République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Département d'Electronique

Projet de Fin d'Etudes
pour l'obtention du diplôme
d'Ingénieur d'Etat en Electronique

Thème :

Codage à large bande de l'enveloppe spectrale

Présenté par :
AOUALI Boudjemaa Reda
TIAIBA Kheir-eddine

Proposé par :
Dr F.MERAZKA

Juin 2006

REMERCIEMENTS

Ce travail a été effectué au sein du laboratoire de signal et communications du département d'électronique de l'Ecole Nationale Polytechnique, sous la direction du Dr F.MERAZKA

Nous tenons à lui exprimer nos plus sincères remerciements pour ses conseils, son aide et sa patience tout au long de ce travail.

Nous exprimons notre plus sincère gratitude et remerciements à Mr. B Bousseksou, et Mr. R Zergui pour avoir accepté d'examiner notre travaille.

Nous tenons à exprimer notre très grande gratitude, et notre profonde affection à nos chers parents pour leur encouragement, leur patience et leur grand soutien, durant toutes ces années d'études.

Nous tenons également à remercier tous nos amis et camarades, pour leur sincère amitié et leur précieux soutien.

REDA** et **ROCHDI.

DEDICACE

A nos chers parents
A nos frères Djamel, Walid et Mami
A nos très chères soeurs
A nos deux familles

A tous mes amis : Moh, Mounir, Khalil, Nawfel, Bilal, Samir et Djamil

A VOUS TOUS, MERCI !

Reda et Rochdi

ملخص

في هذه الدراسة نهتم بالتشفير ذات الشريط الواسع للغلاف الطيفية, لهذا الهدف قمنا بمقارنة 10 ,16, 18 و20 جذر .لذلك قمنا بتكميم خطوط التوترات الطيفية, اخيرا قارنا الفعاليات بواسطة التعوجات الطيفية.

كلمات مفتاحية :تشفير الكلام,خطوط التوترات الطيفية,التكميم الشعاعي,التكميم الخطي

Résumé

Dans cette thèse on s'intéresse au codage large bande de l'enveloppe spectrale, pour cela nous avons comparé plusieurs ordres de prédiction : 10, 16,18 et 20 pôles.

Nous avons procédé à la quantification des LSF pour les différents ordres, par la suite nous avons comparé leurs performances au moyen des distorsions spectrales.

Mots clefs

Codage de la parole, excitation, lignes de fréquences spectrales LSF, quantification vectorielle, quantification scalaire.

Abstract

In this thesis we are interested in wide band coding of the spectral envelope, for that we compared several orders of prediction: 10, 16, 18 and 20 poles.

We proceeded to the quantification of the LSF for the various orders, there after we compared their performances by means of the spectral distortions.

Key words

Speech coding, lines of spectral frequencies LSF, vectorial quantification, scalar quantification.

Sommaire

Liste des figures

Liste des tableaux

Lexique

ACBK	Adaptatif Code BooK
ADPCM	Adaptive Differential Pulse Code Modulation.
AR	Auto-Regressif.
ARMA	Auto-Regressif Moving Average
CELP	Code Excited Linear Prediction.
CS-ACELP	Conjugate Structure Algebraic Code Excited Linear Prediction
DAM	Diagnostic Acceptability Measure
DPCM	Differential Pulse Code Modulation.
DRT	Diagnostic Rhyme Test
EMBSD	Enhanced Modified Bark spectral Distortion
FCBK	Fixed Code BooK
FEC	Forward Error Correction.
IP	Internet Protocol.
ITU	International Telecommunication Union
LP	Linear Prediction.
LPC	Linear Prediction Coding.
LSP	Line Spectrum Pairs.
MIPS	Million d'operations par seconde
MOS	Mean Opinion Score
PCM	Pulse Code Modulation.
PESQ	Perceptual evaluation of Speech Qualité
PLC	Packet Loss Concealment
RTP	Real time Protocol
RTCP	Real time Control Protocol
SNR	Signal to Noise Ratio
SD	Spectral Distortion.
RSB	Rapport Signal sur Bruit
RSBseg	Rapport Signal sur Bruit segmenté
SQ	Scalar Quantization.
SVQ	Split Vector Quantization.
VoIP	Voice cover IP network.
VQ	Vector Quantization.

Tableau I Table des Abréviations

Introduction

Ce projet de fin d'étudie traite le codage des signaux à large bande. La transmission en large bande correspond à l'élargissement de la bande passante utilisée pour la transmission du signal de parole.

En effet, la bande passante utilisée habituellement en téléphonie est 200-3400 Hz, cependant les nouvelles technologies liées aux réseaux permettent une utilisation plus flexible de la transmission de la parole, grâce à un choix assez large de codeur et de bande passante, qui a facilité l'apparition d'une nouvelle bande passante 50-7000 Hz, améliorant la qualité du signal de parole transmis. L'étude de la qualité de la parole est un domaine important de la psycho-acoustique pour ses applications dans la synthèse sonore, la médecine ou comme ici, pour la téléphonie.

Le développement des applications multimédia sur l'Internet ainsi que les systèmes de conférence téléphonique feraient bon usage d'un système adaptatif permettant de régler le niveau de qualité du codage selon le débit disponible. Cette étude propose une solution destinée à répondre à ce besoin.

L'objectif de notre étude est tout d'abord d'élargir la bande passante tout en optimisant le quantificateur, et avoir une distorsion spectrale minime afin d'avoir une bonne transmission de signaux audio dans la gamme de fréquence voulue.

Nous avons organisé notre travail en trois chapitres :

Le premier chapitre Est consacré au codage de la parole : la prédiction linéaire, le modèle de production de la parole humaine et sa distorsion.

Le deuxième chapitre Donne des notions de bases et principes du codage à large bande.

Le troisième chapitre Nous avons comparé les résultats du codage de l'enveloppe spectrale pour différents ordres de prédiction m=10,16 ,18 et 20 pôles , à l'aide de la distorsion spectrale.

Enfin une conclusion générale résume et cols notre travail.

Chapitre I

Codage de la parole

Le traitement de la parole est aujourd'hui une composante fondamentale des sciences de l'ingénieur. Située au croisement du traitement du signal numérique et du traitement du langage. Cette discipline scientifique a connu depuis les années 60 une expansion fulgurante, liée au développement des moyens et des techniques de télécommunications.

Ce chapitre regroupe des généralités sur les notions fondamentales de la production du signal parole, ses propriétés ainsi que sa perception. Cet aspect est utile à la bonne compréhension de l'évolution des techniques de codage de la parole.

I.1 Signal vocal

La parole peut être décrite comme étant le résultat de l'action volontaire et coordonnée d'un certain nombre d'organes. Cette action se déroule sous le contrôle du système nerveux central qui reçoit en permanence des informations par rétroaction auditive et par les sensations kinesthésiques[4].

I.2 Mécanisme de phonation

Les principaux organes composant l'appareil phonatoire sont [1]: les poumons, la trachée artère, le pharynx, les cavités buccales et nasales qui sont schématisés par la Figure I.1.

L'appareil respiratoire fournit l'énergie nécessaire à la production de sons, en poussant de l'air à travers la trachée-artère. Au sommet de celle-ci se trouve le ***larynx*** où la pression de l'air est modulée avant d'être appliquée au conduit vocal. Le larynx est un ensemble de muscles et de cartilages mobiles qui entoure une cavité située à la partie supérieure de la trachée.

Les ***cordes vocales*** sont en fait deux lèvres symétriques placées en travers du larynx. Ces lèvres peuvent fermer complètement le larynx et en s'écartant progressivement, déterminer une ouverture triangulaire appelée ***glotte***. L'air y passe librement pendant la respiration et la voix chuchotée ainsi que pendant la phonation des sons non voisés.

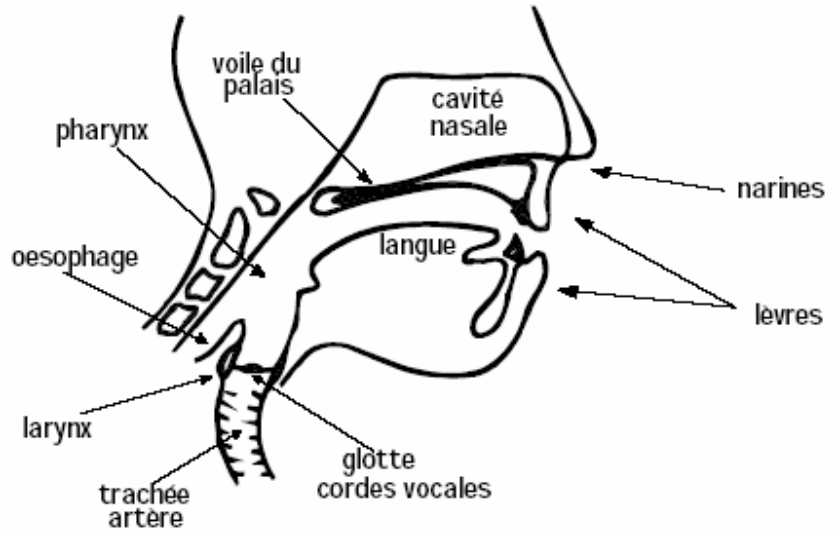


Fig. I.1 Appareil phonatoire

Les sons voisés résultent, au contraire, d'une vibration périodique des cordes vocales. Le larynx est d'abord complètement fermé, ce qui accroît la pression en amont des cordes vocales et force ces dernières à s'ouvrir, ce qui fait tomber la pression en permettant aux cordes vocales de se refermer. Des impulsions périodiques de pression sont ainsi appliquées au conduit vocal composés des cavités pharyngienne et buccale pour la plupart des sons. Lorsque la *luette* est en position basse, la cavité nasale vient s'y ajouter en dérivation. Notons pour terminer le rôle prépondérant de la langue dans le processus phonatoire. Sa hauteur détermine la hauteur du pharynx : plus la langue est basse, plus le pharynx est court. Elle détermine aussi le *lieu d'articulation*, région de rétrécissement maximal du canal buccal, ainsi que l'aperture qui représente l'écartement des organes au point d'articulation. L'intensité du son émis est liée à la pression de l'air en amont du larynx. Sa hauteur est fixée par la fréquence de vibration des cordes vocales, appelée fréquence du fondamental ou pitch. La fréquence du fondamental peut varier [2][3]

- De 80 à 200 *Hz* pour une voix masculine.
- De 150 à 450 *Hz* pour une voix féminine.
- De 200 à 600 *Hz* pour une voix d'enfant.

Un *son voisé* est un signal quasi périodique dont le spectre est tracé à la Figure I.2. On y observe les raies qui correspondent aux harmoniques du fondamentale F0 (pitch).
L'enveloppe de ces raies présente des maximums appelés *formants* et qui correspondent aux fréquences propres *Fi* du conduit vocal (structure formantique). Les trois premiers formants sont essentiels pour caractériser le spectre vocal; les formants d'ordre supérieur ont une influence plus limitée.

Un son *non voisé* ne présente pas de structure périodique. Il peut être considéré comme un bruit blanc filtré par la transmittance de la partie du conduit vocal situé entre la constriction et les lèvres comme le montre la Figure I.3; son spectre ne présente donc pas de structure de pitch.
La classification ainsi exposée est forcément un peu sommaire et concerne surtout la production normale de la parole. Ainsi, une voyelle peut être chuchotée, c'est-à-dire produite avec la glotte largement ouverte; dans ce cas, le spectre du signal résulte de l'excitation du conduit vocal par une source aléatoire : c'est un spectre continu qui présente une structure formantique semblable à celle d'une voyelle voisée mais ne possède pas de structure de pitch (raies dues aux harmoniques

du fondamental).

De nos jours, il reste très difficile de dire comment l'information auditive est traitée par le cerveau. On a pu, par contre, étudier comment elle était finalement perçue dans le cadre d'une science spécifique appelée *psychoacoustique*. Sans vouloir entrer dans trop de détails sur la contribution majeure des *psychoacousticiens* dans l'étude de la parole, il est intéressant d'en connaître les résultats les plus marquants. Ainsi, l'oreille ne répond pas également à toutes les fréquences. Le seuil d'audition de l'oreille est non linéaire par rapport aux fréquences. L'oreille atteint sa sensibilité maximale entre 3 et 4 kHz.

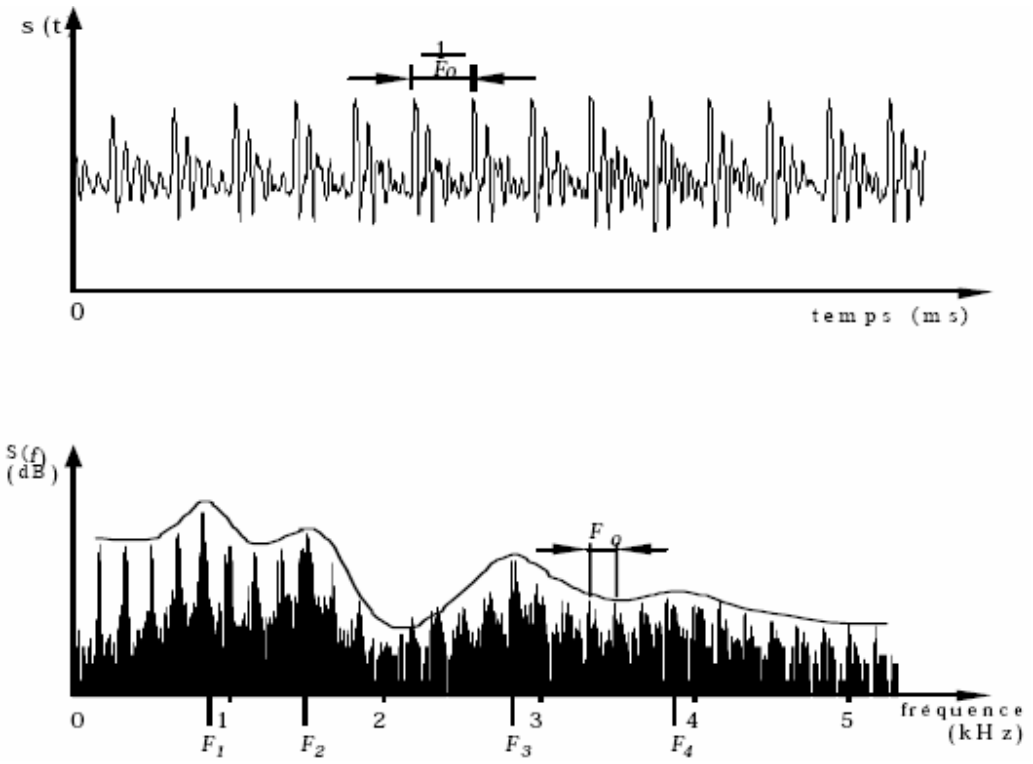


Fig. I.2 Un signal vocal voisé et son spectre [3][4]

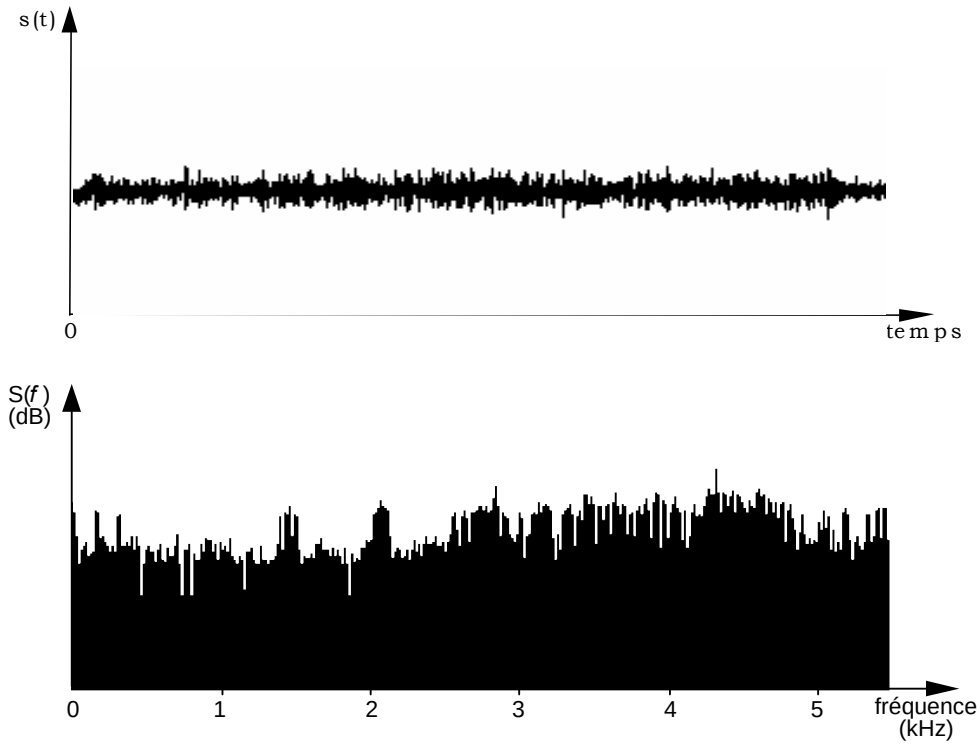


Fig. I.3 Un signal vocal non voisé et son spectre [3][4]

I.3 La redondance du signal vocal

Telle que définie par Shannon, la redondance est la partie du signal parole qui, si elle est éliminée, n'affecte pas le contenu du message ou du signal information.

Le signal vocal est caractérisé par une très grande redondance, condition nécessaire pour résister aux perturbations du milieu ambiant, cette redondance sera mise à profit par les techniques de codage de la parole, pour réduire le débit binaire nécessaire au stockage ou à la transmission de la parole, sans, pour autant nuire à son intelligibilité.

On définie l'information associée à un message constitué par des éléments discrets x_i appartenant à un ensemble donné X, et si $p(x_i)$ est la probabilité a priori d'occurrence du symbole x_i , on a donc l'information moyenne associée à l'occurrence du message $X=[x_1,x_2,\ldots,x_l]$ qui vaut :

$$H(X) = -\sum_i p(x_i) \log_2 p(x_i) \qquad \text{(I.1)}$$

C'est l'*entropie* de la source exprimée en bits.

Dans la conversation courante, environ dix phonèmes [(1)] sont prononcés chaque seconde; l'information moyenne est donc inférieur à 50 bits/s [2].Or, on sait que pour un canal continu sans erreurs, le débit maximum d'information est donné par l'équation (I.2) :

$$C=B\ log_2[1+S/N] \tag{I.2}$$

Avec B est la longueur de la bande passante en Hz, et *S/N* est le rapport signal sur bruit en dB.

Par exemple, pour un canal téléphonique, supposé continu et sans erreurs, de bande passante B=3000 Hz et avec un rapport signal sur bruit *S/N*=30 dB, on trouve C=30000 bits/s, il y a apparemment une redondance énorme dans ce canal. La suppression partielle des redondances permet une représentation plus efficace des données.

La compression des données peut se faire sans pertes d'information ou avec pertes en exploitant dans ce cas la tolérance de l'organe récepteur (l'oreille). La compression du signal consistera à réduire les redondances du signal parole.

I.4 Modèle de production de la parole

L'analyse de la parole est une étape indispensable à toute application de synthèse, de codage ou de reconnaissance.

Le modèle électrique linéaire a été proposé par Fant [3] en 1960, qui spécifie qu'un signal voisé peut être modélisé par le passage d'un train d'impulsions *u(n)* à travers un filtre numérique récursif de type tous-pôles (*Auto Régressif*). On montre que cette modélisation reste valable dans le cas des sons non voisés, à condition que *u(n)* soit cette fois un bruit blanc. Le modèle final est illustré à la Figure I.4. Il est souvent appelé modèle auto régressif (*AR*), parce qu'il correspond dans le domaine temporel à une régression linéaire de la forme :

$$s(n)=G.u(n)+\sum_{i=1}^{p}-a_i s(n-1) \tag{I.3}$$

Où *u(n)* est le signal d'excitation et *p* l'ordre du système.

