à partir de l'intercorrélation à l'étape k.

On obtient l'**ALGORITHME ITERATIF STANDARD**

Pour $j = 1, K, T$

- $\alpha^j = \left\langle f^j, f^j \right\rangle$ et $\beta_1^j = \left\langle f^j, p \right\rangle$

Pour $k = 1, K, K$

- $j(k) = \arg\max \frac{(\beta_k^j)^2}{\alpha^j}$ et $g_{j(k)} = \frac{\beta_k^{j(k)}}{\alpha^{j(k)}}$
- Pour $j = 1, K, T$ (si k<K)
  1. $r(k, j) = \frac{\left\langle f^{j(k)}, f^j \right\rangle}{\alpha^{j(k)}}$
  2. $\beta_{k+1}^j = \beta_k^j - r(k, j)\beta_k^{j(k)}$

Cet algorithme est applicable quel que soit le contenu du dictionnaire d'excitation.

3.8 ALGORITHMES RAPIDES

A partir des différents travaux qui ont été réalisés depuis le modèle de Shroeder deux tendances se dégagent :

- réduction de la complexité rendant possible une implantation proche voire temps réel ;
- amélioration du rapport qualité / débit.

La charge principale de calcul est essentiellement liée au calcul du dictionnaire filtré et à la recherche des vecteurs permettant de modéliser le signal perceptuel p^1.

3.8.1 Modification de l'algorithme de filtrage du dictionnaire d'excitation

3.8.1.1 Le dictionnaire linéaire

On peut calculer la sortie du filtre $1/A(z/\gamma)$ pour le $j^{ème}$ mot de code en utilisant le produit de convolution :

$$f_n^j = \sum_{i=0}^{n} h_i c_{n-i}^j \tag{3.42}$$

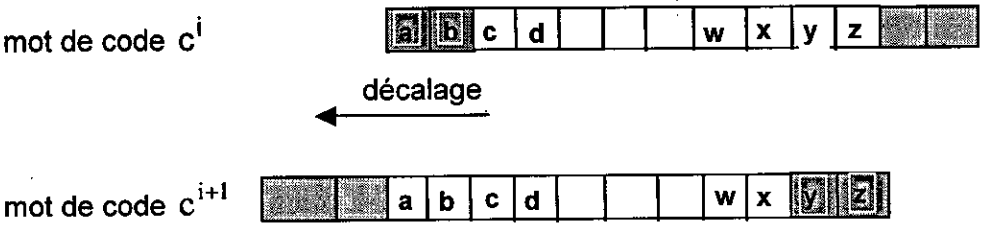
Une réduction substantielle de calculs peut être obtenue en utilisant un dictionnaire linéaire ou de mots de code non disjoints [Lin, 86] (deux mots de code successifs sont décalés de k échantillons). Pour un dictionnaire de taille T mots de N' échantillons, le stockage mémoire est donc limité à k (T-1)+ N' à comparer avec T N' pour le dictionnaire standard.

Soit **v** le vecteur contenant l'ensemble des échantillons du dictionnaire. Le $i^{ème}$ mot de code c^i du dictionnaire est tel que :

$$c_n^i = v(n + k(T-1) - ki) \quad \text{où} \quad i = 0, \Lambda, T\text{-}1 \tag{3.43}$$

On sait que

$$f_n^i = \sum_{j=0}^{n} c_j^i h_{n-j} \tag{3.44}$$



Pour un décalage k = 1, on a

$$f_n^i = \sum_{j=0}^{n} v(T-1-i+j) h_{n-j} \tag{3.45}$$

De même, le vecteur filtré suivant s'écrit :

$$\begin{aligned} f_n^{i+1} &= \sum_{j=0}^{n} v(T-1-(i+1)+j)h_{n-j} \\ &= \sum_{j=1}^{n} v(T-1-(i+1)+j)h_{n-j} + v(T-1-(i+1))h_n \\ &= \sum_{j=0}^{n-1} v(T-1-(i+1)+(j+1))h_{n-1-j} + h_n c_0^{i+1} \\ &= f_{n-1}^{i} + h_n c_0^{i+1} \end{aligned} \qquad (3.46)$$

Nous obtenons finalement les équations récurrentes pour f_j

$$f_n^{i+1} = f_{n-1}^{i} + h_n c_0^{i+1} \quad \text{pour } n = 1,\Lambda\ ,N'-1 \qquad (3.47)$$

$$f_0^{i+1} = h_0 c_0^{i+1} \qquad (3.48)$$

Pour un décalage de k>1, on calcule le vecteur filtré en répétant k fois l'algorithme. Kleijn a montré que pour un décalage supérieur à 2 (k>2) la qualité est équivalente à celle d'un dictionnaire constitué de mots de codes indépendants [Kleijn, 88].

Une réduction supplémentaire peut être obtenue en tronquant la réponse impulsionnelle du filtre, [Kleijn, 90].

3.8.1.2 Dictionnaires spéciaux

Divers dictionnaires ont été proposés par différents auteurs :

- dictionnaire center-clipped (binaire - ternaire),
- dictionnaire déterministe,
- dictionnaire construit,
- dictionnaire contenant des excitations du type multipulse.

En règle générale, les qualités obtenues sont relativement similaires. Par contre, certains types de dictionnaires permettent de simplifier grandement l'algorithme de calcul de la séquence filtrée.

- Des ***dictionnaires pauvres en échantillons*** (sparse code-book) contenant typiquement jusqu'à Np = 3 échantillons non nuls pour une sous trame de longueur N' = 40(95.5% de zéros) permettent d'obtenir une qualité de synthèse subjectivement équivalente à des

dictionnaires contenant moins de zéros [Davidson, 86] [Laflamme, 90].[Taniguchi, 91] La complexité de calcul de la convolution est réduite à $N_p(N'+1)/2$.

- Des ***dictionnaire center - clipping*** :[Lin, 86][Davidson, 88]. Les échantillons gaussiens sont transformés en (-1), 0, (+1) selon le signe et selon qu'ils sont ou non supérieurs à un seuil (généralement 1.96 σ) [Lin, 86]. Le gain en complexité pour le calcul de la convolution est lié au seuil. Le calcul d'une convolution nécessite $N'(N'+1)/2$ opérations.

  Lorsque le seuil est grand, la plupart des échantillons sont nuls. Pour 1.96σ, le gain du dictionnaire est de l'ordre de 20. Des études [Lin, 86] [Kleijn, 88] [Ribbum, 91] montrent que la qualité subjective n'est pas dégradée. Certains bruits haute fréquence disparaissent.
- Des ***dictionnaires déterministes*** ont été proposés. Ils sont construits à partir :

  des codes correcteurs d'erreur : Nordstom - Robinson, Reed - Muller, Schroeder - Sloane, Golay et des codes sphériques [Adoul, Lamblin, 87] [Adoul, Mabilleau, 87] [Laflamme, Su, 90]. Ils permettent de trouver la meilleure séquence dans le dictionnaire avec une complexité de $N\log_2(N)$.

  d'algorithmes basés sur des codes algébriques : Dictionnaires binaires spécialement construits [Le Guyader, 88], [Salami, 89] ternaires [Ireton, 89] ; [Di Francesco, 92]

3.8.2 Suppression du dictionnaire filtré

3.8.2.1 Méthode de covariance

En notant H la matrice impulsionnelle, le vecteur filtré s'écrit :

$$f^j = Hc^j$$

On obtient alors :

$$\begin{aligned}\left\langle f^j, f^j \right\rangle &= \left(c^j\right)^t H^t H c^j \\ &= \left(c^j\right)^t \Gamma c^j\end{aligned} \tag{3.49}$$

où Γ est la matrice de covariance.

$$\begin{aligned}\left\langle p, f^j \right\rangle &= p^t H c^j \\ \left\langle f^j, f^j \right\rangle &= \left(c^j\right)^t \Gamma c^j\end{aligned} \tag{3.50}$$

La procédure utilisée pour la recherche de la meilleure séquence du dictionnaire stochastique est illustrée par la figure 3.7.

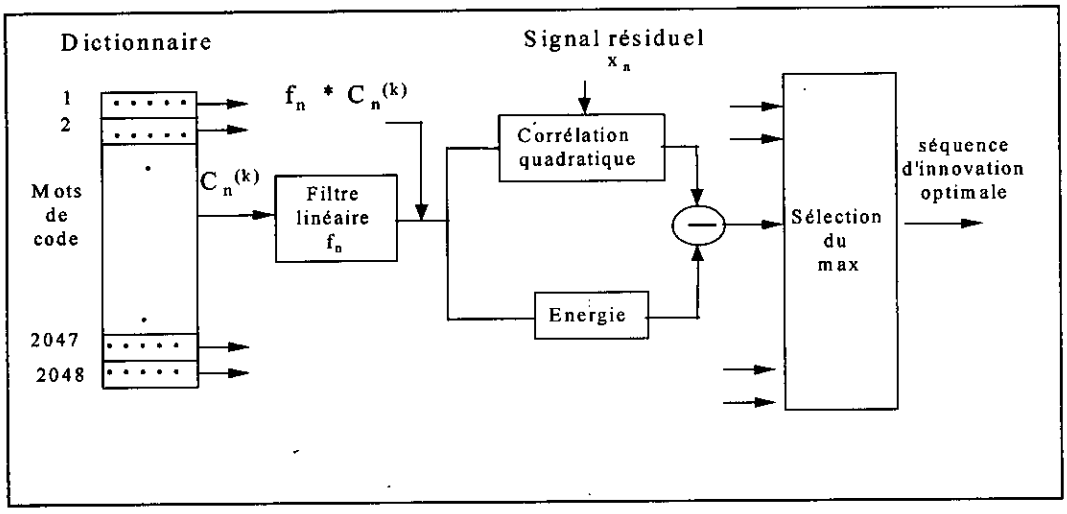


Fig.3.7 : Procédure de recherche pour la détermination du meilleur code stochastique

On distingue 3 étapes :

- calcul du vecteur filtré ;
- calcul de l'intercorrélation et de l'énergie ;
- recherche du maximum.

La première de ces étapes est celle qui nécessite le plus de calculs.

On peut supprimer le calcul du dictionnaire filtré. Dans un premier temps, on remarque que l'intercorrélation s'écrit :

$$\begin{aligned} \langle p, f^j \rangle &= \langle p, Hc^j \rangle \\ &= \langle H^t p, c^j \rangle \\ &= \langle q, c^j \rangle \end{aligned} \tag{3.51}$$

Le vecteur q est calculé une fois pour toute et ne dépend pas du mot de code choisi. Cette technique est identique à celle du "backward filtering" introduite par [Adoul, 87]. On remarque que la complexité du calcul de l'intercorrélation est celle d'un produit scalaire.

3.8.2.2 Principe du filtrage régressif " Backward Filtering "

La procédure de recherche consiste à trouver le mot de code k qui minimise

l'énergie du signal d'erreur (**Fig. 3.8.a.**)

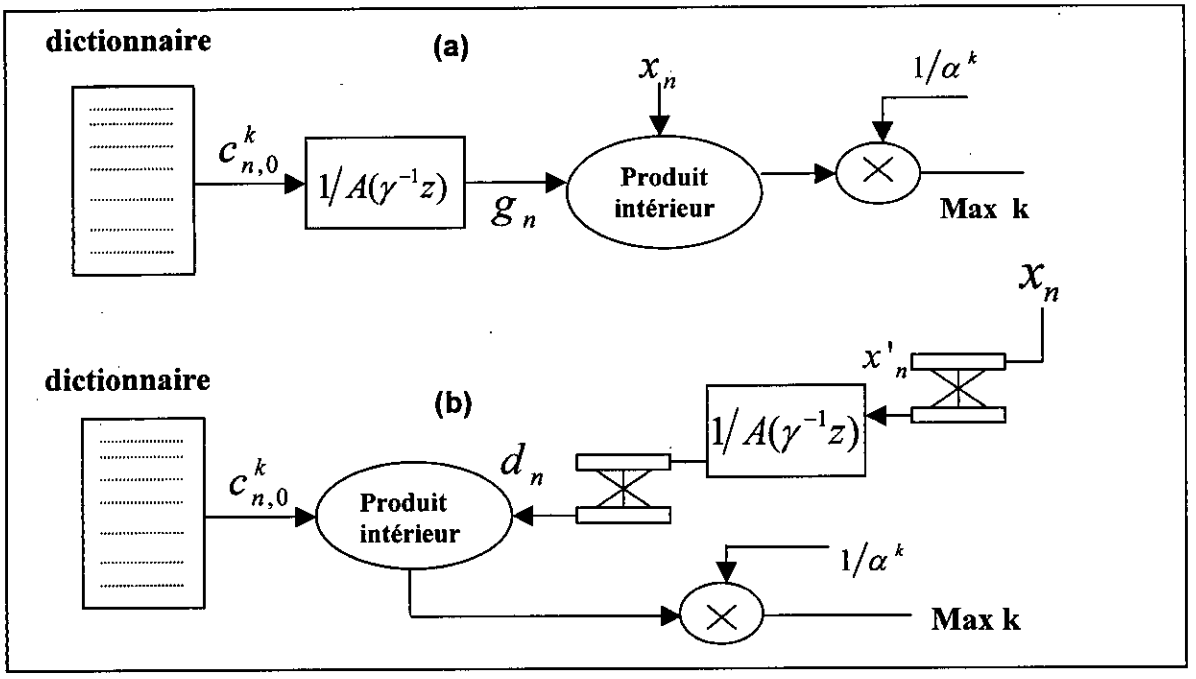


Fig. 3.8 : Principe du " backward filtering "

Puisque g_n est la réponse du filtre inverse sans mémoire $1/A(z/\gamma)$ pour le mot de code c testé couramment, il peut être écrit comme le produit de convolution entre c et la réponse impulsionnelle f_n de $1/A(z/\gamma)$:

$$g_n = \sum_i c_i f_{n-i} \tag{3.51}$$

Soit P le produit intérieur entre x_n et g_n définit par :

$$P = \sum_n x_n g_n \tag{3.52}$$

En utilisant les équations 3.51 et 3.52 , ceci devient :

$$p = \sum_{n=0}^{N-1}\left(x_n \sum_{i=0}^{N-1} c_i f_{n-i} \right) = \sum_{i=0}^{N-1}\left(c_i \sum_{n=0}^{N-1} x_n f_{n-i} \right) \tag{3.53}$$

or

$$p = \sum_{i=0}^{N-1} c_i d_i \tag{3.54}$$

avec

$$d_i = \sum_{n=0}^{N-1} x_n f_{n-i} \tag{3.55}$$

En considérant temporairement les séquences inversées x'_n et d'_n définies par $x'_n = x_{N-n}$ et $d'_n = d_{N-n}$, l'équation 3.55 sera équivalente à :

$$d'_n = \sum_i x'_i f_{n-i} \tag{3.56}$$

La séquence d'_n apparaît dans l'équation 3.56 comme le produit de convolution entre x'_n et la réponse impulsionnelle de $1/A(z/\gamma)$. Ainsi, la séquence d'_n est la réponse de $1/A(z/\gamma)$ à la séquence x'_n, comme montrée en **(Fig 3.8.b.)**.

