

M0023/97B

ECOLE NATIONALE POLYTECHNIQUE
DE GENIE ELECTRIQUE ET INFORMATIQUE

THESE

المدرسة الوطنية المتعددة التخصصات
المكتبة — BIBLIOTHEQUE
Ecole Nationale Polytechnique

Pour l'obtention du grade de

MAGISTER EN ELECTRONIQUE

OPTION : ACQUISITION ET TRAITEMENT DE L'INFORMATION

Présentée par

Melle Fatiha MERAZKA

Ingenieur d'état en ELECTRONIQUE

QUANTIFICATION DES PARAMETRES "LSF"

Soutenue en juin 1997 devant la commission d'examen composée de:

M. M.S. BOUCHERIT,	Maître de conférences à L'E.N.P.	Président
M. D.BERKANI,	Maître de conférences à L'E.N.P.	Rapporteur
M. A.FARAH,	Maître de conférences à L'E.N.P.	Examineur
Melle M.GUERTI,	Maître de conférences à L'E.N.P.	Examineur
M. B. BOUDRAA,	Chargé de cours à L'U.S.T.H.B.	Examineur

Remerciements

Il m'est difficile d'exprimer mes vifs remerciements et ma profonde gratitude à Monsieur D. BERKANI, Maître de conférences à l'Ecole Nationale Polytechnique, Directeur des études et de la recherche du Génie Electrique et Informatique, pour m'avoir dirigé tout au long de ce travail, pour son aide toujours renouvelée, ses encouragements, sa disponibilité et ses conseils avisés.

Je tiens à exprimer mes sincères remerciements à Monsieur M.S BOUCHERIT, Maître de conférences à l'Ecole Nationale Polytechnique, Directeur adjoint de la recherche du Génie Electrique et Informatique, pour m'avoir donné la chance de travailler au sein de l'équipe du laboratoire de commande des processus, pour ses conseils, ses suggestions judicieuses. Je lui suis reconnaissante d'avoir accepté la présidence de ce Jury.

Mes remerciements vont à Monsieur A.FARAH et Melle M.GUERTI, tous les deux Maître de conférences à l'Ecole Nationale Polytechnique et Monsieur B. BOUDRAA enseignant à l'USTHB, pour avoir bien voulu participer au jury d'examen.

Je saisi cette occasion pour exprimer ma profonde reconnaissance à Monsieur M.KOURGLI pour ses conseils, ses qualités humaines et son soutien permanent. Mes sincères remerciements vont particulièrement à Monsieur B.Hemici et toute l'équipe du laboratoire de commande des processus et d'électronique de puissance sans oublier Monsieur O.Mehkaldi.

Je ne manquerai pas de remercier Monsieur L.Saadaoui membre de l'équipe du laboratoire signal et communications, pour son aide.

Table des matières

INTRODUCTION.....	1
Chapitre 1: NOTIONS SUR LA THEORIE DE LA DISTORSION ET QUANTIFICATION... 4	4
Introduction.....	4
1.1- Entropie d'une source.....	4
1.2- La fonction de distorsion.....	6
1.3- Principe de la quantification.....	10
1.4- La quantification scalaire.....	10
1.5- La quantification vectorielle.....	12
1.6- Mesures de performances d'un quantificateur.....	14
1.6.1- Performances asymptotiques des QV.....	15
a- Bornes asymptotiques des performances.....	16
1.7- Algorithmes généralisés des quantificateurs vectoriels.....	18
1.7.1- Approche statistique.....	19
1.7.1.1- Dictionnaires organisés en arbre.....	21
1.7.1.2- Quantification à étages.....	22
1.7.1.3- Quantification par Split.....	23
1.7.2- Approche algébrique.....	23
1.7.2.1- Quantification vectorielle par réseaux.....	24
Conclusion.....	26
Chapitre 2 : LE CODAGE DE LA PAROLE.....	27
Introduction.....	27
2.1- Numérisation d'un signal.....	27
2.2- Le signal parole.....	27
2.3- La prédiction linéaire.....	31
a- Méthode d'autocorrélation.....	33
b- Méthode de covariance.....	34
2.4- Techniques de codage de la parole.....	35
2.4.1- Les codeurs temporels.....	35
2.4.2- Les codeurs paramétriques.....	36
2.4.2.1- Les vocodeurs LPC.....	36
2.4.2.2- Les codeurs à analyse par synthèse.....	36
a- Codeur à excitation par impulsions.....	36
b- Codeur à excitation par codes (CELP).....	37
2.4.3- Les codeurs hybrides.....	37
2.4.3.1- Codeur à excitation par le résidu.....	38
2.4.3.2- Codeur à excitation vocale.....	38
2.4.3.3- Vocodeur à bande de base.....	38
2.5- Evaluation de la qualité des codeurs.....	38
2.5.1- Mise en forme du bruit de codage.....	39
Conclusion.....	40

Chapitre 3 : LA QUANTIFICATION DES PARAMETRES LSF.....	41
Introduction.....	41
3.1- Extraction des paramètres LSF.....	41
3.1.1- Les conditions d'analyse.....	41
3.1.2- Constitution des trames.....	42
3.2- Calcul des Line Spectral Frequencies (LSF).....	44
3.3- Distributions statistiques des LSFs.....	50
3.4- Conversion des LSF au coefficients de prédiction.....	52
3.5- Quantification des paramètres LSFs.....	54
3.5.1- Mesures de distances des LSFs.....	54
3.5.2- Génération de dictionnaires.....	55
3.5.3- Méthodes de quantification des LSF.....	65
3.5.3.1- Quantification directe de chaque LSF.....	66
3.5.3.2- Quantification différentielle des LSFs.....	67
3.5.3.3- Quantification vectorielle par SPLIT.....	68
3.5.3.4- Quantification vectorielle à étages.....	76
Conclusion.....	79
 Chapitre 4 : INTERPRETATION DES RESULTATS.....	 80
Introduction.....	80
4.1- Les paramètres LSF.....	80
4.2- Les dictionnaires utilisés.....	80
4.3- Quantification directe de chaque LSF.....	81
4.4- Quantification différentielle.....	81
4.5- Performances de la quantification vectorielle par Spli.....	82
4.6- Performances de la quantification vectorielle à étages.....	86
Conclusion.....	87
 CONCLUSION.....	 88
 BIBLIOGRAPHIE.....	 90
 ANNEXE A: ALGORITHME DE WIENER-LIVINSON-DURBIN.....	 95

Introduction

La transmission de l'information parlée est un des plus importants domaines des communications. En effet, le téléphone est probablement le moyen de communication le plus répandu de notre époque.

Dans les années 70, l'introduction du codage PCM (Pulse Code Modulation), a ouvert la voie à la numérisation de la parole téléphonique. Cette transmission numérique des signaux présente, par rapport à la transmission analogique, les avantages suivants:

- de pouvoir véhiculer l'information indépendamment de la source qui l'a produite, celle-ci n'est plus alors caractérisée que par son débit.
- de présenter une sensibilité réduite au bruit grâce, d'une part à une possibilité de régénération du signal numérique, et d'autre part à l'utilisation éventuelle de codes correcteurs d'erreurs.
- d'offrir en outre, des possibilités de multiplexage temporel des signaux, sans oublier les facilités de cryptage dans le cas de messages à caractère confidentiel et d'être d'un coût d'utilisation inférieur à l'analogique.

Enfin, le principal intérêt que procure la numérisation du signal, vient de la possibilité de considérer les signaux comme des données informatiques que l'on peut stocker et traiter par ordinateur.

Les progrès technologiques rapides, réalisés depuis quelques années, ont permis la fabrication en série de circuits numériques, de plus en plus puissants, de plus en plus fiables, d'une rapidité accrue, pour un coût en constante diminution. Dès lors, il devient intéressant de pouvoir utiliser les possibilités qu'offre la technologie numérique à la transmission de la parole. Néanmoins, il est nécessaire de se servir du réseau téléphonique existant pour transporter les signaux entre diverses sources qui peuvent être des données informatiques, des images ou des données de télécopie.

La quantité d'informations numériques à transmettre se heurte aux limites technologiques des équipements qui vont véhiculer ces informations. Il est donc important de réduire cette quantité d'informations, nous dirons le *débit numérique*, associé à chaque communication téléphonique. Ainsi, la compression numérique du message parlé est un domaine de recherche d'actualité.

L'extension des services offerts en matière de communication parlée engendre de nouvelles contraintes sur la réduction du débit de l'information et sur la qualité de ces services.

Introduction

Nous citons l'exemple de la messagerie vocale ou le téléphone mobile où l'exigence d'une compression spectrale des signaux est évidente.

L'étude d'algorithmes de réduction du débit numérique associé à la transmission de la parole est un problème dont la solution répond immédiatement aux besoins des utilisateurs de la communication parlée. De nombreuses techniques existent, mais elles ne sont pas toutes réalisables avec la technologie actuelle. De plus, le coût de développement des équipements qui vont réaliser la compression numérique est un facteur important. Il importe donc que, non seulement la technique proposée soit réalisable, mais en plus qu'elle puisse être réalisée à un coût raisonnable.

La *quantification vectorielle* est une méthode théorique qui peut être utilisée en compression numérique du signal de parole. Bien qu'elle ne se présente pas, au premier abord comme une technique débouchant sur une réalisation simple et économique, une étude plus approfondie montre qu'elle correspond bien aux besoins d'une certaine classe de transmission: les bas débits numériques.

La quantification vectorielle a connu un essor important depuis les travaux de Linde Buzo et Gray en 1980 dans le domaine de codage de la parole. Une des premières réalisations de codeur à quantification vectorielle a été proposée par Wong et al en 1981. La quantification vectorielle avait indépendamment été utilisée dans des codeurs pour des débits numériques plus élevés par Adoul et al en 1980.

L'objet principal de cette thèse est l'utilisation de la quantification vectorielle pour le codage numérique du signal de parole représenté par les paramètres LSF (Line Spectral Frequencies). Ces paramètres, issus des coefficients de prédiction, ont été introduit pour la première fois par Itakura en 1975.

Les paramètres LSF, qui représentent le spectre de la parole à temps court (short time), sont complètement équivalents dans un sens mathématique aux autres représentations telles que les coefficients de prédiction LPC et les coefficients de réflexion. Cependant, les paramètres LSF possèdent quelques propriétés qui les rendent plus intéressants que les coefficients de prédiction LPC et les coefficients de réflexion.

Les résultats expérimentaux ont montré que la bonne qualité de la parole synthétisée peut être obtenue en utilisant la représentation dans le domaine des LSF à bas débit. Les tests de performances subjectifs confirment la bonne qualité de la parole synthétisée.

Le premier chapitre de cette thèse présente les notions fondamentales de la théorie de la distorsion et le problème de la quantification. Nous étendrons la notion de la quantification scalaire à celle de la quantification vectorielle. La rétrospective des travaux de recherche présentée fait ressortir les solutions préconisées par les chercheurs et situe le problème de design des quantificateurs. Elle situe également le niveau actuel des recherches dans le domaine.

Le second chapitre de cette étude, sera consacré à une modélisation du processus de production de la parole. Nous y rappellerons les mécanismes de production de la parole. L'analyse par prédiction linéaire sera introduite pour représenter le conduit vocal par un filtre autoregressif.

Nous rappellerons les procédés de codage de la parole, plus particulièrement les codeurs à base de prédiction linéaire qui présentent les meilleures caractéristiques de qualité pour les débits considérés, inférieurs à 16 kbits/s.

Dans le troisième chapitre, nous consacrerons un paragraphe détaillé à l'extraction des paramètres LSF par prédiction linéaire. Ensuite, on abordera le problème de leur quantification en réduisant le nombre de bits aussi que possible pour une bonne qualité de la parole reconstruite.

On étudie plusieurs méthodes de quantification des paramètres LSF. Parmi celles ci, la quantification scalaire, la quantification différentielle introduite par Soong et Juang en 1984, la quantification vectorielle par split et la quantification vectorielle à étages.

La mesure de distorsion spectrale est utilisée pour une comparaison objective entre les différentes méthodes de quantifications. En fin, les commentaires sur les résultats obtenus par l'application de différentes méthodes de quantification des paramètres LSF sont présentés dans la dernière partie de ce travail.

Chapitre 1

Notions sur la théorie de la distorsion et quantification

Introduction

La quantification est actuellement un mécanisme par lequel l'information est réduite en gardant uniquement ce qui est réellement nécessaire pour permettre une reconstitution du signal original avec une certaine exactitude mesurée par des critères donnés de fidélité [1].

Dans ce chapitre, nous allons présenter les idées de base de la quantification. On verra successivement l'essence de la quantification vectorielle, la recherche par arbre (tree searched), la quantification vectorielle à étages (Multistage) et la quantification vectorielle par Split.

Auparavant, il serait intéressant de rappeler quelques notions sur la théorie de la distorsion et de montrer que la fonction de la distorsion est l'expression mathématique de la meilleure performance que peut atteindre un quantificateur [2]. On introduit également les différents développements des idées et questions afférentes à la quantification.

1.1-Entropie d'une source

Considérons une source X à temps discret, stationnaire, ergodique et sans mémoire.

$$X = (x(0), x(1), \dots, x(N-1))$$

Chaque échantillon temporel $x(i)$ est indépendant des échantillons $x(i-1), x(i-2), \dots, x(N-1)$ et qu'il ne peut prendre que N valeurs différentes dans l'ensemble alphabet de lettres $\{a_0, a_1, \dots, a_{N-1}\}$.

Chaque échantillon de la source est associé à une probabilité $p(x(i) = a(i)) = P_x(a_i)$. On définit ainsi les probabilités des différents éléments de X par : $P(a_0), P(a_1), \dots, P(a_{N-1})$.

Cette probabilité permet de définir l'*auto-information* de la source ou surprise exprimée en bits par:

$$I(X) = -\log_2 P(X) \quad (1.1)$$

Plus un événement est rare ($P(x)$ est petit), plus la surprise qui lui est associée est grande, chaque lettre de la source a donc une surprise qui lui est propre.

On peut définir l'entropie de la source, notée par $H(X)$, comme étant la surprise moyenne de cette source:

$$H(X) = E[I(X)] \quad (1.2)$$

ou encore

$$H(X) = -E[\log_2 P(X)] = -\sum_{i=0}^{N-1} P_X(a_i) \log_2 P_X(a_i) \quad (1.3)$$

Sachant que les probabilités discrètes doivent obéir à:

$$0 \leq P_X(a_i) \leq 1 \text{ et } \sum_{i=0}^{N-1} P_X(a_i) = 1 \quad (1.4)$$

On peut montrer que l'entropie d'une source est toujours positive et limitée par:

$$0 < H(X) < \log_2 N \quad (1.5)$$

Lorsque $H(X)$ est nulle, cela se traduit par une seule probabilité non nulle d'un élément de l'alphabet. La source dans ce cas est totalement *prédictible*.

Lorsque $H(X) = \log_2 N$, on obtient une probabilité uniforme $P_X = \frac{1}{N}$. Dans ce cas la source est totalement *non prédictible*.

Dans le cas deux variables aléatoires X et Y associées aux alphabets de valeurs $\{a_0, a_1, \dots, a_{N-1}\}$ et $\{b_0, b_1, \dots, b_{M-1}\}$, avec une distribution de probabilité conjointe $P_{XY}(a_i, b_j)$. On peut définir l'entropie conditionnelle déduite de la notion de probabilité conditionnelle par:

$$H(X/Y) = -E[\log_2 P_{X/Y}(X/Y)] = -\sum_{n=0}^{N-1} \sum_{m=0}^{M-1} P_{XY}(a_n, b_m) \log_2 P_{X/Y}(a_n/b_m) \quad (1.6)$$

La relation (1.6) exprime l'incertitude moyenne qui reste à propos de X après la connaissance de la valeur de Y, ou encore, l'information d'une source connaissant une autre source.

L'information mutuelle s'écrit:

$$I(X; Y) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} P_{XY}(a_n, b_m) \log_2 \frac{P_{XY}(a_n, b_m)}{P_X(a_n) P_Y(b_m)} \quad (1.7)$$

$$\text{ou bien } I(X; Y) = H(X) - H(X/Y) \geq 0 \quad (1.8)$$

Ainsi, l'information mutuelle peut être considérée comme la quantité d'incertitudes soustraites de X ou la quantité d'information apprises sur X en observant les valeurs de Y.

La notion d'entropie d'une source discrète telle que définie par les relations précédentes peut être étendue à des variables vectorielles en considérant les probabilités d'apparition des lettres de l'alphabet comme des probabilités conjointes des composantes des vecteurs.

1.2-La fonction de distorsion

Considérons le schéma d'un quantificateur Q de la figure 1.1.

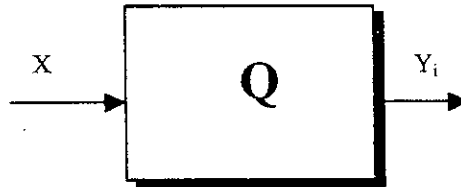


Figure 1.1: Quantificateur type, x est la grandeur d'entrée à quantifier et y_i est la valeur de sortie

L'entrée X est un vecteur de dimension k, dont les composantes sont indépendantes (source discrète sans mémoire),

$$X = (x(0), x(1), \dots, x(k-1))$$

La sortie Y est une séquence ayant des réalisations dans l'alphabet $\{b_0, b_1, \dots, b_{M-1}\}$

$$Y = (y(0), y(1), \dots, y(k-1))$$

L'approximation de la séquence X par le vecteur Y, introduit une distorsion cette dernière est mesurée par $\rho_k(x, y)$, qui exprime *le critère de fidélité* où encore le coût de reproduction de X par Y. Ce critère permet d'établir une comparaison entre les systèmes de communications.

Le quantificateur Q peut être vu comme un canal de transmission. Il va permettre d'associer à chaque lettre de l'alphabet de X une lettre de l'alphabet de Y. Il constitue donc une matrice de taille $M \times N$ avec en pratique $M \leq N$.

Donc, il est préférable en théorie de la distorsion de passer du modèle de Q comme fonction $Y = Q(X)$, à un modèle de Q comme probabilité conditionnelle de Y connaissant X pour un critère de fidélité ρ .

$$\rho_k(x, y) = \frac{1}{k} \sum_{i=0}^{k-1} \rho_k(x(i), y(i)) \quad (1.9)$$

Où ρ est une mesure de distorsion non négative entre les composantes correspondantes des séquences X et Y que l'on appelle souvent distorsion à "lettre unique" ou "sans mémoire" [1,3,4,5].

On peut définir la distorsion moyenne, pour des fonctions de probabilité et de probabilité conjointe de transition P_X et $P_{Y/X}$ données par l'expression:

Chapitre 1

$$d(P_{Y/X}) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} P_X(a_n) P_{Y/X}(b_m / a_n) \rho(a_n, b_m) \quad (1.10)$$

L'expression de l'information mutuelle s'écrit:

$$I(P_{Y/X}) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} P_X P_{X/Y}(b_m / a_n) \log_2 \frac{P_{X/Y}(b_m / a_n)}{P_Y(b_n)} \quad (1.11)$$

$$\text{où } P_{Y/X} = \sum_{n=0}^{N-1} P_x(a_n) P_{X/Y}(b_m / a_n) \quad (1.12)$$

pour une distorsion donnée D , on définit la fonction de taux de distorsion $R(D)$, pour un critère de fidélité ρ par la relation:

$$R(D) = \min_{P_{X/Y} \in P_D} I(P_{Y/X}) \quad (1.13)$$

$$\text{avec: } P_D = \{P_{Y/X} : d(P_{Y/X}) \leq D\} \quad (1.14)$$

Cette équation définit une fonction $R(D)$ qui donne pour une distorsion donnée, un débit minimal R ou réciproquement pour un débit donné, la plus faible distorsion possible D .

Ce qui nous amène à parler du codage d'une source ou d'un vecteur:

On appelle un code C de taille N et de longueur de block k un ensemble de N séquences $Y = \{y_0, y_1, \dots, y_{N-1}\}$ de dimension k chacune.

Chaque séquence de C est appelée mot de code qui est une séquence binaire de longueur $\log_2 N$ au moins. Pour coder un vecteur X dans C , il faudra choisir le mot de code qui minimise la distorsion $\rho(x, y)$, notée:

$$\rho_k(x/C) = \min_{y \in C} \rho_k(x, y) \quad (1.15)$$

On définit ainsi la distorsion moyenne:

$$\rho(C) = E[\rho_k(X/C)] = \sum_x P(x) \rho_k(x/C) \quad (1.16)$$

$$\text{avec: } P(x) = \prod_{i=0}^{k-1} P_x(x_i) \quad (1.17)$$

Pour un vecteur X de k échantillons, on définit le débit par échantillon du code C par la relation:

$$R = \frac{\log_2 N}{k} \text{ bits} \quad (1.18)$$

La théorie sur le codage nous apprend que $R(D)$ est une limite théorique des performances des systèmes de codage, et qu'il est possible d'atteindre une performance assez proche de cette limite.

Ce qui est dicté encore par l'équation ci-dessous:

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log_2 N = R(D) \quad (1.19)$$

Il est à souligner qu'il suffit de considérer, à la place des distributions P_X les fonctions de densité de probabilité avec quelques arrangements mathématiques des relations ci-dessus pour étendre leur validité au cas des sources continues.

Cependant il nous semble utile de noter certaines remarques:

- L'alphabet de reproduction des sources continues est infini.