Chaque échantillon est obtenu en ajoutant un terme d'excitation à une prédiction obtenue par combinaison linéaire des *p* échantillons précédents.

[(1)]Phonème : c'est la plus petite unité présente dans la parole et susceptible par sa présence de changer la signification d'un mot [2].

Les coefficients du filtre $\{a_i\}$ sont appelés coefficients de prédiction et le modèle AR est souvent appelé modèle de prédiction linéaire.

les paramètres du modèle *AR* sont : la période du train d'impulsions (sons voisés uniquement), la décision Voisé/Non Voisé (V/NV), le gain *G* et les coefficients du filtre *1/A(z)*, appelé ***filtre de synthèse***.

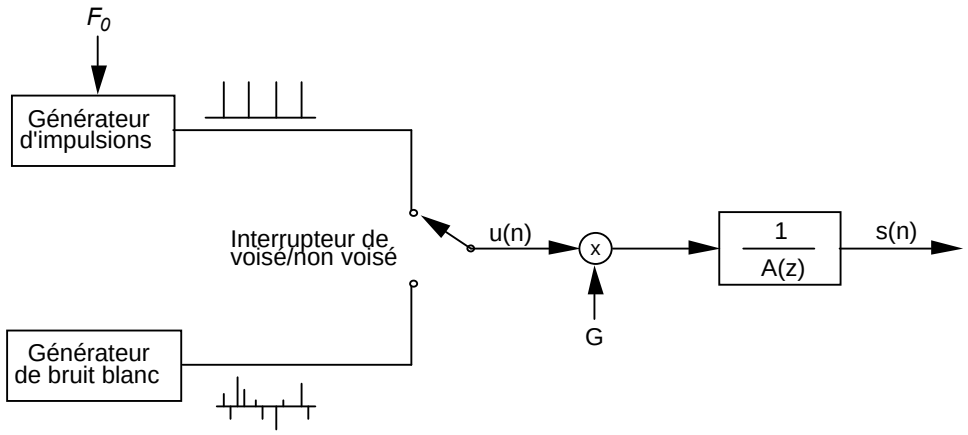


Fig. I.4 Modèle simplifié de production de la parole [3][4]

Les relations d'équivalences entre le modèle physique et le modèle mathématique [4] peuvent être données comme suit :

Conduit vocal	⟺	*H(z)*, le filtre LPC
Le flux d'air	⟺	Le signal d'excitation *u(n)*
Vibration des cordes vocales	⟺	Voisé
Fricatives et plosives	⟺	Non Voisé/Voisé
Volume d'air	⟺	*G,* le gain

Le problème de l'estimation d'un modèle AR, souvent appelée analyse *LPC* revient à déterminer les coefficients d'un filtre tous-pôles dont on connaît le signal de sortie, mais pas celui de l'entrée. Il est par conséquent nécessaire d'adopter un critère, afin de faire un choix parmi

l'ensemble infini de solutions possibles. Le critère généralement utilisé est celui de la minimisation de l'énergie de l'erreur de prédiction.

I.5 Prédiction Linéaire

La prédiction linéaire est assez bien utilisée dans les systèmes de codage et de compression [6][7][8]. Cette méthode est considérée comme une technique prédominante pour l'estimation des paramètres de la parole. Son succès est dû au fait qu'elle représente une solution linéaire au problème de l'estimation des paramètres du modèle de la production de la parole.

Le principe fondamental de la prédiction linéaire est qu'un échantillon donné peut être prédit à partir d'une combinaison linéaire des échantillons finis qui le précèdent. Un seul jeu de coefficients du prédicteur sont déterminés en minimisant les différences entre les échantillons actuels et ceux prédits. La technique de prédiction linéaire est basée sur le modèle de la production de la parole représenté à la figure I.4.

Le signal parole *s(n)* peut être modélisé comme la sortie d'un système *auto régressif à moyenne ajustée* (ARMA) avec une entrée *u(n)* [3][5][9]. Son expression est alors :

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + G\sum_{i=0}^{q} b_i u(n-i)\,,\ b_0=1, \qquad \text{(I.4)}$$

où le gain *G*, les coefficients $\{a_k\}$ et $\{b_i\}$ sont les paramètres du système, et p et q sont les ordres des polynômes. L'équation (I.4) prédit la sortie courante en utilisant une combinaison linéaire des sorties précédentes et les entrées courantes et précédentes.

Dans le domaine fréquentiel, la fonction de transfert du modèle de prédiction linéaire de la parole est de la forme :

$$H(z) = \frac{B(z)}{A(z)} = \frac{G[1+\sum_{i=1}^{q} b_i z^{-i}]}{1-\sum_{k=1}^{p} a_k z^{-k}} \qquad \text{(I.5)}$$

H(z) est le modèle pôle-zéro dans lequel les racines du dénominateur et de numérateur sont, respectivement, les pôles et les zéros du système.

Si a_k=0 pour $1 \le k \le p$, *H(z)* devient un modèle tous-zéros ou modèle à *moyenne ajustée* (*MA*).

Si pour b_i=0, pour 1≤i≤q, *H(z)* devient un modèle tous-pôles ou modèle *auto régressive* (*AR*), exprimé par :

$$H(z) = \frac{1}{A(z)} \tag{I.6}$$

L'analyse spectrale montre que les pôles correspondent aux résonances du conduit vocal, c'est-à-dire aux *pics* du spectre, les *formants* ; tandis que les zéros correspondent aux antirésonances, c'est-à-dire aux *vallées*.

Dans l'analyse de la parole, les classes de phonèmes comme les fricatives et les nasales contiennent des vallées spectrales qui correspondent aux zéros dans *H(z)*.

Par contre, les voyelles contiennent des résonances qui peuvent être modélisées par le modèle tous-pôles; pour des raisons de simplicité, ce modèle est préféré pour l'analyse par prédiction linéaire de la parole. Ainsi, le signal prédit est égal à :

$$\tilde{s}(n) = \sum_{k=1}^{p} a_k s(n-k) \tag{I.7}$$

La différence entre l'échantillon original $s(n)$ et l'échantillon prédit $\tilde{s}(n)$ est appelée *erreur de prédiction* (ou *résidu*) et elle est définie par:

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^{p} a_k s(n-k) \tag{I.8}$$

Le problème de l'analyse par prédiction linéaire se réduit donc à trouver un ensemble de coefficients a_k de façon à minimiser l'erreur de prédiction $e(n)$ dans un certain intervalle. Les méthodes d'estimation des coefficients a_k sont nombreuses [10].

Deux grandes approches sont utilisées pour l'analyse par prédiction linéaire LPC court-terme : La méthode d'autocorrélation et la méthode de covariance.

I.5.1 Méthode d'Autocorrélation

La méthode d'autcorrélation garantit la stabilité du filtre LP. Les hypothèses de cette méthode sont les suivantes :

Le signal est défini pour toutes les valeurs du temps ; il est identiquement nul en dehors d'une séquence de *N* échantillons, où *N* est un entier; ceci est équivaut à multiplier le signal de parole

s(n) par une fenêtre *w(n)* de longueur finie correspondant à *N* échantillons pour obtenir un segment du signal de parole fenêtré $s_w(n)$ [11].

$$s_w(n) = \begin{cases} w(n).s(n) & pour \quad 0 \le n \le N-1 \\ 0 & ailleurs \end{cases} \tag{I.9}$$

La fonction de pondération la plus courante est la fenêtre de *Hamming* :

$$w(n) = \begin{cases} 0.54 - 0.46\cos\dfrac{2n\pi}{N-1} & pour \quad 0 \le n \le N-1 \\ 0 & ailleurs \end{cases} \tag{I.10}$$

Chaque échantillon peut être prédit approximativement à partir des échantillons précédents. Ceci est valable pour toutes les valeurs du temps; ($-\infty<n<+\infty$).

L'erreur quadratique totale entre le signal fenêtré $s_w(n)$ et le modèle (signal prédit) est minimisée sur l'ensemble des échantillons.

La fonction d'autocorrélation du signal fenêtré $s_w(n)$ est :

$$R(i) = \sum_{n=1}^{N-1} s_w(n).s_w(n-i) \qquad 1 \le i \le p \tag{I.11}$$

La fonction d'autocorrélation est une fonction paire: *R(i)= R(-i)*.

Pour trouver les coefficients du filtre LPC, l'énergie du résiduel de prédiction doit être minimisée sur l'intervalle fini : $0 \le n \le N-1$

$$E = \sum_{n=-\infty}^{\infty} e^2(n) = \sum_{n=-\infty}^{\infty} [s_w(n) - \sum_{k=1}^{p} a_k s_w(n-k)]^2 \tag{I.12}$$

Cette erreur peut être minimisée en annulant les dérivées partielles par rapport aux coefficients du filtre :

$$\frac{\partial E}{\partial a_k} = 0 \qquad 1 \le k \le p \tag{I.13}$$

On obtient p équation linéaire avec p coefficient inconnus a_k :

$$\sum_{k=1}^{p} a_k \sum_{n=-\infty}^{\infty} s_w(n-i)s_w(n-k) = \sum_{n=-\infty}^{\infty} s_w(n-i)s_w(n). \qquad tq: 1 \le i \le p \tag{1.14}$$

Alors, les équations linéaires peuvent être écrites sous la forme :

$$\sum_{k=1}^{p} R(|i-k|)a_k = R(i) \qquad 1 \le i \le p \tag{1.15}$$

La forme matricielle de l'ensemble des équations linéaires (I.14) est représenté par ***R.a=v*** et peut être réécrite comme suit :

$$\begin{bmatrix} R(0) & R(1) & \dots & R(p-1) \\ R(1) & R(0) & \dots & R(p-2) \\ & & \dots & . \\ . & . & \dots & . \\ . & . & \dots & . \\ R(p-1) & R(p-2) & \dots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ . \\ . \\ . \\ a_p \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ . \\ . \\ . \\ R(p) \end{bmatrix} \tag{I.16}$$

La matrice d'autocorrélation *pxp* obtenue est symétrique dont tous les éléments de la diagonale sont égaux, c'est une matrice de *Toeplitz*. Ce qui nous permet de trouver les coefficients de prédiction minimisant la moyenne quadratique de l'erreur de prédiction par l'algorithme de *Levinson-Durbin* (Annex A).

I.5.2 Méthode de Covariance

Les méthodes d'autocorrélation et de covariance diffèrent dans l'emplacement de la fenêtre d'analyse.

Dans cette méthode c'est le signal erreur qui est fenêtré au lieu du signal parole, de façon à ce que l'énergie à minimiser soit :

$$E = \sum_{n=-\infty}^{\infty} {e_w}^2(n) = \sum_{n=-\infty}^{\infty} e^2(n)w^2(n) \tag{I.17}$$

En annulant les dérivées partielles en utilisant l'équation (I.13) on obtient p équations linéaires :

$$\sum_{k=1}^{p} \Phi(i,k) = \Phi(i,0) \qquad 1 \le i \le p \tag{I.18}$$

Où la fonction de covariance :

$$\Phi(i,k) = \sum_{n=-\infty}^{\infty} w(n)s(n-1)s(n-k) \tag{I.19}$$

On peut exprimer les p équations, sous la forme : $\Phi.a=\Psi$

$$\begin{bmatrix} \Phi(1,1) & \Phi(1,2) & ... & \Phi(1,p) \\ \Phi(2,1) & \Phi(2,2) & ... & \Phi(2,p) \\ \Phi(3,1) & \Phi(3,2) & ... & \Phi(3,p) \\ & & ... & \\ & & ... & \\ \Phi(p,1) & \Phi(p,2) & ... & \Phi(p,p) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \\ \\ a_p \end{bmatrix} = \begin{bmatrix} \Psi(1) \\ \Psi(2) \\ \Psi(3) \\ . \\ . \\ \Psi(p) \end{bmatrix} \qquad \text{(I.20)}$$

Tel que; $\Psi(i)=\Phi(i,0)$ pour $1 \leq i \leq p$

La matrice Φ n'est pas une matrice Toeplitz, et ne garantie pas la stabilité du filtre LPC, elle est symétrique et définie positive. Donc, la matrice de covariance peut être décomposée en deux matrices, l'une triangulaire inférieure L et l'autre triangulaire supérieure U.

$$\Phi=L.U \qquad \text{(I.21)}$$

La décomposition de Cholesky peut être utilisée pour convertir la matrice de covariance sous la forme :

$$\Phi=C.C^T \qquad \text{tq; } C=L \text{ et } C^T=U$$

Le vecteur a est obtenu en résolvant d'abord l'équation (I.22) :

$$L.y=\Psi \qquad \text{(I.22)}$$

Puis :

$$U.a=y \qquad \text{(I.23)}$$

I.5.3 Considération Pratiques

Pour bien mener l'analyse LPC, il faut choisir :

- La fréquence d'échantillonnage f_e.
- La méthode d'analyse et l'algorithme correspondant.
- L'ordre p de l'analyse LPC.
- Le nombre d'échantillons par tranche N et le décalage entre tranches successives L.

Le choix de la fréquence d'échantillonnage est fonction de l'application visée et de la qualité du signal à analyser :

- 8 kHz pour les signaux téléphoniques.
- 10 kHz pour les applications de reconnaissance.
- 16 kHz pour les applications de synthèse.

L'ordre de prédiction *p* est choisi de façon à ce qu'il permette de bien représenter toute la séquence du signal parole; l'ordre *p* est fonction de la fréquence d'échantillonnage, on estime en général qu'une paire de pôles est nécessaire par 1Khz de bande passante.

Lorsque la fréquence d'échantillonnage est f_e (exprimée en échantillons/sec), une période de 1ms correspond à f_e/1000 échantillons.

A la fréquence d'échantillonnage de 8 kHz, la valeur correspondante de *p* doit être au moins égale à 8. Elle trouve d'ailleurs une justification expérimentale dans le fait que l'énergie de l'erreur de prédiction diminue rapidement lorsqu'on augmente *p* à partir de 1, pour tendre vers une asymptote au voisinage de ces valeurs : il devient inutile d'augmenter encore l'ordre, puisqu'on ne prédit rien de plus.

De plus la durée des trames d'analyse et leur décalage sont souvent fixés inférieur à 30ms.Les valeurs choisies sont liées au caractère quasi-stationnaire du signal parole.

Enfin, comme vu précédemment dans la méthode d'autocorrélation, pour compenser les effets de bord, on multiplie en général préalablement chaque tranche d'analyse par une fenêtre de pondération *w(n)*, la plus souvent utilisées est celle de *Hamming* (équation (I.10)).

I.5.4 Représentation des paramètres de prédiction

Les coefficients de prédiction linéaire (*LP*) sont calculés à base de "bloc par bloc", généralement sur des trames de 5-40ms [12]. Pour une transmission efficace de la parole, les coefficients *LP* sont sujets à une **quantification** et une **interpolation**. L'interpolation rend possible la transmission de l'information sur les coefficients *LP* moins souvent, ainsi réduisant le débit binaire. Cependant, une simple quantification ou une interpolation des coefficients *LP* est problématique parce que de petits changements dans les coefficients peuvent induire un grand changement dans le spectre de puissance et causer l'instabilité du filtre de synthèse *LP*. Par

conséquent, un nombre de représentations des coefficients *LP* été considéré pour essayer de trouver la représentation qui minimise ses limitations.

Les représentations les plus utilisées sont les coefficients de réflexion, les LAR (log-area ratios) [12] et les LSPs (Line Spectrum Pairs) [13].

Cependant la représentation la plus répandue et la plus prisée pour ses performances reste la représentation en paires de rais spectrales LSP.

Elles seront détaillées dans ce qui va suivre.

I.5.4.1 Paires de raies spectrales

Connus aussi sous le nom de fréquences de raies spectrales.

La représentation LSP a été introduite par ***Itakura*** [13].

Les LSPs sont les solutions des deux équations suivantes :

$$\begin{cases} P(z) = A(z) + z^{-(p+1)} A(z) \\ Q(z) = A(z) - z^{-(p+1)} A(z) \end{cases} \tag{I.24}$$

Ce qui nous donne :

$$A(z) = \frac{1}{2}[P(z) + Q(z)] \tag{I.25}$$

Soong et ***Juang*** [14] ont montrés que si *H(z)* est stable, où *A(z)* est à phase minimale, alors les zéros des polynômes *P(z)* et *Q(z)* sont appels les LSP. Ces polynômes ont les propriétés suivantes [4]:

- Tous les zéros de *P(z)* et *Q(z)* se trouvent sur le cercle unité.
- Les zéros de *P(z)* et *Q(z)* sont entrelacés les uns aux autres, les LSP sont dans un ordre croissant.

Il a été montré [15] que le filtre LPC *A(z)* est à phase minimale si et seulement si les LSP satisfont les deux propriétés citées plus haut, donc la stabilité du filtre de synthèse est facilement vérifiable. De plus, les caractéristiques suivantes ont été relevées

1. comme illustré sur la figure il y a une relation évidente entre les LSP et le spectre du filtre LPC. Une concentration des LSP dans une certaine bande de fréquences correspond approximativement à une résonance dans cette bande.

2. sensibilité spectrale; un changement d'une LSP cause seulement un changement dans la forme du filtre d'analyse dans une petite gamme de fréquence autour de cette LSP[17].

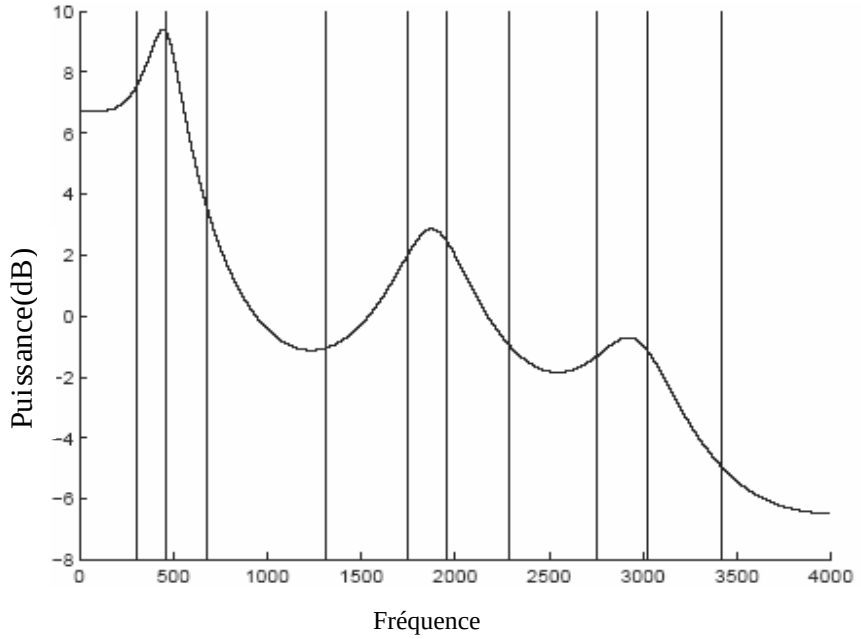


Fig. I.5 Spectre LPC avec LSF superposé [17]

I.6 La quantification

La quantification est le processus de substitution des échantillons d'un signal analogique par des valeurs arrondies prises parmi un nombre fini de valeurs possibles [4].

La quantification peut être *scalaire* ou *vectorielle* selon que les signaux sont à une ou plusieurs dimensions. La quantification vectorielle peut être de deux types soit statistique ou algébrique.

I.6.1 Quantification scalaire

Dans la quantification scalaire (*QS),* chaque échantillon du signal d'entrée est quantifié séparément des autres échantillons. Comme l'illustre la figure I.6, un échantillon x du signal d'entrée est spécifié par l'indice k s'il se trouve dans l'intervalle suivant :

$$I_k : \{x_k \ \pi \ x \le x_{k+1}\} \qquad k = 1,2,....., N \qquad \text{(I.26)}$$

Les valeurs x_k et x_{k+1} sont appelées niveaux de décision ou seuils.