La séquence x_n est temporairement inversée, inversement filtrée à travers $1/A(z/\gamma)$ et temporairement inversée de nouveau, donnant la séquence d_n, qui ne dépend pas du mot de code couramment testé. Cette procédure est appelée le "Backward Filtering", [17].

Le Backward Filtering est strictement équivalent à la procédure précédente et par conséquent les résultats sont inchangés. Mais au lieu d'un filtrage par mot de code, seulement ici un filtrage est exigé pour tous les mots de code, ce qui diminue la charge de calcul par un facteur approximativement égal à l'ordre de prédiction linéaire.

Maintenant la procédure de recherche consiste à maximiser le produit entre P (calculé comme le produit intérieur entre c_n et d_n) et le facteur $1/\alpha_k$.

3.8.2.3 Calcul du terme α_k

En ne gardant que N_h échantillons de la réponse impulsionnelle, la matrice F s'écrit :

$$F = \begin{bmatrix} h_0 & 0 & 0 & & & 0 & 0 \\ h_1 & h_0 & & & & 0 & 0 \\ M & & & & & & M \\ h_{N_h-1} & & & h_0 & & \Lambda & 0 \\ 0 & h_{N_h-1} & & & & & \\ 0 & 0 & O & & & h_0 & 0 \\ 0 & 0 & \Lambda & 0 & 0 & 0 & h_0 \end{bmatrix} \tag{3.57}$$

La matrice $\boldsymbol{F}'\boldsymbol{F}$ est une matrice de Toeplitz.

Le calcul de α_k est fortement réduit [Trancoso, 86] [Trancoso, 90] en utilisant la propriété que la somme des carrés de la convolution de la réponse impulsionnelle et d'une séquence du dictionnaire est égale à l'intercorrélation de ces séquences.

$$\begin{aligned}\alpha_k &= \|Fc_k\|^2 \\ &= \sum_{n=0}^{N_h-1}\left[\sum_{i=0}^{N_h-1} h_{n-i}c_i^k\right]^2 \\ &= \sum_{i=0}^{N_h-1}\sum_{j=0}^{N_h-1} c_i^k c_j^k \sum_{n=0}^{N_h-1} h_{n-i}h_{n-j}\end{aligned} \tag{3.58}$$

On suppose que la réponse impulsionnelle est nulle pour $n > N_{h-1}$.

On écrit alors :

$$\alpha_k = R_0^{(h)}R_{0,k}^{(c)} + 2\sum_{i=1}^{N_h-1} R_i^{(h)}R_{i,k}^{(c)} \tag{3.59}$$

où les $R_i^{(h)}$ et $R_i^{(c)}$ sont les fonctions d'autocorrélation de h_n et de c_n respectivement :

$$\begin{aligned}R_i^{(h)} &= \sum_{n=0}^{N_h-i-1} h_n h_{n+i} \\ R_{i,k}^{(c)} &= \sum_{n=0}^{N_h-i-1} c_n^k c_{n+i}^k\end{aligned} \qquad \text{pour } i = 0,\ldots,N_h-1 \tag{3.60}$$

Le terme $\beta_k{}^2$ se calcule comme pour la méthode de covariance.

Cette méthode ne fait aucune hypothèse quant à la nature du dictionnaire.

3.9 PROCEDURE DE RECHERCHE DU MEILLEUR CODE

SYNOPTIQUE DU CODEUR/DECODEUR CELP

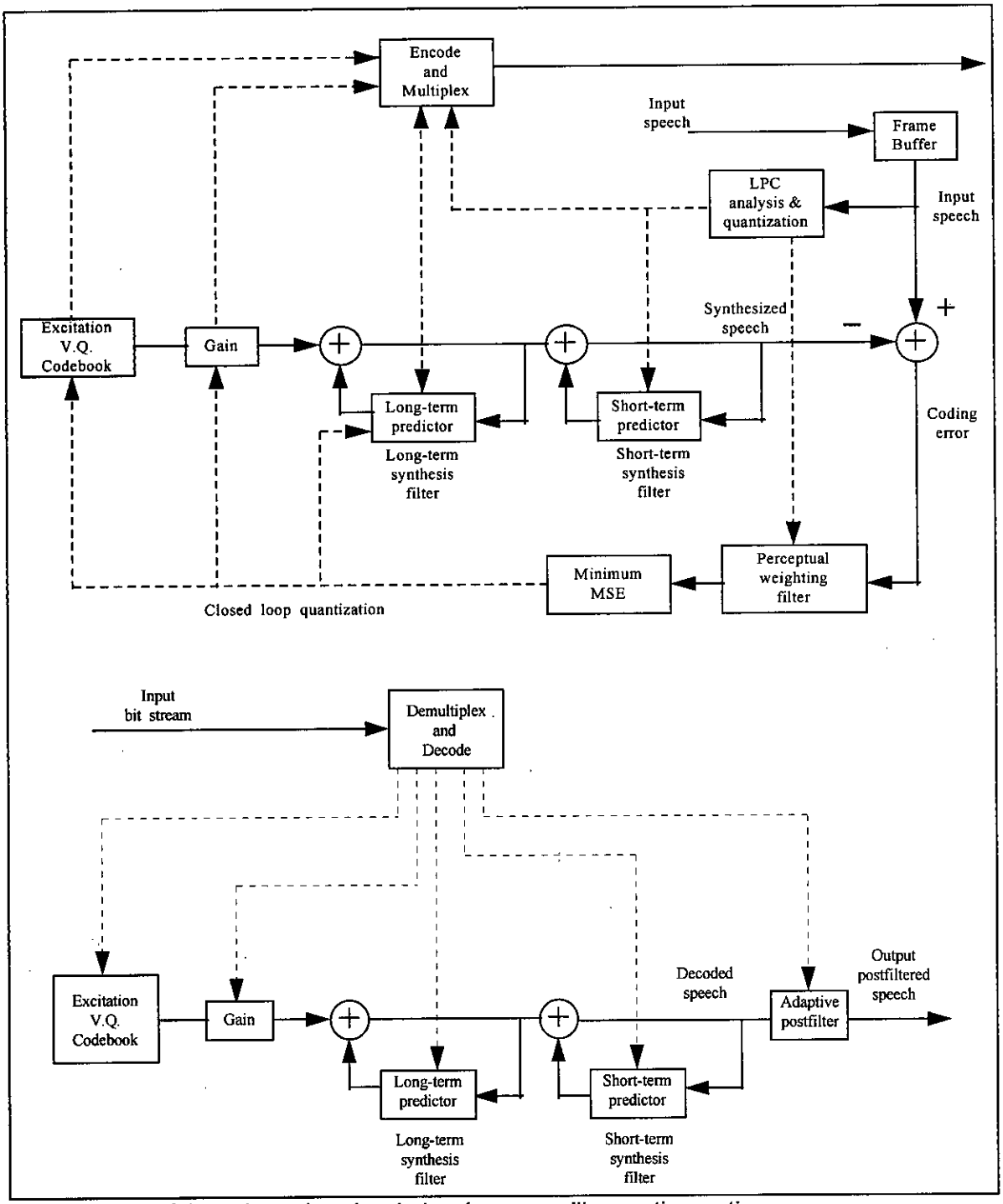


Fig. 3.9 : Procédure de recherche de la séquence d'innovation optimum

Soit **S'** le signal à la sortie du filtre perceptuel $W(z) = \frac{A(z)}{A(z/\gamma)}$. Soit P_0 la mémoire du filtre. Le vecteur $X = S' - P_0$ est le vecteur cible.

Le signal synthétique $\hat{X}$, produit à partir d'un mot de code particulier $C = (c_0, c_1, \Lambda, c_{L-1})$ de dimension L, est exprimé par :

$$\hat{X} = gCH^T$$

où H est une matrice de Toeplitz ($L \times L$) formée à partir de la réponse impulsionnelle des deux filtres en cascade (long terme et court terme)

La séquence d'innovation optimale est le mot de code qui minimise la mesure de distorsion

$$\Delta = \left\| X - gCH^T \right\|^2 \tag{3.61}$$

En prenant $\frac{d\Delta}{dg} = 0$, l'expression suivante donne le gain qui doit être choisi pour un mot de code **C** :

$$g = \frac{X(HC^T)}{\left\| CH^T \right\|^2} \tag{3.62}$$

et le Δ résultant est donné par :

$$\Delta = \|X\|^2 - \frac{\left\| (X(HC^T) \right\|^2}{\left\| CH^T \right\|^2} \tag{3.63}$$

Ainsi, la minimisation de Δ peut être réalisée par maximisation de la quantité absolue suivante :

$$\underset{k}{Max} \left| \frac{X(HC^T)}{\sqrt{\left\| C_k H^T \right\|^2}} \right| \tag{3.64}$$

La recherche se fera sur toutes les entrées du dictionnaire. Le mot de code particulier, C_k, qui maximise l'équation 3.64 est la séquence d'innovation optimale pour le bloc courant du signal de parole d'entrée.

Cette procédure de recherche du mot de code demande beaucoup de calcul, puisque chaque mot de code serait filtré au moins une fois par sous trame de l'analyse LPC.

Plusieurs approches ont été proposées pour réduire la complexité de calcul. Nous avons utilisé la technique du Backward Filtering, qui consiste à réécrire l'expression (3.64) comme :

$$\underset{k}{\text{Max}}\left|\frac{(XH)C_k^T}{\alpha_k}\right| \tag{3.65}$$

où $\alpha = \sqrt{\left\|C_k H^T\right\|^2}$ (le terme "Backward Filtering" vient de l'interprétation de (**XH**) comme le filtrage de X inversé). Le terme α^2, représente l'énergie du mot de code filtré. A cause de ce terme d'énergie, la technique du backward filtering ne réduit pas la complexité de calcul quand on utilise un dictionnaire stochastique.

La méthode du backward filtering n'est efficace que pour un dictionnaire pauvre en échantillons (sparse codebook) [Johnson, 92]

3.10 CONCLUSION

Le codeur CELP a beaucoup évolué depuis le premier modèle de Schroeder et Atal entraînant des modifications de structure et de dictionnaires.

Pour la conception d'un codeur, on doit ou bien employer un algorithme très performant au prix d'un temps de calcul élevée, ou bien on part d'un codeur dont certains paramètres tels le dictionnaire, la longueur des trames ...etc sont imposés.

L'utilisation de la méthode du backward filtering et de la méthode dite réponse à entrée nulle ont permis la réduction de la complexité de notre codeur.

Chapitre 4
Quantification Scalaire et Vectoriel

4.1 INTRODUCTION

Dans ce chapitre, nous exposons les différentes méthodes qui encodent les paramètres spectraux de la parole sur la base de trame par trame. La quantification scalaire code chaque paramètre spectral indépendamment des autres. La quantification vectorielle est alors introduite comme une extension multidimensionnelle de la quantification scalaire dans laquelle le codage s'effectue sur un ensemble entier de paramètres.

Les paramètres LPC sont largement utilisés dans les applications du codage de la parole pour la représentation de l'information de l'enveloppe spectrale de la parole [Kroon, Atal, 91]. Dans ces applications, les paramètres sont obtenus à partir du signal de parole, typiquement à un débit de 50 trames/s, en utilisant l'analyse LPC d'ordre 16 et sont quantifiés pour une éventuelle transmission. Pour les applications à bas débit, il est important de quantifier ces paramètres en utilisant le moins de bits possible.