- L'entropie différentielle des sources continues peut être positive ou négative; elle est une mesure relative au repère alors que dans le cas discret elle est absolue. Ainsi on définit l'entropie différentielle par la relation:

$$h(X) = -E[\log_2 f_X(X)] = - \int f_X(x) \log_2 f_X(x) dx \quad (1.20)$$

où $f_X(x)$ est la densité de probabilité.

L'entropie conditionnelle est défini comme suit:

$$h(X/Y) = - \iint f_{XY}(x,y) \log_2 f_{X/Y}(x/y) dx dy \quad (1.21)$$

et l'information mutuelle par:

$$I(X;Y) = \iint f_{XY}(x,y) \log_2 \frac{f_{XY}(x,y)}{f_X(x)f_Y(y)} dx dy \quad (1.22)$$

A partir des relations précédentes, si l'on considère que $\rho(x,y)$ est une mesure de distorsion, et $f_{Y/X}(y/x)$ la densité de probabilité conditionnelle, on définit la distorsion moyenne par:

$$d(f_{Y/X}) = \iint f_X(x) f_{Y/X}(y/x) \rho(x,y) dx dy \quad (1.23)$$

et l'information mutuelle moyenne par:

$$I(f_{Y|X}) = \iint f_X(x) f_{Y|X}(y/x) \log_2 \frac{f_{X|Y}(y/x)}{f_Y(y)} dx dy \quad (1.24)$$

$$\text{où } f_Y(y) = \int f_X(x) f_{Y|X}(Y/X) dx \quad (1.25)$$

On obtient finalement comme pour le cas discret, la limite théorique des performances ou la fonction $R(D)$:

$$R(D) = \min_{f_{Y|X} \in f_D} I(f_{Y|X}) \quad (1.26)$$

$$\text{où } f_D = \{f_{Y|X} : d(f_{Y|X}) \leq D\} \quad (1.27)$$

Les particularités de chaque type de cas (discret ou continu) créent une légère différence à l'origine de la fonction $R(D)$ entre les représentations graphiques de cette fonction pour les cas des sources continues et discrètes (figure 1.2).

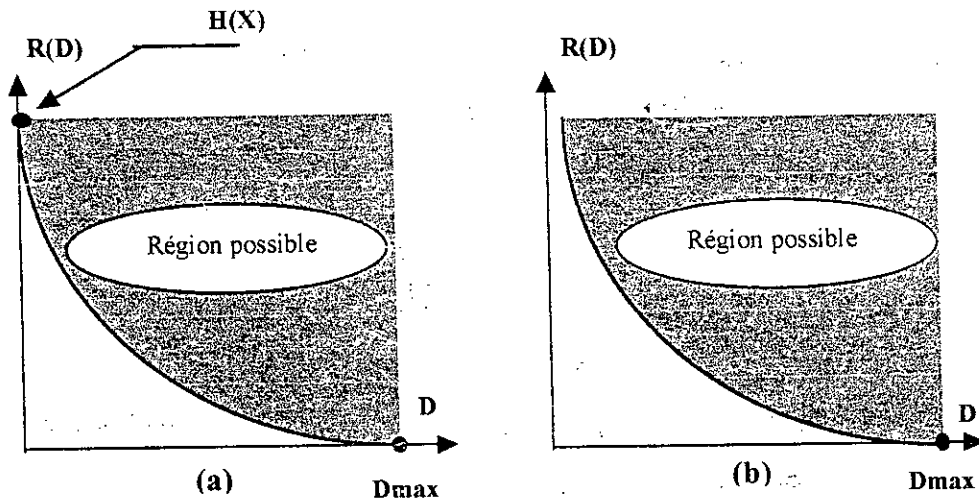


Figure 1.2: Comparaison de la fonction $R(D)$ des sources discrète (a) et continue (b).

On remarque certaines propriétés de la fonction $R(D)$ [1,3,4]

- $R(D)$ est une fonction convexe sur l'intervalle $]0, D_{\max}[$.
- D_{\max} est la distorsion maximale, elle existe toujours.
- La valeur extrême D_{\max} est obtenue comme suit:

pour le cas discret,

$$D_{\max} = \min_{y \in C} \sum_x P(x) \rho_k(x, y) \quad (1.28)$$

et le cas de source continue,

$$D_{\max} = \min_y \int f_x(x) \rho(x, y) dx \quad (1.29)$$

- R(D) est continue, monotone, décroissante.

La méthode utilisée pour la mesure de la distorsion ou le choix du critère de fidélité reste primordiale dans la conception des quantificateurs. Celle de l'erreur quadratique moyenne demeure la plus utilisée malgré quelques cas où elle s'avère inadéquate. Appelée encore MSE ou "mean squared error", elle permet d'exprimer pour une variable aléatoire gaussienne de variance σ_X^2 , la fonction de distorsion R(D) [4]:

$$R(D) = \frac{1}{2} \max \left(0, \log_2 \frac{\sigma_X^2}{D} \right) = \begin{cases} \frac{1}{2} \log_2 \frac{\sigma_X^2}{D} & 0 \leq D \leq \sigma_X^2 \\ 0 & D \geq \sigma_X^2 \end{cases} \quad (1.30)$$

La limite inférieure de la fonction R(D) est appelée la borne inférieure de Shannon (Shannon Lower Bound ou SLB). Cette caractéristique (SLB) peut être obtenue d'une façon explicite dans bien des cas ce qui ne se réalise pas pour la fonction R(D) [1,5].

La borne inférieure de Shannon s'identifie à cette dernière pour des sources gaussiennes lorsque le MSE est utilisé comme estimateur.

1.3- Principe de 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.

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.

1.4- Quantification scalaire

Dans la quantification scalaire (QS), chaque échantillon du signal d'entrée est quantifié séparément des autres échantillons. Comme l'illustre la figure 1.3, un échantillon x du signal

d'entrée est spécifié par l'indice k s'il se trouve dans l'intervalle suivant [4]:

$$R_k : \{x_k < x \leq x_{k+1}\}; \quad k=1,2,\dots,N \quad (1.31)$$

Les valeurs x_k et x_{k+1} sont appelés niveaux de décision ou seuils et N leur nombre.

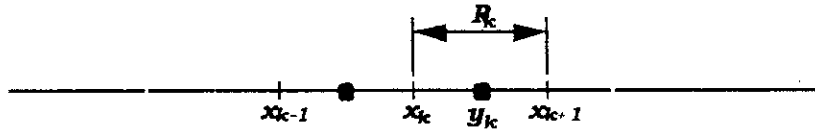


Figure 1.3: Quantification scalaire

Tous les échantillons se trouvant dans l'intervalle R_k seront remplacés par une seule valeur Y_k appelée niveau de reconstruction ou représentant.

la fonction de transfert d'un tel quantificateur est en marche d'escalier. Elle présente deux formes différentes à l'origine, selon que le nombre de niveaux de reconstruction N est pair ou impair comme le montre la figure 1.4.

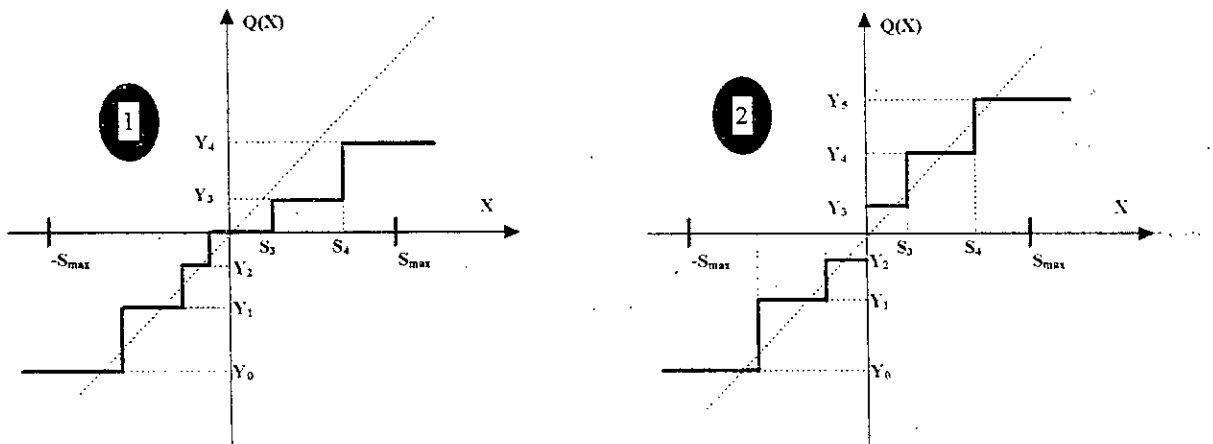


Figure 1.4: Caractéristique type d'un quantificateur scalaire pour un nombre de niveaux (1) impair et (2) pair.

Le quantificateur établit, donc une règle de correspondance entre la variable d'entrée X et une valeur de sortie approchée ou arrondie choisie dans un ensemble fini de N valeurs prédéterminées.

La formulation du problème peut être faite en considérant un ensemble B de N valeurs arrondies:

$$B = \{ Y_i \mid i = 0, 1, \dots, N-1 \}$$

La fonction de quantification s'écrirait alors :

$$Y_i = Q(X)$$

Un quantificateur fournit une seule information à la sortie mais celle-ci est disponible sous deux formes : la valeur arrondie Y et l'indice i .

Plusieurs variantes de conception des quantificateurs ont été présentées dans [6], nous les reprenons sur la figure 1.5.

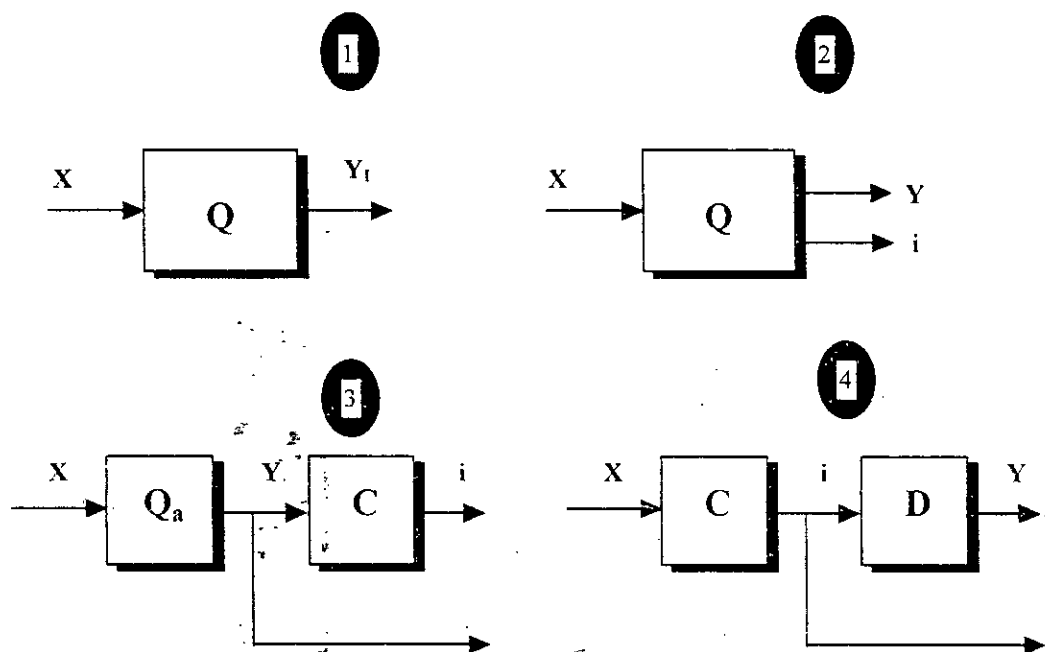


Figure 1.5: Schéma type de quantificateurs

Dans le premier cas (figure 1.5-1), le quantificateur fournit, sous une forme intégrée, une même information, dans le second cas (figure 1.5-2) celle-ci est présentée sous deux formes distinctes. Le troisième schéma décrit par la figure 1.5-3 illustre le point de vue adopté en DPCM ("Differential Pulse Code Modulation ") où la valeur arrondie Y est réinjectée dans la boucle de rétroaction alors que l'indice i est directement transmis, enfin la quatrième alternative fournit tout d'abord l'indice i , tandis que la valeur arrondie s'obtient par une table de décodage.

1.5- La quantification vectorielle

Contrairement à la quantification scalaire, le quantificateur vectoriel (QV) s'applique sur des vecteurs. Il fait correspondre à tout vecteur d'entrée x de dimension k décrit comme suit :

$x = (x(0), x(1), \dots, x(k-1))$ un vecteur de même dimension

$y_i = (y_i(0), y_i(1), \dots, y_i(k-1))$ choisi parmi un ensemble fini B de N vecteurs de reproduction ou mots de code [6]. Dans la littérature, on attribue à cet ensemble B l'appellation de dictionnaire.

Le débit binaire R d'un QV utilisant un dictionnaire B est défini par l'équation suivante:

$$R = \frac{1}{k} \log_2 N \quad (1.32)$$

Le débit représente le nombre de bits par échantillon à coder. $(\log_2 N)$ bits sont nécessaires pour représenter l'indice d'un vecteur de dimension k . Cette formulation du débit permet de faire des études comparatives de quantificateurs opérant sur des vecteurs de dimensions différentes. La figure 1.6 illustre le principe de la quantification vectorielle.

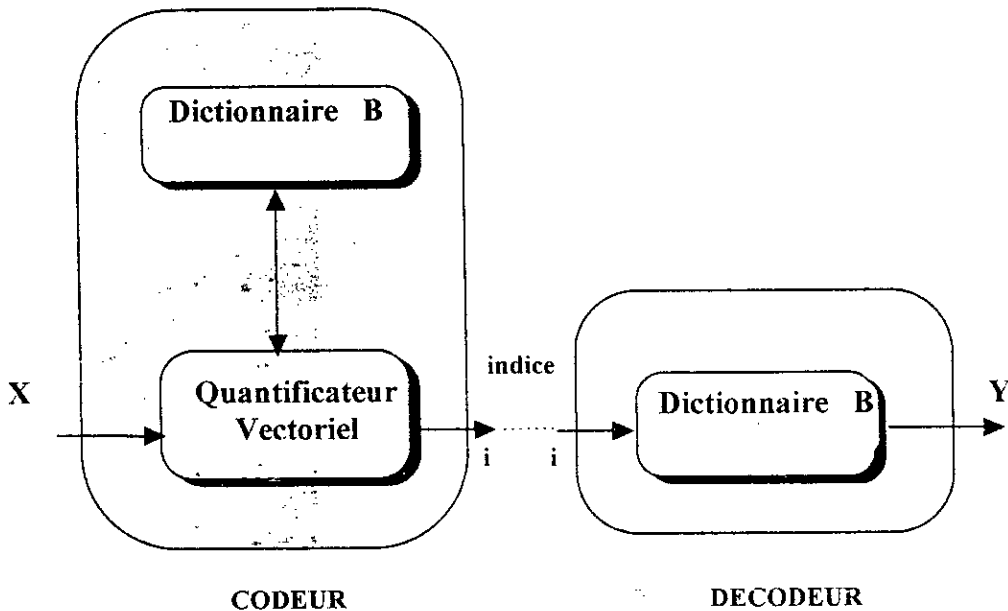


Figure 1.6: Schéma d'un quantificateur vectoriel, x est la séquence à coder, i l'indice à transmettre et y est la séquence reproduite

Dans le codeur, on associe au vecteur d'entrée x un mot de code Y_i du dictionnaire selon le critère du plus proche voisin. Seul l'indice i est transmis au décodeur. On a N indices correspondant aux N mots de code du dictionnaire B . Le même dictionnaire est utilisé au décodage. Le mot de code y_i , représentant du vecteur x , est retrouvé à partir de l'indice i reçu. La figure 1.7 illustre un exemple de quantification d'un vecteur x à deux dimensions [6].

Il est clair que la QS est une forme particulière de QV , pour une dimension k du vecteur égale à 1.

Dans la QV , on utilise le fait que les composantes du vecteur d'entrée ont généralement une corrélation et qu'elles se répartissent principalement dans une région donnée de l'espace \mathfrak{R}^k . On peut atteindre une plus petite valeur de distorsion avec le même nombre de vecteurs de reproduction qu'en QS en construisant convenablement le dictionnaire. Autrement dit, on obtiendra la même distorsion avec ces deux types de quantification en utilisant un plus petit nombre de vecteurs de reproduction.

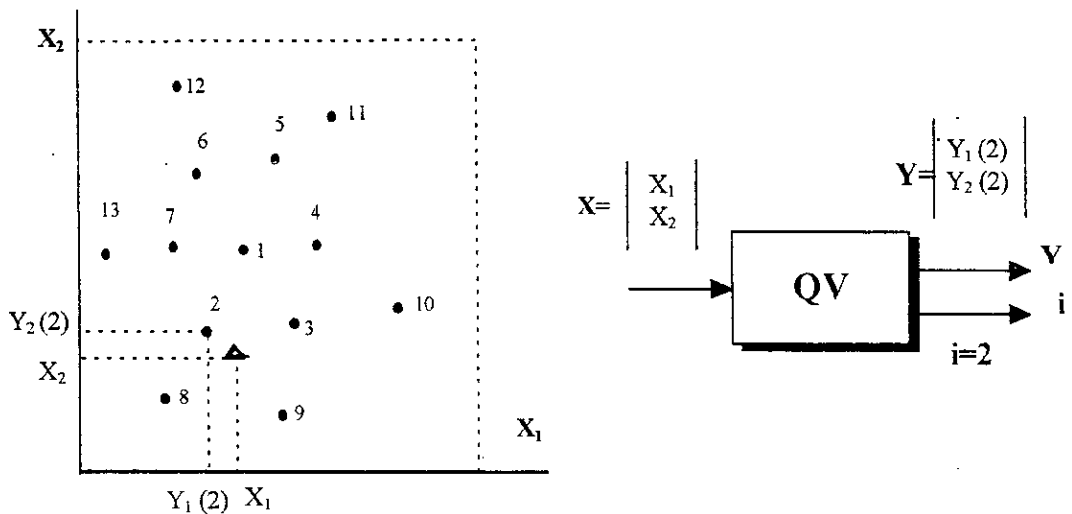


Figure 1.7: Principe de la quantification vectorielle. L'entrée x est représentée dans l'espace par un triangle. Le vecteur arrondi y le plus proche parmi les treize possibles est le vecteur d'indice $i=2$.

1.6-Mesures de performances d'un quantificateur

Un quantificateur vectoriel est complètement défini par un ensemble de vecteurs arrondis et une distance $d(x, y_i)$ définissant le degré de distorsion qui s'introduit lors d'une approximation d'un vecteur x par un vecteur approché y_i . Cette distance est une fonction réelle, non négative et doit satisfaire au principe suivant: la distance d'un point à lui même est toujours plus petite que la distance de ce point à un autre point. La forme la plus utilisée est la suivante [7]:

$$d(x, y_i) = \|x - y_i\|^r \tag{1.33}$$

Cette distance devient le carré de la distance euclidienne lorsque $r=2$. Elle exprime la puissance de l'erreur de quantification. Le principal inconvénient de la quantification vectorielle réside dans sa complexité qui croît de façon exponentielle avec la dimension et la taille du dictionnaire. La recherche exhaustive à travers les N vecteurs arrondis ou mots de code du dictionnaire nécessite une comparaison de la séquence d'entrée avec tous les N mots de code, ce qui requiert le calcul de N distances sachant que: k multiplications et $(2k-1)$ additions sont nécessaires pour estimer une seule distance euclidienne.

Cependant, les performances en terme de rapport signal à bruit obtenus par les QV s'améliorent sensiblement lorsque la dimension k augmente. De plus, la théorie de la distorsion garantit, k suffisamment grand, l'existence d'un ensemble fini de N vecteurs arrondis pour un débit R comme défini par la relation (1.17), tel que la distorsion moyenne par échantillon est proche de la distorsion minimale idéale $D(R)$. C'est pour des raisons de complexité de conception que l'on fait recours aux méthodes sous-optimales de quantification, que nous verrons plus loin.

Par ailleurs, si l'on connaît le nombre N de vecteurs arrondis y_i , la distance $d(x, y_i)$ et la distribution du vecteur d'entrée $P(x)$ on peut reconnaître le quantificateur optimal. C'est de tous

les quantificateurs, il est celui qui minimise l'espérance mathématique de la distance soit:

$$E(d(x, Q(x))) = \int d(x, Q(x)) P(x) dx. \quad (1.34)$$

Cependant approcher les performances d'un tel quantificateur n'est pas une chose aisée, aussi, pour des cas particuliers en quantification scalaire, Lloyd [8] a proposé un algorithme itératif appelé algorithme de Lloyd-Max pour construire un quantificateur optimal et Max a tabulé les valeurs arrondis [9]. Paez et Glisson ont étendu ces résultats aux densités de Laplace et Gamma [10] et Pearlman à la distribution de Rayleigh [11].

Nous donnons ci-dessous quelques aspects sur les limites des performances que peut atteindre un quantificateur sur la base des résultats théoriques obtenus par Gray et al. [12, 13] ainsi que d'autres chercheurs [14].

1.6.1-Performances asymptotiques des QV

On s'intéresse à l'étude des performances des quantificateurs vectoriels en fonction de la dimension k des vecteurs [12, 14].