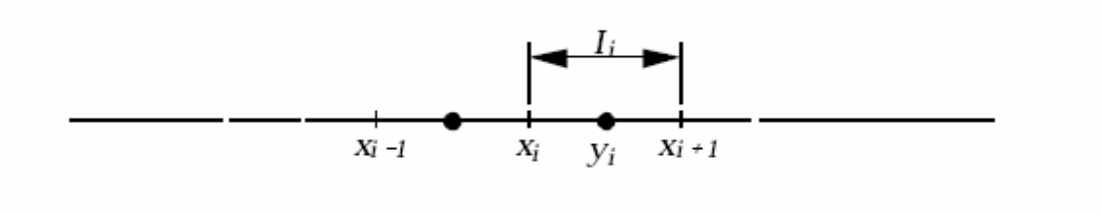


Fig. I.6 Quantification scalaire

Tous les échantillons situés dans l'intervalle I_i seront remplacés par une valeur y_i appelée *niveau de reconstruction* ou *représentant*.

I.6.2 Quantification vectorielle

La quantification vectorielle (*VQ*) est l'extension de la quantification scalaire à un espace multidimensionnel.

Nous appellerons quantificateur vectoriel de dimension *m* à *N* niveaux une application *Q* qui, à un vecteur d'entrée $x = \{x_1, x_2, ..., x_m\}$, fait correspondre une valeur approchée *y* choisie dans un ensemble fini de *N* éléments $y = \{y_i, i = 0,1,....., N-1\}$.

L'ensemble *y* est un dictionnaire de *N* représentants. En posant $R = \log_2(N)$, nous dirons que les vecteurs d'entés sont quantifiés sur *N* niveaux et codés avec *R* bits.

Contrairement à la quantification scalaire, un quantificateur vectoriel peut fonctionner avec un débit fractionnaire (R < 1) [5].

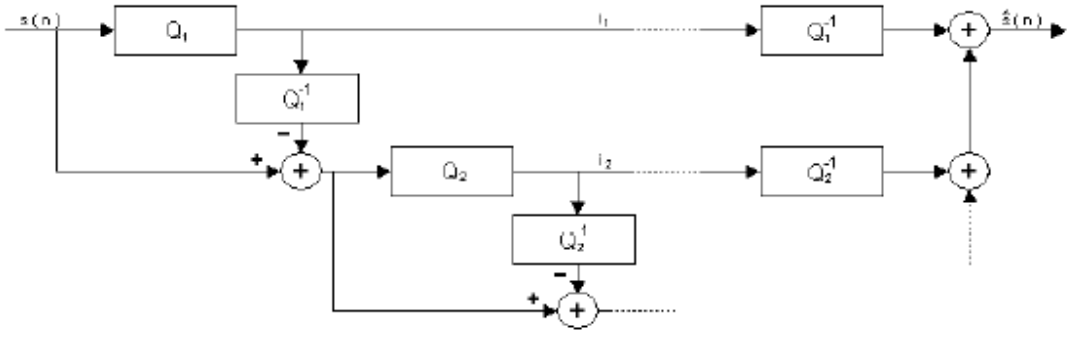


Fig. I.7 Quantification vectorielle multi étages.

I.7 Techniques de codage de la parole

Un système de codage de la parole comprend deux parties: le codeur et le décodeur (codec). Le codeur analyse le signal pour en extraire un nombre réduit de paramètres pertinents qui sont représentés par un nombre restreint de bits pour archivage ou transmission. Le décodeur utilise ces paramètres pour reconstruire un signal de parole synthétique.

Les algorithmes de codage de la parole peuvent être divisés en trois catégories [19]

- Codage de forme d'onde (waveform coding).
- Codage paramétrique (parametric coding).
- Codage hybride (hybrid coding).

La figure 1.7 montre la différence de qualité de parole qui existe entre les codecs.

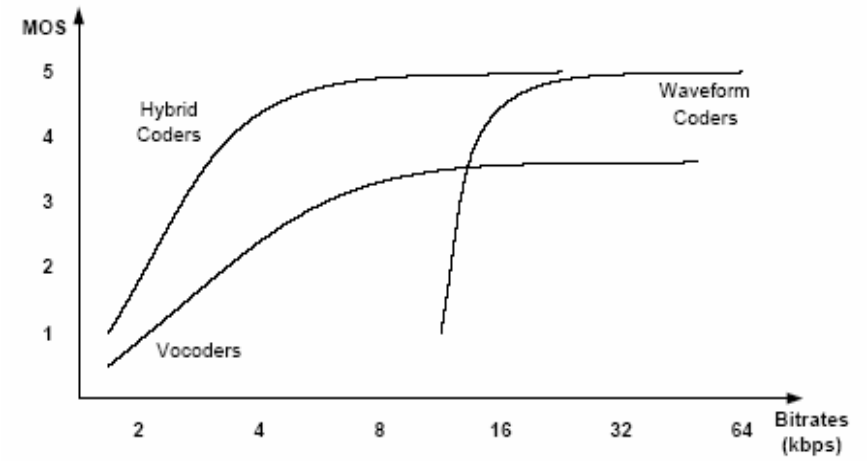


Fig. I.8 Comparaison de la qualité de codage de parole [19]

I.7.1 Le codage de forme d'onde

Les codeurs de formes d'ondes sont relativement simples à mettre en œuvre, ils produisent une qualité acceptable jusqu'à des débits de 16 Kbits/s. En deçà, la qualité du signal reconstruit se dégrade rapidement.

L'algorithme de codage le plus simple est celui qui revient seulement à échantillonner un signal analogique et à quantifier les échantillons (c'est à dire à les convertir des valeurs réelles en valeurs de précision finie) ; ce codage est appelé PCM (*Pulse Coded Modulation*).

Le codage PCM est à la base d'une famille de codages différentiels qui est basé sur l'observation que des échantillons successifs d'une source audio sont fortement corrélés. Il semble donc

judicieux d'encoder non pas les échantillons eux même mais la différence entre des échantillons successifs. On peut citer:

- Le codage DPCM (*Differential PCM*).
- Le codage ADPCM (*Adaptive Differential PCM*).
- Le codage ADM (*Adaptive Delta Modulation*).

I.7.2 Le codage paramétrique

Connu aussi sous le nom de codage de source ou vocodeurs (voice coders), ces codeurs sont destinés à fonctionner pour des bas débits et sont destinés à maintenir l'intelligibilité de la parole. La plupart de ces codeurs sont basés sur le codage linéaire prédictif LP. La performance de ce type de codage dépend du modèle de production de la parole.

Le codage LP consiste à synthétiser des échantillons à partir d'un modèle d'un système de production vocal et d'une excitation. Pour la voix humaine, le système de production vocal est l'ensemble poumons-cordes vocales -trachée -gorge -bouche -lèvres. En pratique, on modélise ce système par un ensemble de cylindres de diamètres différents, 10 dans le cas de LP-10, excités par un signal qui est soit une sinusoïde, soit un bruit blanc. Le choix de la fonction d'excitation (sinusoïde ou bruit blanc) dépend des caractéristiques, voisée ou non voisée, du signal.

I.7.3 Le Codage Hybride

La qualité des codeurs de formes d'ondes chute rapidement pour des débits inférieurs à 16 kbits/s, et comme les vocodeurs apportent une amélioration négligeable dans la qualité à des débits supérieurs à 4 kbits/s, Les codeurs hybrides sont alors utilisés pour combler ce vide, donnant ainsi une qualité de la parole à des débits moyens. Cependant, ces codeurs ont tendance à nécessiter un nombre d'opérations plus élevé. Virtuellement, tous les codeurs hybrides reposent sur l'analyse LPC pour l'obtention des paramètres du modèle de synthèse. Les techniques de formes d'ondes utilisées pour coder le signal d'excitation et les modèles de production du pitch peuvent être incorporés pour améliorer les performances.

A partir des années 80, l'intérêt pour les codeurs CELP (Code-Exited Linear Prediction) ne cesse d'augmenter. Ces codeurs sont basés sur les algorithmes de codage de la parole les plus actuellement utilisés dans la téléphonie sans fil. Dans les codeurs CELP, l'analyse LP est utilisée pour obtenir le signal d'excitation. La modélisation du pitch est utilisée pour coder efficacement le signal d'excitation. Le standard G.729 de l'ITU est un codeur CELP qui produit une qualité téléphonique (toll quality) de la parole à 8 kbits/s [5].

I.8 Mesure de qualité

L'estimation de la qualité d'un codeur est un problème complexe. Une première approche consiste à utiliser une mesure objective de la ressemblance qui existe entre le signal original et le signal reconstitué. Cette méthodologie se situe dans le domaine des tests dits "objectifs". Ils s'appliquent très bien aux codeurs de bonne qualité et font plutôt appel à la théorie du signal qu'aux connaissances sur la parole.

Lorsque l'on cherche une évaluation plus fine des codeurs, il faut faire appel à la dimension subjective de la qualité de la parole. Étant donné la part de subjectivité qui est présente dans l'appréciation d'un individu, il faut utiliser des procédures de test très élaborées. L'évaluation d'un codeur à l'aide de tests subjectifs est une opération délicate qui est généralement confiée à des laboratoires spécialisés.

I.8.1 Mesure de distorsion subjective

L'évaluation subjective est obtenue par des tests d'écoutes; dans ces tests, la qualité de la parole est mesurée par l'intelligibilité spécifiquement définie par le pourcentage de mots ou phonèmes correctement écoutés et avec une sonorité naturelle (naturalness).

Il existe trois types de mesures subjectives [4] de la qualité généralement utilisées.

- Le test DRT (Diagnostic Rhyme Test)
- Le test DAM (Diagnostic Acceptability Measure)
- Le test MOS (Mean Opinion Score)

MOS	Qualité
1	Mauvais
2	Médiocre
3	Passable
4	Bon
5	Excellant

Tableau I.1: Qualité avec la mesure MOS.

I.8.2 Mesure de distorsion objective

Le système auditif de l'être humain est l'estimateur le plus adéquat de la qualité et des performances d'un codeur de la parole. Il permet de préciser l'intelligibilité et la sonorité naturelle des sons. Bien que, Les tests d'écoute subjectifs donnent une bonne évaluation pour les codeurs de la parole, ils peuvent exiger beaucoup de temps et sont non conformé. Les mesures objectives peuvent donner une estimation immédiate de la qualité perceptuelle de la parole [16].

Les mesures objectives de distorsions peuvent être calculées aussi bien dans le domaine temporel que fréquentiel [4].

Les performances d'une mesure objective résident dans sa corrélation avec la mesure subjective correspondante (qualité ou intelligibilité).

Les mesures de distorsions sont classifiées en trois domaines [2] [4] :

- Domaine temporel (RSB et RSBseg)
- Domaine fréquentiel (distorsion spectrale)
- Domaine perceptuel (EMBSD)

I.8.2.1 Domaine temporel

- Rapport Signal sur Bruit

Si $\{S(n)\}_{n=0.Nt}$ sont les N_t échantillons du signal parole original et $\{\check{S}(n)\}_{n=0.Nt}$ sont les N_t échantillons du signal parole codé dans le *RSB* à la forme suivante :

$$RSB = 10\log\left(\frac{\sum_{n=0}^{Nt-1} S(n)^2}{\sum_{n=0}^{Nt-1}\left[S(n)-\tilde{S}(n)\right]^2}\right) \quad (dB) \tag{I.26}$$

Le *RSB* donne une valeur après avoir traiter tout le fichier, donc il n'y a pas moyen de retrouver les instants ou les divergences ont été enregistrées. De plus le *RSB* est dominé par la portion de forte énergie (tranches voisées), alors que le bruit a un effet perceptuel plus important sur les portions de faibles énergies.

- Rapport Signal sur Bruit segmenté

Le RSB_{seg} mesuré en dB, est la moyenne du *RSB* calculé sur de courts intervalles de temps du signal parole. Le RSB_{seg} calculé sur N_F trames de longueur Ns est donné par :

$$RSB_{seg} = \frac{1}{N_F}\sum_{i=0}^{N_F-1} 10\log\left(\frac{\sum_{j=0}^{Ns-1} S(N_s i+j)^2}{\sum_{j=0}^{Ns-1}\left[S(N_s i+j)-\tilde{S}(N_s i+j)\right]^2}\right) \quad (dB) \tag{I.27}$$

Le RSB_{seg} est meilleur que le *RSB*. Cependant, les tranches de silences renvoient de grandes négatives, biaisant de la sorte le résultat final. Ce problème peut être résolu en éliminant dans le calcul de la distorsion les trames de silence.

I.8.2.2 Domaine fréquentiel

La distorsion spectrale est définie comme étant la racine carrée de la moyenne au carrée des différence entre le logarithmique décimale du spectre LPC original et le logarithme décimale du spectre LPC quantifier. La définition mathématique est comme suit :

$$DS_i = \sqrt{\frac{1}{F_e}\int_0^{F_e}\left[10\,Log_{10}\frac{S_i(f)}{\tilde{S}_i(f)}\right]^2 df} \qquad (dB) \tag{I.28}$$

où F_e est la fréquence d'échantillonnage, $S_i(f)$ et $\tilde{S}_i(f)$ sont les spectres de la trame i donnés par :

$$S_i(f) = \frac{1}{A_i(e^{j2\pi f/F_e})} \tag{I.29}$$

$$\tilde{S}_i(f) = \frac{1}{\tilde{A}_i(e^{j2\pi f/F_e})} \tag{I.30}$$

où, $A_i(z)$ *et* $\tilde{A}_i(z)$ sont respectivement, les polynômes PL original et quantifié vus plus haut, pour la trame i, au lieu de l'intégration, une sommation des cœfficients obtenus après application de la TFD (transformée de Fourier Discret) aux cœfficients LPC, peut utilisée pour calculer DS_i. La distorsion devient donc :

$$DS_i = \sqrt{\frac{1}{n_1 - n_0}\sum_{k=n_0}^{n_1-1}\left[10\log\frac{S_i(e^{j2\pi k/N})}{\tilde{S}_i(e^{j2\pi k/N})}\right]^2} \qquad (dB) \tag{I.31}$$

Dans notre travail, les signaux d'entrés sont échantillonnés à F_e=8 KHz et nous avons calculé la distorsion sur une bande allant de 0 KHz à 3 KHz avec une TFD sur N+256 points. Ce qui donne $n_0 = 0$ et $n_1 = 95$. La distorsion fréquentielle est de 31.25 Hz (8000/256).

Une distorsion spectrale moyenne (la moyenne des distorsions spectrales calculées pour toutes les trames) de 1 dB est habituellement acceptée. Cependant, selon ***Atal*** et ***Paliwal*** les conditions de transparence spectrale (pas de distorsion audible) établies expérimentalement sont les suivantes :

- La moyenne DS inférieur à 1dB
- Le nombre de trames ayant DSi dans l'intervalle 2-4 dB est inférieur a 2%
- Pas de trames ayant DSi supérieur a 4 dB

I.8.2.3 Mesure de distance euclidienne LSP pondérée

Cette distance a été développée dans le but d'optimiser le quantification des paramètres LP, elle a la forme suivante :

$$d_{LSF} = \sum_{i=1}^{p} \left[c_i w_i (\omega_i - \tilde{\omega}_i) \right]^2 \tag{I.32}$$

Ou c_i et w_i sont les poids du i^{eme} coefficients LSP ω_i, et p est l'ordre du filtre LP. Pour un filtre d'ordre 10, les poids fixes c_i sont donnés par :

$$c_i = \begin{cases} 1.0 & pour \quad 1 \le i \le 8 \\ 0.8 & pour \quad i = 9 \\ 0.4 & pour \quad i = 10 \end{cases} \tag{I.33}$$

Ces poids sont utilisés pour donner plus d'importance aux basses fréquences par rapport aux hautes fréquences. Ceci est justifié par le fait que l'oreille humaine est plus sensible aux basses fréquences qu'aux hautes fréquences. Les poids adaptatifs ω_i sont utilises pour accentuer les régions de l'enveloppe spectrale $S(e^{j\omega})$ à forte énergie (formants). Ces poids sont données par :

$$w_i = \left[S(e^{j\omega}) \right]^r \tag{I.34}$$

Ou, r est une constante empirique qui contrôle le degré de la pondération, empiriquement r=0.15. Une pondération plus simple a été proposée par [18], elle a la forme suivante :

$$w_i = \frac{1}{\omega_i - \omega_{i-1}} + \frac{1}{\omega_{i+1} - \omega_i} \qquad ou \quad \omega_0 = 0 \, et \quad \omega_{p+1} = \pi \tag{I.35}$$

Les mesures dans le domaine perceptuel sont basées sur les modèles d'audition humaine. Le signal est transformé vers un domaine adéquat de telle manière qu'on puisse exploiter effets de masquage psycho-acoustique. Parmi les mesures perceptuelles les plus utilises nous pouvons

citer : Perceptuel Evaluation of Speech Quality (*PESQ*) et Enhanced Modified Bark Spectrum Distorsion (*EMBSD*).

L'EMBSD estime la distorsion perceptuel d'un signal en le comparant au signal original dans le domaine des sons forts (loudness domain) tout en tenant compte du seuil de masquage de bruit modifié et du modèle cognitif basé su le post-masquage.

Conclusion

La prédiction linéaire exploite la redondance dans le signal parole et extrait des coefficients (paramètres LPC) qui caractérisent le comportement du signal. La simplicité de son concept, la linéarité dans la résolution des systèmes et ses performances dans le codage de la parole, la rendent la plus admise et la plus largement utilisée dans le codage du signal de parole.

Chapitre II

Codage de la parole à large bande

Dans ce chapitre on va étudier le codage de signaux large bande. La transmission en large bande correspond à l'élargissement de la bande passante utilisée pour la transmission du signal de parole. En effet, la bande passante utilisée habituellement en téléphonie est 300-3400 Hz, elle

défini le débit de base d'une ligne téléphonique qui est de 64 kbits/s. Cependant les nouvelles technologies liées aux réseaux permettent une utilisation plus flexible de la transmission de la parole, grâce à un choix assez large de codeur et de sa bande passante, qui a facilité l'apparition d'une nouvelle bande passante 50-7000 Hz, améliorant la qualité du signal de parole transmis.

Le développement des applications multimédia sur l'Internet ainsi que les systèmes de conférence téléphonique feraient bon usage d'un système adaptatif permettant de régler le niveau de qualité du codage selon le débit disponible. Cette étude propose une solution destinée à répondre à ce besoin.

II.1 Intérêts de l'évolution vers le codage large bande

La téléphonie peut être décrite comme un système de communication permettant de transmettre de la parole. Elle peut, en premier lieu être vue comme un système créant une liaison entre deux personnes et permettant la transmission d'un message. Son but (comme celui de tout système comportant de la parole) devrais être de maximiser la compréhension de ce message. Le message étant défini comme un signal de parole transmis de la bouche d'un locuteur jusqu'à l'oreille d'un auditeur, le rôle des deux personnes changeant au cours de la conversation.
Ainsi, la compréhension du message délivré par un locuteur dépend de : [2], [3]

- La compréhensibilité du message, liée directement au locuteur ou au système et à sa capacité à donner une information, de transmettre les phonèmes (souvent en fonction de son articulation, tout dépendant du contexte de locution).
- L'intelligibilité, qui correspond à la possibilité d'établir un sens au message transmis avec l'ensemble des phonèmes du message.
- La communicabilité, qui est la compréhension de l'ensemble des messages, dans les deux sens de la liaison.

Il est important de remarquer que, la compréhension du message dépend du locuteur, du contexte de locution, ainsi que des connaissances du sujet sur le message.

L'étude de la large bande va principalement porter sur le premier facteur. Or, pour connaître en quoi l'augmentation de la bande passante permet d'améliorer la qualité globale de la téléphonie,

il est nécessaire d'étudier le message vocal transmis. Le prochain paragraphe porte donc sur l'étude de la parole : sa production par un locuteur et sa perception par un auditeur.