Des travaux considérables ont été faits dans le passé pour déterminer les quantificateurs optimaux, scalaire et vectoriel, afin de représenter l'information de l'enveloppe spectrale avec le moins de bits possible. Dans l'étude sur la quantification scalaire, différentes représentations des paramètres LPC ont été utilisés. Par exemple, Viswanathan et Makhoul ont utilisés les LAR (Log Area Ratio) pour une quantification scalaire des paramètres LPC. Gray et Markel ont utilisés les coefficients de réflexion arcsine Itakura a proposé la représentation des paires de fréquences spectrales (LSF) qui a montré son efficacité par rapport aux autres représentations [Soong, 84], [Sugamura, Itakura, 86]. Depuis, les paramètres LSF ont été utilisé dans de nombreux études pour la représentation de l'information spectrale [Soong,Juang, 88] [Hagen,Hedelin, 90].

4.2 QUANTIFICATION SCALAIRE

La Quantification Scalaire (QS) assigne à une valeur d'entrée x sa valeur approximée à partir d'un ensemble fini prédéterminé, ou dictionnaire, de N valeurs de sortie acceptables $C=\{y_k / k=1,\ldots,N\}$. Le quantificateur partitionne la sortie en N intervalles I_k de telle manière que l'entrée x est encodée avec la sortie y_k si elle appartient à l'intervalle I_k. Quand un vecteur **x** de dimension m est codé en utilisant la QS, chaque élément x_i du vecteur **x** est indépendamment quantifié tel

$$\hat{x}_i = y_{i,k} = Q_i(x_i), \quad i=1,\ldots,m \tag{4.1}$$

où chaque vecteur du quantificateur $Q_i(x_i)$ serait désigné séparément. La QS a l'avantage de n'avoir besoin que d'une capacité mémoire minimale et nécessite une faible complexité.

La QS peut être vue comme l'introduction d'un bruit $\varepsilon = Q(x) - x$ à l'échantillon d'entrée x. Il y a deux types de quantification : bruit granulaire et bruit de surcharge ou de dépassement. Le bruit granulaire est la différence entre x et $Q(x)$ où ε est compris à l'intérieur d'un intervalle fini définie par les niveaux de décision du quantificateur. Le bruit de surcharge se produit lorsque la valeur d'entrée se situe à l'extérieur de l'intervalle de quantification (Fig. 4.1).

La mesure de distorsion la plus courante dans la conception d'un QS est l'erreur quadratique éntre la valeur originale et la valeur quantifiée :

$$d(x,\hat{x}) = |x - \hat{x}|^2 = |\varepsilon|^2 \tag{4.2}$$

La performance d'un quantificateur scalaire est souvent évaluée en utilisant l'erreur quadratique moyenne (EQM) :

$$D = E\left[d(X,\hat{X})\right] \tag{4.3}$$

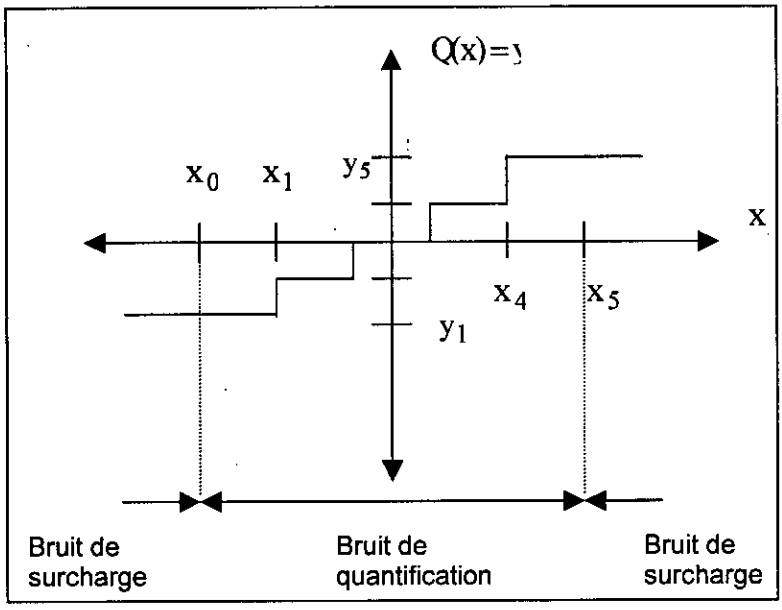


Fig. 4.1 : Exemple d'un QS uniforme pour L = 5

4.2.1 Quantification Uniforme

La *quantification scalaire uniforme* a été la plus utilisée dans la conversion analogique numérique due à sa faible complexité. Les intervalles de décision I_k sont tous

égaux espacés d'une longueur Δ et les niveaux de sortie y_k sont à mi chemins des intervalles de décision, tel que :

$$\begin{aligned}
\Delta &= \frac{x_{max} - x_{min}}{N} \\
I_k &= \{x / x_k < x \leq x_{k+1}\} \\
y_k &= x_{min} + (k - 0.5)\Delta, \quad k=1,...,N \\
Q(x) &= \{y_k / x \in I_k\}
\end{aligned} \tag{4.4}$$

Où x_{min} et x_{max} sont, respectivement, les niveaux d'entrée minimum et maximum observés. Les opérations de troncature et arrondis dans l'approximation des nombres réels en valeurs entières sont des exemples de la quantification uniforme.

Plusieurs codeurs emploient la quantification scalaire pour coder les paramètres spectraux Le choix propre de la représentation paramétrique des coefficients du filtre LPC est déterminée en fonction de leur sensibilité à la quantification. Les coefficients de réflexion ont été souvent utilisés pour la représentation spectrale parce qu'ils étaient moins sensible aux erreurs de quantification que les coefficients prédicteurs [O'Shaugnessy, 87] [Gerson, 90]. Plus récemment, ce sont les paramètres LSF qui sont utilisés.

4.2.2 Quantification Différentielle

Dû à la propriété d'ordonnancement des paramètres LSF, il était prévisible que la différence entre deux des paramètres LSF consécutifs possède un rang dynamique plus petit que celui des paramètres LSF eux mêmes. C'est ce qui a motivé Soong et Juang de quantifier les différences des paramètres LSF consécutifs à la place des paramètres LSF.

Algorithme de Quantification Différentielle

1. quantifier ω_1 en $\vec{\omega}_1$ et poser i =1 ;
2. calculer la différence entre ω_{i+1} et $\vec{\omega}_i$, $\Delta_{\omega_i} = \omega_{i+1} - \hat{\omega}_i$;
3. quantifier $\Delta\omega_i$ en $\Delta\vec{\omega}_i$;
4. reconstruire ω_{i+1} comme $\vec{\omega}_{i+1} = \vec{\omega}_i + \Delta\vec{\omega}_i$;
5. si i = p -1, arrêter ; sinon poser i = i+1 et revenir à 2).

4.2.3 Quantification Adaptative

Nous présentons un algorithme qui utilise la propriété d'ordonnancement des paramètres LSF dans la conception des quantificateurs [Farvardin, 92] L'algorithme utilise le fait que la connaissance de la valeur de ω_i ou sa version quantifiée $\tilde{\omega}_i$ fournie une information utile sur le rang des valeurs possibles de ω_{i+1}; cette information peut être utilisée pour concevoir un meilleur quantificateur pour le $(i+1)^{ième}$ paramètre LSF. Deux versions de cet algorithme sont décrites dans ce qui suit:

A) *Quantification Adaptative séquentielle progressive ou AQFW (Adaptive Quantification ForWard)* :

Algorithme AQFW

1. quantifier ω_1 en $\tilde{\omega}_1$ avec b_1 bits en utilisant un quantificateur uniforme avec un pas, $\Delta = \left(\omega_{1,max} - \omega_{1,min}\right) / 2^{b_i}$; poser i = 1.
2. comparer $\tilde{\omega}_i$ et $\omega_{i+1,min}$,
   - si $\tilde{\omega}_i \leq \omega_{i+1,min}$, alors quantifier ω_{i+1} en $\tilde{\omega}_{i+1}$ avec b_{i+1} bits en utilisant un quantificateur uniforme avec un pas, $\Delta_{i+1} = \left(\omega_{i+1,max} - \omega_{i+1,min}\right) / 2^{b_{i+1}}$; sur l'intervalle $\left[\omega_{i+1,min}, \omega_{i+1,max}\right]$
   - si $\tilde{\omega}_i > \omega_{i+1,min}$, alors quantifier ω_{i+1} en $\tilde{\omega}_{i+1}$ avec b_{i+1} bits en utilisant un quantificateur uniforme avec un pas, $\Delta_{i+1} = \left(\omega_{i+1,max} - \hat{\omega}_i\right) / 2^{b_{i+1}}$; sur l'intervalle $\left[\hat{\omega}_i, \omega_{i+1,max}\right]$,
3. si i = p -1, stop; sinon, poser i = i+1 et revenir à 2)

Il est important de noter que les quantificateurs utilisés dans ce schéma sont adaptatifs. Nous n'avons pas besoin de transmettre une information additionnelle sur le premier LSF. C'est à cause du fait que le quantificateur utilisé pour le codage du $(i+1)^{ième}$ paramètre LSF est uniquement déterminé par la valeur quantifiée de ω_i, qui est disponible et utilisable dans le décodeur.

B) *Quantification Adaptative séquentielle regressive ou AQBW (Adaptive Quantification BackWard)* :

L'idée utilisée dans l'AQFW peut aussi être appliquée dans le sens régressif.

Algorithme AQBW

1. quantifier ω_p en $\hat{\omega}_p$ avec b_p bits en utilisant un quantificateur uniforme avec un pas, $\Delta_p = (\omega_{p,max} - \omega_{p,min}) / 2^{b_p}$; poser i = p
2. comparer $\bar{\omega}_i$ et $\omega_{i-1,max}$
   - si $\bar{\omega}_i \geq \omega_{i-1,max}$, alors quantifier ω_{i-1} en $\hat{\omega}_{i-1}$ avec b_{i-1} bits en utilisant un quantificateur uniforme avec un pas $\Delta_{i-1} = (\hat{\omega}_{i-1,max} - \omega_{i-1,min}) / 2^{b_{i-1}}$;sur l'intervalle $[\omega_{i-1,min}, \hat{\omega}_{i-1,max}]$
   - si $\bar{\omega}_i < \omega_{i-1,max}$, alors quantifier ω_{i-1} en $\hat{\omega}_{i-1}$ avec b_{i-1} bits en utilisant un quantificateur uniforme avec un pas, $\Delta_{i-1} = (\hat{\omega}_i - \omega_{i-1,min}) / 2^{b_{i-1}}$; sur l'intervalle $[\omega_{i-1,min}, \hat{\omega}_i]$,
3. si i = 2, stop; sinon, poser i = i-1 et revenir à 2)

Dans la Quantification Adaptative Progressive et Régressive, puisque l'ordonnancement des paramètres LSF est préservé après quantification, la stabilité du filtre est garantie.

Notons que dans la Quantification Adaptative progressive nous commençons par quantifier le premier paramètre LSF, tandis que dans la Quantification Adaptative Régressive nous commençons par quantifier le dernier paramètre LSF.

4.3 QUANTIFICATION VECTORIELLE (QV)

La quantification vectorielle est une généralisation de la quantification scalaire. Elle concerne la représentation d'un vecteur ***x*** dont les k composantes sont à valeurs réelles continues ($\boldsymbol{x} \in R^k$) par un vecteur appartenant à un ensemble fini{ $\boldsymbol{y_i} \in R^k$, i = 1,2,...,M}

La quantification vectorielle permet d'avoir une constellation qui minimise l'Erreur Quadratique Moyenne pour un dictionnaire de taille M donnée.

La quantification vectorielle peut fournir un décodage rapide en utilisant une simple table d'identification.

La Fig. 4.2 illustre la structure de base d'un quantificateur vectoriel. Le quantificateur vectoriel, ou encodeur, code le vecteur d'entrée x à k dimensions à un symbole canal, ou indice i, qui est transmis. L'encodeur partitionne le vecteur d'entrée multidimensionnel en N régions tel que :

$$P=\{R_1,R_{2,\Lambda},R_N\}$$

où

$$R_i=\{x|d(x,y_i)\le d(x,y_j),j\neq i\} \tag{4.5}$$

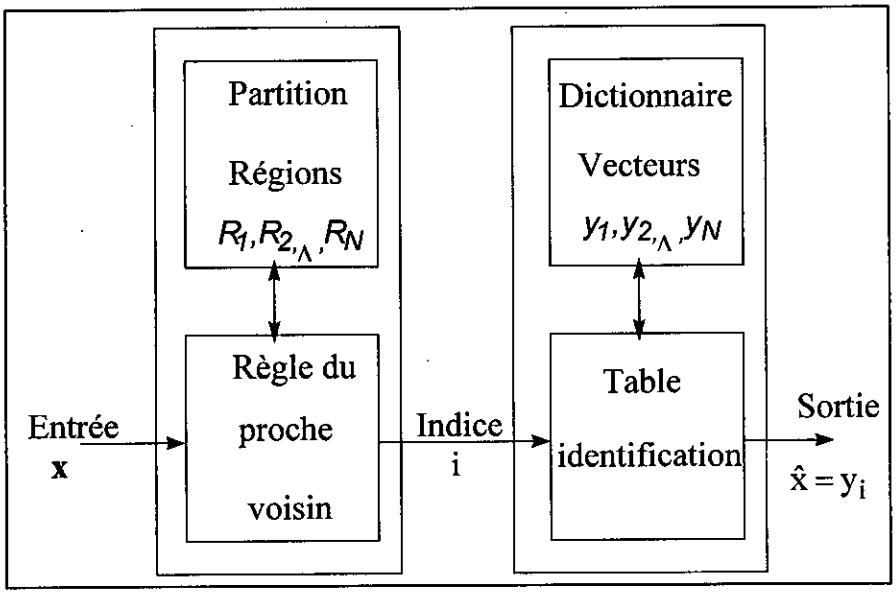


Fig. 4.2 : Modèle d'un quantificateur vectoriel

Le vecteur y_i est le vecteur code associé avec la région R_i. L'indice i est choisi de manière que **x** fasse partie de la cellule de partition R_i dans l'espace de dimension k. Un quantificateur inverse, ou décodeur, reconstruit le symbole " i " en un vecteur code de sortie appropriée $\hat{x}=y_i$ en utilisant une procédure d'identification dans la table.