Soit x une variable aléatoire dans l'espace euclidien de dimension k (\mathcal{R}^k) avec $P(x = (x_1, x_2, \dots, x_k))$ comme densité de probabilité:

$$x = (x(0), x(1), \dots, x(k-1)) \in \mathcal{R}^k$$

On définit un quantificateur vectoriel QV par l'ensemble de N vecteurs arrondis $(y_0, y_1, \dots, y_{k-1})$ et par la partition $\{S_0, S_1, \dots, S_{N-1}\}$ de \mathcal{R}^k ou

$$S_i = Q^{-1}(y_i) \subset \mathcal{R}^k$$

$$S_i \cap S_j = \emptyset \text{ pour } i \neq j$$

$$\text{et } \bigcup_{i=1}^N S_i = \mathcal{R}^k$$

On définit un quantificateur vectoriel par :

$$Q(x) = y_i \quad \text{si } x \in S_i \quad \text{pour } i = 1, 2, \dots, N. \quad (1.35)$$

La performance d'un tel QV est mesurée par la distorsion moyenne donnée par l'expression suivante:

$$D = E[d(x, Q(x))] = \sum_{i=0}^{N-1} \int_{S_i} p(x) \|x - y\|^2 dx \quad (1.36)$$

Et le débit du quantificateur Q est mesuré par son entropie H_Q :

Chapitre 1

$$H_Q = - \sum_{i=0}^{N-1} P_i \log_2 P_i \quad (1.37)$$

$$\text{Avec} \quad P_i = \int_{S_i} P(x) dx = P_X(x \in S_i) \quad (1.38)$$

soit la région $G \subset \mathfrak{R}^k$. On définit le volume $V(G)$ de G par :

$$V(G) = \int_G dx \quad (1.39)$$

On définit aussi le volume V_k de la sphère unité de dimension k par:

$$V_k = \text{Vol}(\{u: \|u\| \leq 1\}) = \frac{2\Gamma(1/2)^k}{k\Gamma(k/2)} \quad (1.40)$$

$$\text{Où} \quad \|u\| = \left(\sum_{j=0}^{k-1} u^2(j) \right)^{1/2} \quad (1.41)$$

et Γ est la fonction Gamma ($\Gamma(1/2) = \sqrt{\pi}$)

On va déterminer à présent, la distorsion moyenne minimale que peut atteindre un quantificateur Q ayant une entropie donnée H_Q sur un ensemble de points N .

a-bornes asymptotiques des performances

On considère ici, par souci de simplicité, le cas asymptotique où la performance des quantificateurs tend vers une meilleure valeur possible. Dans ce cas, le nombre de niveaux ou de vecteurs arrondis N , sera assez grand (H_Q grand). Par conséquent nous pouvons considérer que la probabilité $p(x)$ varie à l'intérieur de la région S_i , on peut donc assimiler cette dernière à p_i , si $x \in S_i$, ($p(x) \approx p_i$).

A partir des relations déjà citées nous obtenons :

$$P_i = \int_{S_i} p(x) dx = p_i V(S_i) \quad (1.42)$$

Donc l'expression de la distorsion moyenne peut être réécrite comme suit:

$$D = - \sum_{i=0}^{N-1} \frac{P_i}{V(S_i)} \int_{S_i} \|x - y_i\|^2 dx \quad (1.43)$$

on montre que [24], en prenant $L(x-y) = (x-y)^2$ que:

$$\int_S d(x, y) dx = \int_{S_i} \|x - y_i\|^2 dx \geq \frac{V(S)}{V_k} \int_{u: \|u\| \leq 1} \left\| \left(\frac{V(S)}{V_k} \right)^{1/k} \cdot u \right\|^2 du \quad (1.44)$$

Soit $F_k(V)$ la fonction:

$$F_k(V) = \frac{1}{V_k} \int_{u: \|u\| \leq 1} \left\| \left(\frac{V(S)}{V_k} \right)^{1/k} \cdot u \right\|^2 du \quad (1.45)$$

L'inégalité précédente peut s'écrire:

$$\int_S d(x, y) dx = \int_{S_i} \|x - y_i\|^2 dx \geq V(S) \cdot F_k(V(S)^{1/k}) \quad (1.46)$$

Et nous obtenons [15]

$$D \geq \sum_{i=0}^{N-1} P_i F_k(V(S_i)^{1/k}) \approx E[F_k(V(x)^{1/k})] = D_L \quad (1.47)$$

D_L est une borne inférieure aux performances asymptotiques du QV obtenues par la distorsion moyenne D .

Sachant que F_k est une fonction convexe, l'application de l'inégalité de Jensen à l'équation précédente donne:

$$D \geq D_L \geq F_k \left[E(V(x)^{1/k}) \right] \quad (1.48)$$

$$\text{où } F_k^{-1}(D) \geq F_k^{-1}(D_L) \geq E(V(x)^{1/k}) \quad (1.49)$$

A l'aide des équations (1.37) et (1.38), nous obtenons une nouvelle expression de l'entropie:

$$H_Q = \sum_{i=0}^{N-1} \int_{S_i} p(x) \log_2 p(x) dx - \sum_{i=0}^{N-1} \int_{S_i} p(x) \log_2 V(S) dx \quad (1.50)$$

On peut en déduire:

$$H_Q = h(p) - E[\text{Log}_2 V(X)] \quad (1.51)$$

Chapitre 1

où $h(p)$ est l'entropie différentielle

$$h(p) = - \int p(x) \log_2 p(x) dx \quad (1.52)$$

L'inégalité de Jensen donne:

$$H_Q = h(p) - E \left[\text{Log}_2 V(X)^{1/k} \right] \geq h(p) - k \text{Log}_2 \left(E \left(V(X)^{1/k} \right) \right) \quad (1.53)$$

En utilisant la relation (1.49) on obtient:

$$H_Q \geq h(p) - k \text{Log}_2 \left(F_K^{-1}(D_L) \right) \geq h(p) - k \text{Log}_2 \left(F_K^{-1}(D) \right) \quad (1.54)$$

Ce qui donne après avoir élevé au carré:

$$D \geq D_L \geq F_k \left(2^{-(H_Q - h(p))/k} \right) \quad (1.55)$$

La relation exprime une borne inférieure asymptotique de la distorsion moyenne obtenue par un quantificateur Q ayant une entropie fixe H_Q . En remplaçant $F_K(v)$ dans la relation ci-dessus, on obtient:

$$D \geq D_L \geq \frac{k V_K^{-2/K}}{k + 2} 2^{-(2/K)(H_Q - h(p))} \quad (1.56)$$

1.7-Algorithmes généralisés des quantificateurs vectoriels

Soit un quantificateur Q défini par son dictionnaire $B = \{ y_0, y_1, \dots, y_{N-1} \}$ et sa partition $S = \{ S_0, S_1, \dots, S_{N-1} \}$ de l'espace \mathcal{R}^k est représenté par la figure ci-dessous (figure 1.8).

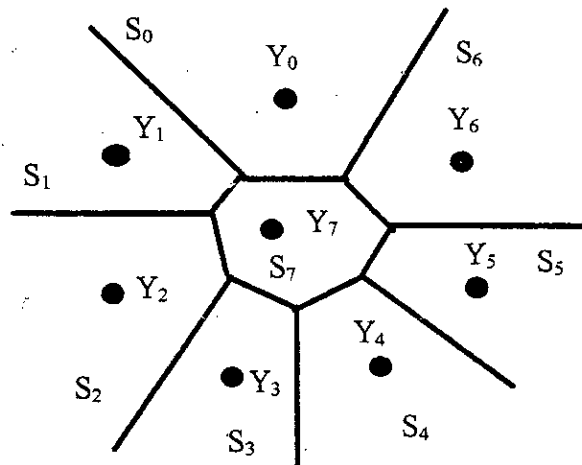


Figure 1.8: Exemple de partition de \mathcal{R}^2 pour $N=8$

Trois opérations sont nécessaires pour quantifier un vecteur donné X appartenant à l'espace \mathfrak{R}^k [16].

1- Retrouver la région de quantification contenant x , c'est à dire la région de Voronoi S_i telle que $x \in S_i$ ce qui revient à déterminer le plus proche voisin de x dans le dictionnaire B .

2- Chercher l'indice i de la région de Voronoi contenant x qui est aussi l'indice du mot de code y le plus proche de x .

3- Régénérer à la réception le mot de code y_i à partir de l'indice i .

Ces trois opérations sont représentées par la figure ci-dessous (figure 1.9).

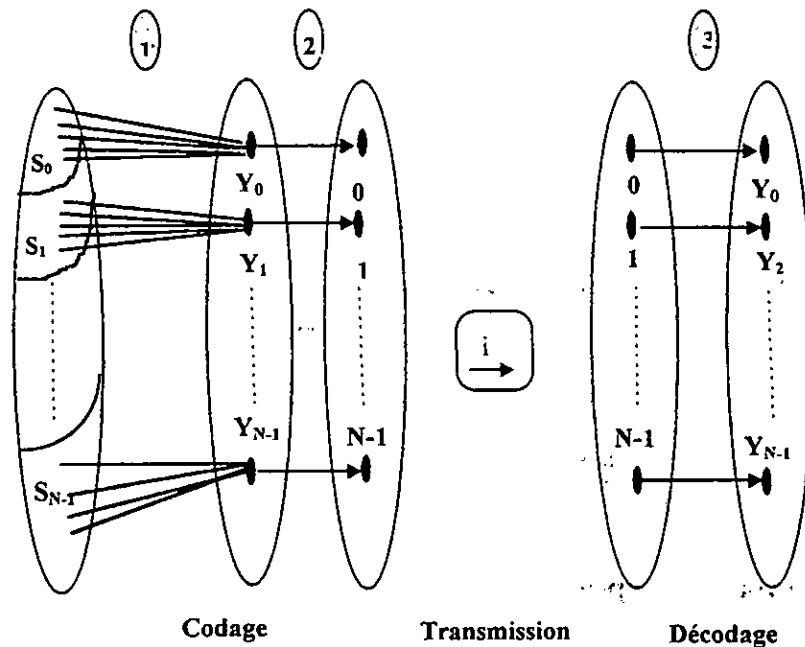


Figure 1.9: Schéma représentant les trois opérations nécessaires pour une quantification vectorielle

Il existe deux approches essentielles en quantification vectorielle, qui sont l'approche statistique et l'approche algébrique.

1.7.1-Approche statistique

Cette approche est basée sur un algorithme itératif désigné par *K-moyenne* ou *LBG* du nom de ces auteurs Lynde, Buzo et Gray [7,17].

Cette approche est recommandée lorsque la distribution des orientations sur la sphère unité est non-uniforme [15].

L'algorithme de la *k-moyenne* est généralement attribué à Mac Queen [18], Adoul et Al l'ont également utilisé dans le contexte du traitement de la parole comme objet de classification [19,20] et enfin, Linde et al le présentent comme une extension de l'algorithme de Lloyd-Max [17] pour des variables vectorielles.

Chapitre 1

Il s'agit de faire l'apprentissage, à l'aide d'un grand nombre de données en entrée, à plusieurs reprises afin de prendre en compte les caractéristiques statistiques du signal d'entrée. Les étapes de construction du dictionnaire sont les suivantes: (figure 1.10).

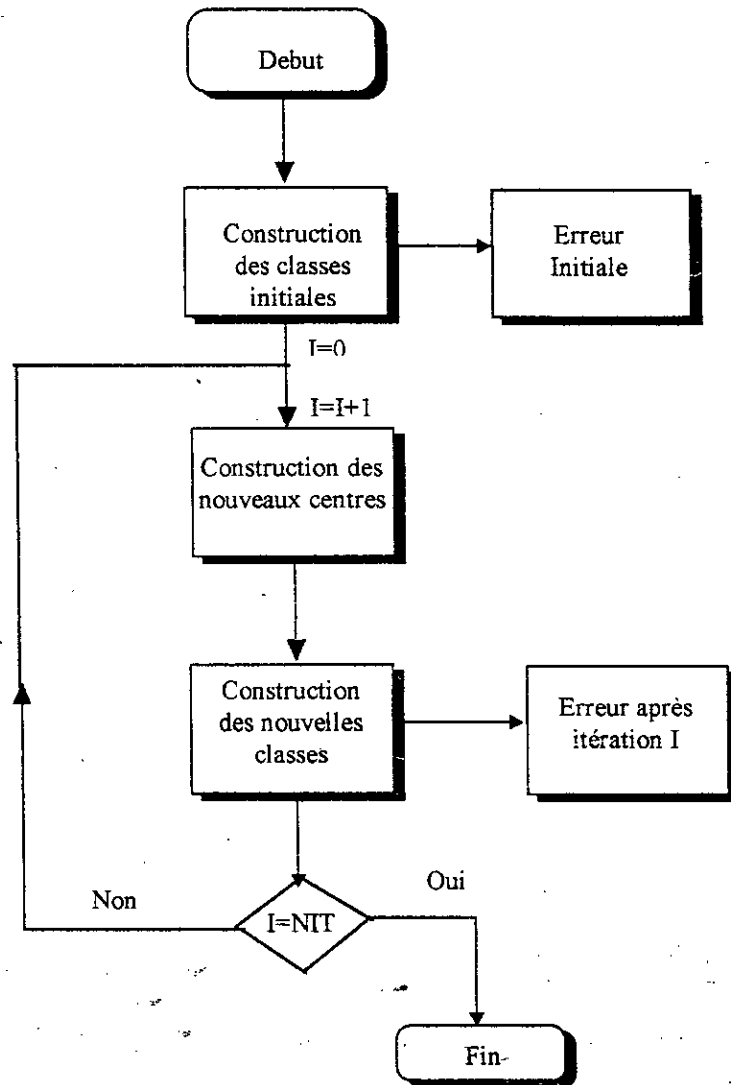


Figure 1.10: Structure de l'algorithme de la K-moyenne. NIT est le nombre d'itérations.

Soit $X = \{x_j ; j = 1, 2, \dots, N\}$ une séquence d'apprentissage composée de N vecteurs de dimension k .

1- Un dictionnaire initial $Y_0 = \{y_i ; i=1, 2, \dots, N\}$ est choisi à priori, les vecteurs qui le composent vont former les centres de départ pour une construction initiale des classes.

2- A chaque vecteur de la séquence d'apprentissage x_j est associé un vecteur le plus proche voisin y_i selon le critère de distorsion choisi. La distorsion moyenne globale est calculée sur toute la séquence d'apprentissage par sommation des distorsions estimées entre les vecteurs d'apprentissage x et y , elle est donnée par l'équation suivante:

Chapitre 1

$$D = \frac{1}{M} \sum_{j=1}^N \min_{y \in Y_0} \|x_j - y\|^2 \quad (1.57)$$

3- pour chaque classe ainsi obtenue, il va falloir calculer le nouveau centre, représentant au mieux tous les vecteurs qui lui sont associés. Le calcul se fait en cherchant à minimiser la distorsion sur la classe étudiée. La nouvelle valeur du centre remplace alors l'ancienne dans le dictionnaire.

Les centres sont calculés par l'expression suivante:

$$C = \frac{1}{N} \sum_{j=1}^N x_j \quad (1.58)$$

4- L'algorithme reprend de façon itérative au niveau de (2) tant qu'un critère d'optimalité n'est pas satisfait.

L'arrêt se produit lorsque la distorsion totale sur le dictionnaire entre deux itérations successives de la K-moyenne est inférieure à un seuil S .

$$(D_m - D_{m-1}) / D_m < S \quad (1.59)$$

La technique du QV statistique permet de bien rendre compte de la distribution du signal à coder, mais la taille du dictionnaire et la dimension du vecteur d'entrée sont sévèrement limités par des contraintes technologiques en matière de stockage et de complexité de calcul. Plusieurs techniques ont été proposées pour contourner le problème de complexité, comme l'organisation des dictionnaires en arbre, les quantificateurs vectoriels en Lattice (Lattice VQ), la quantification vectorielle à étages, la quantification vectorielle par split et quantificateurs vectoriels Gain et Forme (gain shape vector quantizers). Nous détaillerons quelques techniques dans les paragraphes suivants.

1.7.1.1-Dictionnaires organisés en arbre (hiérarchique)

Les quantificateurs à recherche par arbre (*tree search vector quantizer*) ou TSVQ ont été introduit pour la première fois pour une utilisation dans le codage par prédiction linéaire [21], pour la réduction de la complexité au prix d'une dégradation de performances en terme de RSB (rapport signal à bruit) et une mémoire plus large comparée à la recherche systématique des dictionnaires. Le processus de recherche est un arbre binaire.

Dans la recherche hiérarchique, une grande partie de l'ensemble des chemins stockés dans le dictionnaire sont utilisés uniquement pour guider la recherche. Par exemple, dans un dictionnaire organisé en arbre binaire de taille 2^{kr} , la recherche consiste en kr décisions binaires uniquement à la dernière étape de décision, le vecteur est comparé avec des chemins qui sont les mots de codé qui représentent le vecteur d'entrée. La complexité de calcul pour les dictionnaires organisés en arbre croît linéairement avec la dimension; la mémoire de stockage peut être de deux fois supérieur à celle de la recherche complète.

Le codage par TSVQ nécessite le même taux de stockage que la recherche systématique mais sa complexité de calcul est proportionnelle uniquement au logarithme de la taille du dictionnaire, ce qui constitue une amélioration considérable.

Chapitre 1

Les performances en terme de débit-distorsion de la *TSVQ* sont bonnes mais inférieures à celles de la recherche exhaustive car d'une part, le dictionnaire possède une structure spécifique. Ce qui permet une recherche rapide; d'autre part, la recherche ne permet pas de trouver nécessairement le vecteur code le plus proche. En terme quantitatif, l'erreur quadratique moyenne de la *TSVQ* est plus grande que celle de la recherche systématique d'environ 0.7 dB pour une dimension de deux, 1 dB pour une dimension de 4 et 2.7 dB pour des dimensions plus grandes [22]. Cependant, la diminution de la complexité de calcul de la *TSVQ* la rend plus intéressante.

1.7.1.2-Quantification à étages

La quantification à étages ou *Multistage* a été proposée pour la première fois pour la quantification des paramètres LPC par Juang [23]. C'est une technique qui permet de réduire la complexité et la mémoire de stockage. Dans ce type de quantificateur, le vecteur courant est quantifié une première fois. Le dictionnaire utilisé est produit selon l'algorithme LBG ou méthode hiérarchique. Le nombre de représentant dans ce dictionnaire est assez faible, et produit une erreur de quantification importante. Le vecteur erreur est calculé par soustraction du vecteur quantifié par le vecteur original. Ce dernier est envoyé à un second étage de quantification. Ceci est illustré par la figure 1.11.

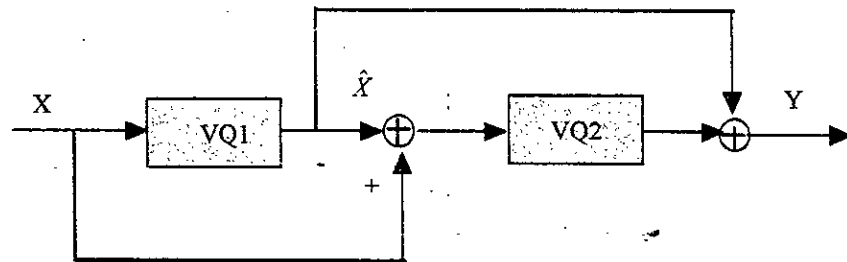


Figure 1.11: Quantificateur vectoriel à 2 étages. La version quantifiée Y du vecteur X est égale à la somme des deux sorties des quantificateurs (VQ1, VQ2).

La forme finale est obtenue par sommation vectorielle des sorties des deux étages.

On peut formuler la quantification vectorielle à étages comme suit [24,25,26]:

Soit X le vecteur de dimension m, contenant les valeurs des paramètres à quantifier et \hat{X} sa version quantifiée qui peut être approximé par l'expression suivante:

$$\begin{aligned}\hat{X} &= Y_0^{(l_0)} + Y_1^{(l_1)} + \dots + Y_{k-1}^{(l_{k-1})} \\ &= B_0^{(l_0)} C_0 + B_1^{(l_1)} C_1 + \dots + B_{k-1}^{(l_{k-1})} C_{k-1} \\ &= B C\end{aligned}$$

Où k est le nombre d'étages et $Y_j^{(k)}$ est le $k^{\text{ème}}$ vecteur code de l'étage j . Le dictionnaire du $j^{\text{ème}}$ étage est composé de L_j vecteurs codes et le vecteur C_j est créé par stockage des vecteurs codes (pour le jème étage).

$$C_j = \left[y_j^{(0)T} y_j^{(1)T} \dots y_j^{(L_j-1)T} \right]^T$$

Le vecteur C_j représente le dictionnaire du $j^{\text{ème}}$ étage.

Avec cette approche, la complexité et le stockage peuvent être considérablement réduits en comparaison à un quantificateur à un seul étage au même débit. Si on utilise k étages, la dimension m des vecteurs augmente proportionnellement au nombre d'étages, ainsi en gardant (m/k) constant, la complexité augmente uniquement de manière quadratique avec la dimension (qui croît en m^2) plutôt que d'une manière exponentielle comme dans le cas de la QV à un seul étage. Cette complexité peut être considérablement réduite par des dictionnaires structurés en arbre au niveau de chaque étage [27].

1.7.1.3-Quantification par Split

Dans la quantification par *Split* (SVQ), le vecteur des paramètres à quantifier x est divisé en plusieurs (k) blocs $[x_0 \ x_1 \ \dots \ x_{k-1}]$ et chaque bloc est quantifié séparément par un quantificateur vectoriel à recherche exhaustive [26].

On sait que la SVQ réduit la complexité au prix d'une dégradation de performance. D'où il y a un compromis entre la complexité et les performances, qui détermine le nombre de blocs à prendre dans le split.