II.1.1 La production vocale

Tout d'abord, un son a une forme physique qui se propage dans un milieu par le biais d'ondes. Ces ondes sont liées au canal de transmission (air, câble) mais surtout au producteur de ce son. Elles peuvent alors être quantifiées sur une échelle de fréquences. Le signal de la parole a donc des caractéristiques temporelles mais aussi fréquentielles. Dans le cas de la parole humaine, les fréquences dépendent de la forme et de la position de certains organes du corps humain. La parole peut alors être vue comme un signal source (corde vocales, glotte) qui est filtré par des tuyaux formés par les conduits vocaux (comme le conduit nasal). La manière dont le son d'origine est filtré dépend de la signification que veut lui donner le locuteur. Ainsi on peut voir que la production vocale est composée de sons ayant des composantes fréquentielles très spécifiques, dont on donne souvent comme valeur la fréquence fondamentale (f_0, correspondant au signal porteur), et les premiers formants (f_i, piques dans l'amplitude spectrale dus aux résonances du conduit vocal). Certains phonèmes comme les voyelles se caractérisent très facilement par ses formants. Les consonnes sont produites de manières différentes [4], elles peuvent être sonores (« l », « r »), nasales (« m », « n ») ou fricatives (« h », « f »). Ces dernières, comme le « s » ou le « f », produisent de l'énergie essentiellement dans les hautes fréquences ainsi que dans un formant très bas, autour de 150 Hz. De même la résonance nasale se situe au alentour de 250 Hz.

Les fréquences utilisées par la parole humaine, peuvent donc être comprises entre 110 et 7 kHz (speech communications). La bande étroite 300-3400 Hz utilisée par la téléphonie permet de faire passer les 3 premiers formants, et ainsi de garantir une intelligibilité de la parole du locuteur, mais ne permet pas de transmettre l'intégralité des fréquences présentes dans un signal de parole. Par exemple, il est très difficile de différencier un « s » d'un « f » prononcé seul, lors d'une conversation téléphonique.

II.2 La perception de la parole

Il est nécessaire de rappeler que les fréquences audibles par une oreille humaine sont habituellement situées entre 20 Hz et 20 kHz : ce qui à première vu semble très loin de la bande

étroite de la téléphonie. De plus, suite à de nombreuses utilisations, le cerveau humain, a créé une référence de la qualité « sonore » de la voix humaine, transmise à travers un système de téléphonie [5]. L'évaluation de la qualité est alors biaisée par la référence de la téléphonie fixe, fortement liée aux fréquences de la bande étroite.

Afin de mieux appréhender le choix de la bande élargie pour la téléphonie, le paragraphe suivant rappelle la notion de « bande critique » ;

Les bandes critiques correspondent à une répartition dans le spectre des fréquences d'un ensemble de bandes de fréquences. Ces bandes sont des regroupements des excitations sonores ayant des fréquences voisines et perceptivement proches au sein de certaines bandes fréquentielles. Il est possible de passer de l'échelle des fréquences à celle des bandes critiques grâce à la fonction suivante [8] :

$$z_{(barks)} = 13.\tan^{-1}.(0{,}76.f_{(khz)}) + 3{,}5.\tan^{-1}(f_{(khz)}/7{,}5)^2 \tag{2.1}$$

Cette échelle en bande critique est une échelle perceptive (correspondant à la Tonie), dont l'unité, le « Bark » regroupe un ensemble variable de fréquences, 1 Bark correspondant à une bande passante de 100 Hz à 3500 Hz. Une bande passante peut alors être obtenue en Barks :

$$z_{bp} = z(f_h) - z(f_b) \tag{2.2}$$

Les fréquences audibles vont de 0 (20 Hz) jusqu'à 24 Barks (16 kHz). La téléphonie bande étroite représente 14 Barks (de 3 à 16 Barks), soit plus de la moitié de l'échelle audible.

L'utilisation des bandes critiques permet de connaître la valeur perceptive d'une bande passante. Par exemple, dans [7] une bande passante de 180-2800 Hz est considérée de même qualité qu'une seconde bande passante de 280-3550 Hz. Un simple calcul suffit pour voir que la différence en Barks de ces deux bandes passantes est 0, car les deux fréquences de coupures subissent une simple translation d'environ 1 Bark.

Cette échelle permet alors d'analyser le choix de la bande étroite en téléphonie classique et l'apport du large bande. La fréquence de coupure basse, baisse de 300 Hz jusqu'à 50 Hz, soit une augmentation de la bande passante de 2,5 Barks, tandis que la fréquence de coupure haute, permet une augmentation de 4 Barks.

La fréquence centrale, permet également de connaître le poids de fréquences basses et hautes dans les deux bandes passantes.

$$f_c = \sqrt{f_b . f_h} \,. \tag{2.3}$$

Pour la bande étroite : $f_C = 1010\,\text{Hz}$, et pour la bande élargie : $f_C = 590\,\text{Hz}$.

Ces deux valeurs montrent que la bande élargie comporte perceptivement plus de basse fréquence que la bande étroite. Cette description des bandes passantes par le biais de la fréquence centrale et de la largeur spectrale en Barks est utilisée dans [1], afin de lier la perception fréquentielle à la qualité d'un signal de parole.

Par ailleurs, il a été vu précédemment que l'intelligibilité était fortement liée aux premiers formants. Dans [7] il est montré que la sensation naturelle de la voix dépend fortement du premier formant et l'intelligibilité du second formant. En effet, les différents phonèmes sont perçus en fonction :

- Du rapport entre les formants.
- De leurs variations dans le temps.

De plus, le premier formant étant approximativement entre 270 et 730 Hz pour les hommes, et entre 310 et 850Hz pour les femmes, le choix de la fréquence de coupure basse à 300 Hz a été choisi judicieusement. La bande étroite permet un bon compromis entre intelligibilité et qualité du son.

Pour autant, celle-ci ne permet pas de transmettre la fréquence fondamentale f_0, qui semble être liée à la sensation naturelle de la voix. Celle-ci comprise entre 110 et 200 Hz pour un adulte, et montant jusque 300 Hz pour un enfant, permet de transmettre la prosodie, comme les intonations ou les émotions.

Théoriquement, la perception humaine permet de reconstruire cette fréquence fondamentale en son absence, et de percevoir tout de même les différentes intonations exprimées par le locuteur. Malgré cela, dans [7] on voit également que la perception de la parole à travers un système de téléphonie dépend énormément de la présence des fréquences basses ; Une petite différence de 45 Hz, de 225 à 180 Hz, sur la fréquence de coupure basse améliore nettement la sensation naturelle

de la voix. La perception de la fréquence fondamentale semble donc avoir une importance dans l'évaluation de la qualité de la parole.

Enfin, lors de l'augmentation de la fréquence de coupure basse de 123 à 208 Hz, une dégradation est ressentie sur la perception de la voix humaine, celle-ci semble moins naturelle. Une dégradation est obtenue également lors d'une diminution de la fréquence de coupure de 5500 à 3500 Hz. Mais il montre également qu'il est nécessaire d'améliorer la bande étroite aux deux extrémités. En effet, il montre que pour une fréquence de coupure basse proche de 300 Hz, le changement de la fréquence de coupure haute (de 7000 à 3500 Hz) n'a que peu d'effets. De même, pour une fréquence de coupure haute à 3500 Hz, un changement de la fréquence basse de 55 à 300 Hz à peu d'effets également. Il est donc impossible de compenser une coupure trop forte, en haut ou en bas spectre, en agrandissant l'autre coté.

La bande étroite introduit donc une dégradation de la sensation naturelle de la voix par :

- L'atténuation du premier formant.
- L'absence de transmission de f_0.
- L'absence de transmission des hautes fréquences.

La parole humaine produit des fréquences qui en partie ne sont pas comprises dans la bande étroite, et qui sont nécessaires pour obtenir une voix humaine naturelle. La téléphonie large bande, qui permet la transmission de la majorité des fréquences produites par la voix, nous espérons que celle-ci permettra de rendre la voix d'un interlocuteur plus naturelle.

L'oreille ne peut percevoir que certains sons. La figure-1- donne une représentation du domaine audible pour un être humain. On remarque tout d'abord que le niveau de perception dépend grandement de la plage de fréquences considérée ainsi que du niveau sonore. On définit alors deux courbes dans le plan fréquence/intensité : un seuil d'audibilité et un seuil de confort. La zone ainsi définie est le domaine dans lequel les sons peuvent être perçus. Tout signal en dehors de cette plage est inaudible, gênant ou même dangereux [2].

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

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

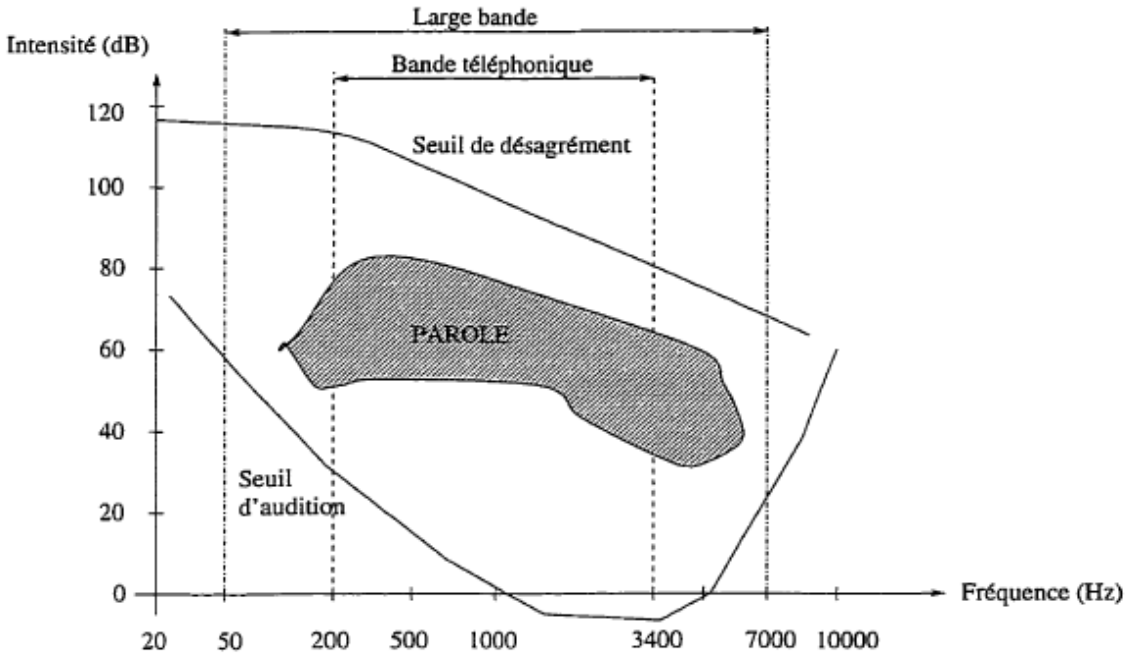


Fig.II.1 perception auditive

II.3 Codage large bande

Le codage large bande considère théoriquement aussi bien les signaux de parole que les signaux de musique. Les techniques de codage peuvent alors différer selon le type d'application du codeur. On sait que le premier système large bande est apparu en 1988, ce type de transmission est aujourd'hui très peu répandu. En effet, il a fallu une suite d'évolution dans les télécommunications pour permettre l'utilisation de CoDec large bande dans l'industrie.

II.3.1 Codage de la parole large bande

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

- Pour les phonèmes voisés, l'addition de la bande de fréquences 50 Hz - 140 Hz donne une meilleure représentation des premières harmoniques. On remarque surtout cela pour un locuteur masculin pour lequel la fréquence de pitch est assez faible. D'une manière plus générale les basses fréquences procurent une sensation de confort et un sentiment de parler « face à face ».
- L'apport des hautes fréquences, supérieures à 3400 Hz, n'a de l'importance que pour les phonèmes plus complexes tels que les fricatives non voisées (ex: « S », « CH », « F »), les fricatives voisées (ex: « Z ») ou encore Les plosives (ex: « T », « D »).

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

II.3.2 Codage de la musique en large bande

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

Comme il vient d'être précisé, il est difficile de prévoir à l'avance l'allure de l'enveloppe spectrale. On peut cependant admettre que la musique est une combinaison de bruit et de ''tons''. Un ton pur est une concentration d'énergie sur une raie spectrale donnée, avec un plancher de bruit faible. A titre d'exemple, une note de musique isolée conduit à un spectre comprenant seulement un ton pur, localisé à la fréquence de la note.

II.3.3 Les codeurs large bande

Idéalement un codeur large bande doit pouvoir traiter sans préférences aussi bien les sources de parole que celles de musique. Deux approches sont alors possibles: soit on améliore un codeur de parole (type ACELP par exemple) pour qu'il traite au mieux la musique, soit on part d'un codeur mieux conçu pour la musique (codage par transformée type TCX par exemple) que 1'on adapte afin de mieux coder la parole. La seconde stratégie semble plus prometteuse dans la mesure où un codeur ACELP est conçu presque exclusivement pour la parole (pour la musique, le débit consacré au pitch est parfois du gaspillage) tandis qu'un codeur par transformée assure toujours une contribution minimale quelle que soit la source considérée.

Il n'existe pas encore beaucoup de normes en large bande pour le moment. La référence à considérer est encore la norme UIT G.722. Ce codeur est un codeur de forme d'onde temporelle de type ADPCM. Il utilise un débit de 45 kbits/s à 64 kbits/s. Une seconde norme devrait bientôt pouvoir remplacer ce dernier. Les nouveaux débits à considérer seront probablement 16 kbits/s à 32 kbits/s. La technique de codage utilisée est cette fois un codage par transformée MLT avec un codage entropique sur les indices de quantification en bout de ligne (un peu comme MPEGl layer 3). Le tableau-1- donne une brève description des deux codeurs large bande qui viennent d'être évoqués.

Codeur	G.722	Nouveau large bande G.7XX
Année Débit Délai	1988 3 modes :48 Kb/s, 56 Kb/s et 64 Kb/s 0,125 ms (+ 1,5 ms lookahead)	1998 3 modes :16 Kb/s, 24 Kb/s et 32 Kb/s 20 ms (+ 20 ms lookahead)
Modèle	• Codage en deux sous-bandes (QMF) • Codage ADPCM d'ordre 4dans chacune des bandes • Hautes fréquences quantifiés à 2 bits par échantillon • Basses fréquences quantifiées à 3 bits, 4 bits ou 5 bits par échantillon selon le debit choisi, les trois modes étant encapsulés	• Codage par transformée MTL • Quantification scalaire des raies de la transformée • Attribution du budget par catégorisation selon les bandes de fréquence • Codage entropique (huffman) sur les indices de quantification

Tableau II.1 Codeurs large bande

Conclusion

Le codage large bande n'est pas un domaine d'étude très récent, le premier CoDec utilisant une bande passante élargie, le ITU G.722 , fut développé dans les années 1980 pour être utilisé sur le réseau RNIS. Mais le codage de la parole en large bande entraîne des techniques différentes de la bande étroite. En effet, il y a une plus forte dynamique spectrale pour la parole en bande large. De plus la voix est plus inharmonique dans les hautes fréquences, comme pour les fricatives, en raison des caractéristiques morphologiques. Mais, plusieurs études, efforts de développement et standardisation ont permis de créer quelques CoDecs de meilleure qualité et moins coûteux en débit, dans la suite de notre travail, on va s'intéresser aux codeurs utilisant le principe du codage de l'enveloppe spectrale, c'est ce qui va être abordé dans le prochain paragraphe.

ChapitreIII

Résultats et simulations

L'objectif de nos tests et simulation est d'étudier le codage large bande. Ces codeurs, du type d'analyse-par-synthèse, doivent transmettre des signaux possédant une gamme de fréquences limitée à 8 kHz et ayant été échantillonnés à 16 kHz. Pour cela on va procéder de la manière suivante :

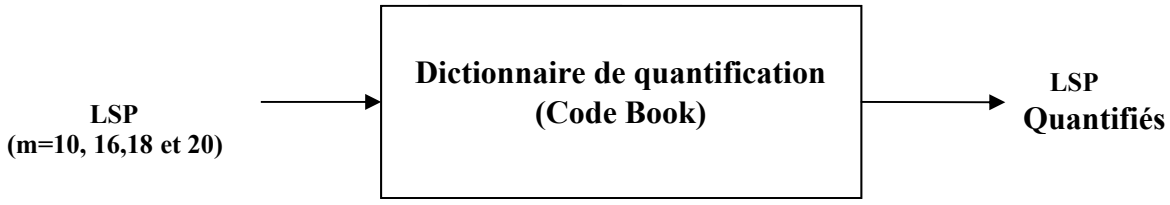
- En premier lieu, on code l'enveloppe spectrale d'un signal possédant une large bande. L'étude se fait à l'aide des Lignes de Prédiction Spectrales (LSP), qui sont obtenus par l'analyse LPC :

  - Extraction des coefficients $\{a_k\}$.
  - Conversion des $\{a_k\}$ en coefficients LSP (Lineare Spectral Prediction)

L'extraction des LSP se fait pour un ordre de prédiction (m) variable entre 10 et 20 pôles (m=10 pôles, m=16 pôles, m=18 pôles, m=20 pôles).

- Application de la méthode LBG sur les Lignes de Prédiction Spectrales (LSP), afin d'obtenir le dictionnaire de quantification (Code- Book) pour m=10, 16, 18, et 20 pôles.

- Faire passer les coefficients LSP par le dictionnaire de quantification, pour extraire les LSP quantifiés pour m=10, 16, 18, et 20.



- Calculer la distorsions spectrale entre les LSP et les LSP quantifiés en fonction du nombre de bite qui vari entre (40bits-80 bits).

- Pour la partie programmation on utilisé le langage C (builder C++ 6.0), et le Matlab pour les représentations des graphes.

III.1 Conditions d'analyse

Le signal utilisé dans notre simulation est signal parole qu'on échantillonnera avec une

fréquence f_e=16 KHz, afin de satisfaire la condition de Schanon ($f_e \geq 2BP$).

Le signal échantillonné ainsi obtenu sera découpé en trames de 80 échantillons, dans notre exemple on aura 1430 trames.

Le signal parole utilisé dans notre simulation est schématisé à l'aide de Matlab comme suit :

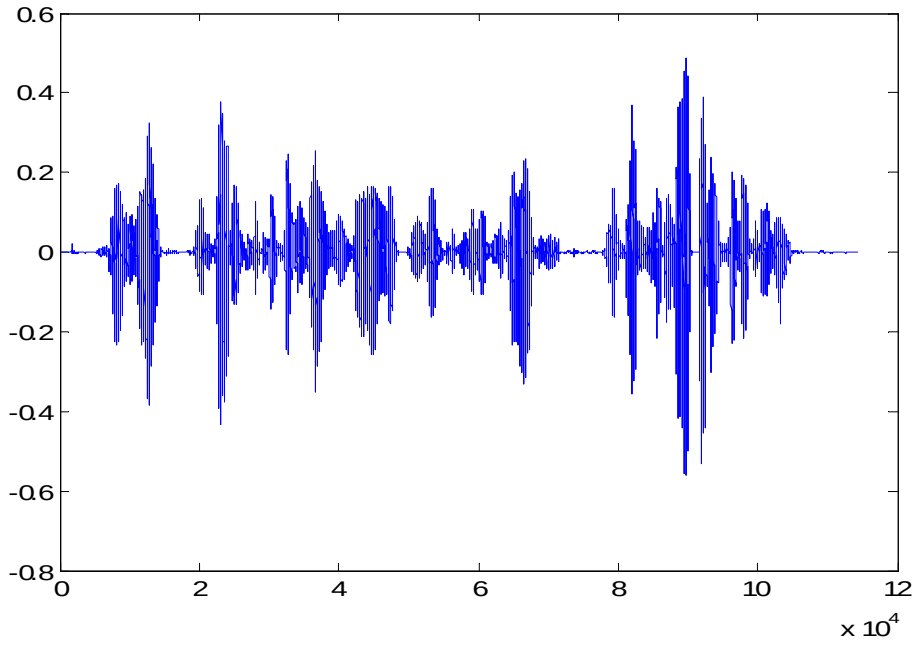


Fig.III.1 Signal parole utilisé.