En traitement de la parole, la quantification vectorielle peut s'effectuer sur des formes d'ondes ou sur des paramètres.

4.3.1 Conditions d'optimalité

La performance d'un quantificateur vectoriel est dépendante de l'espace de partition de l'encodeur et des vecteurs de reproduction, ou vecteurs code, du décodeur. Un

quantificateur vectoriel est optimal quand la distorsion moyenne $E[d(X,\hat{X})]$ est minimisée pour la séquence du vecteur d'entrée **X**. Puisqu'il n'y a pas de méthode directe pour le concept de la QV, les méthodes itératives sont facilement disponibles.

Un quantificateur se décompose en deux applications :un codeur et un décodeur. L e quantificateur optimal est alors celui qui réunis les points suivantsl :

- Un codage optimal (pour un dictionnaire fixé), celui-ci respecte " la règle du plus proche voisin "
- Le décodage optimal (pour une partition R_i), le vecteur représentant y_i doit minimiser la distorsion associée au Voronoï R_i, y_i est donc le centroïde de cette cellule : $y_i = \text{cent}(R_i)$.

4.3.1.1 Condition du proche voisin

Pour un décodeur donné et son ensemble fini de vecteurs code de sortie **C**, les cellules de partition $\{R_i\}$ du codeur optimal satisfont

$$R_i \subset \{x \mid d(x,y_i) \le d(x,y_j), \forall j \ne i\} \tag{4.6}$$

Cela veut dire que les régions de partition sont définies par les vecteurs code $\{y_i\}$ dans **C** :

$$Q(x) = y_i \quad \text{seulement si} \quad d(x,y_i) \le d(x,y_j), \forall j \tag{4.7}$$

4.3.1.2 Condition sur le centroide

Pour une partition donnée de l'encodeur $P = \{R_i \mid i = 1,K,N\}$, les vecteurs code optimaux y_i dans **C** sont les centroides dans chaque cellule de partition R_i :

$$y_i = \text{cent}(R_i) \tag{4.8}$$

$$= \underset{y}{\text{argmin}}\, E[d(x,y) \mid x \in R_i] \tag{4.9}$$

Quand la mesure de distorsion d'erreur quadratique est utilisée pour la conception d'un QV, les centroides sont définis comme les centres de masse des cellules de partition.

4.3.2 Algorithme de Lloyd généralisé

Un algorithme itératif est utilisé pour concevoir un quantificateur vectoriel optimal (dictionnaire). Un ensemble de vecteurs représentatifs de la source est compilé pour une

séquence d'entraînement, et le dictionnaire est optimisé en utilisant une mesure de distorsion appropriée. L'Algorithme de Lloyd Généralisé (ALG), aussi appelé algorithme LBG [Linde, Buzo et Gray, 80], est peut être l'algorithme itératif le plus communément utilisé pour la conception d'un quantificateur vectoriel optimal basé sur des vecteurs d'apprentissages :

- **Etape 1** Commencer avec un dictionnaire initial C_1. Mettre m = 1.
- **Etape 2** Pour un dictionnaire C_m donné, exécuter l'itération de Lloyd pour produire le nouveau dictionnaire C_{m+1}.
- **Etape 3** Calculer la distorsion moyenne pour C_{m+1}. Si la différence de distorsion est inférieure à un certain seuil , stop. Sinon faire m = m+1 et répéter les étapes 2 et 3

La distorsion moyenne d'un quantificateur vectoriel décroît d'une façon monotone ou reste pratiquement inchangée à chaque itération de l'ALG par optimisation alternative de l'encodeur (pour un décodeur donné) et du décodeur (pour un codeur donné). L'étape 2 dans L'ALG est l'extension vectorielle de l'itération de Lloyd qui a été définie pour la conception d'un quantificateur scalaire optimale non uniforme :

- **Etape 2a** Pour un dictionnaire donné $C_m = \{y_i\}$, partition de la séquence d'entraînement en un ensemble de régions R_i en utilisant la condition du plus proche voisin, où $R_i = \{x \in T \mid d(x,y_i) \le d(x,y_j), \text{ pour tout } j \neq i\}$
- **Etape 2b** En utilisant la condition sur le centroide, calculer les centroides pour l'ensemble des régions déjà trouvées dans l'étape **1** pour obtenir le nouveau dictionnaire $C_{m+1} = \{cent(R_i) \mid i = 1, \ldots, N\}$. Si une cellule vide a été générée dans l'étape (a), un éventuel code vecteur code est créé pour cette cellule.

La dimension de la séquence d'entraînement et le nombre d'itérations de l'ALG sont des facteurs critiques durant le processus d'entraînement. Puisque cet algorithme n'est que localement optimal, le choix du dictionnaire de départ est important. Une variante très utilisée de l'algorithme de Lloyd-Max est celui de Linde, Buzo et Gray; il procède hiérarchiquement , et réalise une sorte d'initialisation itérative au cours de la construction.

Linde, Buzo et Gray ont proposé un algorithme appelé algorithme LBG pour construire un dictionnaire tel que chacun des vecteurs de ce dernier soit indexé lors de l'opération de codage; c'est l'index du vecteur le plus proche au sens d'un certain critère (distance euclidienne par exemple) de la séquence à coder qui servira à le représenter. Pour reconstituer la séquence, le décodeur doit être muni du même dictionnaire.

En résumé trois opérations successives sont nécessaires pour quantifier un vecteur donné ***X*** :

- trouver la région de quantification contenant **X**, c'est à dire la région S_i telle que $X \in S_i$; ce qui revient à déterminer le plus proche voisin de **X** dans le dictionnaire
- Transmettre l'indice i de la région contenant **X** qui est l'indice du mot de code Y_i le plus proche de **X**.
- Régénérer à la réception le mot de code Y_i à partir de l'indice i .

En ce qui concerne la transmission, chaque vecteur y_i est codé dans un dictionnaire de digits binaires C_i, de longueur B_i bits. En général, les différents mots de code ont des longueurs différentes.
Le débit de transmission T est alors donné par :

$$T = BF_c \text{ bits/s} \tag{4.10}$$

où $B = \lim_{M\to\infty} \frac{1}{M}\sum_{n=1}^{M} B(n)$, est la longueur moyenne du mot de code.

B(n) est le nombre de bits utilisés pour coder le vecteur *x (n)* au temps n,
F_c est le nombre de mots de code transmis par seconde. Il devrait être aussi intéressant de définir le nombre moyen de bits par paramètre ou par dimension

$$R=B/d \quad \text{bits/dimension} \tag{4.11}$$

Pour un dictionnaire de dimension L, le nombre maximum de bits dont on a besoin pour coder chaque vecteur est :

$$B_{max} = \log_2(L) \tag{4.12}$$

Le but poursuivi dans l'élaboration d'un système de codage est de minimiser la distorsion moyenne D pour un débit donné ou réciproquement minimiser le débit pour une distorsion donnée.

4.4 RESULTATS DE PERFORMANCE DE LA QUANTIFICATION SCALAIRE.

Plusieurs codeurs CELP et VSELP emploient la quantification scalaire pour encoder chaque paramètre spectral de la parole indépendamment de chaque autre. Le choix propre de la représentation paramétrique des coefficients du filtre LP est influencée par sa performance de quantification. Les coefficients de réflexion sont souvent utilisés pour la quantification parce qu'ils sont moins sensibles aux erreurs de quantification que les coefficients prédicteurs [D. O'Shaugnessy, 87] ; [A. Gerson, 90].

Ces dernières années, les LSF sont devenus très populaire pour l'utilisation dans la quantification spectrale [G. S. Kang, 85] ; [G. S. Kang, 87] ; [R.P. Ramachandran,92].

Le **tableau 4.1** présente l'allocation de bit pour la QS différentielle adaptative pour un débit de 54 bits/trame.

Bits par trame	**Allocation de bit pour les LSF**															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
54	**4**	**4**	**4**	**4**	**4**	**4**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**

Tableau 4.1 : Allocation de bit pour la Quantification Scalaire des LSF.

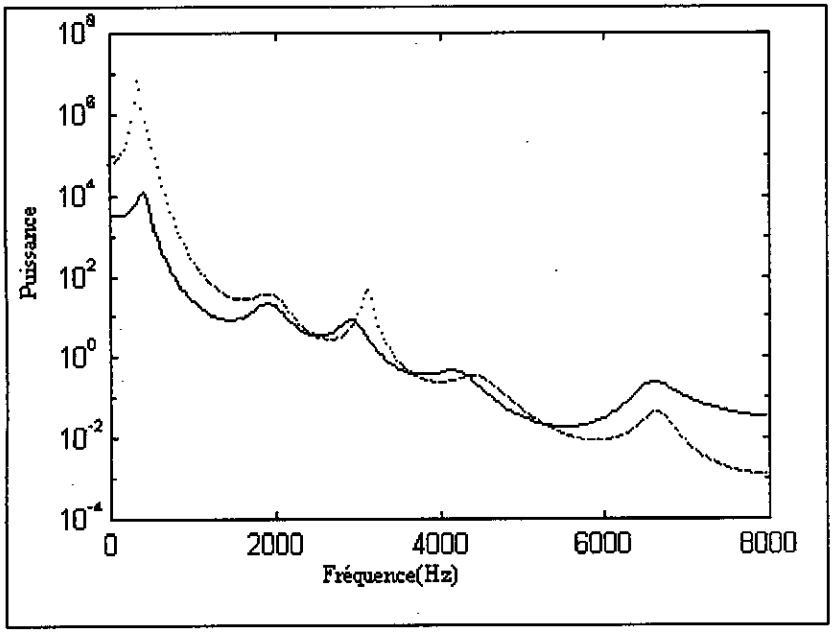


Fig.4.3 : Spectre LP:—— LSF non quantifié .....LSF quantifié.

4.5 CONCLUSION :

Les LSF offrent plusieurs avantages pour leur utilisation comme des paramètres de codage spectrale. Ils approximent les positions des formants, et exposent les distributions distincts localisées. En outre, les LSF d'odre supérieures sont moins significatifs que les LSF d'ordre inférieure ; d'où, ces LSF d'ordre supérieures peuvent être grossièrement quantifiés.

Une des propriété des LSF qui garantie la stabilité du filtre de synthèse d'ordre p est la propriété d'ordonnancement des LSF.

La Quantification Vectorielle a été utilisée lors de la sélection des paramètres du dictionnaire (Gain et code vecteur optimaux).

Chapitre 5
Tests et Résultats

5.1 INTRODUCTION

Le travail a consisté à concevoir et réaliser un codeur/ décodeur pour les signaux de parole à large bande.

Des dictionnaires d'excitation très large sont nécessaires dans le but de maintenir une haute qualité de la parole. Ceci induit une augmentation de la complexité qui est difficile à être maniée par les algorithmes CELP existants. L'une des approches qui est traditionnellement suivie est de diviser la largeur de bande en deux bandes (fréquences basses et fréquences hautes) et coder indépendamment chaque bande.

Du à l'efficacité de l'algorithme CELP en bande étroite nous avons utilisé la *même* approche , comme déjà mentionné auparavant, pour l'encodage de la bande totale.
Il faut pour cela préaccentuer le signal et utiliser les bons dictionnaires d'excitation (algébrique par exemple).

La préaccentuation du signal d'entrée réduit le rang dynamique spectral et assure que les fréquences hautes sont efficacement codées.

5.2 ANALYSE PAR PREDICTION LINEAIRE

L'analyse par prédiction linéaire est exécutée pour obtenir les paramètres du filtre de synthèse (ou le prédicteur court terme). Ce filtre décrit l'enveloppe spectrale court terme du signal de parole et il est actualisé chaque 15 ms. Le filtre de synthèse est donné par

$$H(z) = \frac{1}{1 - \sum_{k=1}^{m} a_k z^{-k}} \tag{5.1}$$

où a_k, k = 1, . . . , m, sont les coefficients prédicteurs, et m est l'ordre du prédicteur.

L'analyse LP est exécutée en utilisant la méthode d'autocorrélation. Dans cette méthode, les premiers (m+1) autocorrélations du signal de parole pondéré par une fenêtre de Hamming sont calculés, et les paramètres LP sont déterminés en résolvant le système de Toeplitz de m équations grâce à l'algorithme de Levinson – Durbin.