1.7.2-Approche algébrique

L'inconvénient majeur de la quantification vectorielle statistique vient de la détermination du représentant le plus proche voisin d'un vecteur à coder qui se fait par une recherche exhaustive dans un dictionnaire stocké en mémoire. Ceci limite l'emploi de dimension importante. Des méthodes de codage accélérés et sous-optimales ont été développées mais conduisant généralement à un accroissement de l'allocation mémoire de stockage. Il serait donc intéressant de construire un dictionnaire virtuel qui nous évite le stockage de la totalité des représentants. Ce dernier doit posséder également une structure permettant un codage rapide des vecteurs. De telles structures existent et exploitent les propriétés des réseaux réguliers de points dans l'espace à plusieurs dimensions.

Dans son article fondamental sur la quantification vectorielle, Gersho [12] entrevoit l'utilisation de réseaux réguliers pour la quantification et met l'hypothèse que dans le cas d'une variable uniformément répartie, le quantificateur optimal est un réseau régulier, mais aucune démonstration ne venait le confirmer avant 1982 pour les dimensions deux et trois.

Les réseaux réguliers de points sont déjà utilisés dans le domaine de codage du canal en codes correcteurs d'erreurs, ils constituent une réponse au problème de l'empilement des sphères ou il s'agit de trouver les coordonnées des centres de sphères tangentes, de rayon unité et qui remplissent au mieux l'espace \mathcal{R}^k [28].

Une importante contribution à l'étude des réseaux est due à Leech, Conway et Sloane [29,30].

1.7.2.1-Quantification vectorielle par réseaux

Soit $\{z_1, z_2, \dots, z_M\}$ un ensemble de vecteurs dans l'espace R^M linéairement indépendants, un réseau est défini par un ensemble de points x tels que [6]:

$$L_M = \left\{ x / x = \sum_{i=1}^M k_i z_i \right\}; \quad k_i \text{ entier.}$$

Les coefficients k_i sont les coordonnées des points x dans le repère défini par $\{k_1, k_2, \dots, k_M\}$.

De cette définition nous pouvons faire quelques remarques:

- un réseau est invariant par translation, ce qui nous permettra d'emmener son origine à l'un quelconque de ses points.
- le réseau est engendré par une formule constructive indépendante des données à quantifier et sera utilisé pour des répartitions essentiellement uniformes; ce dernier point est important, pour une quantification efficace, il permet d'adapter les données au réseau par une compression appropriée, ou à l'inverse de transformer le réseau pour l'adapter à la répartition des données.

Plusieurs méthodes de construction de réseaux font appel aux codes correcteurs d'erreurs.

Un des réseaux réguliers les plus performants est le réseau de Gosset en 8 dimensions E_8 .

Le réseau D_n contient les points de Z_n , qui est l'ensemble des vecteurs de R^n à coordonnées entières et dont la somme des coordonnées est paire. En deux dimensions, D_2 a la forme d'un damier et en trois dimensions, D_3 correspond au réseau cubique à face centrées. Ces réseaux sont montrés par la figure 1.12.

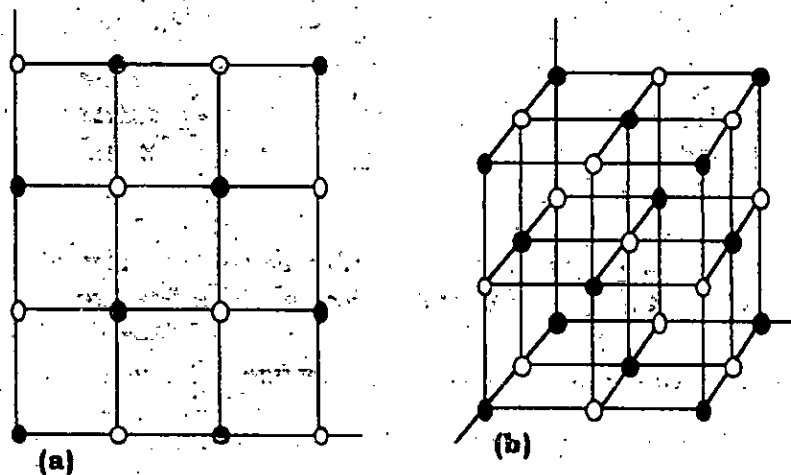


Figure 1.12: Réseau D_n pour 2 et 3 dimensions.

car l'indice se déduit directement de l'adresse du représentant du dictionnaire. Conway et Sloane [30] ainsi que Schalkwijk [31] ont décrit des algorithmes permettant de trouver rapidement l'indice du représentant dans la collection.

Conclusion

Dans ce chapitre, nous avons présenté l'essentiel de la théorie de la quantification. Cette partie permet de situer le problème de design des quantificateurs. Les limites théoriques des performances des quantificateurs restent la référence.

Nous avons effectué une revue des méthodes de quantification vectorielle, cette revue n'est pas exhaustive et des efforts considérables sont faits dans ce domaine. Notre but n'est pas ici de toutes les tester.

Chapitre 2

Le codage de la parole

Introduction

Pour des raisons économiques, on cherche toujours à réduire le débit des informations vocales transmises sur le canal sans trop affecter la qualité des messages. L'objectif principal du codage de la parole consiste à atteindre le débit le plus faible possible pour une qualité donnée de la parole codée [4].

Dans ce chapitre, nous allons nous intéresser au signal parole, nous verrons quelles sont ses propriétés. Nous décrirons les techniques utilisées pour transmettre la parole à bas débits.

Les méthodes de codage qui seront exposées, utilisent largement la prédiction linéaire.

En prédiction linéaire, on suppose que la valeur du signal à un instant donné peut être approximativement prédite à partir d'un certain nombre de valeurs précédentes.

Pour utiliser cet outil mathématique, nous devons auparavant avoir numériser le signal que nous étudions.

2.1- Numérisation d'un signal

L'origine physique des signaux informationnels tel que la parole ou l'image est intrinsèquement analogique et continue dans le temps. D'où une conversion analogique numérique (A/D) est indispensable pour une représentation numérique de ces formes d'ondes. L'opération de conversion (A/D) peut être vu comme étant composée de quatre opérations: le préfiltrage, l'échantillonnage, la quantification et le codage.

2.2- Le signal parole

La parole est le mécanisme de communiquer la pensée par un système de sons articulés. Le signal parole est produit par un ensemble d'organes, qu'on regroupe sous le terme d'appareil vocal. La figure 2.1 représente une section transversale de la tête humaine.

Ce système vocal se compose des poumons, de la trachée artère, du larynx et du conduit vocal formé des cavités buccale et nasale.

Dans le processus de production de la parole, les poumons se comportent comme un soufflet insufflant de l'air à travers la trachée artère jusqu'au larynx. Leur rôle est de fournir l'énergie nécessaire à la phonation.

Le larynx est un organe complexe, dont le rôle essentiel est de régler la tension des cordes vocales.

L'élément principal dans la production des sons est constitué des cordes vocales. Une transformée de Fourier appliquée sur le signal glottal, donne un spectre de raies harmoniques espacées de F_0 qui correspond à la fréquence de vibration des cordes vocales, appelée *fréquence fondamentale* ou *pitch*. La fréquence fondamentale peut varier:

- de 80 à 200 Hz pour les hommes,
- de 150 à 450 Hz pour les femmes,
- de 200 à 600 Hz pour les enfants.

Cette différence est due à la longueur et à la masse des cordes vocales.

Le conduit vocal est constitué des cavités se situant au-dessus du larynx.

La partie orale comprend le pharynx et la cavité buccale qui est de forme et de volume très variables suivant la position des différents articulateurs que sont: la langue, la mâchoire inférieure et les lèvres.

La partie nasale, quant-à-elle, est formée des deux fosses nasales de géométrie fixe placées en parallèle au conduit oral et couplées de façon variable avec celui-ci par abaissement du voile du palais.

Ces cavités jouent le rôle de résonateur, privilégiant certaines zones fréquentielles, qui sont les formants. Ces derniers représentent les pôles de la fonction de transfert associée au conduit vocal. Les anti-formants représentent les zones affaiblies dans le cas d'un couplage entre le conduit oral et nasal, ce sont les zéros de la fonction de transfert.

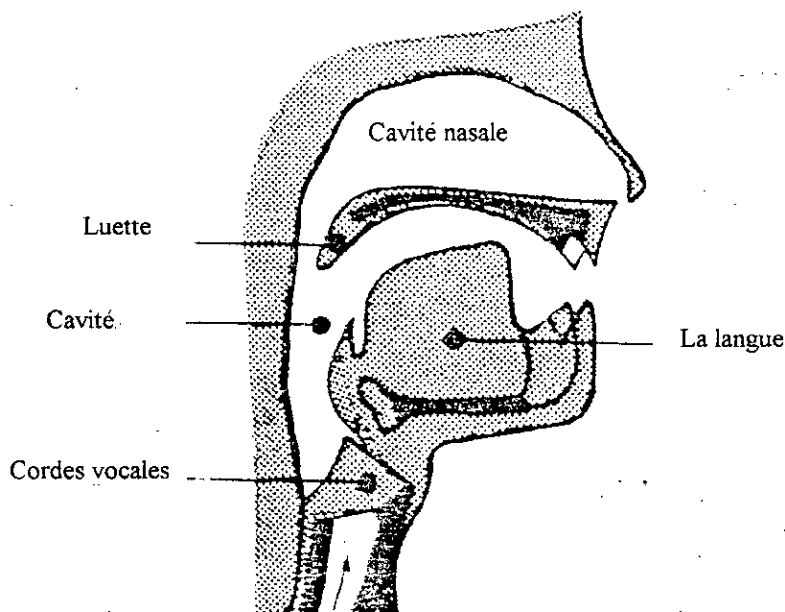


Figure 2.1: Section transversale de la tête humaine

Le système vocal présente trois modes de fonctionnement différents et qui s'entrelacent lors de la phonation.

Dans le premier mode de fonctionnement, un train quasi-périodique et riche en harmoniques, appelé *son voisé* est produit par les cordes vocales. C'est le modèle de production des voyelles, des consonnes nasales, ainsi que certaines consonnes fricatives ([v],[z],...) et occlusives ([b],[d],...). La figure 2.2 montre la forme d'onde et le spectre d'un son voisé.

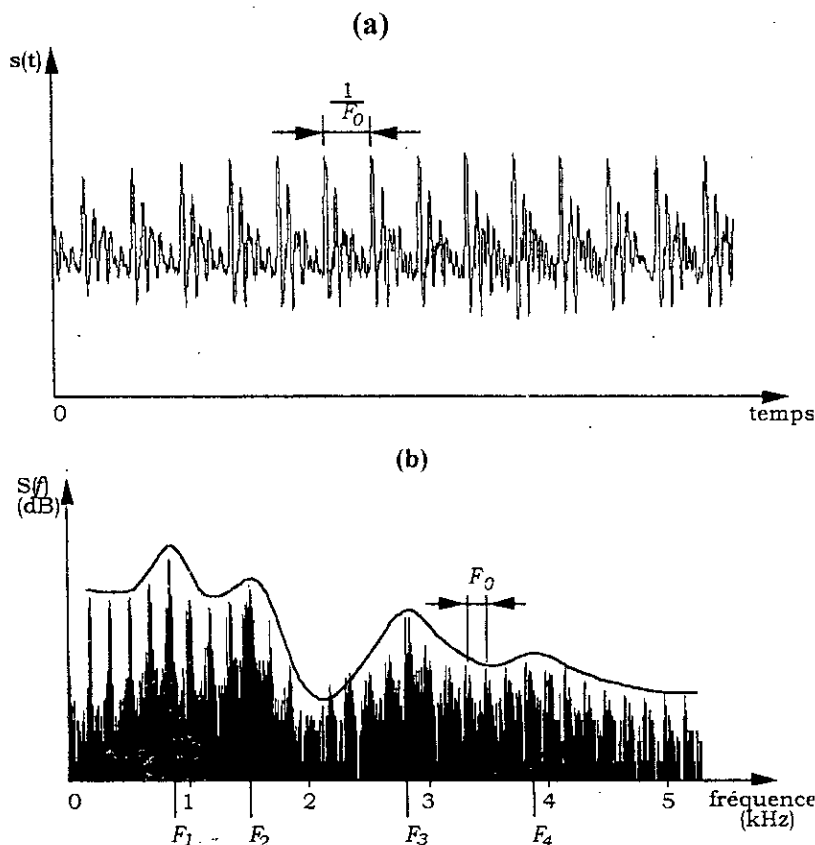


Figure 2.2 : Forme d'onde (a) et spectre (b) du son voisé.

Dans le second mode, le flux d'air venant des poumons ne sert pas à exciter les cordes vocales mais produit un bruit d'écoulement à un point de constriction du conduit vocal. Le spectre de ce son est riche en composantes aiguës et devient plus compliqué à mesure que le point de constriction se déplace vers l'arrière mettant en jeu une plus grande longueur du conduit vocal. Ce mode de production correspond aux *sons non voisés*, tels que les consonnes fricatives ([f] et [s]), les consonnes occlusives ([p] et [t]), etc. La figure 2.3 décrit la forme d'onde et le spectre d'un son non voisé.

Le dernier mode est celui des consonnes plosives. Le signal est produit par une brusque ouverture du conduit vocal permettant d'égaliser la pression avant et arrière de l'occlusion. La séquence de production de ce son commence par un court silence pendant lequel le conduit vocal est fermé; puis vient ensuite une impulsion au moment où l'air recommence à circuler, l'ouverture du conduit vocal reste faible succède alors un bruit de constriction, dans la dernière étape les cordes vocales peuvent alors être excitées.

Durant la production de la parole, les formes du conduit vocal et de l'excitation varient avec le temps. En général, ces variations se produisent d'une façon relativement lente. Pour simplifier l'analyse, on suppose que le signal de parole est stationnaire pour de courts intervalles compris entre 10 ms et 30 ms.

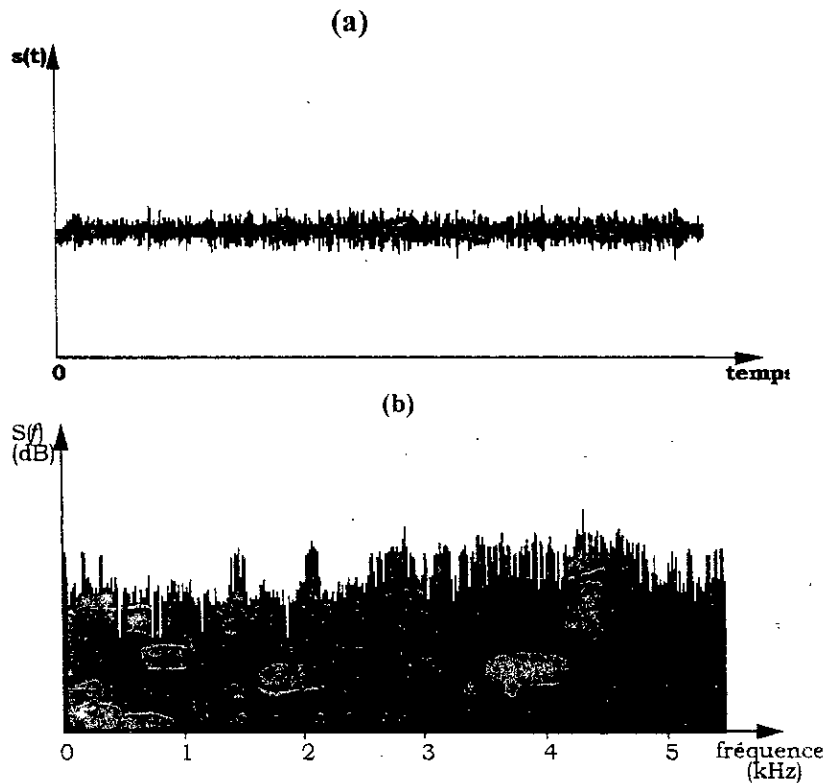


Figure 2.3: Forme d'onde (a) et spectre (b) du son non voisé.

L'étude du mécanisme de la phonation montre que le modèle du conduit vocal se comporte comme un filtre tous pôles dont la fonction de transfert est [4,32]:

$$H(z) = \frac{G}{A(z)} = \frac{G}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (2.1)$$

ou p est l'ordre du modèle, les coefficients a_k sont les paramètres du modèle, G représente le facteur de gain. La fonction de transfert $H(z)$ est appelée filtre de synthèse et $A(z)$ est appelée filtre inverse ou filtre d'analyse.

Le signal de parole est donc modélisé par un signal d'excitation attaquant ce filtre dans les paramètres varient dans le temps. Bien que la nature du signal d'excitation soit variable, elle présente deux situations extrêmes: une série d'impulsions quasi-périodiques, pour les sons voisés et un bruit blanc pour les sons non voisés. La figure 2.4 et 2.5 représentent le modèle de production de la parole.

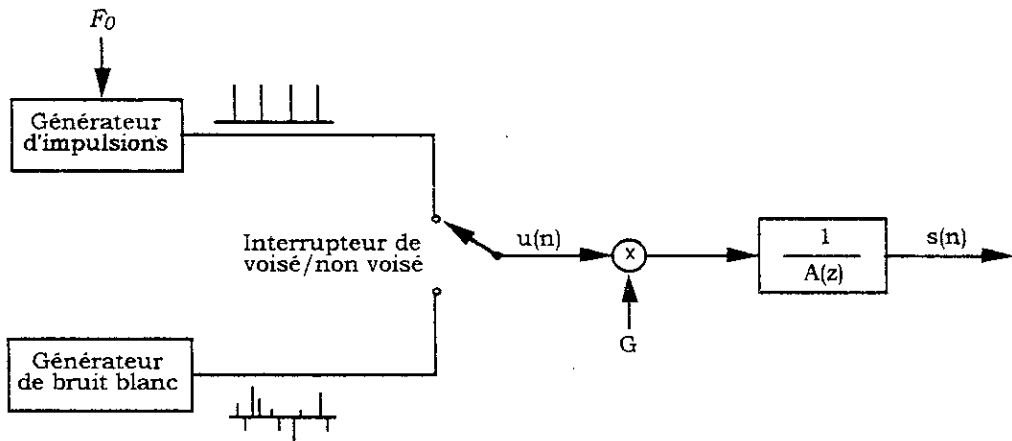


Figure 2.4: Modèle simplifié de production de la parole.

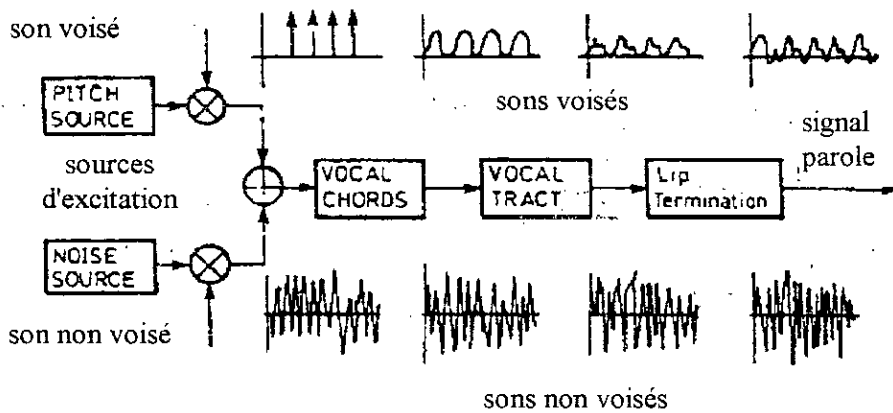


Figure 2.5: Modèle de production de la parole

2.3-La prédiction linéaire

La prédiction linéaire est assez bien utilisée dans les systèmes de codage et de compression [33,34,35]. 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 prédicateur 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 2.4.

La fonction de transfert du modèle de la production de la parole est décrite par (2.1). Dans le domaine temporel, cette équation devient :

$$s(n) = \sum_{k=1}^p a_k s(n-k) + Gu(n) \quad (2.2)$$

Ainsi, chaque échantillon de la parole $s(n)$ est constitué par une combinaison linéaire de p échantillons passés de la parole. Le prédicteur est défini comme un système dont la sortie est:

$$\tilde{s}(n) = \sum_{k=1}^p a_k s(n-k) \quad (2.3)$$

Où les a_k sont appelés coefficients de prédiction. La fonction de système d'un prédicteur linéaire d'ordre p peut s'exprimer sous la forme suivante:

$$P(z) = \sum_{k=1}^p a_k z^{-k} \quad (2.4)$$

La différence $e(n)$ 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:

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^p a_k s(n-k) \quad (2.5)$$

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 [32].