III.2 L'analyse LPC

Apres avoir échantillonnée le signal parole, ce dernier va être sujet de l'analyse LPC (comme vu dans le chapitre 1).afin d'extraire les coefficients (a_i) et cela pour un ordre de prédiction de 10, 16,18et20 pôles.

Le tableau suivant donne les valeurs des (a_i) pour les deux première trame de notre signal parole déjà échantillonné dans la première partie de notre simulation.

M=10 pôles	Trame 1	Trame 2
a1	1.000000	1.000000
a2	-0.287399	-0.423111
a3	-0.240764	-0.315719
a4	0.233074	0.451945
a5	0.123018	0.028644
a6	-0.024625	-0.232478
a7	0.018431	0.193160
a8	-0.017185	0.058880
a9	0.109583	-0.097218
a10	-0.031213	0.055709

Tableau III.1 Valeurs des (a_i) pour les deux premières trames pour m=10 pôles.

M=16 pôles	Trame 1	Trame 2
a1	1.000000	1.000000
a2	0.024929	0.141256
a3	-0.235596	-0.246808
a4	-0.265476	-0.372025
a5	-0.011110	-0.184260
a6	-0.125437	-0.327174
a7	-0.205724	-0.332292
a8	0.155493	0.294994
a9	0.092839	0.218936
a10	-0.127162	-0.021126
a11	0.200685	0.278801
a12	0.149240	0.287703
a13	0.014280	0.026931
a14	-0.181261	-0.282291
a15	-0.007304	-0.072346
a16	0.043103	-0.019470

Tableau III.2 Valeurs des (a_i) pour les deux premières trames pour m=16 pôles.

M=18 pôles	Trame 1	Trame2
a1	1.000000	1.000000
a2	0.024799	0.145503
a3	-0.243996	-0.294186
a4	-0.260993	-0.382204
a5	-0.011367	-0.218635
a6	-0.142888	-0.414753
a7	-0.204687	-0.317894
a8	0.169360	0.389123
a9	0.112353	0.306492
a10	-0.139902	-0.019424
a11	0.209183	0.354939
a12	0.164631	0.370822
a13	-0.005106	-0.082710
a14	-0.193058	-0.387982
a15	-0.007475	-0.142877
a16	0.018111	-0.142638
a17	-0.111816	-0.228903
a18	-0.000542	0.072744

TableauIII.3 Valeurs des (a_i) pour les deux premières trames pour m=18 pôles.

M=20 pôles	Trame 1	Trame 2
a1	1.000000	1.000000
a2	-0.038247	-0.020277
a3	-0.259045	-0.299815
a4	-0.257825	-0.343385
a5	-0.066038	-0.004006
a6	-0.013869	-0.093052
a7	-0.172994	-0.200917
a8	0.071197	0.234078
a9	0.134452	0.189854
a10	-0.079627	-0.065573
a11	0.110242	0.061451
a12	0.043871	0.101066
a13	0.011563	0.003676
a14	-0.130037	-0.222242
a15	-0.017304	-0.033874
a16	0.054208	0.060283
a17	-0.029663	-0.040396
a18	-0.022421	-0.004453
a19	0.032027	0.064833
a20	0.017209	0.061762

Tableau III.4 Valeurs des (a_i) pour les deux premières trames pour m=20 pôles.

III.2.1 Représentation des LSP

Afin d'obtenir une précision optimale, on utilise les coefficients LSP (les paires de raies spectrales) qui est la représentation la plus répondue des coefficients (a_i), afin de stabiliser le filtre de synthèse LP.

Pour plus de détails sur la détermination des coefficients LPS, Veuillez voir *Annexe A.*

Dans les tableaux suivants nous avons donné les valeurs des LSP des deux premières trames de notre signal pour m=10, 16,18et 20 pôles:

M=10poles	Trame 1	Trame 2
LSP1	0.949780	0.954089
LSP2	0.865249	0.874741
LSP3	0.701371	0.713419
LSP4	0.525716	0.542672
LSP5	0.247921	0.273811
LSP6	-0.088602	0.013174
LSP7	-0.418647	-0.402625

LSP8	-0.663515	-0.715121
LSP9	-0.875379	-0.867702
LSP10	-0.956497	-0.963347

Tableau III.5 Valeurs des LSP pour les deux premières trames pour m=10 pôles.

M=16poles	Trame 1	Trame 2
LSP1	0.991562	0.991949
LSP2	0.959491	0.965410
LSP3	0.868061	0.882498
LSP4	0.776024	0.798075
LSP5	0.702251	0.733351
LSP6	0.421479	0.563979
LSP7	0.258875	0.259223
LSP8	0.152963	0.229245
LSP9	-0.012120	0.009498
LSP10	-0.204207	-0.030638
LSP11	-0.391750	-0.358140
LSP12	-0.489356	-0.374369
LSP13	-0.756081	-0.781249
LSP14	-0.856430	-0.797229
LSP15	-0.933906	-0.961217
LSP16	-0.988969	-0.997561

Tableau III.6 Valeurs des LSP pour les deux premières trames pour m=16 pôles.

M=18 pôles	Trame 1	Trame 2
LSP1	0.991148	0.993572
LSP2	0.963702	0.969338
LSP3	0.921106	0.944805
LSP4	0.814724	0.814724
LSP5	0.665668	0.625122
LSP6	0.566103	0.557409
LSP7	0.400089	0.337803
LSP8	0.228094	0.225597
LSP9	0.093596	0.094783
LSP10	-0.042830	-0.001938
LSP11	-0.348309	-0.409156
LSP12	-0.460598	-0.474432
LSP13	-0.567045	-0.590027
LSP14	-0.660782	-0.667544
LSP15	-0.768034	-0.770416
LSP16	-0.890212	-0.885984
LSP17	-0.948920	-0.953699
LSP18	-0.982302	-0.985065

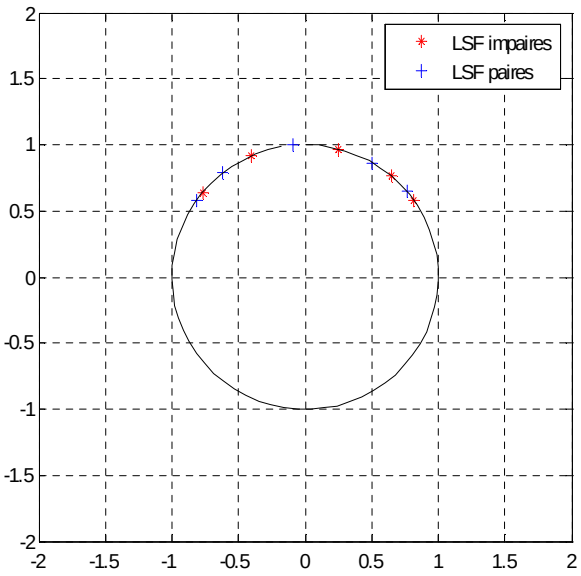
.

Tableau III.7 Valeurs des LSP pour les deux premières trames pour m=18pôles

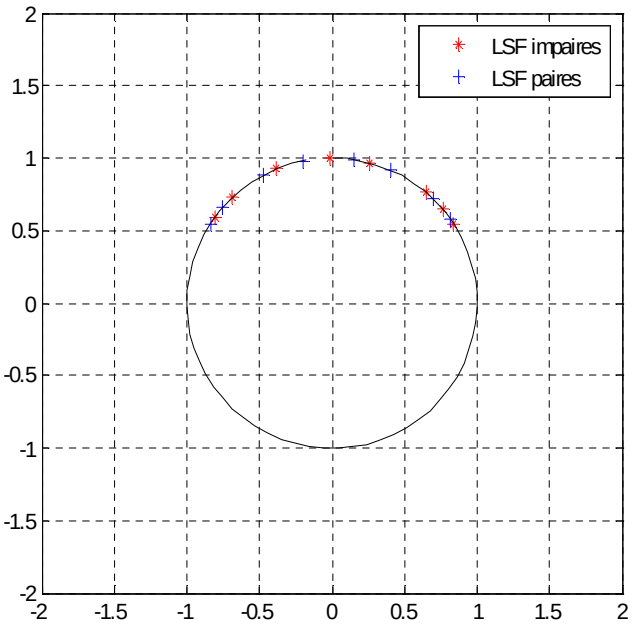
M=20 pôles	Trame 1	Trame 2
LSP1	0.991672	0.992739
LSP2	0.967090	0.971297
LSP3	0.926047	0.934127
LSP4	0.861874	0.879951
LSP5	0.746488	0.750545
LSP6	0.633356	0.624512
LSP7	0.522155	0.508934
LSP8	0.385774	0.363716
LSP9	0.221405	0.218464
LSP10	0.085516	0.097043
LSP11	-0.078037	-0.061637
LSP12	-0.290370	-0.313980
LSP13	-0.415835	-0.423169
LSP14	-0.518835	-0.526835
LSP15	-0.618321	0.620907
LSP16	-0.709264	-0.711167
LSP17	-0.823632	- 0.810908
LSP18	-0.908868	-0.907895
LSP19	-0.954786	-0.956996
LSP20	-0.985187	-0.987562

Tableau III.8 Valeurs des LSP pour les deux premières trames pour m=20pôles.

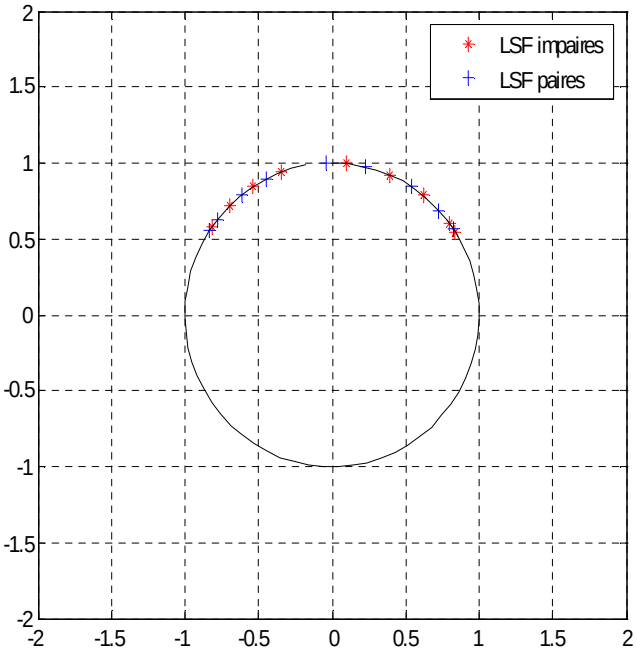
Cependant, comme vu dans le paragraphe II.1, tous les coefficients LSF se trouvent sur le cercle unité et sont entrelacés, ce qui limite notre travaille à coder les phases seulement.



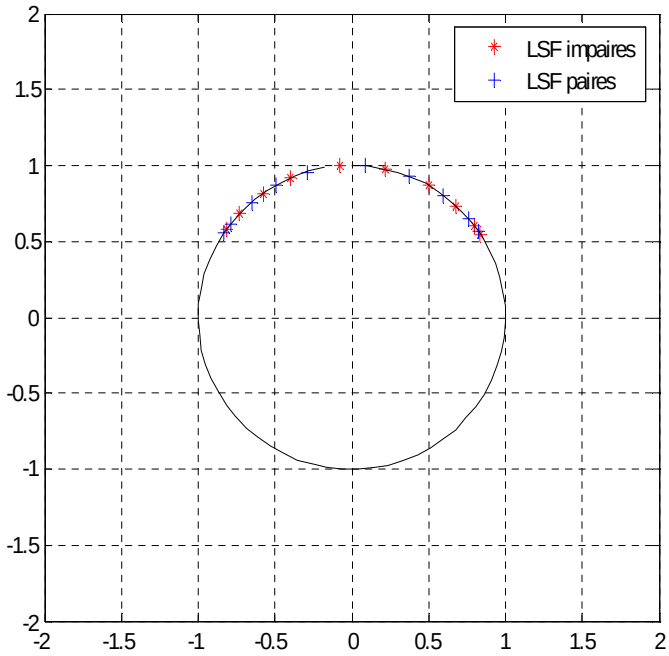
(a) m=10 pôles



(b) m=16 pôles



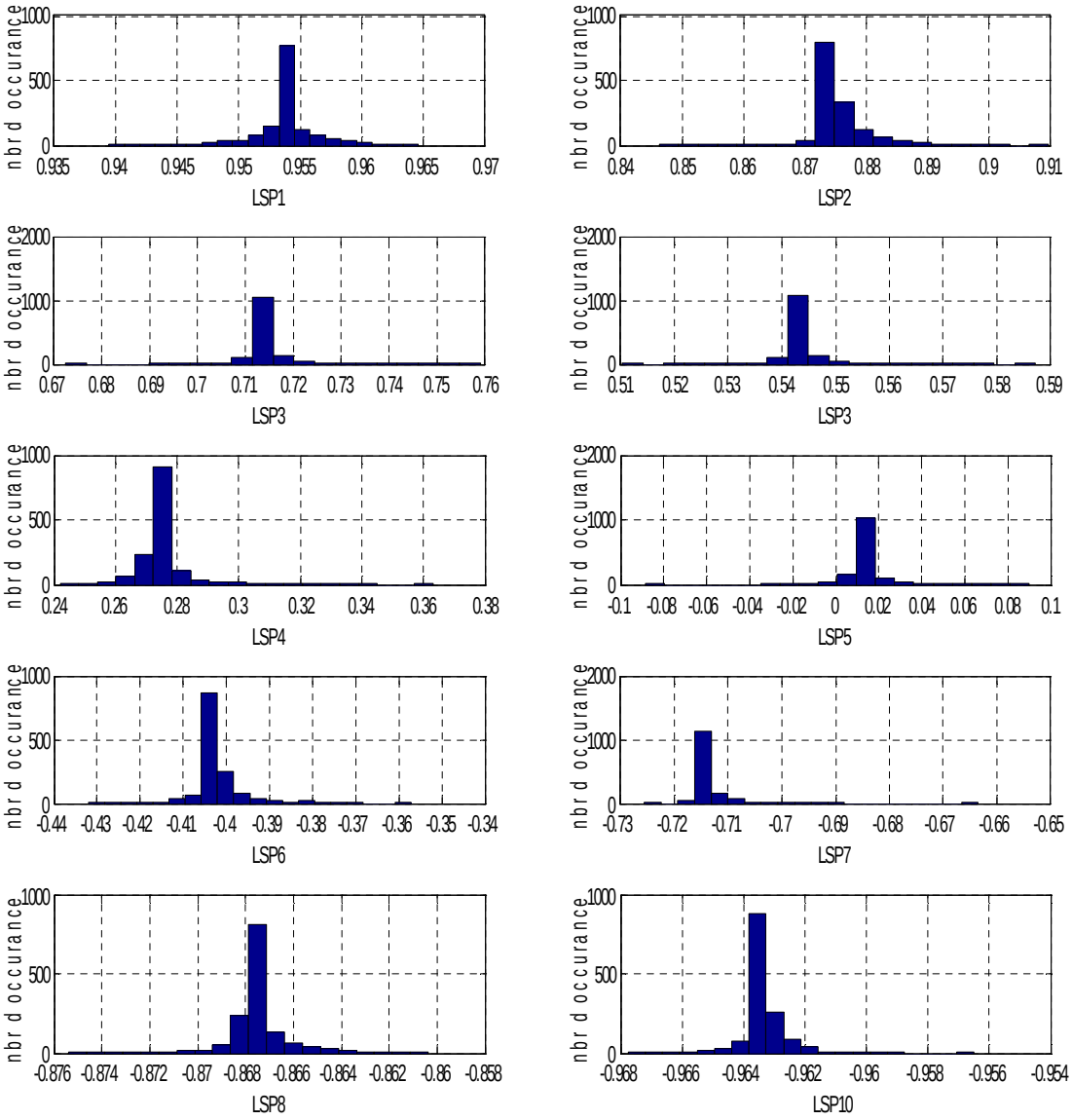
(c) m=18 pôles



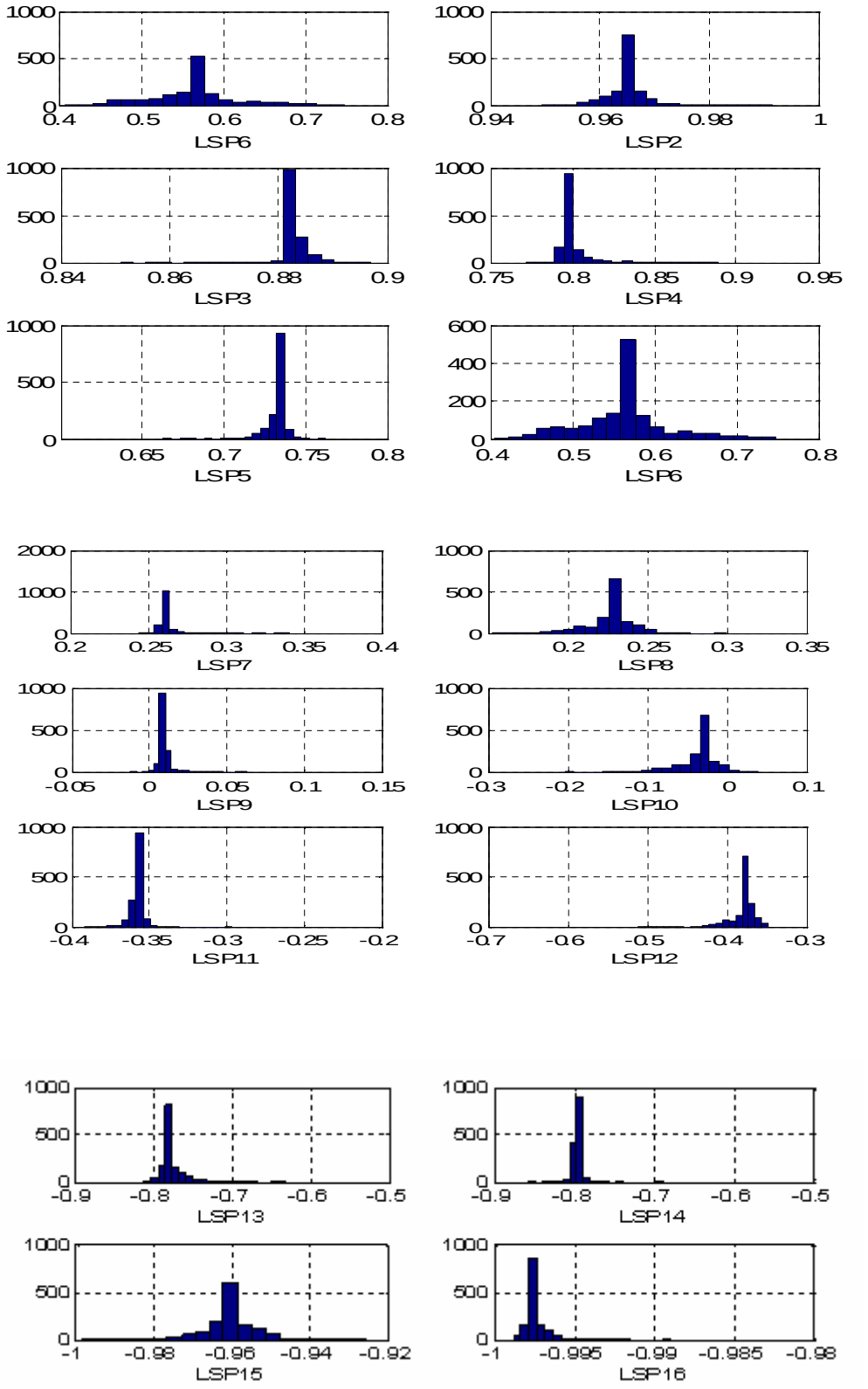
(d) m=20 pôles

Fig.III.2 Représentation des LSF de la première trame sur le cercle unité pour 10, 16, 18,20 pôle.