En général, la méthode d'autocorrélation assure la stabilité du filtre de synthèse (c-à-d les pôles de H(z) sont à l'intérieur du cercle unité). Cependant, comme la fréquence d'échantillonnage est de 16000 échantillons/s dans le cas du signal de parole à large bande, plusieurs problèmes de stabilité peuvent être rencontrés. Cette haute fréquence d'échantillonnage et l'augmentation de l'ordre du filtre qui sont nécessaires pour la parole à large bande nous donnent des gains de prédiction très grands (ou très faibles). Il existe plusieurs procédures pour remédier à ce problème.

- La première est de préaccentuer le signal de parole d'entrée. Ceci est accomplit par filtrage du signal de parole d'entrée par le filtre à un seul zéro $1-\mu z^{-1}$. La préaccentuation a deux avantages. Elle réduit le rang dynamique du signal d'entrée et accentue ces fréquences hautes du signal de parole d'entrée. Ainsi les fréquences hautes peuvent être prises en compte par le modèle d'ordre fixé de l'algorithme CELP algébrique.
- La seconde procédure utilisée pour améliorer l'analyse LP est de faire un " lag windowing " sur les coefficients de prédiction de la parole pour résoudre le système d'équations. Le " lag windowing " a pour effet d'élargir les largeurs de bande des formants de la parole, et ainsi éviter la sous estimation de la largeur de bande qui est manifestée par des pics extrêmement aigus dans l'enveloppe spectrale. Une fenêtre binomiale est utilisée où les coefficients de prédiction sont modifiés par

$$a'(i) = a(i)\,(\exp(-\pi f_0/f_s))^i, \qquad i = 1, \ldots, m, \tag{5.2}$$

où f_0 est l'expansion de la largeur de bande et f_s est la fréquence d'échantillonnage.

Le premier coefficient d'autocorrélation r(0) (les éléments de la diagonale dans la matrice de Toeplitz) est augmentée par 0.003%, qui est équivalente à ajouter un bruit de fond qui est de 45 dB en dessous de la puissance du signal. Ceci élimine le mal conditionnement occasionnel de la matrice des autocorrélations. Concernant l'ordre du filtre, nous avons trouvé qu'un ordre de 16 était suffisant pour modéliser l'enveloppe spectrale court terme.

5.3 ANALYSE PITCH

L'analyse pitch est exécutée chaque 3.5 ms, et consiste en la détermination du retard "pitch" et le gain. Le filtre de synthèse pitch modélise la structure fine du spectre de parole, et est donné par

$$P(z) = \frac{1}{1-\beta z^{-D}} \tag{5.3}$$

où β est le gain pitch et D est le retard pitch. Pour la parole humaine, la période du fondamentale peut s'étendre de 2.5 à 20 ms. Les locuteurs féminins ont des périodes basses (fréquences pitch hautes). Dans notre cas, les retards sont pris dans l'intervalle 40-295 échantillons et le délai est ainsi codé sur 8 bits. Les paramètres pitch sont déterminés par la méthode d'analyse par synthèse et en boucle fermée.

5.4 DICTIONNAIRE D'EXCITATION

Le signal d'excitation est sélectionné à partir d'un large dictionnaire de séquences d'innovation en minimisant l'erreur perceptuelle pondérée entre le signal original et le signal synthétique. La trame d'excitation est de 3.75 ms (60 échantillons). Les trames larges demandent des dimensions énormes du dictionnaire. L'algorithme CELP utilise un dictionnaire algébrique efficace qui peut atteindre 2^{20} entrées. Le dictionnaire n'a pas besoin d'être stocké, et le code vecteur est très efficacement trouvé en utilisant une stratégie minutieuse de recherche.

Un code vecteur d'excitation constitue un petit nombre d'impulsions non nulles, avec des amplitudes fixées (-1 ou +1) et d'un ensemble des positions prédéfinis. La recherche dans le dictionnaire est pour effet de trouver la position de chaque impulsion qui minimise le critère d'erreur. Ainsi le vecteur code d'excitation sélectionné aurait, dans une certaine mesure, les positions optimums des impulsions, avec la contrainte d'amplitudes fixés et d'un ensemble limité de positions pour chaque impulsion. Cette structure a un avantage sur d'autres structures CELP dans le sens que le vecteur sélectionné n'a pas de composantes non nécessaires. Le second avantage est que la structure est robuste aux erreurs dans le canal car une erreur affecterait seulement une position d'impulsion et non le vecteur d'entrée.

Une autre caractéristique importante du CELP algébrique est que le dictionnaire d'excitation fixé peut être rendu variable (modélisation dynamique dans le domaine fréquentiel). Ceci est accomplit en passant les vecteurs d'excitation à travers un filtre de la forme

$$F(z) = \frac{A(z)}{A(z/\gamma)} \tag{5.4}$$

où A(z) est le filtre LP inverse, et γ et sont des facteurs de pondérations tels que :

$0 < \gamma < 1$

5.5 CELP LARGE BANDE BONNE QUALITE (~13 kb/s)

Dans cette section nous décrivons l'allocation de bits du codeur CELP large bande à 13 kb/s qui se situe dans la gamme " bonne qualité ". L'allocation de bits est donnée dans le **tableau 5.1**

Paramètres	Intervalle d'actualisation(ms)	Nombre de bits
Filtre LP	15	54
Période pitch	3.75	8
Gain pitch	3.75	4
Indice dictionnaire	3.75	15
Gain dictionnaire	3.75	(5+1)
Total		12.4 kb/s

Tableau.5.1 : Allocation de bit pour le codage CELP algébrique à 12.4 kb/s

La trame de parole est de 15 ms (240 échantillons). Les paramètres LP sont calculés en utilisant une fenêtre de Hamming de 15 ms centré à la fin de la trame. Un filtre d'ordre 16 est utilisé et les coefficients sont quantifiés en utilisant la représentation LSF.

Les paramètres LSF sont quantifiés avec 54 bits (4,4,4,4,4,4,3,3,3,3,3,3,3,3,3,3). La quantification est exécutée en utilisant une approche adaptative forward où la propriété d'ordonnancement des LSF est efficacement utilisée pour réduire le rang de l'intervalle de quantification.

Dans la conception du quantificateur, les statistiques du signal de parole sont tirées en utilisant une base de donnée d'une minute de phrases prononcées par des locuteurs masculin et féminin en deux langues Française et Anglaise.

La trame de parole de 15 ms est divisée en 4 sous trames de 60 échantillons (3.75 ms). Les paramètres pitch et excitation sont actualisés à chaque sous trame. Le gain du pitch est limité entre 0 et 1.4 et quantifié avec 4 bits.

Pour la conception du dictionnaire d'excitation, deux types d'excitation ont été utilisés et combinés :

- Excitation formée d'impulsions ternaires (-1, 0, 1)
- Excitation formée d'impulsions régulièrement espacées (Regular pulse).

Le dictionnaire est modélisé comme une série de m impulsions d'amplitudes $\beta_0,\ldots,\beta_{m-1}$ à des positions $\eta_0,\eta_1,\ldots,\eta_{m-1}$.

- $\eta_{m-1} = \eta_0 + 8(m-1) \quad 0 < \eta_0 < 13$
- $\eta_{m-1} = \eta_0 + 7(m-1) \quad 0 < \eta_0 < 12$
- $\eta_{m-1} = \eta_0 + 9(m-1) \quad 0 < \eta_0 < 7$

Les vecteurs d'excitation de taille 60 échantillons contiennent 7 impulsions codées en (–1), 0, et (+1).

Le premier sous bloc contient 1093×12 = 13116 combinaisons.

Le deuxième sous bloc contient 1093×11 = 12023 combinaisons.

Le troisième sous bloc contient 1093×6 = 6558 combinaisons.

L'amplitude du gain d'excitation est quantifiée logarithmiquement avec 6 bits dont 1 bit est utilisé pour le signe.

Les tests ont été faits sur un extrait de phrases à large bande phonétiquement équilibrée échantillonnées à 16 kHz (trois locuteurs masculins et trois locuteurs féminins).

5.5.1 Phrases tests :

Locuteurs féminins

Phrase 1 : " I must have reread that article three times before I realized what was bothering me."

Phrase 2 : " Quand il s'est réveillé il était trop tard, huit satellites ont été mobilisés."

Phrase 3 : " Là bas il y a de mauvaises vagues très hautes. C'est la question que tout le monde se Pose. "

Locuteurs masculins

Phrase 4 : " La voiture s'est arrêté au feu rouge, la vaisselle propre est mise sur l'évier."

Phrase 5 : " The other memorable event in that conference was the worst presentation I ever heard."

Phrase 6 : " Je ne peux atteindre les bocaux de confiture, dans cette crémerie on vend du fromage fort."

Le **tableau 5.2** donne l'évaluation du rapport signal à bruit (RSB) et RSB segmental sur des phrases à large bande phonétiquement équilibrée échantillonnées à 16 kHz.

Phrase	RSB(dB)	RSBseg (dB)
1	**15.64**	**11.93**
2	**13.97**	**10.75**
3	**15.36**	**11.05**
4	**11.34**	**7.72**
5	**14.65**	**10.53**
6	**11.19**	**9.23**

Tableau. 5.2 : Rapport Signal à Bruit et RSB segmental

Les tests d'écoute ont été faits par des sujets non prévenus et inéxperimentés.