Pour trouver les coefficients a_k , on prend la transformée Z de l'équation (2.5), on obtient alors:

$$E(z) = S(z)A(z) \quad (2.6)$$

La moyenne de l'erreur de prédiction est défini par:

$$\begin{aligned} E &= \sum_n e^2(n) \\ &= \sum_n \left[s(n) - \sum_{k=1}^p a_k s(n-k) \right]^2 \end{aligned} \quad (2.7)$$

Pour trouver les valeurs des $\{a_k\}$ qui minimisent E , on met:

$$\frac{\partial E}{\partial a_i} = 0 \quad \text{pour } i = 1, \dots, p. \text{ Alors}$$

$$\frac{\partial E}{\partial a_i} = -2 \sum_n \left\{ \left[s(n) - \sum_{k=1}^p a_k s(n-k) \right] s(n-k) \right\} = 0 \quad (2.8)$$

Cette dernière équation nous conduit à écrire:

$$\sum_n s(n)s(n-i) = \sum_n \sum_{k=1}^p a_k s(n-k)s(n-i) \quad (2.9)$$

En faisant changer l'ordre de sommation de l'équation (2.8), on obtient:

$$\sum_n s(n)s(n-i) = \sum_{k=1}^p a_k \sum_n s(n-k)s(n-i) \quad i = 1, 2, \dots, p \quad (2.10)$$

Si on admet que:

$$\phi(i, k) = \sum_n s(n-i)s(n-k) \quad (2.11)$$

Alors l'équation (2.10) peut être écrite comme suit:

$$\sum_{k=1}^p a_k \phi(i, k) = \phi(i, 0), \quad i = 1, 2, \dots, p. \quad (2.12)$$

Cet ensemble de p équations à p inconnus peut être résolu d'une manière efficace pour les coefficients de prédiction inconnus $\{a_k\}$.

On doit calculer d'abord $\phi(i, k)$ pour $i = 1, \dots, p$ et $k = 0, \dots, p$. Pour calculer $\phi(i, k)$ à partir de l'équation (2.12), on doit spécifier les limites de sommation. Pour une analyse à temps court (short time), les limites doivent être sur un intervalle fini. Deux méthodes pour l'analyse par prédiction linéaire ont émergé sans considération des limites de sommation: la méthode d'autocorrélation et la méthode de covariance [36,37].

a-La méthode d'autocorrélation

Dans cette approche, on suppose que l'erreur dans équation (2.7) est calculée sur une durée infinie $-\infty < n < +\infty$, puisque ceci est irréalisable en pratique, on suppose que le segment de la parole est nul en dehors de l'intervalle $0 < n < L_a - 1$, ou L_a est la longueur de la fenêtre de l'analyse LPC. Ceci est équivalent à multiplier le signal parole d'entrée par une fenêtre de longueur finie.

Considérons l'équation (2.7), $e(n)$ est non nulle uniquement sur l'intervalle $0 < n < L_a + p - 1$. Ainsi

$$\phi(i, k) = \sum_{n=0}^{L_a+p-1} s(n-i)s(n-k) \quad \begin{array}{l} i = 1, \dots, p, \\ k = 0, \dots, p. \end{array} \quad (2.13)$$

On pose $m = (n - i)$, l'équation (2.13) peut être exprimée par:

$$\phi(i, k) = \sum_{m=0}^{L_a-1-(i-k)} s(m)s(m+i-k) \quad (2.14)$$

Donc, $\phi(i, k)$ est l'autocorrélation de $s(m)$ évaluée sur $(i-k)$. D'où

$$\phi(i, k) = R(i-k) \quad (2.15)$$

$$\text{Où } R(j) = \sum_{n=0}^{L_a-1-j} s(n)s(n+j) = \sum_{n=j}^{L_a-1} s(n)s(n-j) \quad (2.16)$$

Donc, l'ensemble des équations dans (2.12) peut être exprimé par la relation suivante:

$$\sum_{k=1}^p a_k R(|i-k|) = R(i), \quad i = 1, \dots, p. \quad (2.17)$$

L'équation (2.17) peut être exprimée sous une forme matricielle comme suit :

$$\begin{pmatrix} R(0) & R(1) & R(2) & \dots & R(p-1) \\ R(1) & R(0) & R(1) & \dots & R(p-2) \\ R(2) & R(1) & R(0) & \dots & R(p-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R(p-1) & R(p-2) & R(p-3) & \dots & R(0) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} R(1) \\ R(2) \\ R(3) \\ \vdots \\ R(p) \end{pmatrix} \quad (2.18)$$

La matrice, de dimension $p \times p$, des valeurs d'autocorrélation est une matrice de Toeplitz symétrique, tous les éléments d'une diagonale donnée sont égaux. Cette propriété peut être exploitée pour obtenir un algorithme efficace de résolution du système équations (2.18).

La solution la plus efficace est une méthode itérative connue sous le nom de l'algorithme de Wiener Livinson Durbin [36], le développement de cet algorithme se trouve dans l'annexe A.

b-La méthode de covariance

Contrairement à la méthode d'autocorrélation, ici, on suppose que l'erreur dans équation (2.7) est minimisée sur un intervalle fini, $0 \leq n \leq L-1$, donc $\phi(i, k)$ de équation (2.11) sera donnée par:

$$\phi(i, k) = \sum_{n=0}^{L-1} s(n-i)s(n-k) \quad \begin{matrix} i = 1, \dots, p \\ k = 0, \dots, P \end{matrix} \quad (2.19)$$

L'ensemble des équations dans (2.12) peut être écrit sous une forme matricielle par:

$$\begin{pmatrix} \phi(1,1) & \phi(1,2) & \phi(1,3) & \cdots & \phi(1,p) \\ \phi(2,1) & \phi(2,2) & \phi(2,3) & \cdots & \phi(2,p) \\ \phi(3,1) & \phi(3,2) & \phi(3,3) & \cdots & \phi(3,p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi(p,1) & \phi(p,2) & \phi(p,3) & \cdots & \phi(p,p) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} \phi(1,0) \\ \phi(2,0) \\ \phi(3,0) \\ \vdots \\ \phi(p,0) \end{pmatrix} \quad (2.20)$$

Puisque $\phi(i,k) = \phi(k,i)$, la matrice de dimension $p \times p$ est symétrique mais pas de forme Toeplitz. Cholesky a proposé une décomposition pour la résolution du système équations (2.20) [38].

Contrairement à la méthode d'autocorrelation, la méthode de covariance n'assure pas toujours la stabilité du filtre tous pôles $H(z)$. Une méthode de covariance stabilisée est employée pour garantir la stabilité du filtre $H(z)$ [39].

De nos jours, les LSF (*Line Spectral Frequencies*) présentent le plus d'intérêt. Ces paramètres issus des coefficients de prédiction s'adaptent mieux à la quantification.

2.4-Techniques de codage de la parole

Nous avons vu aux paragraphes précédents comment étaient calculés les coefficients de prédiction linéaire à partir d'une trame du signal de parole.

Nous allons maintenant rappeler les techniques de codage de la parole. Après un bref rappel sur les codeurs d'onde, nous nous intéressons plus particulièrement aux codeurs LPC.

Les travaux sur le codage de la parole ont donné naissance à une multitude de codeurs. Mais, on distingue trois familles:

- le codage des formes d'onde (*Waveforms coding*).
- les codeurs paramétriques qui extraient du signal un ensemble de paramètres plus propices à une transmission économique, le codage par prédiction linéaire en est un exemple.
- les codeurs hybrides qui associent à la modélisation par prédiction linéaire du signal, le codage temporel du résidu de l'analyse LPC.

2.4.1-Les codeurs de forme d'onde.

Les codeurs de forme d'onde sont les plus anciens. La version la plus simple est le codeur PCM (Pulse Code Modulation) dont le principe repose entièrement sur la quantification des échantillons d'un signal.

La seconde génération des codeurs PCM, utilise la corrélation à court terme en quantifiant la différence entre les échantillons successifs. Dans cette catégorie, on trouve la modulation delta; dans une troisième approche, le signal à un instant est prédit en fonction des

M échantillons de signal précédents, l'erreur de prédiction est alors quantifiée: PCM-Différentiel (DPCM).

Pour tenir compte de la non-stationnarité importante du signal de parole, le quantificateur ainsi que le prédicteur peuvent être adaptés dans le temps. C'est ainsi qu'a été mis au point le codeur ADPCM (Adaptif Différentiel Pulse Code Modulation).

L'aspect prédictible du signal de parole affecte deux domaines différents. Le premier, à long terme est dû à un caractère pseudo-périodique de la parole alors que le second, à court terme, provient du caractère résonant de celui-ci et qui est exploitée par les codeurs ADPCM ainsi que dans l'analyse par prédiction linéaire.

Dans les codeurs à base de prédiction linéaire, les modèles d'analyse et de synthèse restent les mêmes que définis précédemment: le conduit vocal est modélisé par un filtre tout pôle, et les différences que nous allons noter entre les codeurs viennent du traitement réservé au signal résiduel qui est obtenu par filtrage inverse du signal de parole et qui représente le signal d'excitation pour la synthèse.

2.4.2-Les codeurs paramétriques

2.4.2.1-Les vocodeur LPC

Ce codeur est destiné à fonctionner pour des débits numériques inférieurs à 5 Kbits/s.

Le filtre de synthèse est réactualisé toutes les 20 ms environ. Le signal d'excitation du filtre de synthèse répond aux hypothèses de départ concernant la modélisation par prédiction linéaire.

Dans le cas des sons voisés le signal d'excitation est composé d'impulsions périodiques, de période égale à la période de voisement.

Dans le cas d'un son non voisé le signal d'excitation du filtre est un bruit blanc. Le problème est alors de détecter le voisement et de déterminer la fréquence fondamentale.

Le problème est alors de détecter si le signal est voisé ou non ensuite de mesurer la valeur de la fréquence fondamentale.

2.4.2.1-Codeurs à analyse par synthèse

Le codeur AVPCM (Adaptive Vector Pulse Code Modulation) fonctionne à des débits proches de 32 kbits / s, pour atteindre des débits inférieurs, il est nécessaire de quantifier le signal résiduel sur un nombre de niveaux moins élevé. Une quantification des échantillons sur au plus 2 bits produit un bruit de codage important, l'erreur de quantification se répercutant sur toute la durée de la réponse impulsionnelle du filtre de synthèse.

Ce phénomène de cumul des erreurs peut être évité en prenant en compte à la fois l'erreur de quantification instantanée, mais également l'erreur produite aux instants suivant de filtrage.

Depuis 1982 sont apparus un grand nombre de codeurs qui tous ont en commun le fait que le signal résiduel n'est plus quantifié mais modélisé par une forme d'onde. La forme d'onde optimale étant celle qui après synthèse minimise l'erreur quadratique moyenne entre le signal à coder et le signal synthèse.

a-Codeur à excitation par impulsions multiples

Atal et Remde ont les premiers montré qu'un signal d'excitation composé d'impulsions en nombre de 5 à 10 par tranche de 10 ms de signal était suffisant pour modéliser par synthèse tous les sons avec une bonne précision.

On montre que, pour un bloc de 16 échantillons, il existe 120 combinaisons différentes de 2 impulsions, dont on doit calculer simultanément les positions et les amplitudes; aussi est-il préférable de procéder de façon séquentielle et sous optimale. Pour chaque impulsion nous aurons 16 positions et 16 amplitudes à déterminer, le signal étant réactualisé par soustraction de la contribution due au filtrage des impulsions précédentes. La charge en calcul devient ainsi proportionnelle au nombre d'impulsions.

Dans le codeur à impulsions multiples, une partie importante du débit est utilisée pour coder les positions des impulsions, plusieurs auteurs ont proposé d'utiliser des codeurs à impulsions régulièrement réparties. Le signal d'excitation est alors périodique. Pour permettre une certaine souplesse du système, la phase de la forme d'onde est variable ce qui permet de la faire coïncider avec les impulsions glottales.

b-Codeur à excitation par codes (CELP)

Dans ce type de codeur, les dictionnaires de formes d'ondes sont stockés en mémoire et peuvent être de natures différentes:

- séquence de bruit Gaussien.
- éléments du réseau E_g .
- codes à permutation.
- séquences de +1 et -1 (code universel)

2.4.3-Les codeurs hybrides

Ces codeurs sont destinés à fonctionner à des débits compris entre 8 et 16 kbits / s, le surcroît de débits par rapport au vocodeur LPC sert à transmettre l'information relative au signal résiduel.

Dans les codeurs que nous allons voir, il n'est plus nécessaire de calculer la fréquence fondamentale ni même procéder à une détection de voisement.

Les coefficients de prédiction linéaire restent calculés par analyse d'une trame de parole, le signal résiduel est obtenu par élimination de la corrélation à court terme du signal par filtrage inverse.

a-Codeur à excitation par le résidu

Le principe du codeur à excitation résiduelle (RELP), découle de façon directe de la prédiction linéaire, le signal résiduel est chargé de l'information concernant la structure fine du spectre et correspond au meilleur signal d'excitation utilisable pour la synthèse, il est donc logique de transmettre celui-ci au décodeur.

Le codage du signal résiduel, qui beaucoup moins corrélé que le signal original, se fait par des méthodes temporelles mais nécessite un nombre de niveaux de quantification moins important que pour le signal original. La quantification peut être rendue adaptative en fonction de la non-stationnarité du résidu. Enfin, pour accroître l'efficacité du codeur il est souhaitable d'éliminer la corrélation à long terme en exploitant la périodicité pour les sons voisés par un prédicteur de pitch de fonction de transfert:

$$P(z) = \beta \cdot z^{-T}$$

Avec β , le gain et T la période d'estimation du pitch.

Un filtrage perceptuel du signal résiduel quantifié permet de limiter le bruit de quantification apparent du codeur en déformant son spectre de telle façon qu'il devienne moins perceptible dans les plages des fréquences ou énergie du signal de synthèse est faible.

b-Codeur à excitation vocale

Dans le codeur précédent la plus grande partie du débit est utilisée pour représenter le signal résiduel. Pour le codeur à excitation vocale (VELP), seule la partie audiofréquence du signal résiduel est transmise par rééchantillonnage à la fréquence de Nyquist.

Au récepteur, la partie haute du spectre, dont l'énergie a été transmise, est reconstruite par passage de la partie basse fréquence dans une non-linéarité qui se fait par une coupure de la zone centrale du signal par un seuil (central clipping), ou encore par un redressement mono ou double alternance.

c-Vocodeur à bande de base

A la différence du vocodeur à excitation vocale, dans le vocodeur à bande de base, c'est la partie basse fréquence du signal original qui est transmise; elle est codée avec précision par un procédé de codage temporel, la partie haute fréquence est modélisée quant-à-elle par une technique de prédiction linéaire.

Au récepteur, la partie haute du spectre est synthétisée à partir de la bande de base et des coefficients de prédiction linéaire.

2.5-Evaluation de la qualité des codeurs

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. Etant 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.

2.5.1-Mise en forme du bruit de codage

La minimisation de l'erreur quadratique moyenne entre le signal original et le signal synthétisé introduit un bruit de codage, de spectre blanc, qui est d'autant perceptible dans les zones de fréquences où l'énergie du signal est faible devant l'énergie du bruit.

Il apparaît donc que le critère d'erreur quadratique moyenne n'est pas significatif d'une qualité perceptuelle.

La qualité apparente du codeur peut être nettement améliorée en pondérant le spectre de bruit de manière telle qu'il soit plus important dans les zones de fréquences où l'énergie du signal est élevée, il s'agit des zones de formants.

Le principe de masquage fréquentiel montre que le bruit situé dans les zones de formant peut être totalement ou partiellement masqué par le signal de parole. Par conséquent, le bruit perçu provient pour la plupart des zones de fréquences où le signal est faible. Une méthode consiste à utiliser un filtre de pondération perceptuelle F et son inverse F^{-1} pour redistribuer la puissance du bruit d'une fréquence à l'autre [40]. La figure 2.6 schématise le principe de cette méthode.

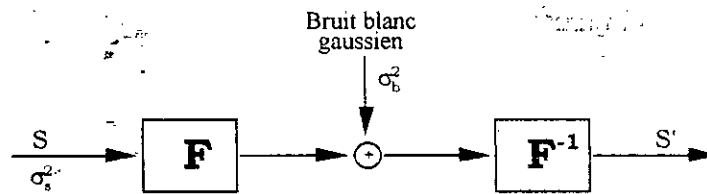


Figure 2.6: Schématisation du filtrage perceptuel.

L'utilisation de filtre F^{-1} a pour objet de modifier le spectre du bruit afin de diminuer la distorsion perceptible introduite par le codeur. On réduit la densité spectrale du bruit dans les zones de fréquences où le niveau du signal est fort et peut masquer efficacement le bruit. Il est important de choisir le filtre F pour que la sortie S' soit le plus près possible de l'entrée S d'un point de vue subjectif.

Une bonne solution au problème est d'utiliser des filtres de la forme suivante:

$$F(z) = \frac{A(z)}{A(z\gamma^{-1})}$$

$$\text{et } F(z) = \frac{A(z\gamma^{-1})}{A(z)}$$

ou $A(z\gamma^{-1})$ est défini à partir de $A(z)$ en ramenant ses zéros vers le centre du cercle unité du plan des Z . Le facteur γ est une fraction comprise entre 0 et 1.

Pour $\gamma = 1$, le spectre du bruit de parole sera plat. Comme le spectre moyen du signal de parole décroît en fonction de la fréquence, le bruit sera surtout audible aux fréquences élevées.

Pour $\gamma = 0$, le spectre du bruit sera cette fois parallèle à celui du signal de parole. Il sera perçu sous forme de bruit basse fréquence, essentiellement au niveau du premier formant. Une bonne répartition fréquentielle du bruit est obtenue pour des valeurs de γ comprises entre 0.7 et 0.9 [39].

Conclusion

Tout au long de ce chapitre, nous avons décrit le signal parole et les méthodes employées pour son codage. Visant un bas débit, il est indispensable d'utiliser une méthode de codage spécifique au signal pour obtenir des résultats acceptables. Parmi ces méthodes, *la prédiction linéaire* correspond au modèle le plus communément admis pour le signal de parole, et c'est celle qui est utilisée avec le plus grand succès.

Chapitre 3

La quantification des paramètres LSF

Introduction

Nous avons vu au chapitre 2, que la prédiction linéaire était une des approches la plus populaire employée pour la description des variations temporelles du spectre à terme court (short term) du signal de parole. De nombreux travaux ont été effectués sur la quantification scalaire et vectorielle des coefficients du filtre LPC.

Il est connu, dans le contexte de la compression de la parole que les coefficients de prédiction sont peu appropriés à la quantification à cause de leur large gamme dynamique et aux possibilités d'instabilités du filtre d'analyse LPC.

Différentes représentations de ces mêmes paramètres comme les coefficients de réflexion, les LAR (Log Area Ratio), les coefficients cepstraux et les LSF (Line Spectral Frequencies) ont été proposées pour pallier à ce problème [37,41,42].

Parmi toutes les représentations citées, les paramètres LSF sont les plus efficaces [43,44,45,46].

Le concept d'un LSF a été introduit pour la première fois par Itakura [47]. Leur efficacité est due principalement à [44,45,48]:

- La relation étroite qui existe entre les LSF et les fréquences des formants.
- La sensibilité spectrale localisée des LSF, en d'autres termes un changement d'un LSF donné produit une variation du spectre LPC uniquement dans son voisinage. Cette dernière caractéristique rend les LSF très utile en quantification par *Split*.

Ce chapitre a pour objet de présenter les LSF, leurs propriétés ainsi que les différentes méthodes employées pour leurs quantification.

3.1- Extraction des paramètres LSF

3.1.1- Les conditions d'analyse

Les considérations à prendre pour le choix des conditions d'analyse par prédiction linéaire sont les suivantes:

- La méthode d'analyse.
- L'ordre de prédiction m .
- La longueur de la fenêtre de mise à jour (update frame).

Comme il a été mentionné au chapitre 2, les méthodes d'autocorrelation et de covariance stabilisée conduisent presque aux mêmes résultats. De ce fait, nous avons utilisé la méthode d'autocorrelation, lors de nos simulations.

Les autocorrelations, ainsi obtenues permettent de calculer les coefficients de prédiction par l'algorithme de Wiener-Livinson-Durbin.

La seconde variable à prendre en compte dans l'analyse LPC est la détermination de l'ordre de prédiction m . Le nombre de pôles, ou nombre de coefficients du prédicteur permet de représenter correctement toute séquence de signal de parole. Il est fonction du nombre de résonances, d'antirésonances du conduit vocal dans la bande de fréquence considérée, de la nature de la fonction d'excitation ainsi que des caractéristiques du rayonnement au niveau des lèvres. Plusieurs méthodes permettent de calculer approximativement la valeur de ce nombre m .

Il a été montré dans [37] que pour donner une représentation correcte des pôles de la fonction de transfert du conduit vocal, la durée de mémorisation du prédicteur linéaire doit correspondre à deux fois le temps mis par l'onde de parole pour se propager depuis la glotte jusqu'aux lèvres. Ainsi, par exemple, si le conduit vocal a une longueur moyenne de 17 cm, le temps équivalent au nombre d'échantillons contenus dans la mémoire doit être égale à 1 ms. Ceci correspond au double du temps mis par l'onde sonore à la vitesse de 340 m/s, pour traverser le conduit vocal.

Lorsque la fréquence d'échantillonnage f_s est exprimée en (échantillon/sec), la période de 1 ms correspondra à $f_s / 1000$ échantillons. A la fréquence d'échantillonnage de 8 kHz, la valeur correspondante de m doit être égale au moins à 8 pour un modèle idéal. Par ailleurs, on peut montrer aussi que deux pôles supplémentaires sont nécessaires pour représenter les influences de la source et du rayonnement, ce qui porte à 12 la valeur moyenne du nombre m dans le cas d'une fréquence d'échantillonnage égale à 10 kHz.

La dégradation de la qualité de la parole devient significative lorsque la valeur de m est réduite à 8 et on a une saturation de la qualité lorsque la valeur de m est supérieur à 12. Le choix de $m = 10$ est raisonnable pour une représentation adéquate du conduit vocal.

La troisième variable à prendre en compte dans l'analyse LPC est l'intervalle de mise à jour. Comme l'ordre de prédiction m ; le choix de l'intervalle L présente un compromis entre la qualité et le débit binaire.