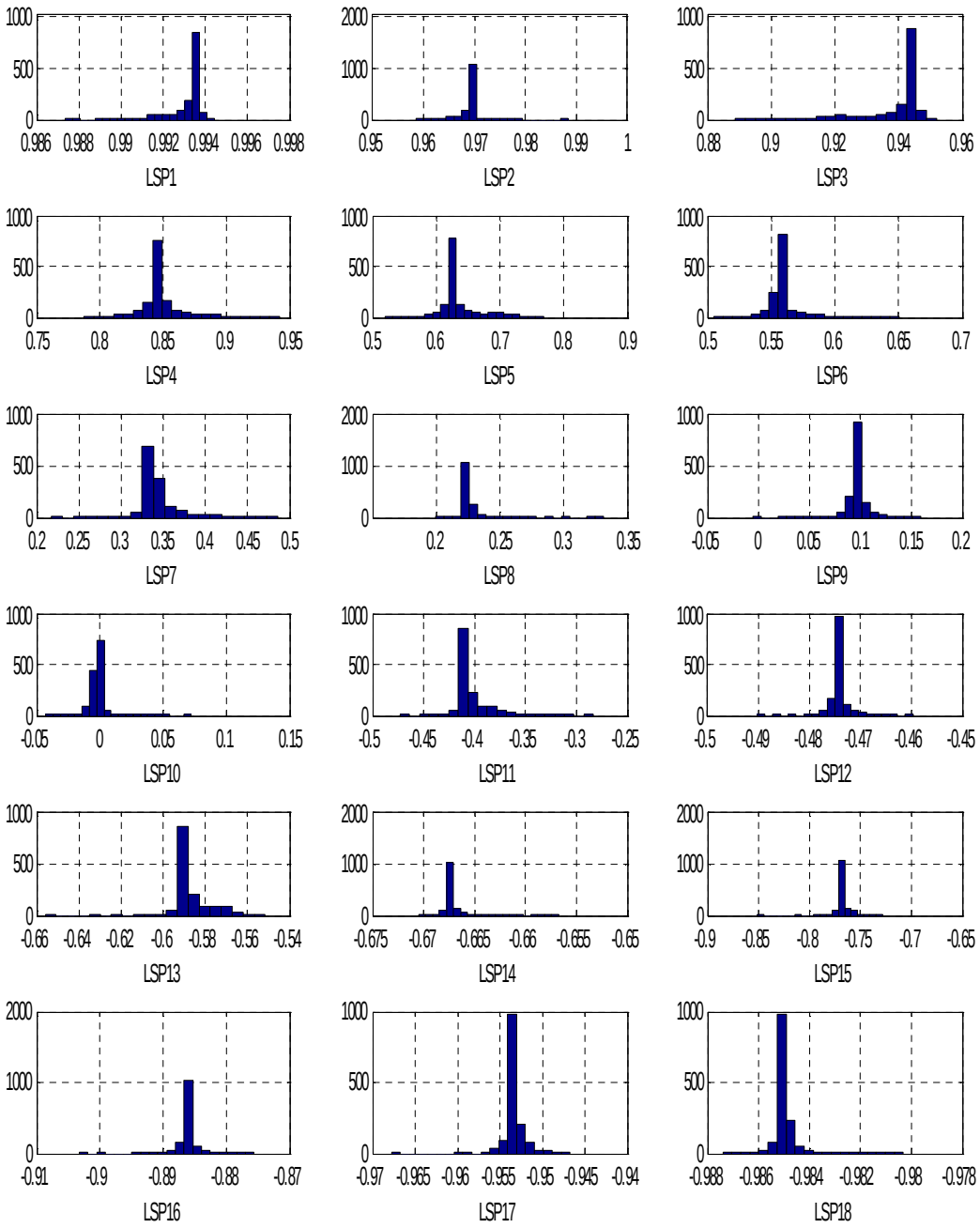
Il y'a aussi une autre de représentation des LSP, a l'aide des histogrammes qui sont représentés dans la figures III.3 :



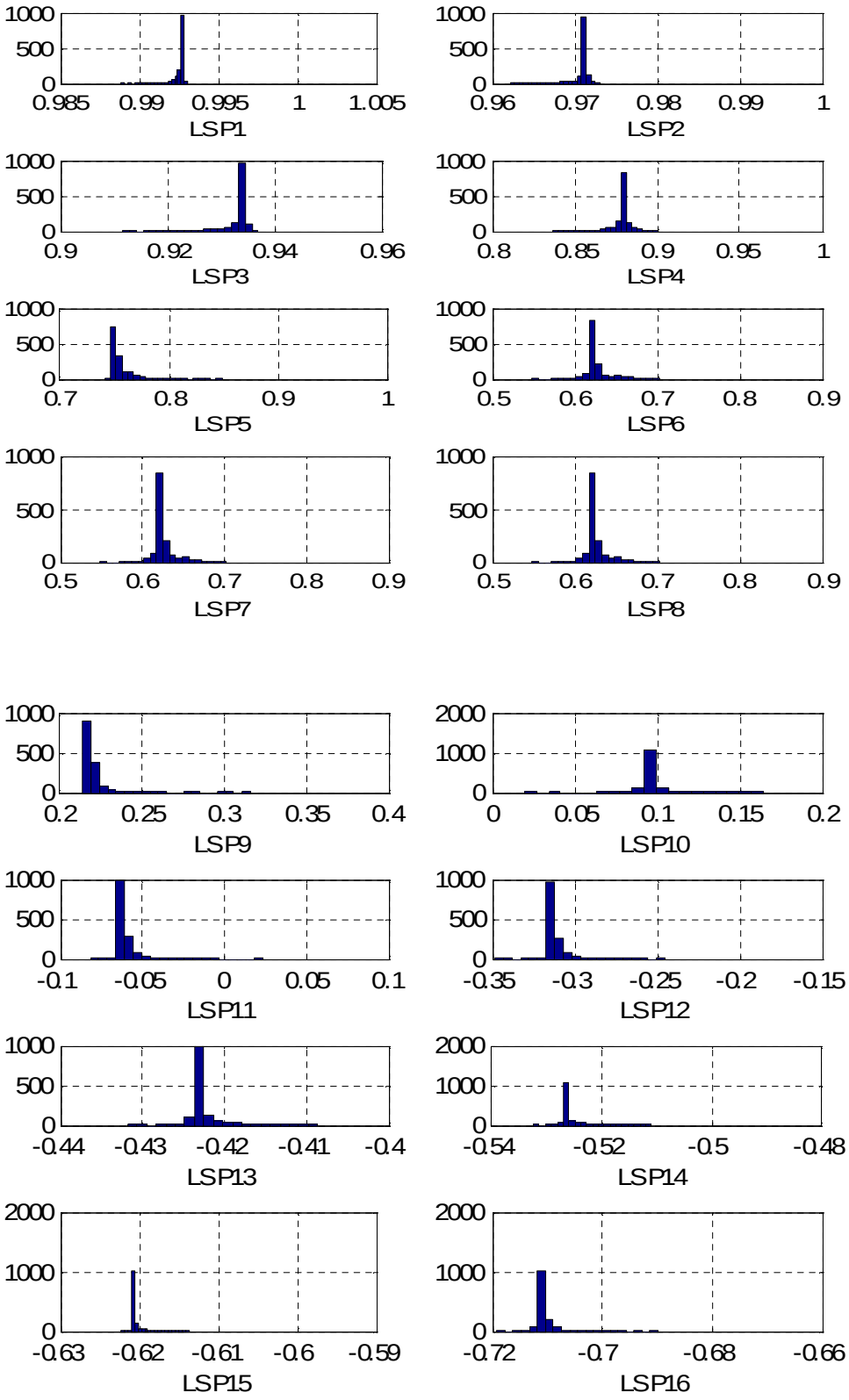
FigIII.3 Histogramme représentant les LSP de la première trame pour m=10 pôles.

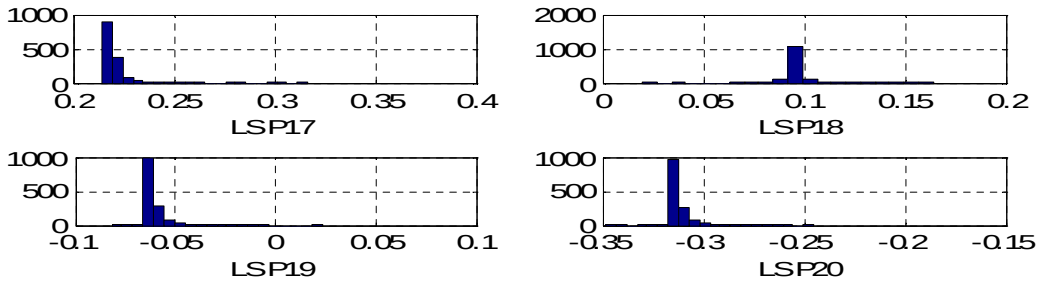


FigIII.4 Histogramme représentant les LSP de la première trame pour m=16 pôles



FigIII.5 Histogramme représentant les LSP de la première trame pour m=18 pôles.


1000
500
0
0.985 0.99 0.995 1 1.005
LSP1
1000
500
0
0.96 0.97 0.98 0.99 1
LSP2
1000
500
0
0.9 0.92 0.94 0.96
LSP3
1000
500
0
0.8 0.85 0.9 0.95 1
LSP4
1000
500
0
0.7 0.8 0.9 1
LSP5
1000
500
0
0.5 0.6 0.7 0.8 0.9
LSP6
1000
500
0
0.5 0.6 0.7 0.8 0.9
LSP7
1000
500
0
0.5 0.6 0.7 0.8 0.9
LSP8
1000
500
0
0.2 0.25 0.3 0.35 0.4
LSP9
2000
1000
0
0 0.05 0.1 0.15 0.2
LSP10
1000
500
0
-0.1 -0.05 0 0.05 0.1
LSP11
1000
500
0
-0.35 -0.3 -0.25 -0.2 -0.15
LSP12
1000
500
0
-0.44 -0.43 -0.42 -0.41 -0.4
LSP13
2000
1000
0
-0.54 -0.52 -0.5 -0.48
LSP14
2000
1000
0
-0.63 -0.62 -0.61 -0.6 -0.59
LSP15
2000
1000
0
-0.72 -0.7 -0.68 -0.66
LSP16

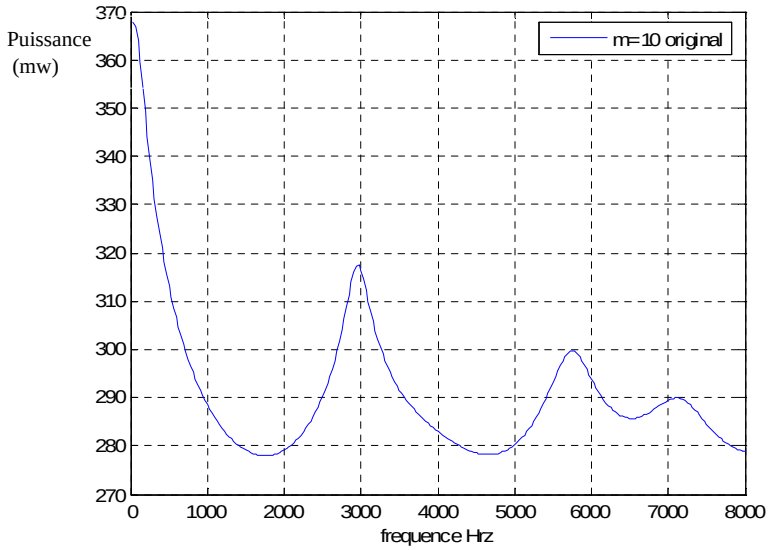


FigIII.6 Histogramme représentant les LSP de la première trame pour m=20 pôles.

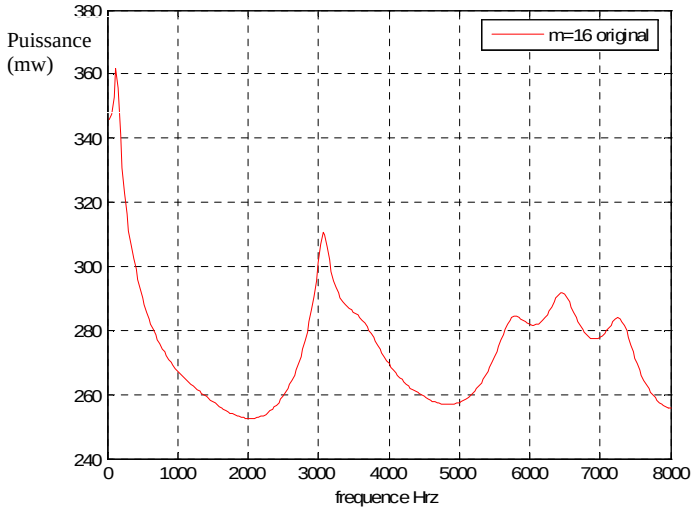
III.3 Codage de l'enveloppe spectral

III.3.1 Principes

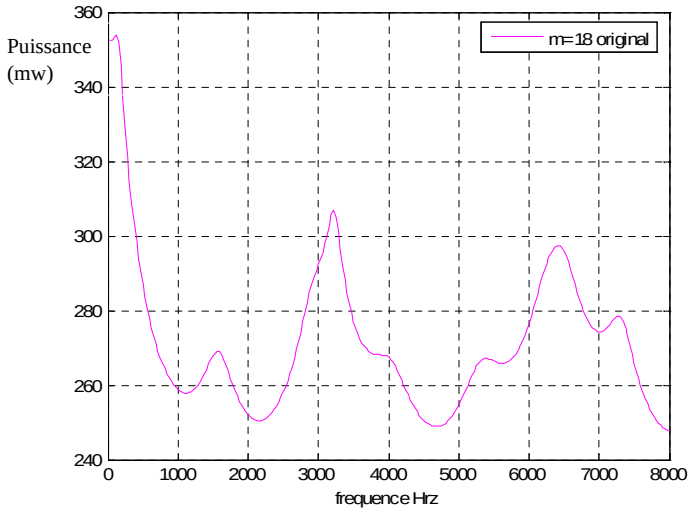
Le principe du codage de l'enveloppe spectrale est basé sur la quantification des lignes de fréquences spectrales (LSF), afin de transmettre l'information de cette dernière à large bande. Pour avoir une idée sur l'enveloppe spectrale, on a va schématiser cette dernière on utilisant les LSF obtenus dans le paragraphe III.2.1 :



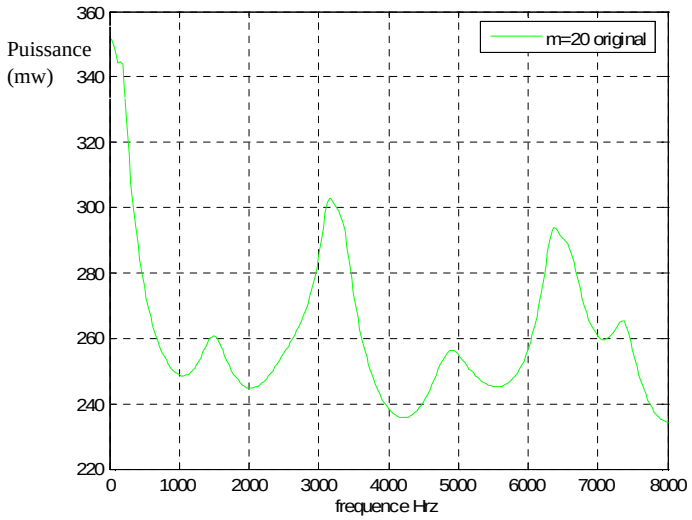
a) m=10 pôles



(b) m=16 pôles.



(c) m=18 pôles.



(d) m=20 pôles

Fig.III.7 L'enveloppe spectrale des LSF pour m=10, 16,18et 20 pôles.

III.3 .2 Création du dictionnaire de quantification

Afin de pouvoir quantifier les coefficients LSP on crée un dictionnaire optimal par l'application de la méthode LBG (Algorithme de **L**INDE **B**USO **G**RAY).

Dans le but de faire une comparaison précise, on fait varier le nombre de bits par trame pour chaque m (10, 16,18 et 20 pôles).

Dans notre cas on a pris pour :

- **m=10**

On trouver le dictionnaire de quantification pour :

- 40bits par trame. (Dictionnaire de 16entrées).
- 60bits par trame. (Dictionnaire de 64entrées).
- 80bits par trame. (Dictionnaire de 256entrées).

On cite comme exemple les deux premières valeurs du code book (dictionnaire) obtenu pour les m=10, 16,18et 20 pôles dans les tableaux suivants :

	1iere valeur	2ieme valeur
Dico 1	0.9521	0.9480
Dico 2	0.8770	0.8768
Dico 3	0.7203	0.7384
Dico 4	0.5420	0.5243
Dico 5	0.2659	0.2604
Dico 6	-0.0147	0.0528
Dico 7	-0.3878	-0.3891
Dico 8	-0.7045	-0.6935
Dico 9	-0.8652	-0.8671
Dico 10	-0.9625	-0.9653

(a) 40 bits /trame

	1iere valeur	2ieme valeur
Dico 1	0.9498	0.9455
Dico 2	0.8652	0.8697
Dico 3	0.7014	0.7512
Dico 4	0.5257	0.5354
Dico 5	0.2479	0.2510
Dico 6	-0.0886	-0.0467
Dico 7	-0.4186	-0.3853
Dico 8	-0.6635	-0.6901
Dico 9	-0.8754	-0.8644
Dico 10	-0.9565	-0.9643

(b) 60 bits /trame

	1iere valeur	2ieme valeur
Dico 1	0.9498	0.9465
Dico 2	0.8652	0.8726
Dico 3	0.7014	0.7475
Dico 4	0.5257	0.5274
Dico 5	0.2479	0.2503
Dico 6	-0.0886	-0.0456
Dico 7	-0.4186	-0.3876
Dico 8	-0.6635	-0.6914
Dico 9	-0.8754	-0.8643
Dico 10	-0.9565	-0.9643

(c) 80 bits /trame

Tableau III.9 Exemples des deux premières valeurs du dictionnaire obtenu par LBG.

On trouver le dictionnaire de quantification pour :

- 48bits par trame. (Dictionnaire de 8entrées).
- 64bits par trame. (Dictionnaire de 16entrées).
- 80bits par trame. (Dictionnaire de 32entrées).

-

	1iere valeur	2ieme valeur
Dico 1	0.9921	0.9906
Dico 2	0.9651	0.9614
Dico 3	0.9341	0.9139
Dico 4	0.7153	0.7664
Dico 5	0.5752	0.6112
Dico 6	0.4278	0.4421
Dico 7	0.2168	0.2329
Dico 8	0.1186	0.1221
Dico 9	-0.0607	-0.0408
Dico 10	-0.4211	-0.4110
Dico 11	-0.5026	-0.4989
Dico 12	-0.6231	-0.6182
Dico 13	-0.7081	-0.7062
Dico 14	-0.8466	-0.8438
Dico 15	-0.9428	-0.9419
Dico 16	-0.9819	-0.9814

(a) 48 Bits/trame

	1iere valeur	2ieme valeur
Dico 1	0.9927	0.9905
Dico 2	0.9648	0.9604
Dico 3	0.9400	0.9153
Dico 4	0.7073	0.7714
Dico 5	0.5577	0.5979
Dico 6	0.4322	0.4466
Dico 7	0.2101	0.2372
Dico 8	0.1172	0.1214
Dico 9	-0.0703	-0.0378
Dico 10	-0.4202	-0.4097
Dico 11	-0.5048	-0.4996
Dico 12	-0.6223	-0.6175
Dico 13	-0.7084	-0.7055
Dico 14	-0.8452	-0.8425
Dico 15	-0.9438	-0.9425
Dico 16	-0.9820	-0.9815

(b) 64 Bits/trame

	1iere valeur	2ieme valeur
Dico 1	0.9928	0.9908
Dico 2	0.9648	0.9615
Dico 3	0.9416	0.9178
Dico 4	0.7041	0.7601
Dico 5	0.5536	0.5970
Dico 6	0.4333	0.4471
Dico 7	0.2082	0.2340
Dico 8	0.1169	0.1213
Dico 9	-0.0733	-0.0400
Dico 10	-0.4200	-0.4119
Dico 11	-0.5053	-0.5000
Dico 12	-0.6221	-0.6183
Dico 13	-0.7086	-0.7059
Dico 14	-0.8449	-0.8436
Dico 15	-0.9440	-0.9424
Dico 16	-0.9820	-0.9814

(c) 80 Bits/trame

Tableau III.10 Exemples des deux premières valeurs du dictionnaire obtenu par LBG.