5.5.2 Formes d'ondes

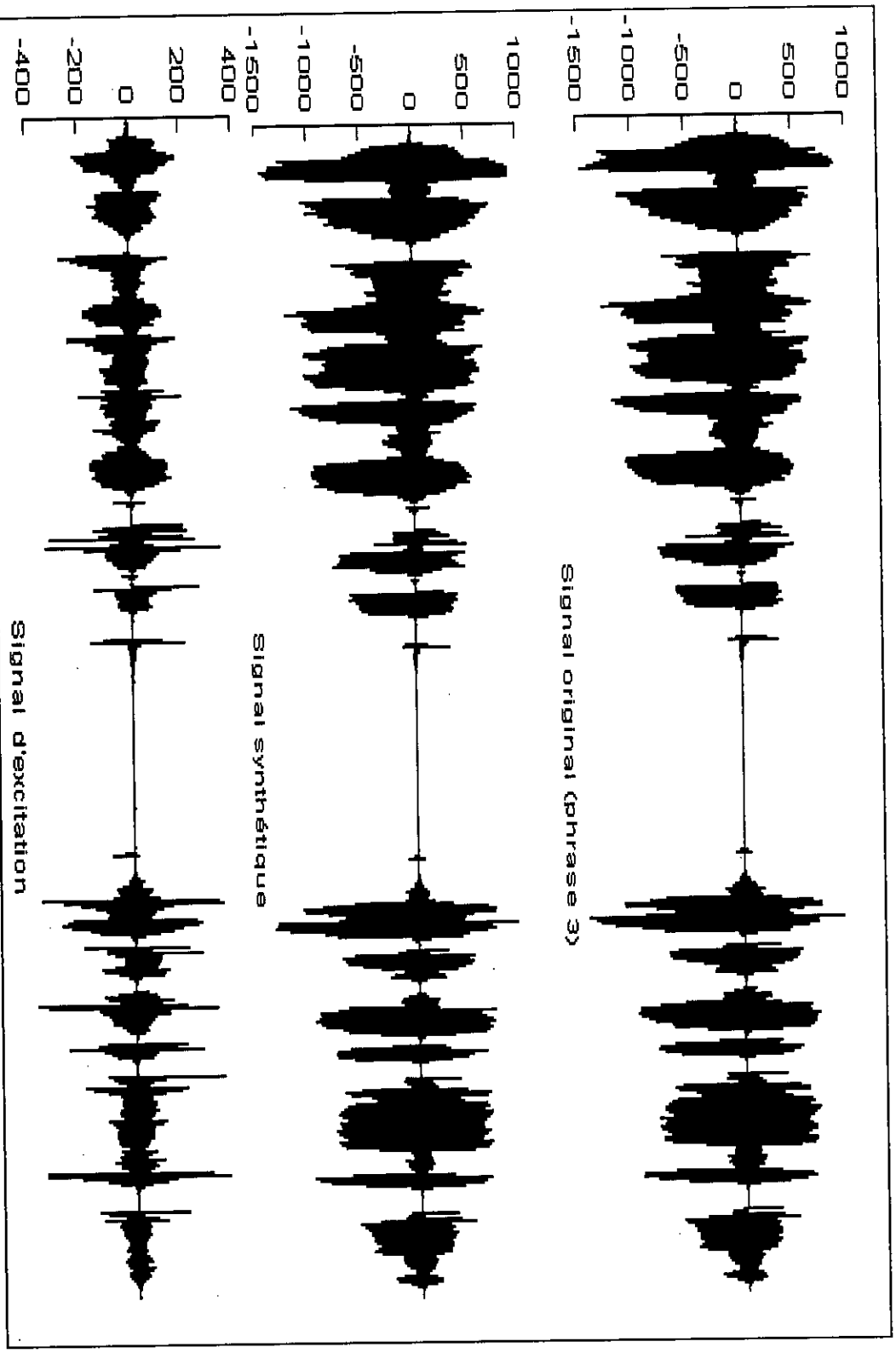


Figure. 5.1. Exemple de phrase large bande codée et les signaux obtenus

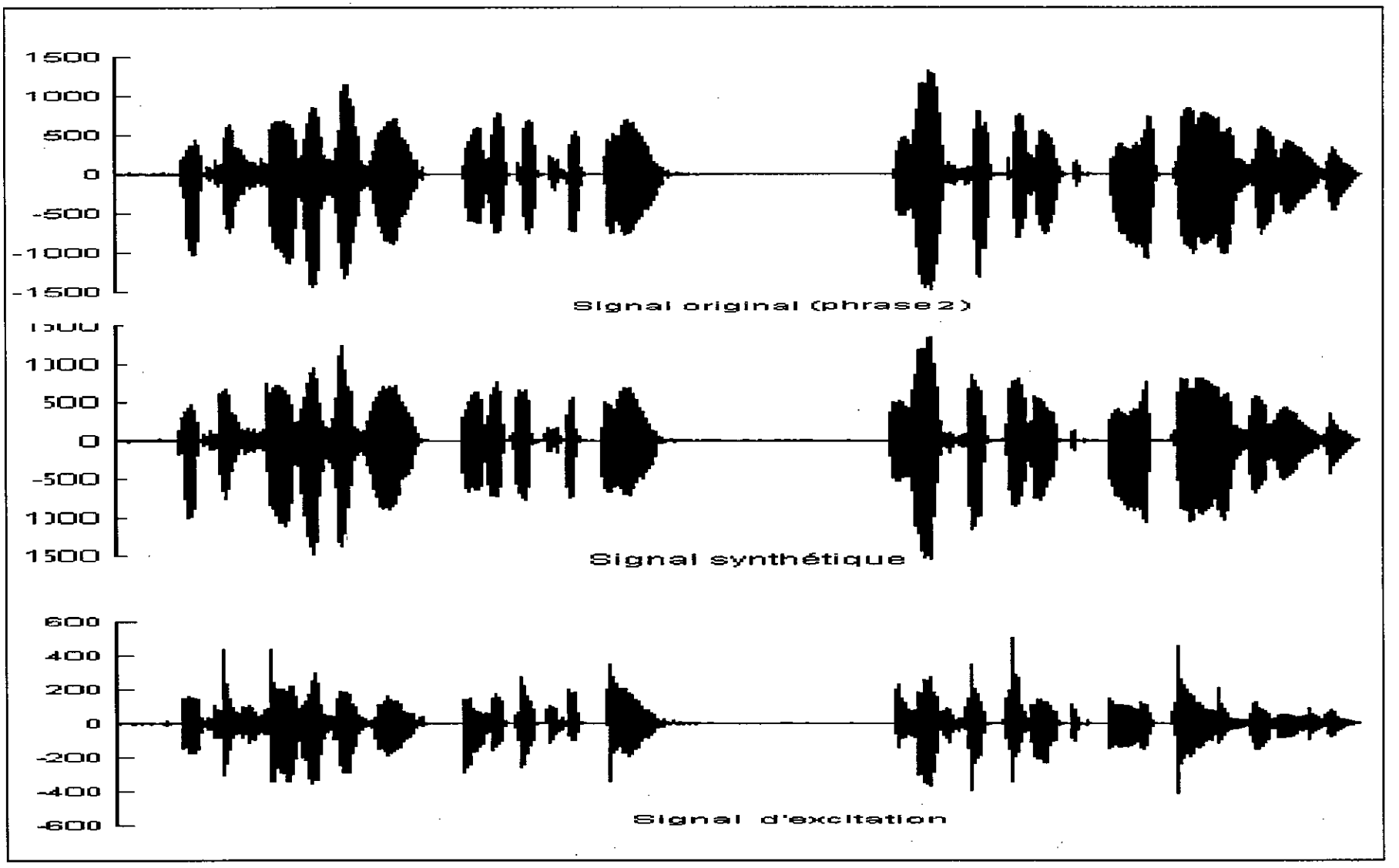


Figure. 5.2. Exemple de phrase large bande codée et les signaux obtenus

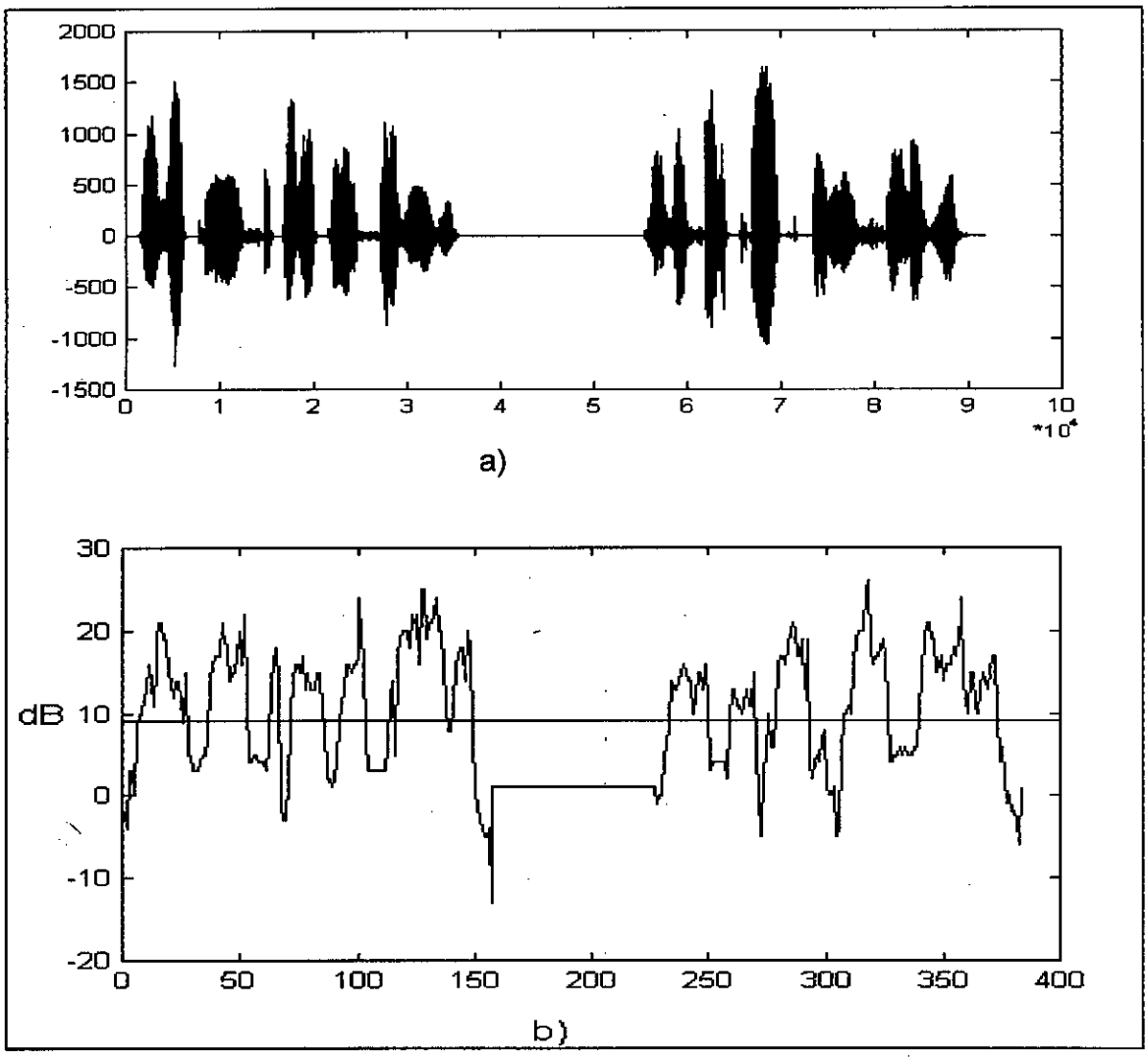


Figure. 5.5. a) Signal de la parole (phrase 4) b) Evolution du RSB obtenu en fonction du temps (évalué sur des trames de 240 échantillons)

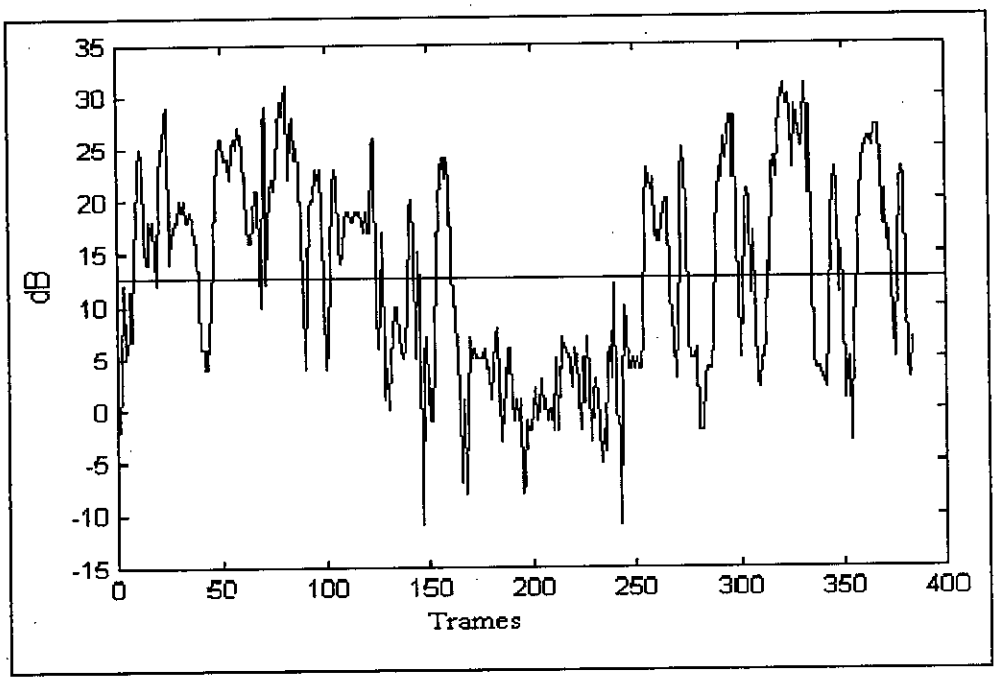


Fig. 5.3 : Evolution du RSB en fonction du temps

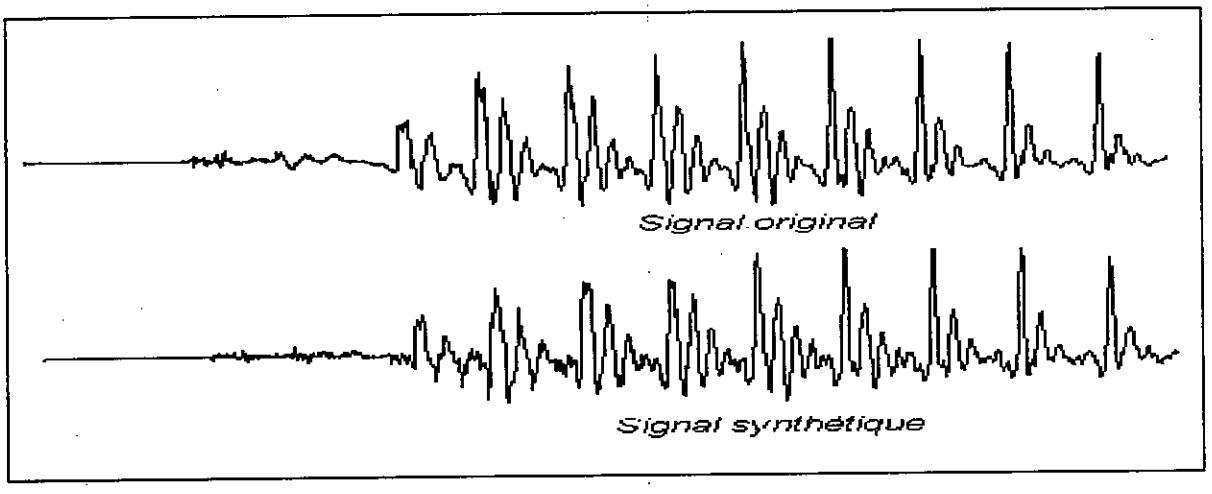


Fig. 5.4 : Comparaison de forme d'onde d'un segment de parole (phrase 1)

Conclusion Générale

CONCLUSION GENERALE

Un Codeur/Décodeur de parole CELP large bande à bonne qualité a été développé pour un débit inférieur à 16 kbits/s.

Le problème du codage à débit moyen consiste à calculer, pour des segments de parole, les coefficients d'un filtre de synthèse puis à rechercher le signal d'excitation qui rend minimale l'énergie de la différence entre le signal à coder et le signal de synthèse.

L'introduction dans l'expression de l'erreur de codage d'une pondération tenant compte du processus de perception du bruit de quantification améliore considérablement la qualité de la parole codée.

La complexité de recherche du meilleur code vecteur a été accéléré grâce à la transformation des opérations de filtrage en des produits matriciels (soustraction du " zéro input response"). On n'omettra pas de signaler la mise en œuvre du filtre de perception pour le masquage du bruit. Ce filtre améliore considérablement l'effet subjectif.

Afin d'accélérer davantage le temps de calcul mis pour la recherche du meileur vecteur d'excitation, nous avons implémenté la technique dite " Backward Filtering " qui permet de réduire la complexité de recherche du vecteur d'excitation d'un facteur d'ordre p (ordre du modèle Autoregressif) dans le cas des dictionnaires pauvres en échantillons (Sparse codebook).