Il est généralement souhaitable d'effectuer l'analyse spectrale dans un intervalle où le mouvement du conduit vocal est négligeable. En codage à bas débit, un intervalle de 20 ms est suffisant pour maintenir une bonne qualité de la parole, cependant ceci introduit une petite dégradation dans quelques sons possédant des caractéristiques spectrales très changeantes.

3.1.2-Constitution des trames

La parole est souvent décomposée par analyse LPC en plusieurs paramètres que l'on code à des fins de transmissions, de rehaussement ou autres.

Pour obtenir les trames de la parole à partir d'un signal continu, une méthode couramment employée consiste à le multiplier par une fenêtre appelée *fonction de pondération*. La pondération effectuée sur le signal annule celui-ci pour les échantillons situés en dehors de l'intervalle d'analyse.

Le choix de *la fonction de pondération* est important dans le domaine spectral, en effet, lorsqu'on multiplie le signal temporel par une fenêtre $w(n)$, le spectre résultant se trouve modifié; dans le domaine fréquentiel le produit simple se transforme en un produit de convolution. La formulation mathématique du problème est comme suit:

Soit $S(n)$ le signal de parole et $s(f)$ son spectre;

$$SW(f) = S(f) \cdot W(f) \quad (3.1)$$

$$\text{Si: } W(n) = 1 \quad \forall n \quad (3.2)$$

Le spectre de $W(n)$ correspond alors à une impulsion de Dirac:

$$W(f) = \delta(f) \quad (3.3)$$

et le produit de convolution donne:

$$SW(f) = S(f) \quad (3.4)$$

Le spectre résultant $SW(f)$ ne subit alors aucune déformation.

Une fonction de pondération sera d'autant meilleure que sa réponse fréquentielle ressemblera davantage à Dirac.

En parole, la fonction de pondération utilisée est celle de *Hamming*, son équation est de la forme suivante:

$$w_h(i) = 0.54 - 0.46 \cos(2\pi i / (n-1)) \quad i=0, 1, \dots, n-1 \quad (3.5)$$

où n est la longueur de la fenêtre.

La longueur de la fenêtre de Hamming est fixée à 250 échantillons, convenant aux longueurs typiques de quasi-stationnarité pour l'analyse de Wiener-Livinson-Durbin.

Nous présenterons dans ce qui suit un tableau récapitulatif des conditions d'analyse employées lors de nos simulations (tableau 3.1).

Méthode d'analyse	Méthode d'autocorrelation
Ordre de prédiction	m=10
Fréquence d'échantillonnage	8 kHz
Période des fenêtres avec recouvrement	20 ms
Fenêtre de pondération	Fenêtre de Hamming
Longueur de la fenêtre de pondération	250 échantillons
Fichier de parole	nombre de fenêtre 565

Tableau 3.1: Conditions expérimentales.

3.2- Calcul des Line Spectral Frequencies (LSF)

En analyse par prédiction linéaire, un segment de parole est supposé être généré comme sortie d'un filtre tous pôles (figure 3.1) $H(z) = 1/A(z)$. Où $A(z)$ est un polynôme en z appelé le *filtre inverse* dont l'expression est donnée par:

$$A(z) = 1 + a_1 z^{-1} + \dots + a_m z^{-m}. \quad (3.6)$$

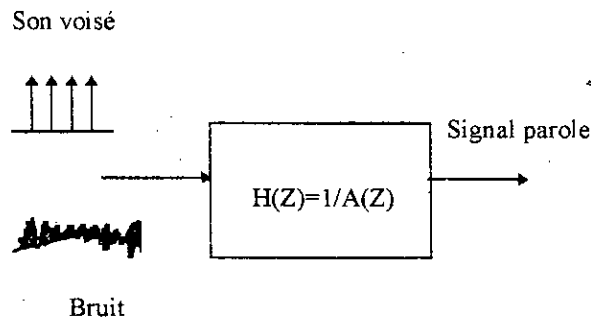


Figure 3.1: Modèle discret de production de la parole employé en prédiction linéaire

L'ensemble $\{a_i\}$ désigne les coefficients de prédiction du filtre et m l'ordre de prédiction. Le polynôme $A_n(z)$ associé à l'ordre n d'analyse LPC, vérifie la relation de récurrence suivante [36, 49,50,51]:

$$A_n(z) = A_{n-1}(z) - r_n z^{-n} A_{n-1}(z^{-1}), \quad n=1, 2, \dots, m, \quad (3.7)$$

Où $A_0(z)=1$.

Les paramètres $\{r_i\}$ $i=1, 2, \dots, m$, sont appelés les coefficients de réflexion ou coefficients PARCOR (PARTIAL CORrelation). Il est important de noter que les coefficients de réflexion sont équivalents aux coefficients LPC dans le sens mathématique. Ils représentent la même information spectrale.

Les coefficients LPC possèdent une large gamme dynamique, cependant les coefficients de réflexion $|r_i|$ sont tous inférieurs à 1, ce qui garantit la stabilité au filtre de synthèse. Les figures (3.2) et (3.3) illustrent la gamme de variabilité des coefficients de prédiction et des coefficients de réflexion, sous forme d'histogrammes.

Les coefficients de PARCOR sont interprétés comme les limites du modèle du tube acoustique du conduit vocal [36].

En posant $n = m+1$ dans l'équation (3.7), on obtient:

$$A_{m+1}(z) = A_m(z) - r_{m+1} z^{-(m+1)} A_m(z^{-1}) \quad (3.8)$$

Dans l'équation (3.8), on considère deux conditions limites artificielles notées par $r_{m+1} = 1$ et $r_{m+1} = -1$. Ces conditions correspondent, respectivement à maintenir le modèle du tube acoustique à la glotte entièrement fermé ou entièrement ouvert. Sous ces conditions le polynôme $A_{m+1}(z)$ permet de former les deux polynômes suivants:

Pour $r_{m+1} = 1$, on obtient le polynôme suivant:

$$\begin{aligned} F_1(z) &= A_m(z) - z^{-(m+1)} A_m(z^{-1}) \\ &= 1 + (a_1 - a_m) z^{-1} + \dots + (a_m - a_1) z^{-m} - z^{-(m+1)}. \end{aligned} \quad (3.9)$$

Pour $r_{m+1} = -1$, on obtient:

$$\begin{aligned} F_2(z) &= A_m(z) + z^{-(m+1)} A_m(z^{-1}) \\ &= 1 + (a_1 + a_m) z^{-1} + \dots + (a_m + a_1) z^{-m} + z^{-(m+1)} \end{aligned} \quad (3.10)$$

Lorsque l'ordre de prédiction m est un entier pair, on peut montrer que les polynômes $F_1(z)$ et $F_2(z)$ peuvent être exprimés par les deux équations suivantes [49,50]:

$$F_1(z) = (1 - z^{-1}) \prod_{i=2,4,\dots,m} (1 - 2z^{-1} \cos w_i + z^{-2}) \quad (3.11a)$$

$$\text{et } F_2(z) = (1 + z^{-1}) \prod_{i=1,3,\dots,m-1} (1 - 2z^{-1} \cos w_i + z^{-2}) \quad (3.11b)$$

On suppose que $w_1 < w_3 < \dots < w_{m-1}$ et $w_2 < w_4 < \dots < w_m$.

Il est clair à partir de l'équation (3.11) que e^{jw_i} pour $i = 1, 2, \dots, m$, sont les racines des polynômes $F_1(z)$ et $F_2(z)$. Les paramètres $\{w_i\}$ $i=1,2,\dots,m$, définissent alors les "Line Spectral Frequencies" (LSF) [32,49]. Il est important de noter que $w_0 = 0$ et $w_{m+1} = \pi$, sont des racines

fixées, des polynômes $F_1(z)$ et $F_2(z)$ respectivement et seront exclus de l'ensemble des paramètres LSF.

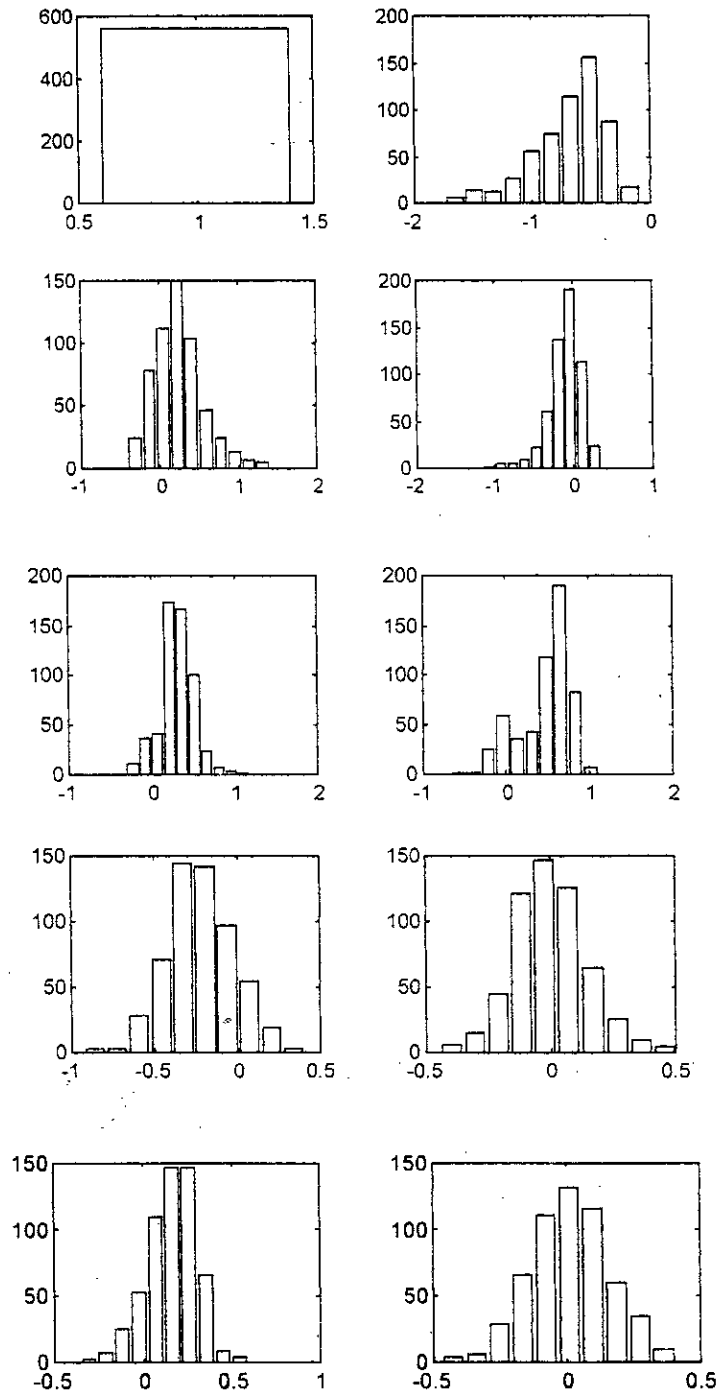


Figure 3.2. Histogrammes des coefficients de prédiction

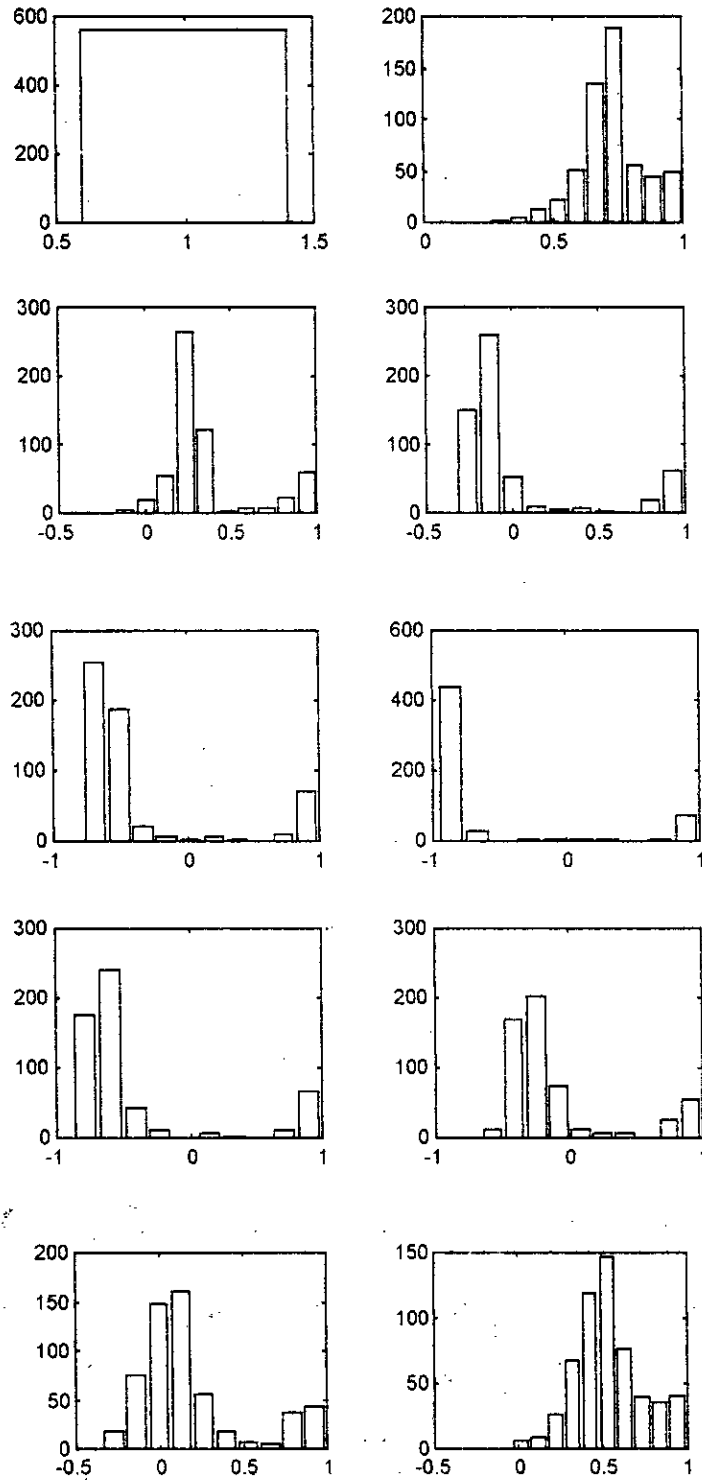


Figure 3.3: Histogrammes des coefficients de réflexion

Chapitre 3

Les paramètres LSF, ainsi définis, peuvent être interprétés comme les fréquences de résonance du conduit vocal sous les deux conditions limites artificielles au niveau de la glotte.

Les polynômes $F_1(z)$ et $F_2(z)$ possèdent quelques propriétés très intéressantes et importantes que nous allons résumer dans ce qui suit [49,50]:

1- les racines des polynômes $F_1(z)$ et $F_2(z)$ sont sur le cercle unité.

2- Les racines des polynômes $F_1(z)$ et $F_2(z)$ sont entrelacées, c'est à dire dans un ordre ascendant et se trouvent dans le premier et le second quadrants du plan complexe Z ce qui se traduit par la relation suivante:

$$0 = w_0 < w_1 < w_2 < \dots < w_m < w_{m+1} = \pi. \quad (3.12)$$

Cette dernière relation (3.12) exprime la propriété *d'ordonnement* des LSF.

Dans ce qui suit, on verra l'interprétation de la représentation spectrale par LSF du signal parole. La figure (3.4) montre un exemple de l'enveloppe spectrale du filtre LPC.

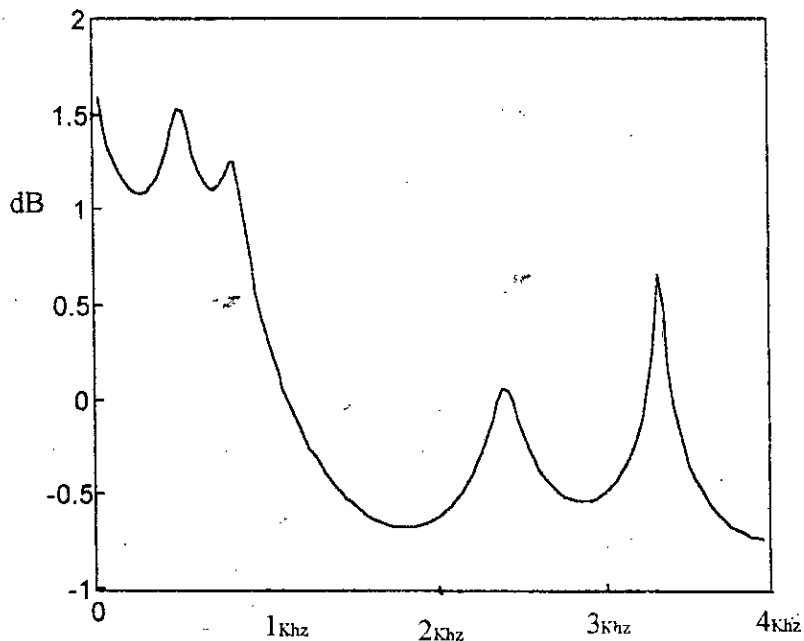


Figure 3.4: Enveloppe spectrale d'un segment de parole

En utilisant les LSF, la puissance associée à la fonction de transfert $H(z)$, peut être calculée comme suit:

$$\begin{aligned} |H(e^{jw})|^2 &= 1/|A_m(e^{jw})|^2 = 4/|F_1(e^{jw}) + F_2(e^{jw})|^2 \\ &= 2^{-m} / \left\{ \sin^2 \frac{w}{2} \prod_{i=2,4,\dots,m} (\cos w - \cos w_i)^2 + \cos^2 \frac{w}{2} \prod_{i=1,3,\dots,m-1} (\cos w - \cos w_i)^2 \right\}. \quad (3.13) \end{aligned}$$

Cette équation implique que, $|H(e^{jw})|^2$ possède une forte résonance à la fréquence w , lorsque deux paramètres LSF sont localisés près de w . D'où, les paramètres LSF peuvent être interprétés comme une représentation du filtre tous pôles par la localisation de la densité de m fréquences discrètes, notées par $\{w_1, w_2, \dots, w_m\}$, dans le domaine fréquentiel.

Il est clair, à partir des définitions précédentes que les paramètres LSF, sont équivalents aux coefficients LPC $\{a_i\}$ $i = 1, 2, \dots, m$, et aux coefficients de réflexion $\{k_i\}$ $i = 1, 2, 1, \dots, m$. Cependant, il est établi expérimentalement que les paramètres LSF possèdent de meilleures propriétés de quantification et d'interpolation comparés aux coefficients LPC et coefficients de réflexion [45,47].

Il a été montré que la propriété d'ordonnement des LSF, est une condition nécessaire et suffisante pour la stabilité du filtre de synthèse basé sur les LSF [43,49].

Dans un système de compression de la parole basé sur le codage des LSF, ces derniers doivent être quantifiés. Donc, pour garantir la stabilité du filtre de synthèse LSF, on doit s'assurer que les versions quantifiées des LSF vérifient toujours la propriété d'ordonnement. Ceci représente un résultat très important dans le design des quantificateurs basés sur les paramètres LSF.

Nous présentons dans les tableaux (3.2 et 3.3) les autocorrelations, les coefficients LPC et les LSF calculés sous les conditions expérimentales du tableau 3.1.

Les paramètres LSF sont calculés par les polynômes de Chebychev [48].

Pour vérifier les propriétés (1 et 2), les paramètres relatifs à quatre fenêtres sont représentés sur la figure 3.5 [52].

Indices	$r_i \times 10^3$ normalisés	(a_i) non normalisés	LSF (radian)	LSF dans le domaine cos
1	1000.000000	1.000000	0.375332	0.930386
2	704.922302	-1.375521	0.481021	0.886523
3	167.818451	1.159798	0.646785	0.798025
4	-133.031754	-0.578293	1.077784	0.473282
5	-137.953232	0.271745	1.221366	0.342362
6	-135.589874	-0.182240	1.344605	0.224268
7	-341.440155	0.208636	1.744627	-0.172956
8	-594.398865	0.174513	2.069958	-0.478689
9	-565.095215	-0.074004	2.356713	-0.707474
10	-212.524048	-0.006863	2.739402	-0.920206

Tableau 3.2: Quelques ensembles de paramètres LPC pour une fenêtre donnée.

Indices	$r_i \times 10^3$ normalisés	(a_i) non normalisés	LSF (radian)	LSF dans le domaine cos
1	1000.00000	1.000000	0.273464	0.962841
2	883.396545	-1.340813	0.320995	0.948922
3	704.885864	1.030209	0.560991	0.846728
4	564.372681	-1.198992	1.033787	0.511568
5	361.652649	1.011294	1.315954	0.252093
6	124.610046	-0.779929	1.421504	0.148738
7	-120.552711	0.670096	1.940039	-0.360909
8	-380.874878	-0.120703	2.049762	-0.460862
9	-543.817749	0.227665	2.273050	-0.645940
10	-599.035583	-0.424687	2.610722	-0.862367

Tableau 3.3 Quelques ensembles de paramètres LPC pour fenêtre donnée.