➢ **m=18 :**

On trouver le dictionnaire de quantification pour :

- 54bits par trame. (Dictionnaire de 8 entrées).
- 72bits par trame. (Dictionnaire de 16 entrées).

	1ière valeur	2ieme valeur
Dico 1	0.9932	0.9921
Dico 2	0.9690	0.9675
Dico 3	0.9415	0.9308
Dico 4	0.8449	0.8597
Dico 5	0.6348	0.6551
Dico 6	0.5584	0.5498
Dico 7	0.3467	0.3723
Dico 8	0.2254	0.2281
Dico 9	0.0973	0.1017
Dico 10	-0.0034	-0.0069
Dico 11	-0.4027	-0.3808
Dico 12	-0.4741	-0.4743
Dico 13	-0.5860	-0.5747
Dico 14	-0.6673	-0.6670
Dico 15	-0.7678	-0.7585
Dico 16	-0.8861	-0.8862
Dico 17	-0.9532	-0.9528
Dico 18	-0.9849	-0.9850

(a) 54 Bits/trame

	1ière valeur	2ieme valeur
Dico 1	0.9932	0.9923
Dico 2	0.9691	0.9677
Dico 3	0.9410	0.9319
Dico 4	0.8402	0.8483
Dico 5	0.6416	0.6630
Dico 6	0.5609	0.5600
Dico 7	0.3482	0.3695
Dico 8	0.2250	0.2277
Dico 9	0.0989	0.1036
Dico 10	-0.0035	-0.0046
Dico 11	-0.4017	-0.3855
Dico 12	-0.4737	-0.4734
Dico 13	-0.5854	-0.5766
Dico 14	-0.6673	-0.6667
Dico 15	-0.7678	-0.7611
Dico 16	-0.8859	-0.8861
Dico 17	-0.9530	-0.9524
Dico 18	-0.9849	-0.9848

(b) 72 Bits/trame

Tableau III.11 Exemples des deux premières valeurs du dictionnaire obtenu par LBG m=18

➢ **m=20 :**

On trouver le dictionnaire de quantification pour :

- 40bits par trame. (Dictionnaire de 4 entrées).
- 60bits par trame. (Dictionnaire de 8 entrées).
- 80bits par trame. (Dictionnaire de 16 entrées).

	1iere valeur	2ieme valeur
Dico 1	0.9927	0.9917
Dico 2	0.9712	0.9676
Dico 3	0.9338	0.9270
Dico 4	0.8809	0.8674
Dico 5	0.7524	0.7783
Dico 6	0.6220	0.6559
Dico 7	0.5092	0.5241
Dico 8	0.3639	0.3812
Dico 9	0.2189	0.2293
Dico 10	0.0953	0.1018
Dico 11	-0.0614	-0.0534
Dico 12	-0.3128	-0.2986
Dico 13	-0.4232	-0.4192
Dico 14	-0.5267	-0.5214
Dico 15	-0.6209	-0.6186
Dico 16	-0.7108	-0.7078
Dico 17	-0.8108	-0.8104
Dico 18	-0.9081	-0.9070
Dico 19	-0.9570	-0.9564
Dico 20	-0.9875	-0.9870

(a) 40 Bits/trame

	1iere valeur	2ieme valeur
Dico 1	0.9925	0.9920
Dico 2	0.9713	0.9683
Dico 3	0.9335	0.9270
Dico 4	0.8866	0.8649
Dico 5	0.7564	0.7707
Dico 6	0.6104	0.6671
Dico 7	0.5096	0.5277
Dico 8	0.3645	0.3774
Dico 9	0.2211	0.2271
Dico 10	0.0895	0.1051
Dico 11	-0.0599	-0.0519
Dico 12	-0.3094	-0.3028
Dico 13	-0.4235	-0.4187
Dico 14	-0.5264	-0.5219
Dico 15	-0.6207	-0.6185
Dico 16	-0.7097	-0.7090
Dico 17	-0.8094	-0.8109
Dico 18	-0.9088	-0.9064
Dico 19	-0.9573	- 0.9563
Dico 20	-0.9875	-0.9871

(b) 60 Bits/trame

	1iere valeur	2ieme valeur
Dico 1	0.9926	0.9922
Dico 2	0.9715	0.9685
Dico 3	0.9340	0.9267
Dico 4	0.8871	0.8607
Dico 5	0.7532	0.7639
Dico 6	0.6102	0.6576
Dico 7	0.5093	0.5375
Dico 8	0.3638	0.3841
Dico 9	0.2207	0.2278
Dico 10	0.0901	0.1058
Dico 11	-0.0597	-0.0488
Dico 12	-0.3108	-0.3009
Dico 13	-0.4238	-0.4184
Dico 14	-0.5268	-0.5207
Dico 15	-0.6208	-0.6181
Dico 16	- 0.7100	-0.7084
Dico 17	-0.8093	-0.8115
Dico 18	-0.9088	-0.9061
Dico 19	-0.9573	-0.9559
Dico 20	-0.9876	-0.9868

(c) 80 Bits/trame

Tableau III.12 Exemples des deux premières valeurs du dictionnaire obtenu par LBG 20 pôles.

III.3.3 Quantification des coefficients LSP

Dans cette étape on introduit les LSP pour (m=10, 16, 18,20 pôles) dans un quantificateur qui a comme code book les dictionnaires obtenus dans III.3.2.

Les coefficients LSP vont subir une quantification vectorielle afin de réduire le nombre de bits à envoyer. A partir d'un dictionnaire déjà trouvé, on cherche des indices qui minimisent l'erreur quadratique de la manière suivante :

- **Pour m=10 :**
  - Dans un dictionnaire de 16, 64, et 256entrées trouvé dans le paragraphe III.4.1et de dimension 10, chercher une entrée qui se rapproche des coefficients LSP « Min $(ErrQuad\{Vlsp, T[indice]\})$ ».
  - Coder l'indice i trouvé sur 4,6, et 8 bits pour l'envoyer au décodeur.

Et voici le schéma qui explique le procédé :

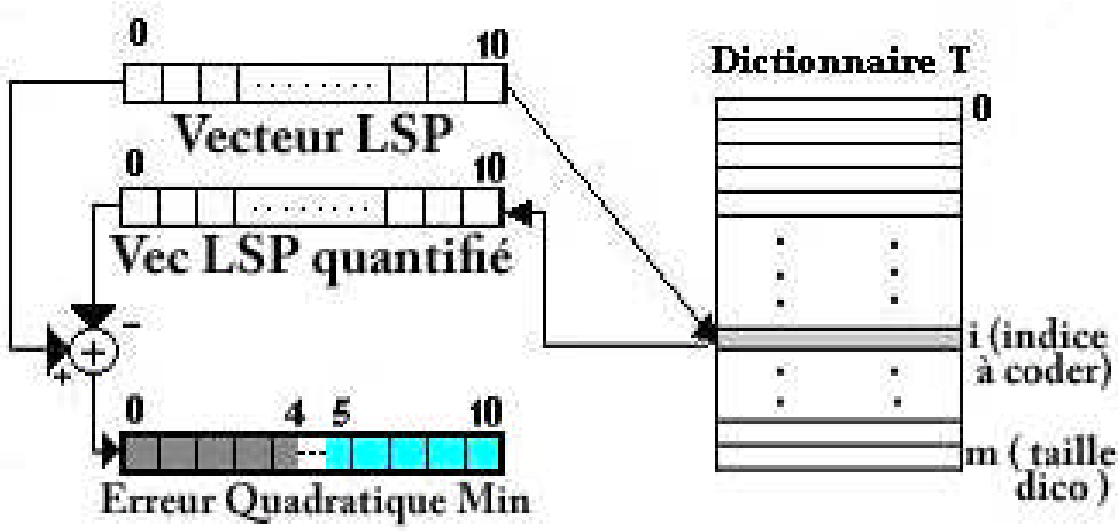


Fig.III.8 Quantification Vectorielle des coefficients LSP pour m=10 pôles.

On refait la même étape pour m=16,18et 20 pôles, on change les dictionnaires utilisés (déjà trouvé) et on introduisant les changements suivants :

- **Pour m=16 :**
  - Dans un dictionnaire de 8, 16, et 32entrées trouvé dans le paragraphe III.4.2et de dimension 16, chercher une entrée qui se rapproche des coefficients LSP « Min $(ErrQuad\{Vlsp, T[indice]\})$ ».
  - Coder l'indice i trouvé sur 3,4 et 5 bits pour l'envoyer au décodeur.

Et voici le schéma qui explique le procédé :

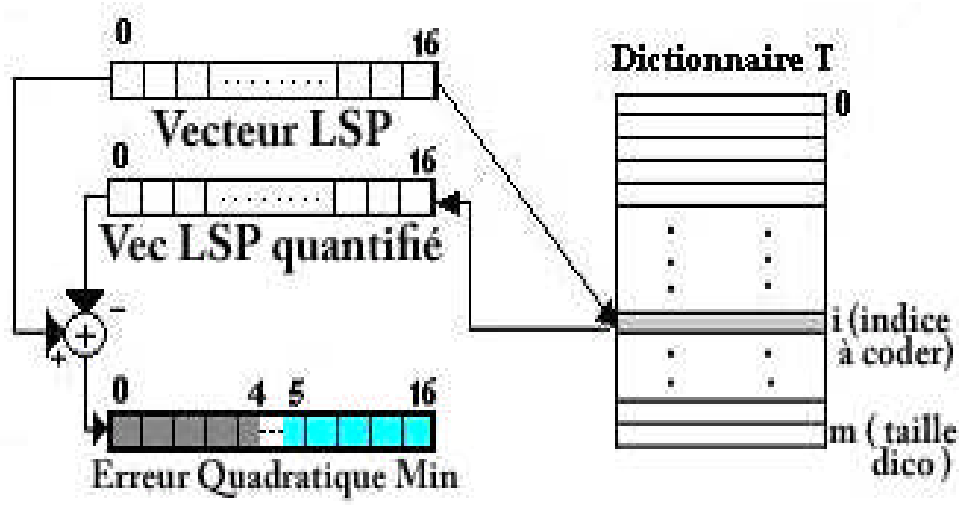


Fig. III.9 Quantification Vectorielle des coefficients LSP pour m=16 pôles.

- **Pour m=18 :**
  - Dans un dictionnaire de 8et16 entrées trouvé dans le paragraphe III.4.3et de dimension 18, chercher une entrée qui se rapproche des coefficients LSP « Min $(ErrQuad\{Vlsp, T[indice]\})$ ».
  - Coder l'indice i trouvé sur 3,4 et 5 bits pour l'envoyer au décodeur

Et voici le schéma qui explique le procédé :

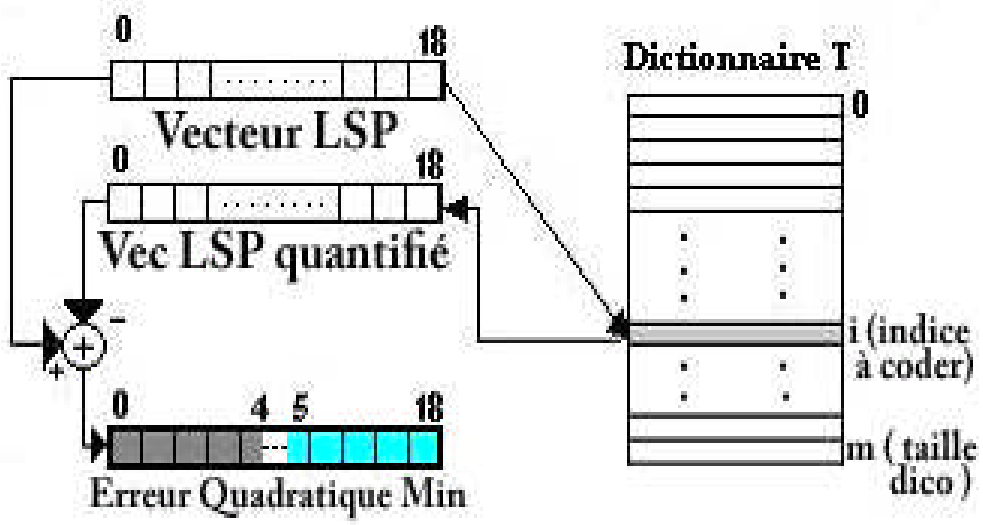


Fig. III.10 Quantification Vectorielle des coefficients LSP pour m=18 pôles.

- **Pour m=20 :**
  - Dans un dictionnaire de 4, 8, et 16entrées trouvé dans le paragraphe III.4.4et de dimension 20, chercher une entrée qui se rapproche des coefficients LSP « Min $(ErrQuad\{Vlsp, T[indice]\})$ ».
  - Coder l'indice i trouvé sur 2,3 et 4 bits pour l'envoyer au décodeur.

Et voici le schéma qui explique le procédé :

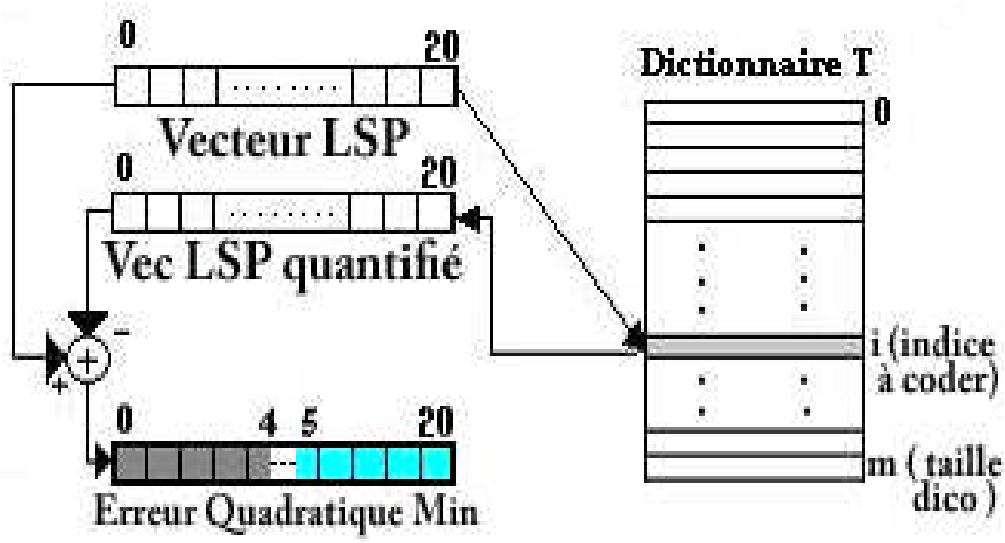


Fig. III.11 Quantification Vectorielle des coefficients LSP pour m=20 pôles.

Le tableau ci dessous, donne à titre d'exemple les deux premières trames des LSF quantifiées par la méthode de quantification énoncée ci dessus en utilisant le premier dictionnaire obtenu pour chaque ordre de prédiction

M=10poles	Trame 1	Trame 2
LSP1q	0.994200	0.914089
LSP2q	0.024788	0.814741
LSP3q	-0.239865	0.813419
LSP4q	-0.267542	0.542672
LSP5q	-0.011459	0.223811
LSP6q	-0.124534	0.073174
LSP7q	-0.205724	-0.352625
LSP8q	0.156676	-0.835121
LSP9q	0.097537	-0.817702
LSP10q	-0.956497	-0.903347

(A) m=10 Pôles.

M=16poles	Trame 1	Trame 2
LSP1q	0.991562	0.911949
LSP2q	0.967922	0.915410
LSP3q	0.923356	0.912498
LSP4q	0.865425	0.848075
LSP5q	0.745355	0.803351
LSP6q	0.636547	0.623979
LSP7q	0.524999	0.139223
LSP8q	0.382963	0.139245
LSP9q	-0.222120	0.007498
LSP10q	-0.084207	-0.070638
LSP11q	-0.071750	-0.458140
LSP12q	-0.299356	-0.254369
LSP13q	-0.416081	-0.841249
LSP14q	-0.516430	-0.847229
LSP15q	-0.613906	-0.921217
LSP16q	-0.708969	-0.927561

(B) m=16 pôles.

M=18 pôles	Trame 1	Trame 2
LSP1q	0.961148	0.967548
LSP2q	0.963702	0.976543
LSP3q	0.961106	0.943216
LSP4q	0.854724	0.884359
LSP5q	0.695668	0.687654
LSP6q	0.508103	0.564328
LSP7q	0.522089	0.398765
LSP8q	0.201094	0.276547
LSP9q	0.099596	0.094783
LSP10q	-0.042830	-0.006132
LSP11q	-0.380989	-0.434565
LSP12q	-0.460598	-0.489659
LSP13q	-0.590544	-0.234589
LSP14q	-0.626547	-0.665434
LSP15q	-0.700439	-0.787654
LSP16q	-0.810988	-0.887660
LSP17q	-0.934549	-0.965434
LSP18q	-0.912345	-0.998765

(C) m=18 pôles.

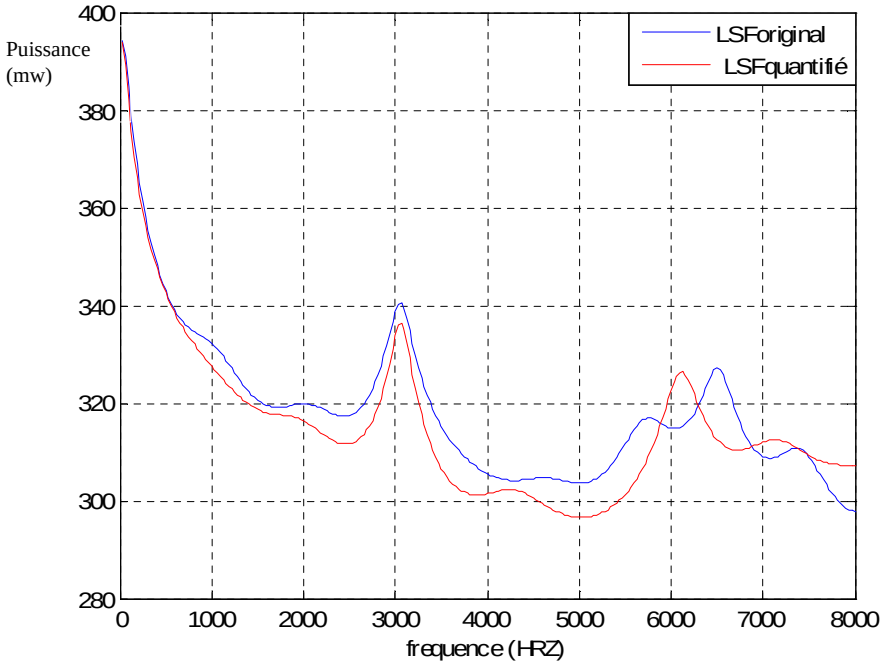
.