La recherche des paramètres du pitch $-\beta$ (gain) et D (délai) – s'est faite en boucle fermée. Il est connu que la recherche en boucle fermée offre de meilleure performance que celle en boucle ouverte. Le codage du délai D a nécessité huit (8) bits.

Pour la conception du dictionnaire d'excitation, deux types d'excitations ont été utilisés et combinés :

- Excitation formée d'impulsions ternaires (-1, 0, 1) ;
- Excitation formée d'impulsions régulièrement espacées (Regular pulse)

La quantification des paramètres du modèle Autorégressif a fait l'objet d'une attention particulière. Les paramètres " Line Spectral Frequencies " ont été finalement choisis pour être quantifiés. Ces paramètres sont moins sensibles au problème de la quantification que

les autres paramètres (coefficients de corrélation partielle coefficients log spectrales, etc...). Une quantification Scalaire Adaptative a été utilisée pour coder les paramètres LSF.

Les mesures objectives et subjectives montrent que la qualité de la parole synthétisée est de bonne qualité.

Un codeur CELP à large bande et à un débit inférieur à 1bit/échantillon (~13 kbits/s) a été conçu et réalisé. Il peut être utilisé comme programme de compression de signaux de parole dans les supports de stockage.

Nous souhaitons que le travail ainsi réalisé soit suivi par une implémentation de l'algorithme de compression CELP à large bande sur une carte DSP (exemple le TMS320C30) pour les applications en temps réel.

BIBLIOGRAPHIE

1) Adoul. J. P, F. Didelot, P. Mabilleau and S. Morisette
"Generalization of the multipulse coding for low bit rate coding purposes: The generalized decimation Proc. of Intern. Conf. on Acoust. Speech and Signal Processing ICASSP 85 pp. 256-259.

2) Adoul. J. P, Mabilleau. P, Delprat. M,Morisette. S
"Fast CELP Coding based on Algebraic Codes." IEEE-ICASSP, pp.1957-1960, 1987.

3) Adoul. J. P. , Lamblin. C.
"A Comparison of some Algebraic Structures for CELP Coding of Speech."
IEEE-1CASSP, pp 1953-1956.1987

4) Atal. B.S. and Schroeder. M.R.
"Adaptive predictive coding of speech signals." Bell System Technical Journal, Vol. 49, October 1970

5) Atal. B.S. and Schroeder. M.R.
"Predictive coding of speech signals and subjective error criteria."
IEEE Transactions on Acoustics Speech and Signal Processing. Vol. ASSP 27. N°3 June 1979.

6) Atal. B. S.
"Predictive Çoding at Low Bit Rates."
IEEE Transaction on Communication,vol. com-30, NO. 4, pp. 600-614, .April 1982.

7) Atal. B. S.
"High quality speech at low bit rates: Multipulse and stochastically excited linear predictive coders"
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing ICASSP 86 Tokyo pp. 1681-1684.

8) Atal. B. S. , Cuperman. V, Gersho. A.
"Advances jn Speech Codig" Kluwer Academic Publishers 1991.

9) Bei. C. D. and Gray0. R.M.
"Simulation of vector treillis encoding systems."
I.E.E.E Trans. on Communications Vol. COM 34 pp 214-218 March 86.

10) Copperi. M. and Sereno. D.
"Celp coding for high quality speech at 8 kbits/s." ICASSP 86 Tokyo pp. 1685-1688.

11) Chen. J. H. and Gersho. A.
"Vector adaptive predictive coding of speech at 9.6 kbits/s." ICASSP 86 Tokyo pp. 1693-1696

12) Chen. J. H. and Gersho. A.
"Real time vector APC speech coding at 4800 bps with adaptive postfiltering."
ICASSP 87 pp. 2185-2188

13) Copperi. M.
"Rule based speech analysis and application to CELP coding." ICASSP 88 pp. 143-146

14) Chmielewski. A., Domaszewicz. J. and Mitek. J.
"Real time implementation of forward gain-adaptive vector quantizer"
Eurocon 88. 8th European Conference on Electrotechnics Conference Proceedings on Area Communications. Stockholm Sweden June 13-17 1988 pp. 40-43.

15) Calliope.
"La Parole et son Traitement Automatique."
Collection Technique et Sciennfique des télécornmunications.NL~SSON 1989

16) Chen. J. H., Jayant. N. and Cox. R.V.
"Improving the performance of the 16 kbits/s LD-CELP speech coder." ICASSP 92 pp. I-69 I-72

17) Chih-Chung. K., Fu-Rong. J. and Hsiao-Chuan. W.
"Low bit rate quantization of LSP parameters using two-dimensional differential coding."
ICASSP 92 pp. I-97 I-100

18) Cheng. Y. M., D.O. Shaughnessy. D. O. and Mermelstein. P.
"Statistical recovery of wideband speech from narrowband speech."
IEEE Transactions on Speech and Audio Processing October 1994. Vol.2 n°4 pp 544-548.

19) Costantinos. Papacostantinou.
"Improved Pitch Modelling for Low Bit-Rate Speech Coders."
Thèse Master, Université de McGill, Montreal, Canada, Aout 1997.

20) Delsarte. P, Genin. Y. V.
"The Split Levinson Algorithm." IEEE on ICASSSP, vol ASP-34, *n°3.* pp 470478, 1985.

21) Davidson. G. and Gersho. A.
"Complexity reduction methods for vector excitation coding." ICASSP 86 Tokyo pp. 3055-3058.

22) Davidson. G, Yong. M. and Gersho. A.
"Real time vector excitation coding of speech at 4800 bits/s." ICASSP 87 pp. 2189-2192.

23) Davidson. G. and Gersho. A.
"Multiple stage vector excitation coding of speech waveforms." ICASSP 88 pp. 163-166.

24) De Brito. G.S.
"Low bit rate speech for the GSM system." EUROCON 88. 8th European Conference.

25) Delprat. M, Lever. M. and Gruet. C.
"Efficient excitation model and fast selection in CELP coding of speech." Eurospeech 89.

26) Di Francesco. R.
"Codage Algébrique de la Parole: Prédiction Linéaire a Excitation par Code Ternaire."
Annales. Télécom. 47, n5-6, 1992.

27) Dymarski. P, Moreau. N.
"Algorithms for the CELP coder with Ternary Excitation." Eurospeech , pp 241-244, 1993.

28) Deller. J. R, Proaskis. J. G. and Hansen. J. H. L.
"Discrete-Time Processing of Speech Signals" New York: MacMillan, 1993.

29) Elroy. C. M, Murray. B. and Fagan. A. D.
"Wideband speech coding in 7.2 kbits/s." ICASSP 93 pp. II-620 II-623.

30) Erzin. E. and Cetin. A.E.
"Interframe differential coding of line spectrum frequencies."
IEEE Transactions on Speech and Audio Processing April 1994. Vol.2 n°2 pp 350-352

31) Fant. G.
"Acoustic theory of speech production." Mouton, The hage, 1960.

32) Feng. G. et Lacoume. J.L.
"Amélioration de l'estimation du spectre de parole par suppression d'impulsions dans le résidu."
Revue Traitement du Signal Volume 3 N°6 1986

33) Foster. J, Gray. R.M. and Dunham. M.O.
"Finite state vector quantization for waveform coding."
IEEE Trans. on Inf. Theory Vol. IT 31 pp 348-359 May 1985

34) Gray. A. H. and Markel. J.D.
"Distance measures for speech processing."
IEEE Acoustic., Speech, Signal Proc., vol. ASSP-24, pp. 380-391, October 1976.

35) Gray. A. H. and Markel. J.D.
"Quantization and bit allocation in speech processing."
IEEE Transactions on Acoustics Speech and Signal Processing .Vol. ASSP 24 N°6 December 1976

36) Galand. C, Esteban. D, Mauduit. D. and Menez. J.
"Codage prédictif du signal de parole à 4800 bps".
7eme Colloque sur le Traitement du Signal et ses Applications, NICE 28 Mai -2 Juin 1979

37) Gray. R. M, Buzzo. A, Gray. A. H. and Matsuyama. Y.
"Distortion measures for speech prooessing."
IEEE Acoustic., Speech, Signal Proc., vol. ASSP-28, pp. 367-376, August 1980.

38) Gray. R. M.
"Vector quantization." IEEE ASSP. Magazine Vol. 1 n° 2 pp 4-29 April 1984.

39) Gueguen. C.
"Analyse de la parole par les méthodes de modélisation paramétrique."
Annales des Télécommunications 40 N°5-6 pp. 253-269 1985.

40) Gerso. I. A, Jasiuk. M. A.
"Vector Sum Excited Linear Prediction (VSELP) Speech Coding at 8kbps."
IEEE-ICASSP, pp. 461464, 1990

41) Gray. R. M.
"Source Coding Theory." Boston: Kluwer Academic Press, 1990.

42) Grass. J.
"Quantization of predictor coefficients in speech coding."
Master's thesis, McGill University, Montreal, Canada, September 1990

43) Grass. J. and Kabal. P.
"Methods of improving vector-scalar quantization of LPC coefficients."
in Proc. Int. Conf Acoust., Speech, Signal Proc., (Toronto), pp. 657-660, May 1991.

44) Gerso. I. A, Jasiuk. M. A.
"Techniques for improving the performance of CELP type speech coders."
IEEE Journal on Selected Areas in Communications Vol. 10 N=5 June 92

45) Galand. C, Menez. J. and Rooso. M.
"Adaptive code excited predictive coding."
IEEE Trans. on Signal Processing Vol . 40 N°6 pp. 1317-1326 June 92.

46) Gersho. A. and Gray. R. M.
"Vector Quantization and Signal Compression." Boston. Kluwer Academic Press, 1992.

47) Hernandez-Gomez. L.A, Casajus Quiros. F.J. and Garcia Gomez. R.
"High quality vector adaptive transform coding at 4.8 kb/s."
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 88 pp. 167-170

48) Hedelin. P.
"A multi-stage perspective on CELP speech coding."
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 92 pp. I-57 I-60

49) Hussain. Y and Farvardin. N.
''Finite state vector quantization over noisy channels and its application to LSP parameters.''
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 92 pp. II-133 II-136

50) Hagen. R.
''Spectral quantization of cepstral coefficients.''
In Proc. Int. Conf. Acoust, Speech, Signal Proc., (Adelaide),pp. 1509-1512, April 1994

51) Itakura. F.
''Line spectrum representation of linear prediction coefficients of speech signals.''
Journal Acoustical Society America, vol. 57, p. 535,1975. (abstract).

52) Jayant. B.S. and Rabiner. L.R.
''The application of dither to the quantization of speech signals.''
Bell System Technical Journal July-August 1972

53) Jayant. N.S.
''Digital coding of speech waveforms: PCM, DPCM and DM quantizers.''
Proceedings of the IEEE. Vol. 62 pp. 611-632 May 1974.

54) Jayant. N. and Noll. P.
''Digital Coding of Wavefrms: Principles and Applications to Speech and Video''
Englewood , New Jersey: Prentice-Hall, 1984.

55) Jain. V. K. and Atal. B. S.
''Robust LPC analysis of speech by extended correlation matching.''
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 85 pp. 473-476

56) Jayant. N.
''Signal compression: Technology targets and research directions.''
IEEE Journal on Selected Areas in Communications Vol. 10 N°5 June 92

57) Jacek Stachurski
''A Pitch Pulse Evolution Model for Linear Predictive Coding of Speech.''
Thèse de Doctorat, Université de McGill, Montreal, Canada , Mai 1997

58) Johnson. M. and Taniguchi. T.
''ON-line and off-line computational reduction techniques using Backward filtering inCELP speech coders.'' IEEE Trans. on Signal Processing Vol . 40 N°8 pp 2090-2093 Aug.1992

59) Kang. G. S. and Fransen. L. J.
''Application of line spectrum pairs to low-bit-rate speech encoders.''
in Proc. Int. Conferen Acoust., Speech, Signal Froc., (Tampa), pp. 244-247, April 1985.