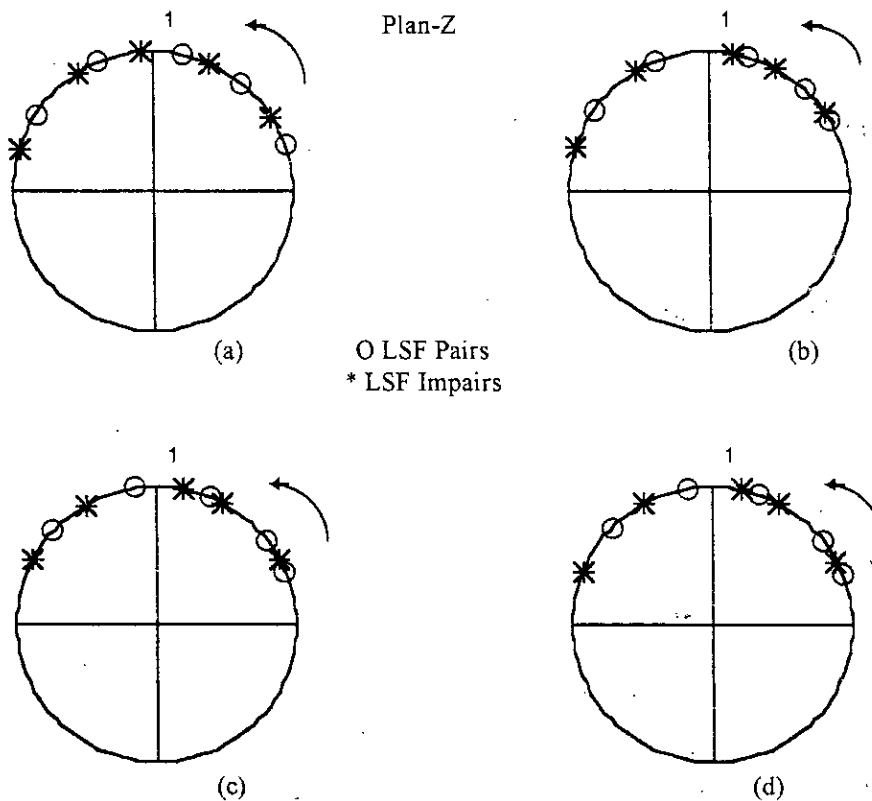


Figure 3.5: Représentation des LSF dans le plan Z

3.3- Distributions statistiques des LSF

Nous avons étudié les propriétés statistiques des LSF sous les mêmes conditions précédentes. Les distributions des LSF sont représentées sur la figure 3.6 en termes d'histogrammes. Les LSF, ainsi représentés ont été normalisés à la fréquence d'échantillonnage f_s par la relation suivante:

$$f_i = (w_i / 2\pi) \cdot f_s \text{ Hz} \tag{3.14}$$

Où w_i représente le $i^{\text{ème}}$ LSF en radian.

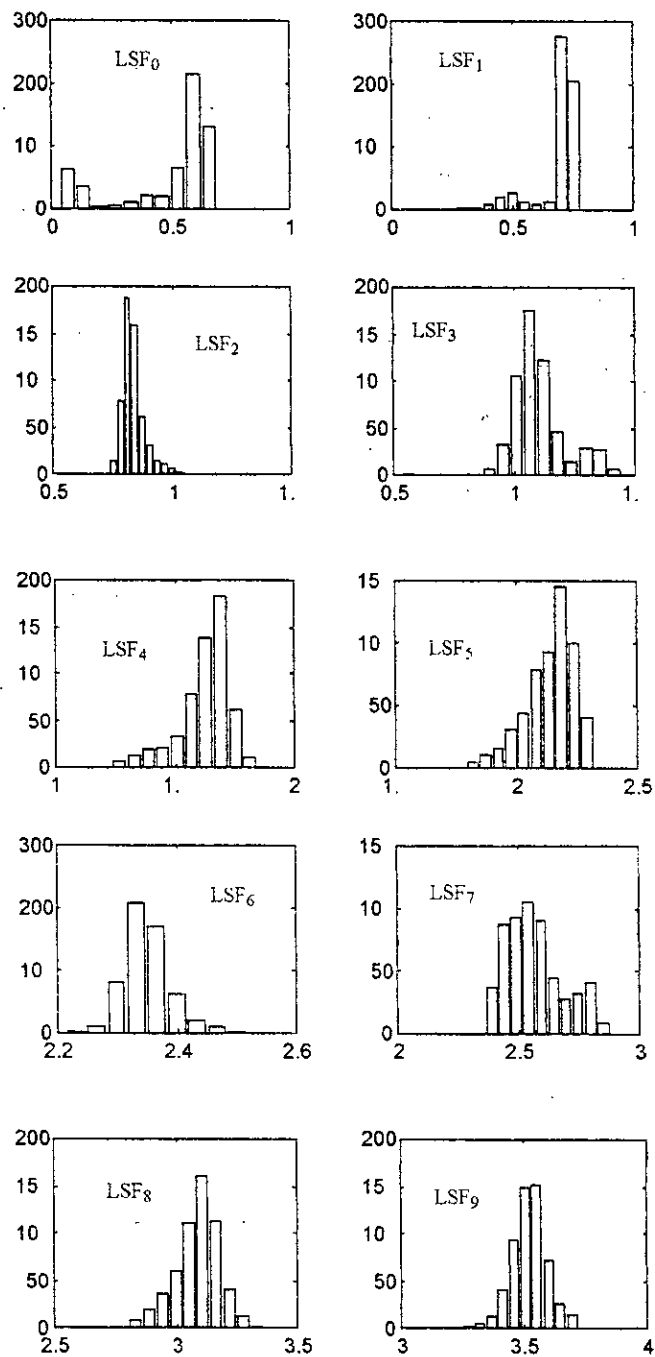


Figure 3.6: Histogrammes des LSF.

Il est clair à partir des histogrammes (figure 3.6) que la gamme de distribution varie d'un LSF à un autre. Il a été observé également par les chercheurs que cette gamme varie avec les caractéristiques du locuteur et les conditions d'enregistrements [44].

La réduction de cette variabilité est obtenue par l'utilisation des différences entre les LSF adjacents qu'on notera par: (figure 3.7)

$$\Delta LSF(i) = LSF(i+1) - LSF(i)$$

où i représente le $i^{ème}$ paramètre LSF.

Les valeurs des ΔLSF sont représentées sous formes d'histogramme par la figure 3.8. La réduction de la gamme de variabilité est apparente par comparaison des histogrammes relatifs aux LSF et ΔLSF et spécialement autour des fréquences du milieu comme LSF_4 et LSF_5 . [53].

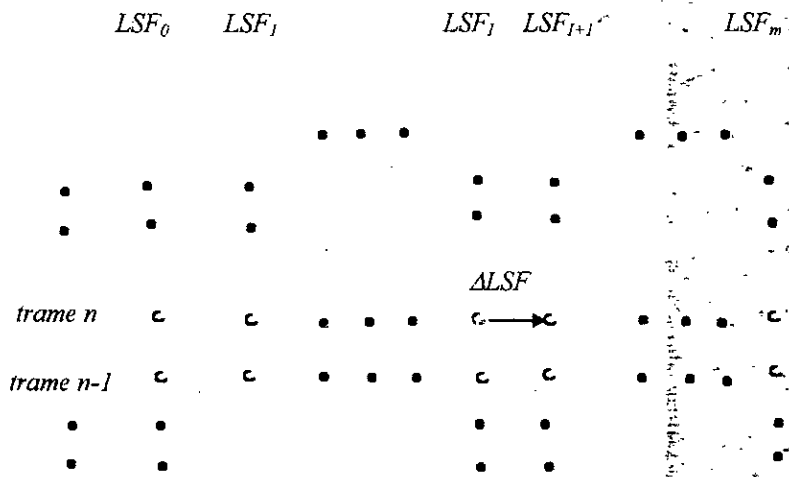


Figure 3.7: Les paramètres LSF pour des trames consécutives d'ordre m.

3.4-Conversion des LSF en coefficients de prédiction

La conversion des paramètres LSF aux coefficients de prédiction est réversible de nature, c'est à dire qu'il est possible de calculer exactement les coefficients de prédiction à partir des LSF [54].

Les nombre d'opérations à calculer pour la conversion des LSF en coefficients de prédiction est moins important que celui de l'extraction des LSF à partir des coefficients de prédiction [54].

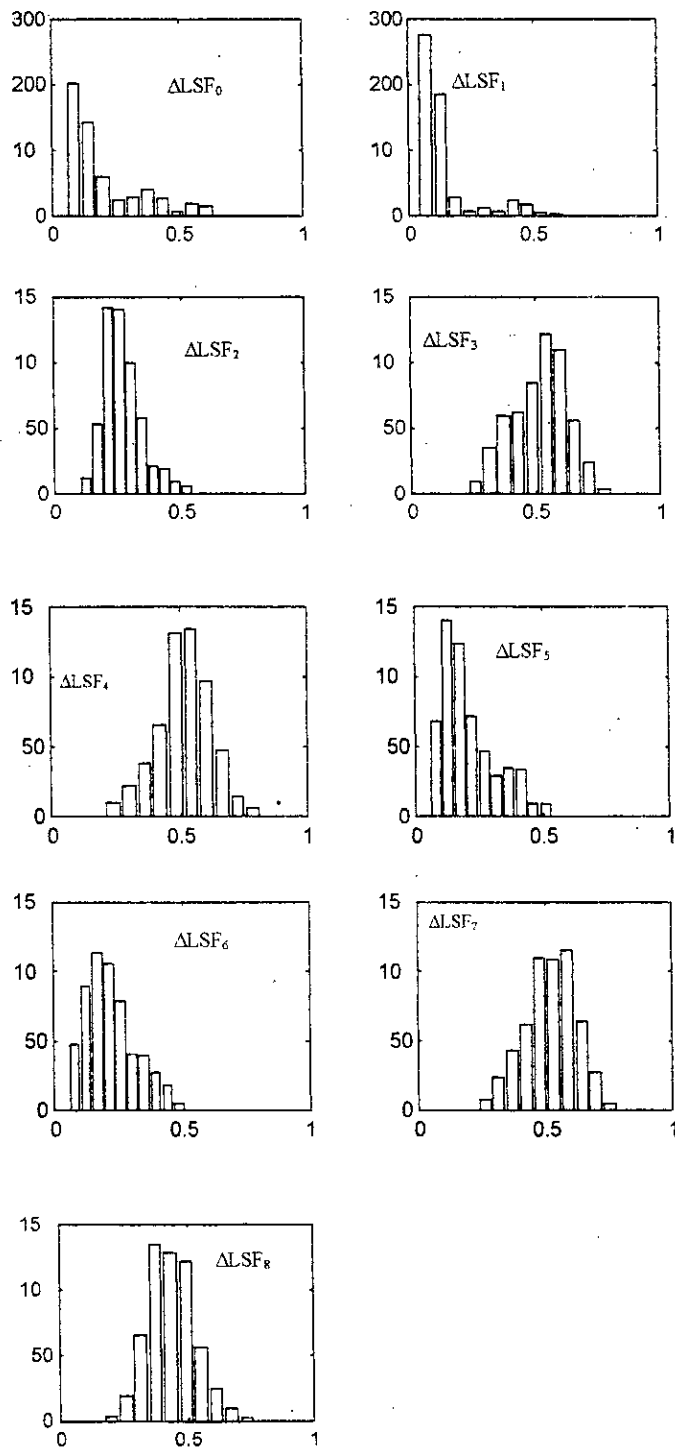


Figure 3.8: Histogrammes des ΔLSF

3.5- Quantification des paramètres LSF

L'objectif principal dans le design des quantificateurs des paramètres LSF est la minimisation de la distorsion de quantification tout en vérifiant la propriété d'ordonnancement pour les LSF quantifiés.

L'objectif primordial de la quantification des paramètres LSF est la réduction, lors de la transmissions des trames d'information, du nombre de bits attribuées aux LSF. Cette étude va servir à réduire le nombre de bits (34 bits) attribués jusqu'alors aux LSF dans le codeur (standard federal US à 4.8 kb/s) [55].

3.5.1-Mesures de distances des LSF

La sélection d'une mesure de distorsion est un choix important dans le design et l'opération de la quantification vectorielle.

La distance euclidienne est largement utilisée. Elle est facile à mettre en oeuvre et possède une propriété intéressante. lors du design du dictionnaire les valeurs de la base d'apprentissage sont associés à leurs centroïdes. Chaque centroïde ou barycentre d'une classe donnée, est représenté par un point dont la somme des distances avec les vecteurs de la classe considérée est minimale. Cette propriété n'est plus rigoureusement vérifiée dans le cas de la plupart des autres distances.

Néanmoins, la distance euclidienne ne constitue pas le critère de fidélité le plus approprié à la quantification des LSF, dans la mesure où les 10 LSF n'ont pas la même importance. En effet, 2 LSF très proches illustrent la présence d'un formant au voisinage de la fréquence; il paraît alors judicieux d'accorder à ces LSF, un degré d'importance plus élevé. C'est pourquoi, on s'intéresse à l'utilisation de distances qui permettent de pondérer les LSF en fonction de la position des formants.

La distance de *Farvardin* consiste à pondérer un LSF en fonction de l'espace qui le sépare de ces voisins [56].

La distance de *Atal* consiste à pondérer un LSF en fonction de la densité spectrale du filtre LPC selon la fréquence du LSF [57,58].

On rappelle que la distance entre un LSF et un LSF quantifié, noté par LSFQ est donnée par l'équation (3.15):

$$d(LSF, LSFQ) = \sum_{i=1}^{10} [w_i (LSF_i - LSFQ_i)]^2 \quad (3.15)$$

Où w_i est le poids assigné au $i^{\text{ème}}$ LSF.

Les poids des différentes distances sont illustrés par les expressions suivantes:

Distance Euclidienne $w_i = 1$

Distance de Farvardin $w_i = 1 / (\text{LSF}(i+1) - \text{LSF}(i)) + 1 / (\text{LSF}(i) - \text{LSF}(i-1))$

Distance de Atal $w_i = [\text{DSP} (\text{Fréquence du LSF})]^{0.15}$

Au regard de ces trois distances, on constate que la distance de Atal est plus complexe, dans la mesure où, elle impose le calcul d'une DFT après l'analyse de Wiener-Levinson-Durbin.

3.5.2-Génération des dictionnaires

Les dictionnaires, que nous avons élaborés ont été générés par l'algorithme LBG décrit au chapitre 1.

Comme nous l'avons rappelé précédemment, l'algorithme LBG n'est pas directement compatible avec les distances proposées au paragraphe précédent; deux possibilités de contourner cet obstacle nous sont offertes:

- La première consiste à exécuter le LBG à l'aide de la distance euclidienne mais d'effectuer la quantification au moyen de la distance choisie; cette méthode semble néanmoins sous-optimale.

- La seconde méthode, utilise la distance considérée non seulement lors de la quantification, mais également au sein de l'algorithme LBG, lors de l'attribution des vecteurs LSF aux différents centroides. Seule la recherche des nouveaux représentants du dictionnaire conserve la distance euclidienne pour le calcul des barycentres.

Les résultats obtenus par LBG pour plusieurs dimensions des vecteurs allant de 3 à 10 à des débits variant de (3 bits/vecteur) à (10 bits/vecteur) sont représentés sur les figures ci dessous.

Il est représenté l'entropie en fonction du nombre d'itérations. (fig.3.9, fig.3.10, fig.3.11 et fig.3.12).

Les valeurs du rapport signal à bruit pour chaque vecteur en fonction du nombre d'itérations sont représentées sur les figures suivantes: (fig.3.13, fig. 3.14 ,fig. 3.15 et fig.3.16).

La convergence est également représentée par la variation en % du RSB en fonction du nombre d'itérations pour un seuil de (0.1 %). (fig.3.17, fig.3.18, fig.3.19 et fig. 3.20).

Algorithme LBG
Variation de l'entropie en fonction du nombre d'itérations

Dimension 3

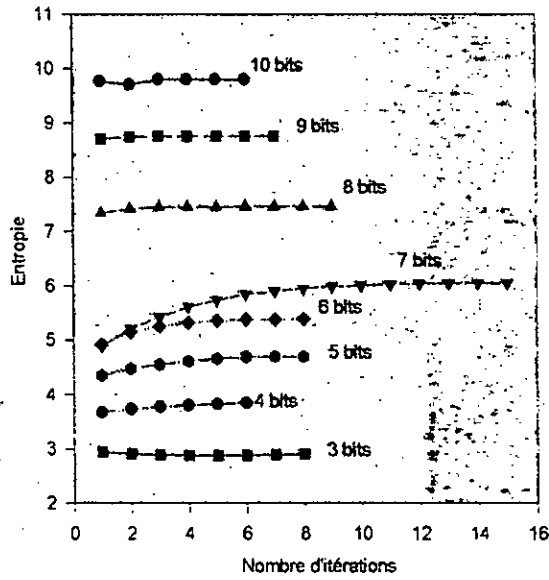


Figure 3.9: Variation de l'entropie en fonction du nombre d'itérations pour des débits fixes.

Dimension 4

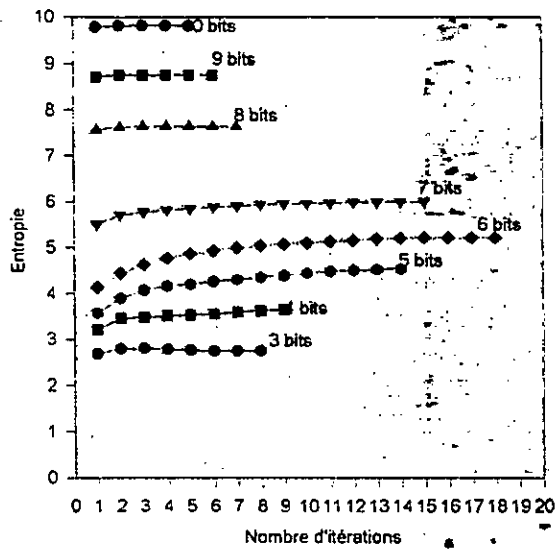


Figure 3.10: Variation de l'entropie en fonction du nombre d'itérations pour des débits fixes.

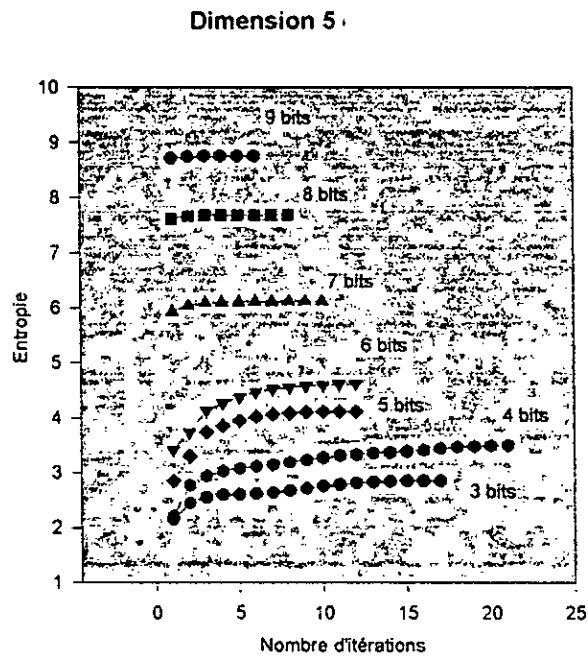


Figure 3.11: Variation de l'entropie en fonction du nombre d'itérations pour des débits fixes.

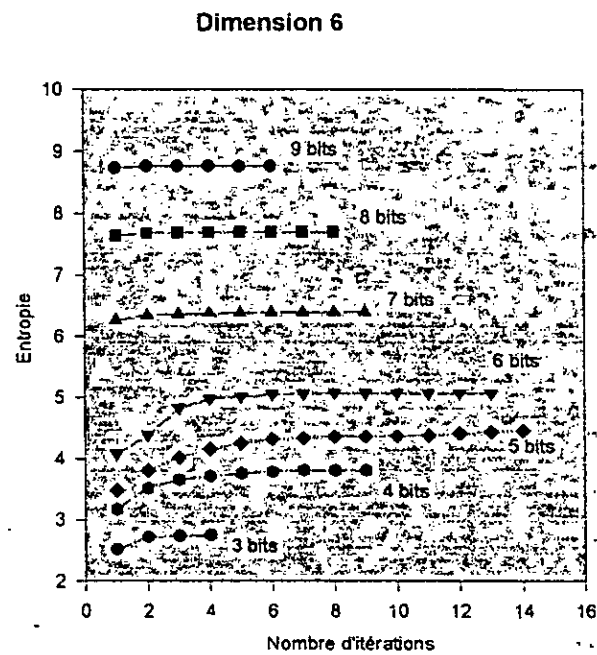


Figure 3.12: Variation de l'entropie en fonction du nombre d'itérations pour des débits fixes.

On remarque à partir de ces figures, que l'allure de la courbe des entropies pour différents débits et différentes dimensions des vecteurs est la même. C'est une fonction croissante du nombre d'itérations.

La valeur de l'entropie se stabilise après environ six itérations pour un débit fixe. La valeur de l'entropie augmente avec le débit pour une dimension fixe. Une meilleure qualité des dictionnaires est donc obtenue pour des débits élevés.