M=20 pôles	Trame 1	Trame 2
LSP1q	0.987654	0.982739
LSP2q	0.909874	0.998765
LSP3q	0.998765	0.934987
LSP4q	0.898765	0.878765
LSP5q	0.796547	0.758765
LSP6q	0.667654	0.623454
LSP7q	0.423247	0.324450
LSP8q	0.398765	0.398765
LSP9q	0.123459	0.876543
LSP10q	0.012345	0.023345
LSP11q	-0.987654	-0.544328
LSP12q	-0.298564	-0.397654
LSP13q	-0.498765	-0.435432
LSP14q	-0.899495	-0.544325
LSP15q	-0.432233	0.609887
LSP16q	-0.987654	-0.797654
LSP17q	-0.854323	- 0.988764
LSP18q	-0.909877	-0.999876
LSP19q	-0.967548	-0.912334
LSP20q	-0.999654	-0.998765

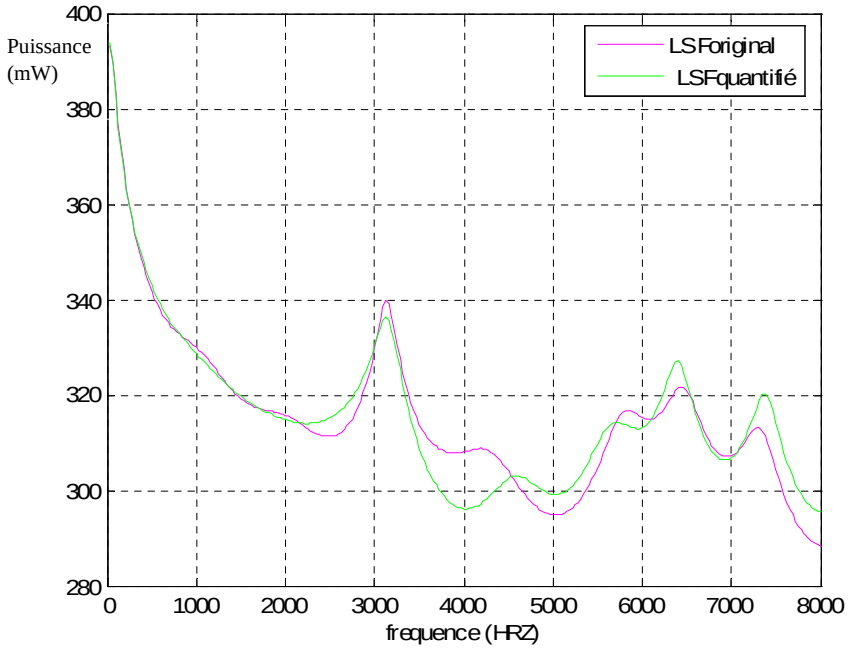
(D) m=20 pôles.

Tableau III.13 Valeurs des LSP quantifiées pour les deux premières trames.

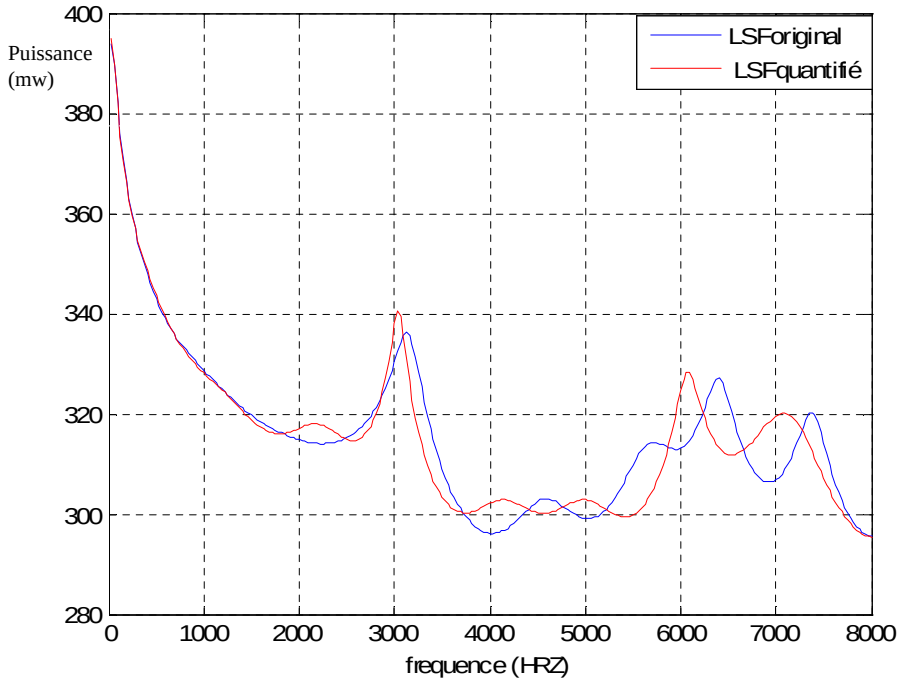
Apres avoir quantifié les LSF on utilisant le dictionnaire du paragrapheIII.3.3,pour m=10,16,18et 20 pôles .On va schématiser l'enveloppe spectrale(figureIII.12) des LSF et des LSF quantifiés dans un même graphe afin de faire une analyse comparative de la performance du codage large bande pour différents ordres de prédiction .



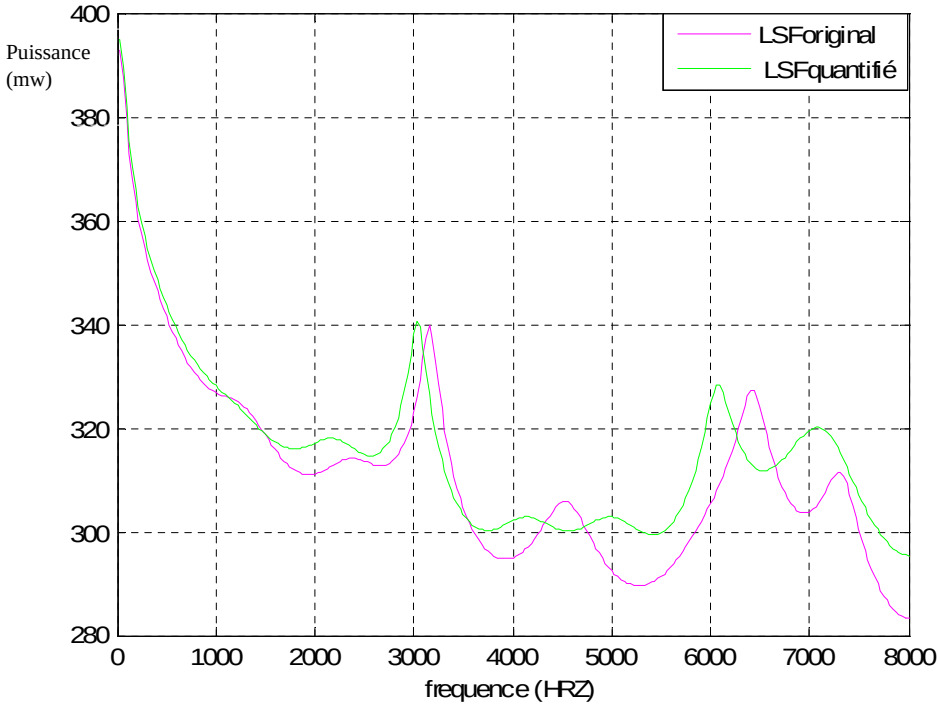
(a) m=10 pôles



(b) m=16 pôles



(c) m=18 pôles



(d) m=20 pôles

Fig.III.12 Enveloppes spectrales des LSF et LSF quantifié pour m=10,16 ,18et 20 pôles.

III.3.4 Interprétations et commentaires

En analysant la figure III.12 on constate que l'enveloppe spectrale des LSF originales et des LSF quantifiées se rapprochent dans la plage de 0-4000 Hz (bande étroite) et cela pour les différents ordres de prédiction, dépassé cette plage là (c.a.d dans la bande élargie :4000-8000 Hz) on remarque que la différence entre l'enveloppe originale et quantifiée augmente pour un ordre de prédiction de 10 et 20 pôles ce qui provoque la perte de l'authenticité du signal original entraînant une diminution de la performance du quantificateur. Quant aux ordres de prédiction de 16 et 18 pôles l'enveloppe spectrale quantifiée garde la même forme que l'enveloppe spectrale originale dans la bande élargie augmentant ainsi la performance du quantificateur.

III.4 Calcul de la distorsion spectrale

Après l'extraction des LSP quantifiés, on a calculé la distorsion spectrale en dB, pour m=10,16 ,18 et 20 en fonction du nombre de bits (variant entre 40 et 80bits).

Afin de faire une comparaison des résultas trouvés entre la variation de la distorsion spectrale pour m=10, 16,18 et 20 pôles, on schématisé les valeurs obtenus sur même graphe, comme le montre la figure suivante 3.10 :

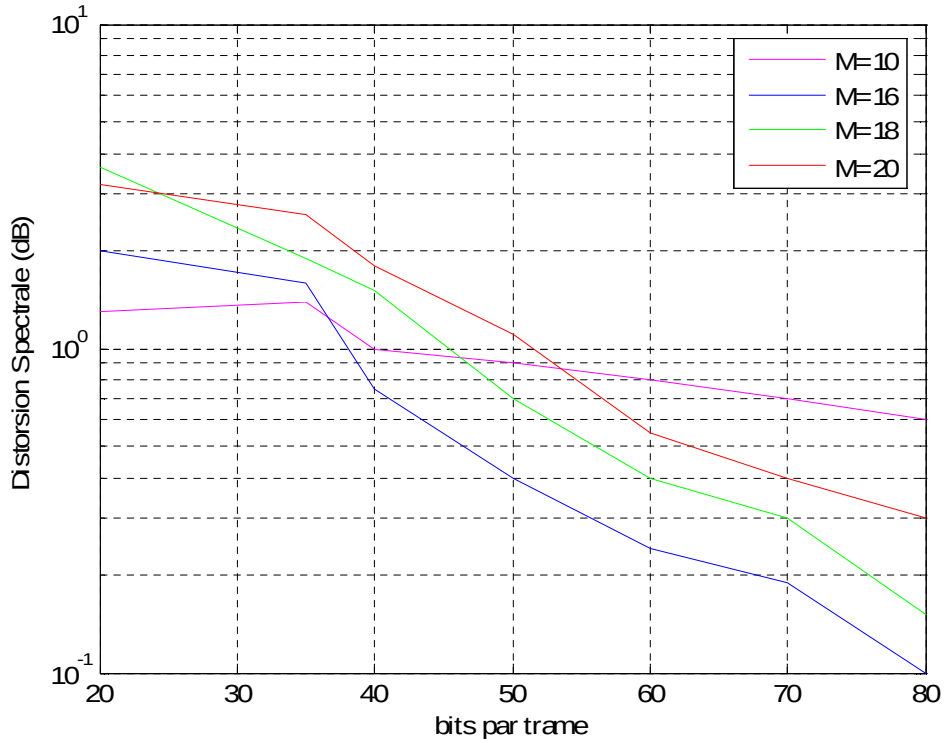


Fig.III.13 distorsions spectrale entre les LSP et les LSP Quantifié pour 10,16, 18 et 20 pôles.

III.4.1 Interprétation et commentaires

La figure III.13 représente la distorsion spectrale calculée en dB pour différents ordres de prédictions en fonction du nombre de bits par trame,on voit que la distorsion spectrale est inversement proportionnel aux nombre de bits par trame car pour un nombre réduit de bits le dictionnaire de quantification (code book) est moins riche et donc le quantificateur est moins performant .

On constate aussi que pour l'ordre de prédiction m=16 pôles la distorsion est moindre par rapport à celles calculées pour m=18,20 pôles, ce qui valorise la quantification pour un ordre de m=16 pôles dans la bande élargie.

Conclusion

La technique standard de codage utilisée dans la téléphonie s'effectue dans une bande de fréquences étroite (200-3400 Hz) utilisant un ordre de prédiction du filtre analyse-synthese égale à 10,cette dernière ne permet pas la transmission de signaux musiques.

Nous avons vu le long des chapitres précédents que pour avoir une gamme de fréquences large s'étalant jusqu'à 8000 Hz (la large bande), et donc avoir plus de flexibilité et de performance, il faut augmenter l'ordre de prédiction des filtre .Pour une analyse comparative fiable nous avons pris un ordre de prédiction de 10, 16, 18, 20 pôles.

En premier lieu nous avons extraits les paires de raies spectrales LSP du signal parole échantillonnée utilisé dans notre simulation,le codage se fait sur des trames de 80 échantillons chacune .

En deuxième lieu nous avons appliqué l'algorithme de **L**inde **B**uso **G**ray (LBG) afin de constituer un dictionnaire de quantification (code book) pour un nombre de bits par trame allant de 20 à 80 bits/trame.

En dernier lieu nous avons quantifié les LSP en utilisant les code books trouvés,par la suite on a schématise les enveloppes spectrales des LSP originale et quantifiée et on a calculé les distorsions spectrales entre celles cis pour des ordres de 10 ,16 ,18 ,20 pôles.

L'étude citée ci dessus nous informe sur l'état du quantificateur pour des ordres supérieurs et nous permet de constater que le codage à large bande est meilleur pour un ordre de prédiction de 16 pôles.

On conclut que pour le codage à large bande il est conseillé d'utiliser un ordre de prédiction égale à 16 pôles permettant d'avoir une quantification performante et une distorsion acceptable.

Annexe A

Algorithme de Levinson-Durbin :

Les coefficients d'autocorrelation R(k), k=0,1,……..,P sont utilisées pour obtenir les coefficients du filtre *LP* après résolution du système linéaire (1.13)

Il s'agit donc d'inverser une matrice d'ordre "p". Les méthodes algébriques classiques exigent pour cela un nombre d'opérations (multiplication+ addition) de l'ordre de p^3, ce que l'on note $O(p^3)$.

L'algorithme qui va être décrit profite de la structure particulière (Toeplitz symétrique) de la matrice d'autocorrélation pour résoudre (1.13) par une récurssion sur l'ordre de prédiction: autrement dit, ils fournissent toutes les solutions d'ordre M=1,2,…p , le nombre d'opérations est seulement $O(p^2)$.

La variance de l'erreur de prédiction α_p sera obtenue également par une récurrence sur l'ordre m.

Rappelons que la fonction d'autocorrélation est supposée connue et que pour un signal stationnaire, on a :

$$R(i,j) = R(|i-j|) = R(k) \tag{A.1}$$

Initialisation:

$$a_m(0) = 1, \qquad (m = 1,2,\ldots,p) \qquad E_0 = R(0) = \sigma_x{}^2$$

Récurssion:

pour: $m = 1,2,\ldots p.$

$$k_m = -\frac{1}{E_{m-1}}\left[R(m) - \sum_{k=1}^{m-1} \alpha_{m-1}(k)R(m-k)\right] \tag{A.2}$$

pour k=1,2,…m-1.

$$\alpha_k(m) = \alpha_k(m-1) - k_m\alpha_{m-k}(m-1) \tag{A.3}$$

$$E_m = E_{m-1}(1-k_m{}^2) \tag{A.4}$$

Les coefficients $a_k(m)$ résultant, quand $m = p$ représentent les coefficients de prédiction d'un prédicteur linéaire d'ordre p :

La valeur de k_m joint à la propriété : $-1 \leq k_m \leq 1$

Cette relation est une condition nécessaire et suffisante pour que le filtre soit stable.

La méthode d'autocorrélation garantit la stabilité du filtre, de plus le calcul de $R(i)$ nécessite un fenêtrage de $S(n)$ par un la fenêtre de Hamming.

Bibliographie

[1] T.Dutoit, "*Introduction au Traitement Automatique de la Parole*", Faculté Polytechnique de Mons 1989.

[2] R.Boite et M.Kunt,"*Traitement de la parole*", Presses Polytechniques Romandes, première édition.

[3] M. Xie et D.Berkani. **"***Amélioration des performances des codeurs de parole***"**Août 1997

[4] F.Merazka, "Techniques *de codage de la parole : applications aux LSPs et aux systèmes VoIP*", Thèse de Doctorat d'État, Présenté a l'École National Polytechnique Alger 2004.

[5] F.Merazka, "*quantification des paramètres LSF*", Thèse de Magistèr, a l'École National Polytechnique Alger 1997.

[6] F.Itakura and S. Saito, "*Analysis synthesis telephoney based upon the maximum likelihood method*" in Rep 6 th Int. Congr. on acoustics, Kohasi, Ed. Tokyo, Japan Aug. 21-28, 1968, C-5-5.

[7] J. D Markel and A. H. Gray, Jr "*A linear prediction vocoder simulation based upon the autocorrelation method*", IEEE Trans Acoust. Speech. Signal Processing vol ASSP622, PP.124-134, Apr. 1974.

[8] P.Kroon and B.S. Atal, "*Predictive coding of speech using analysis-by-synthesis techniques*", in Advances in *Speech Signal Processing* S. Furui and M.M. Sondhi, Eds New York: Markel- Dekker, pp 141-164. *1991.*

[9] A. H. Gray, and J. D. Markel "*Quantization and bit allocation in speech processing*", IEEE *Trans, on Acoustic, Speech Signal Processing,* vol. ASSP-24, pp. 459-473, Oct. 1976.

[10] P. E. Papamichlis, "*Practical Approaches to Speech Coding*", Prentice-Hall, Englewood Cliffs, N. J. 1987.

[11] D.O'Shaugnessy,"*speech communication, Humanand machine.* Reading",MA:Addison-Wesley, 1987.

[12] J.D. Markel and A. H. Gray, Jr., "Linear prediction of speech*",* New York: Springer-Verlag, 1976.

[13] F.Itakura, "Line *spectrum representation of linear predictive coefficients of speech signals*" *J. Acoust. Soc. Amer.,* vol. 57, suppl. 1 p. S35(A), 1975.

[14] F. K. Soong and B. H. Juang, "*Line spectrum pair (LSP) and speech data compression*", in *Proc. IEEE Int Conf. Acoust. Speech, Signal Processing*, San Diego, CA, pp.1.10.1-1.10.4, Mar.1984.

[15] B.S Atal, R, V Cox and P.Kroon,"*Spectral quantization and interpolation for CELP Coders*", in Proc. IEEE in t. Conf. On Acoustics, speech and signals.

[16] S. Wang, A. Sekey, and A. Gersho, "*An objective measure for predicting subjective*"

[17] G.A. Mian and G.Riccardi" A localisation property of line spectrum frequencies",IEEE Trans.Speech and audio processing,vol.2,pp.536-539,Oct 1994.

[18] R. Laroia, N Phambo, and N,Favardin,"*Robust abs=d efficient quantization of speech LSP parameter using structured vector quantizer*", in Proc.IEEE Int. Conf on acoustics, speech, and Sig.processing(Toronto, Canada) ,may 1991 pp 641-644.

[19] Alexis Pascal Bernard, **"***Source-Channel Coding of Speech***"**, Master of Science in Electrical Engineering University of California Los Angeles,1998.

[20]. **Raake, A.**, *Assessment and Parametric Modelling of Speech Quality in Voice-over-IP. Networks*. Doctoral dissertation, Institut für Kommunikationsakustik, Ruhr-Universität, DEBochum. 2005.

[21]. **Möller, S**. *Assesment and Prediction of Speech Quality in Telecommunications*. Kluwer Academic Publishers, USA-Boston. 2000.

[22]. Jekosch, U. ***Spache Hören und Beurteile,*** *Ein Ansatz zur Grundlegung der Sparchqualitätsbeurteilung*. Habilitation thesis, UniversitÄt/Gesamthochschule, DE-Essen.2000

[23]. Jekosch, U. *Voice and Speech Quality Perception – Assessment and Evaluation*. Springer, DE-Berlin. 2000.

[24]. **ITU-T Rec.** G.722.2. *Wideband Coding of Speech at Around 16 kbits/s Using Adaptative Multi-Rate Wideband (AMR-WB).* International Telecommunication Union, CHGeneva. 2002.

[25]. Gleiss, N. *The effect of Bandwidth Restriction on Speech transmission Quality in Telephony*. Proc. 4th Int. Symp. On Human Factors in Telephony, 1-6, VDE-Verlag, DEBerlin. 1970.

[26]. Zwicker, E. and Fastl, H. *Psychoacoustics: Facts and Models*. Springer, DE-Berlin. 1999.