60) Kroon. P. , Deprettere. ED F. and Sluyter. ROB J.
''Regular Pulse Excitation. A novel approach to effective and efficient Multipulse coding of speech.''
IEEE Transactions on Acoustics Speech and Signal Processing. Vol. ASSP 34 N°5 pp.1054-1063 Oct.1986

61) Kabal. P. and Ramachandran. R. P.
''The computation of line spectral frequencies using Chebyshev polynomials.''
IEEE. Transactions on Acoustics Speech and Signal Processing.
Vol. ASSP 34 N°6 pp.1419-1426 Dec. 1986

62) Kroon. P. and Atal. B. S.
''Quantization procedures for the excitation in CELP coders.''
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 87 pp. 1649-1652

63) Kroon. P. , Deprettere. ED F.
''A class of Analysis by Synthesis predictive coders for high quality speech coding at rates between 4.8 and 16 kbits/s''. IEEE Journal on Selected Areas in Communications Vol. 6 N°2 Feb. 88

64) Kondoz. A. M, Lee. K.Y. and Evans. B.G.
"Speech coding at 9.6 Kb/s and below using vector quantized transform coder."
Eurocon 88. 8th European Conference on Electrotechnics Conference Proceedings on Area Communications. Stockholm Sweden June 13-17 1988 pp. 36-39

65) Kondoz. A. M. and Evans. B.G.
"A robust vector quantized sub band coder for good quality speech coding at 9.6 kb/s."
Eurocon 88. 8th European Conference on Electrotechnics Conference Proceedings on Area Communications Stockholm Sweden June 13-17 1988 pp. 44-47

66) Krasinski. and Ketchum. R.H.
"An efficient stochastically excited linear predictive coding algorithm for high quality low bit rate transmission of speech." Speech Communications Vol 7 N°3 pp. 305-316 October 1988

67) Kabal. P , Moncet. J.L. and Chu. C.C.
"Synthesis filter optimization and coding: Applications to CELP."
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 88 pp. 147-150

68) Kroon. P. and Atal. B. S.
"Strategies for improving the performance of CELP coders at low bit rates."
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 88 pp. 151-154

69) Kleijn. W.B., Krasinski. D.J. and Ketchum.R.H.
"Improved speech quality and efficient vector quantization in SELP."
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 88 pp. 155-158

70) Kondoz. A. M. and B.G. Evans
"CELP Base-Band Coder for high quality speech coding at 9.6 to 2.4 kbps."
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 88 pp. 159-162

71) Kabal. P. and Ramaachandran. R.P.
"Joint optimization of linear predictors in speech coders."
IEEE Transactions on Acoustics Speech and Signal Processing. Vol. ASSP 37 N°5 pp. 642-650 May 1989

72) Kroon. P., Atal. B. S.
"Pitch Predictors with High Temporal Resolution."
Proc ICASSP, pp 661-664, 1990.

73) Kleijn. W. B. and Hagen. J.
"Transformation and decomposition of the speech signal for coding."
IEEE Signal Processing Letters Vol .1 n° 9 pp 136-138 Sept. 1994

74) Kleijn. W.B.
"On the periodicity of speech coded with linear prediction based analysis by synthesis coders."
IEEE Transactions on Speech and Audio Processing October 1994 Vol.2 n°4 pp 539-54

75) Kroon. P. and Kleijn. W. B.
"Linear predictive analysis by synthesis coding, in Modern Method of Speech Processing."
(R. P. Ramachandran and R. J. Mammone, eds.). Kluwer Academic Press, 1995.

76) Leroux. J. and Gueguen. C.
"A fixed point computation of partial correlation coefficients."
IEEE Transactions on Acoustics Speech and Signal Processing June 77

77) Lim. J.S. and Oppenheim. A.V.
"Effect of quantization noise in PCM speech coding."
IEEE Transactions on Acoustics Speech and Signal Processing. Vol. ASSP 28 N°1 February 1980.

78) Lloyd. S.P.
"Least squares quantization in PCM."
I.E.E.E Trans. on Information Theory Vol. IT 28 N°2 March 1982.

79) Leguyader. A. et Gilloire. A.
"Codage différentiel de la parole: algorithmes de prédiction adaptative et performances."
Annales des Télécommunications 38 N°9-10 1983

80) Lever. M. and Delprat. M.
"RPCELP: A High quality and low complexity scheme for narrowband coding of speech."
Eurocon 88. 8th European Conference on Electrotechnics Conference Proceedings on Area Communications. Stockholm Sweden June 13-17 1988 pp. 24-27

81) Laroja. R., Phmdo. N. and N. Farvardin.
"Robust and efficient quantization of speech LSP parameters using structured vector quantizers,"
in Proc Int. Conf. Acoust., Speech, Signal Proc., (Toronto), pp. 641-644, May 1991.

82) Laflamme. C. , Adoul. J. P., Salami. R. , Morisette. S. , Mabilleau. P.
"16 kbps Wideband Speech Coding Technique based on Algebraic CELP."
IEEE, CASSP, pp. 13-lb, 1991.

83) Liu. Y.J.
"On reducing the bit rate of a Celp-based speech coder."
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 92 pp. I-49 I-52

84) Leblanc. W. P., Bhattacharva. B. , Mahmoud. S. A. and Cuperman. V.
"Efficient search and design procedures for robust multi-stage VQ of LPC parameters for 4 kb/s speech coding." IEEE Tro.ns. Speech and Audio Proc., vol. 1, pp. 373-385, October 1993.

85) Loo. J. H. Y., Chan. W.-Y. and Kabal. P.
"Classified non linear predictive vector quantization of speech spectral parameters."
in Proc. Int. Conf. Aco~t., Speech, Signal Proc., (Atlanta, GA), pp. 11-761-11-764, May 1996.

86) Loo. J. H. Y.
Intraframe and Interframe Coding of Speech Spectral Parameters
Thèse de Master , Université de McGill Montreal, Canada Septembre 1996.

87) Makhoul. J.
"Linear prediction: A tutorial review." Proceedings of the I.E.E.E. Vol. 63 N°4 April 1975

88) Makhoul. J., Roucos. S. and Gish. H.
"Vector quantization in speech coding" Proceedings I.E.E.E. Vol. 73 n°11 pp 1551- 1588 Nov. 1985.

89) Mermelstein. P.
" G.722 : A new CCITT coding standard for digital transmission of wideband audio signals"
IEEE Commun. Mag., pp. 8-15, Jan. 1988.

90) Marcellin. M. W. and Fisher. T.R.
"Trellis coded quantization of memoryless and Gauss Markov sources."
IEEE Trans. on Communications Vol . 38 pp 82-93 Jan. 1990

91) Mauc. M. and Baudoin. G.
"Reduced complexity celp coder."
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 92 pp. I-53 I-56

92) Moreau. N. and Dymarsky. P.
"Successive orthogonalizations in the multistage CELP coder."
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 92 pp. I-61 I-64

93) Macree. A. V. and Barnwell III. T.P.
"Improving the performance of a mixed excitation LPC vocoder in acoustic noise."
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 92 pp. II-137 II-140

94) Mauc. M
"Réduction de la complexité des algorithmes de codage de la parole de type CELP. Application au standard FS-1016." Thèse.de Docteur en sciences, Université Paris XII Novembre 1993.

95) Nielsen. H., Mikkelsen. K. B., Hansen. H. B., Larsen. K. J., Sorenson. J. A.
"Comparative study of error correction coding schemes for the GSM half rate channel."
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 92 pp. II-129 II-132

96) O'shaugnessy. D.
"Speech Communication: Human and Machine. Reading, MA."
Addison-Wesley, 1987.

97) Ono. S. and Ozawa. K.
"2.4 kbits/s pitch prediction multipulse speech coding."
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 88 pp. 175-178

98) Paez. M.D. and Glisson. T.H.
"Minimum mean-squared error quantization in speech PCM and DPCM systems."
IEEE Trans. on Communications Vol . COM. 20 pp. 225-230 April 1972

99) Pan. J. and Fisher. T.R.
"Vector quantization of speech LSP parameters using trellis codes and L_1 -norm constraints."
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 93 pp. II-17 II-20

100) Paliwal. K. and Atal. B.S.
"Efficient vector quantization of LPC parameters at 24 bits/frame."
IEEE Transactions on Speech and Audio Processing January 1993. Vol.1 n°1 pp. 3-14

101) Quackenbush. S. R.
" A 7 kHz bandwidth, 32 kbps speech coder for ISDN"
presented at Proc. ICASSP, 1991.

102) Rabiner. L.R., Atal. B.S. and Sambur. M.R.
"LPC prediction error - Analysis of its variation with the position of the analysis frame."
IEEE Transactions on Acoustics Speech and Signal Processing. Vol. ASSP 25 N=5 October 1977

103) Ramachandran. R. P. and kabal. P.
"Pitch prediction filters in speech coding."
IEEE Transactions on Acoustics Speech and Signal Processing. Vol. ASSP 37 N°4 pp. 467-478 April 1989

104) Rabiner. L. R. and Juang. B. H.
"Fundamentals of Speech Recognition." Englewood , New Jersey: Prentice-Haîl, 1993.

105) Ramabadran. T. V. and Lueck. C.D.
"Complexity reduction of CELP speech coders through the use of phase information."
IEEE Transactions on Communications Vol.42 N°2/3/4 pp 248-251 Feb./March/April 1994

106) Ramabadran. T. V. and Sinha. D.
"Speech data compression through sparse coding of innovation."
IEEE Transactions on Speech and Audio Processing April 1994. Vol.2 n°2 pp 274-284

107) Ricordel. V.
"Etude de schémas de quantification vectorielle algébrique et arborescente. Appplication à la compression de séquences d'images numériques."
Thèse de Docteur de l'Université de Rennes I Décembre 1996.

108) Singhal. S. and Atal. B. S.
''Improving performance of multipulse LPC coders at low bit rates.''
Proc. of Inte. Conf. on Acoust. Speech and Signal Processing. ICASSP 84 Vol. 1 N°13 March 1984

109) Soong. F. K. and Juang. B. H.
''Line Spectrum Pair (LSP) and speech data compression.''
Proc. of Inter. Conf. on Acoust. Speech and Signal Processing. ICASSP84 pp.1-10-1,1-10-4 March 1984

110) Schroeder . M. R. and Atal. B. S.
''Code excited linear prediction (CELP) : High quality speech at very low bit rates.''
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 85 pp. 937-940

111) Schroeder . M. R. and Atal. B. S.
''High quality speech at very low bit rates.''
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 85 pp. 937-940

112) Schroeder . M. R. and Atal. B. S.
''Stochastic coding of speech signals at very low bit rates: The importance of speech perception
Speech Communications Vol 4 pp. 155-162 August 1985.

113) Singhal. S.
''On encoding filter parameters for stochastic coders.''
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 87 pp. 1633-1636

114) Shoham. Y.
''Vector predictive quantization of the spectral p&ameters for low rate speech coding.''
in Proc. Int. Conf. Aco~t., Speech, Signal Proc., (Dallas), pp. 2181-2184, April 1987

115) Sreenivas. T.V.
''Modelling LPC residue by components for good quality speech coding.''
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 88 pp. 171-174

116) Saoudi. S., Le Guyader. A. and Boucher. J.M.
''Optimal scalar quantization of the Parcor coefficients for speech coding.''
Eurocon 88. 8th European Conference on Electrotechnics Conference Proceedings on Area
Communications. Stockholm Sweden June 13-17 1988 pp. 32-35

117) Sugamura. N. and Farvardin. N.
''Quantizer design in LSP Speech Analysis Synthesis.''
IEEE Journal on Selected Areas in Communications Vol. 6 N=2 Feb. 88

118) Singhal. S. and Atal. B. S.
''Amplitude optimization and pitch prediction in multipulse coders.''
IEEE Transactions on Acoustics Speech and Signal Processing. Vol. ASSP 37 N°3 pp.317-327 March 1989

119) Soheili. R., Kondoz. A.M. and Evans. B.G.
''Techniques for improving the quality of LD-CELP coders at 8 kb/s.''
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 92 pp. I-41 I-44

120) Su. H.Y. and Mermelstein. P.
''Improving the speech quality of cellular mobile systems under heavy fading.''
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 92 pp. II-121 II-124

121) Shoham. Y.
''High quality speech coding at 2.4 to 4.0 kbps based on time frequency interpolation.''
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 93 pp. II-167 II-170

122) Soong. F.K. and Juang. B.H.
''Optimal quantization of LSP parameters.''
IEEE Transactions on Speech and Audio Processing January 1993 Vol.1 n°1 pp. 15-24

123) Trancoso. I.M. and Atal. B. S.
''Efficient procedures for finding the optimum innovation in stochastic coders.''
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 86 Tokyo pp. 2375-2378

124) Ungerboeck. G.
''Trellis coded modulation with redundant signal sets.'' Parts I and II
IEEE Communications Magazine Vol. 2 pp 5 -21 February 1987.

125) Viswanathan. R. and Makhoul. J.
''Quantization properties of transmission parameters in linear predictive systems.''
IEEE Transactions on Acoustics Speech and Signal Processing. Vol. ASSP 23 N=3 June 1975

126) Wang. T.K., Foster. J. and Ardalan. S.
''Adaptive vector quantization for waveform coding.''
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 92 pp. I-101 I-104

127) Wang. S., Sekey. A. and Gersho. A.
''An objective measure for predicting subjective quality of speech coders.''
IEEE Journal on Selected Areas in Communications Vol. 10 N=5 June 92

128) Xiongwei. Z. and Xianzhi. C.
A new excitation model for LPC vocoder at 2.4 kb/s.
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing. ICASSP 92 pp. I-65 I-68

129) Xydeas. C.S. and So. K.K.M.
''A long history quantization approach to scalar and vector quantization of LSP coefficients.''
Proc. of Intern. Conf. on Acoust. Speech and Signal Processing ICASSP 93, pp. II-1 II-4

130) Yong. M., Davidson. G. and Gersho. A.
''Encoding of LPC spectral parameters using switched-adaptive interframe vector prediction.''
in Proc. Int. Conf Acoust., Speech, Signal Froc., (New York), pp. 402-405, April 1988.

131) Yu-Hung Kao
''Low Complexity CELP Speech Coding at 4.8 kbps.''
Thèse de Master, Universté de Maryland USA, 1990

132) Zeger. K., Bist. A. and Linder. T.
Universal source coding with codebook transmission
IEEE Transactions on Communications. Vol.42 N°2/3/4 pp 336-346 Feb./March/April 1994

133) CCITT Study Group XVIII, '' 7kHz audio coding within 64 kb/s,''
CCITT Draft Recommendation G.722, Report of working Party XVIII/8, July 1986.