Variation du RSB en fonction du nombre d'itérations

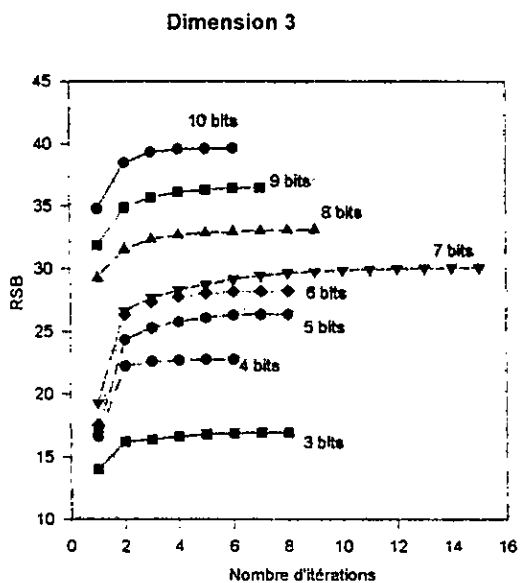


Figure 3.13: Variation du RSB en fonction du nombre d'itérations pour des débits fixes.

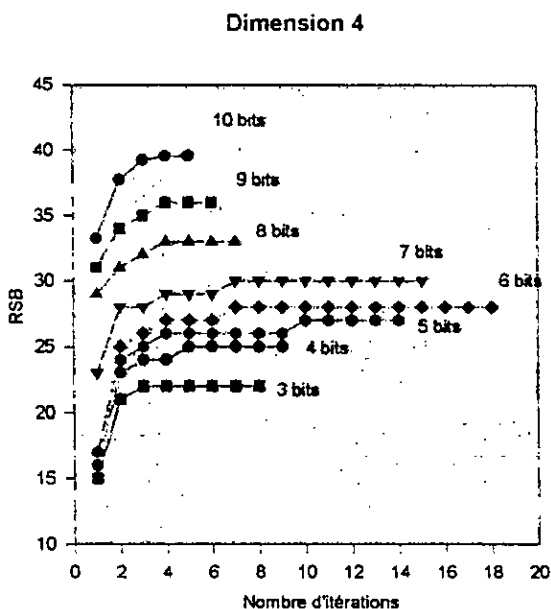


Figure 3.14: Variation du RSB en fonction du nombre d'itérations pour des débits fixes.

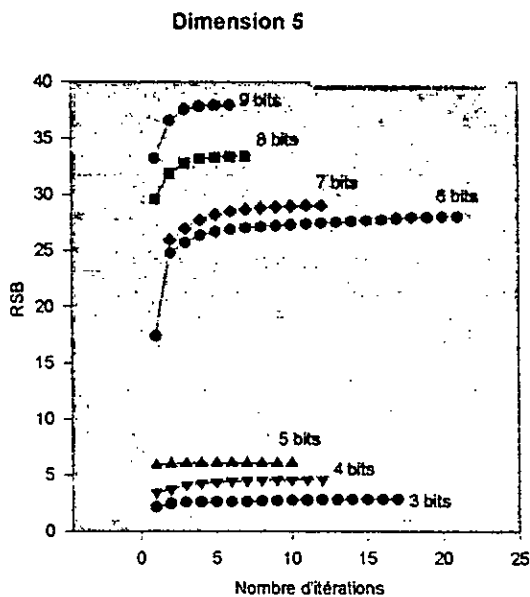


Figure 3.15: Variation du RSB en fonction du nombre d'itérations pour des débits fixes.

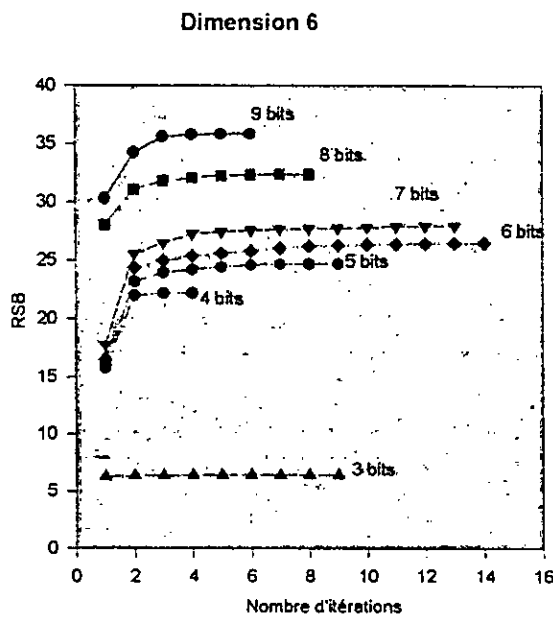


Figure 3.16: Variation du RSB en fonction du nombre d'itérations pour des débits fixes.

Il est important de remarquer que l'allure des courbes du rapport signal à bruit en fonction du nombre d'itérations est une fonction croissante. La valeur du rapport signal à bruit pour un débit fixe se stabilise après environ six itérations sauf pour des débits de 5, 6 et 7 bits, cette valeur augmente pour une dimension fixe des vecteurs.

Variation en % du RSB en fonction du nombre d'itérations

Dimension 3

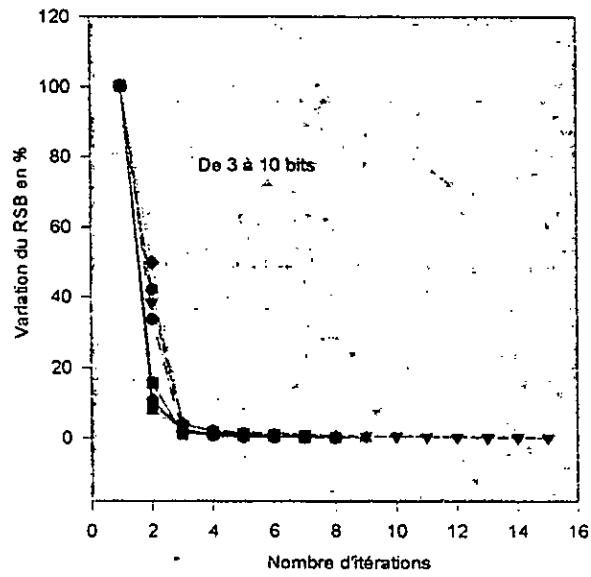


Figure 3.17: Variation du RSB en % en fonction du nombre d'itérations pour des débits fixes.

Dimension 4

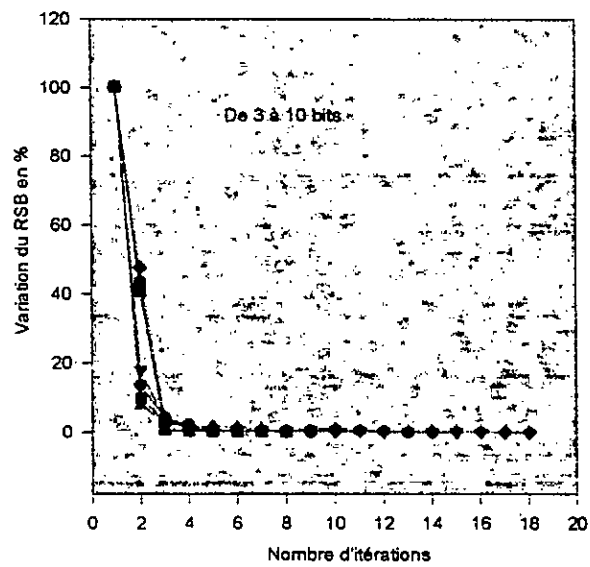


Figure 3.18: Variation du RSB en % en fonction du nombre d'itérations pour des débits fixes.

Dimension 5

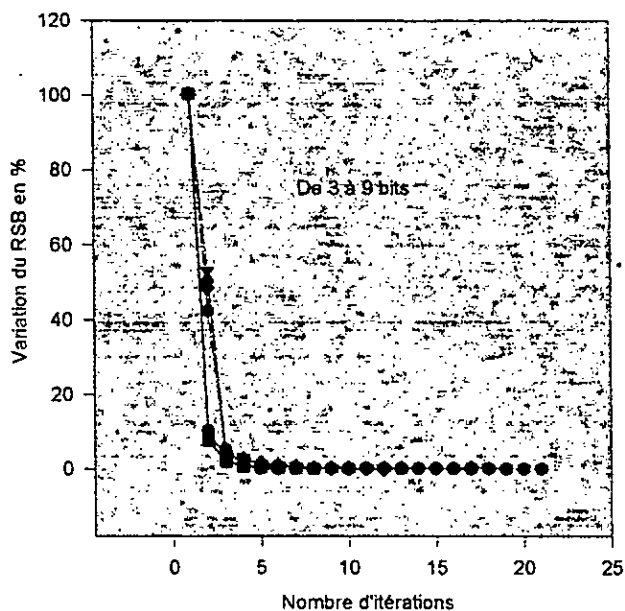


Figure 3.19: Variation du RSB en % en fonction du nombre d'itérations pour des débits fixes.

Dimension 6 (wd101)

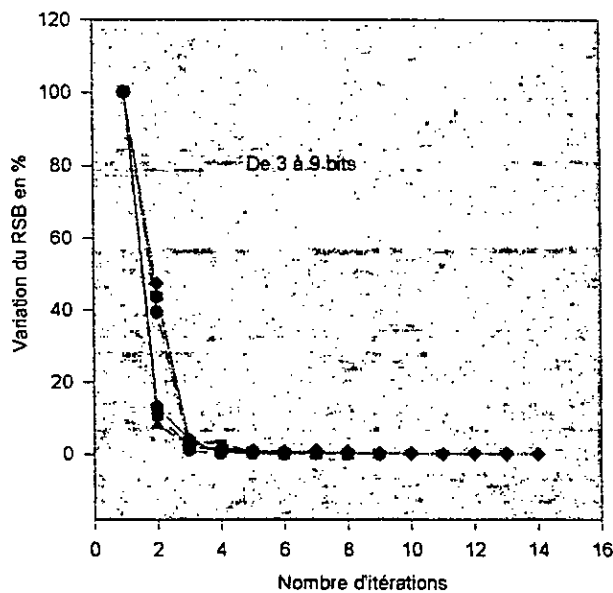


Figure 3.20: Variation du RSB en % en fonction du nombre d'itérations pour des débits fixes.

On remarque que la convergence de l'algorithme K-moyenne pour les différents vecteurs est obtenue après six itérations.

On remarque également que le nombre d'itérations est maximum pour les différentes dimensions lorsque le nombre de bits attribués aux vecteurs est de 5, 6 et 7 bits.

Nous avons testé également la convergence des dictionnaires des Δ LSF, les résultats relatifs sont représentés sur les figures (3.21 à 3.26).

Vecteur de dimension 1 (3 bits) de Δ LSF

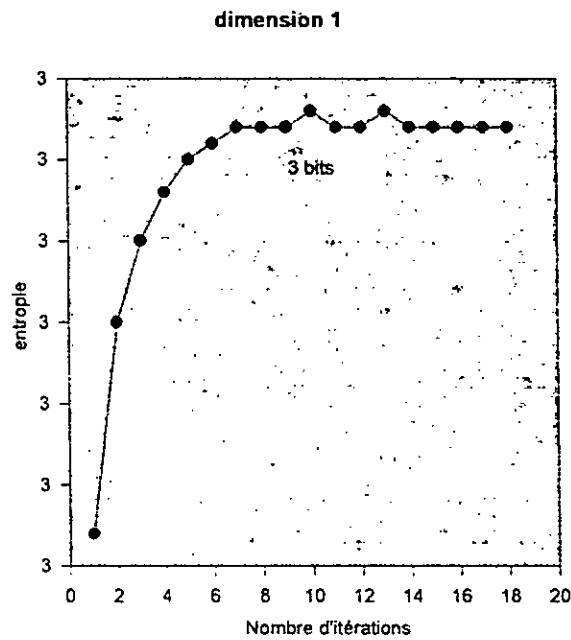


Figure 3.21: Variation de l'entropie en fonction du nombre d'itérations pour 3 bits/vecteur

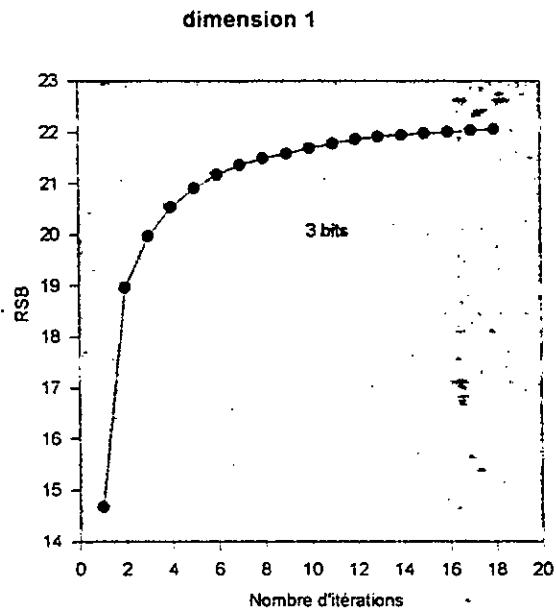


Figure 3.22: Variation du RSB en fonction du nombre d'itérations pour 3 bits/vecteur

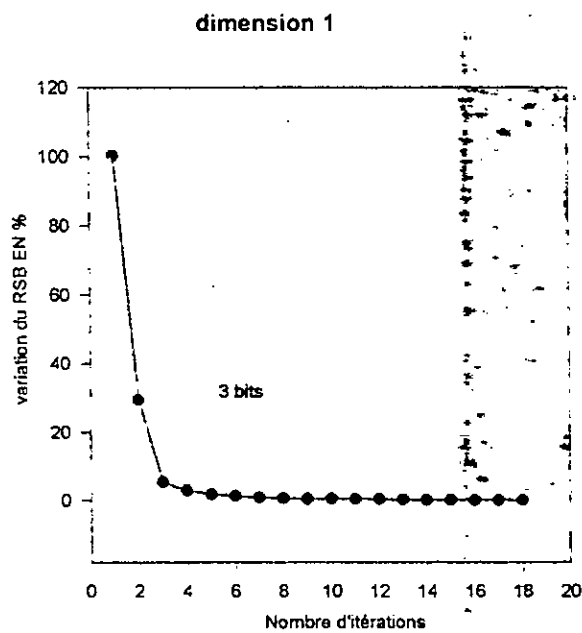


Figure 3.23: Variation du RSB en % en fonction du nombre d'itérations pour 3 bits/vecteur

Vecteur de dimension 1 (3 bits) de LSF

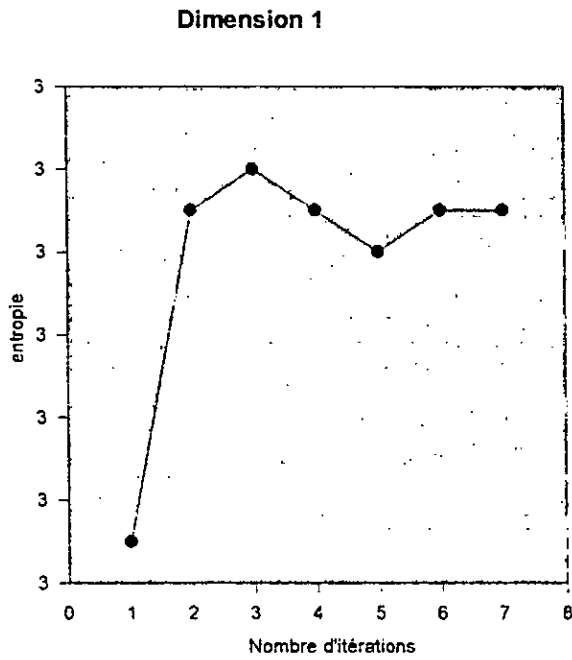


Figure 3.24: Variation de l'entropie en fonction du nombre d'itérations pour 3 bits/vecteur

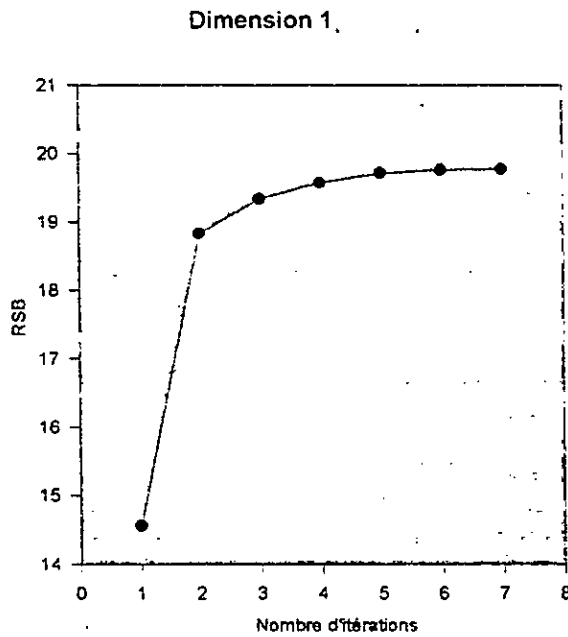


Figure 3.25: Variation du RSB en fonction du nombre d'itérations pour 3 bits/vecteur

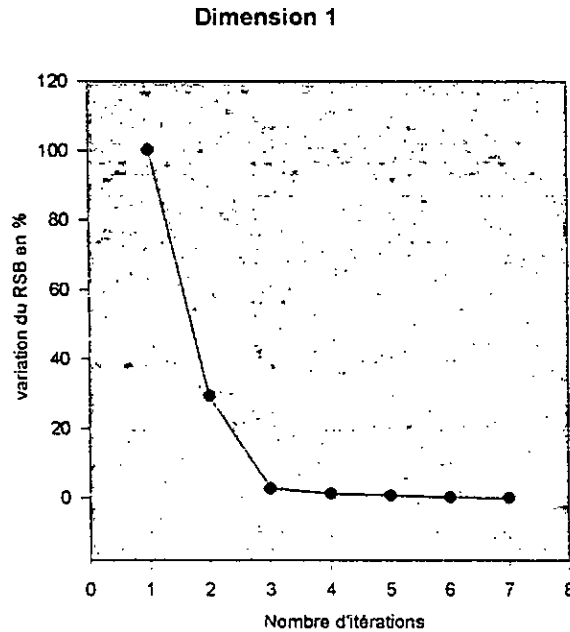


Figure 3.26: Variation du RSB en % en fonction du nombre d'itérations pour 3 bits/vecteur

On remarque que l'allure des courbes pour les entropies n'est pas croissante comme pour les vecteurs à des dimensions supérieures à 3.

La convergence pour le vecteur des Δ LSF est obtenue après plusieurs itérations comparée à celle d'un vecteur de même dimension de LSF.

3.5.3- Méthodes de quantification des LSF

Pour la quantification des paramètres LSF, on testera plusieurs méthodes, on commencera par la quantification directe, la quantification différentielle (Δ LSF), la quantification par Split et en dernier lieu la quantification à étages.

En ce qui concerne la mesure des performances, on utilise la distorsion spectrale exprimée en dB.

La mesure de distorsion spectrale est utilisée pour une comparaison objective entre les différentes méthodes de quantifications.

La distorsion spectrale entre un vecteur LSF et un vecteur LSF quantifié est donnée par l'expression suivante:

$$SD = \int_{-\infty}^{\infty} [10 \log s(f) - 10 \log \hat{s}(f)]^2 \quad (3.16)$$

Où $S(f)$ et $\hat{S}(f)$ correspondent respectivement aux spectres sans et avec quantification. La distorsion moyenne est exprimée à travers un ensemble de fenêtres par:

$$SD = \frac{1}{N_f} \sum_{i=1}^{N_f} \left(\int_{-\infty}^{\infty} [10 \log S(f) - 10 \log \hat{S}(f)]^2 \right) dB \quad (3.17)$$

Avec N_f le nombre de fenêtres. La mesure de distorsion spectrale est connue d'avoir une bonne correspondance avec les mesures subjectives [59].

Les conditions suivantes sont nécessaires pour avoir une bonne qualité de la parole synthétisée (reconstruite), c'est à dire la quantification LPC n'introduit aucune distorsion audible dans le signal parole synthétisé. En d'autre terme, les deux versions de la parole codée obtenues par les paramètres LPC non quantifiés et par les paramètres LPC quantifiés ne sont pas distinguables à l'écoute (listening). Il est généralement admis que la bonne qualité de l'information LPC est obtenue en maintenant les trois conditions suivantes [57,58,60]:

- La distorsion spectrale moyenne doit être d'environ 1 dB avec moins de 2% ayant une distorsion spectrale entre 2 et 4 dB.
- Aucun pourcentage n'est permis pour une distorsion spectrale supérieure à 4 dB.

3.5.3.1- Quantification directe de chaque LSF

Au cours de notre étude de la quantification, nous avons testé cette dernière pour des débits allant de 34 à 30 bits attribués aux 10 LSF. Les bits supplémentaires sont alloués aux premiers LSF selon la forme suivante (4 4 4 4 3 3 3 3 3 3) pour 34 bits et la seconde de 32 bits (4 4 3 3 3 3 3 3 3 3) et la dernière de 30 bits (3 3 3 3 3 3 3 3 3 3).

Les figures suivantes illustrent ces résultats.(fig.3.27, fig.3.28, fig.3.29).

Ces courbes illustrent le pourcentage le nombre de vecteurs LSF présents à une distorsion spectrale donnée.

Quantification directe avec la base d'apprentissage non variée distance euclidienne et Farvardin

Quantification sur 34 bits

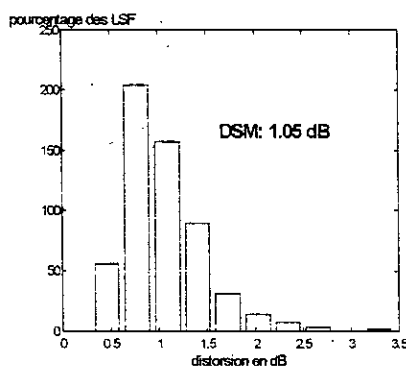


Figure 3.27

Quantification sur 32 bits

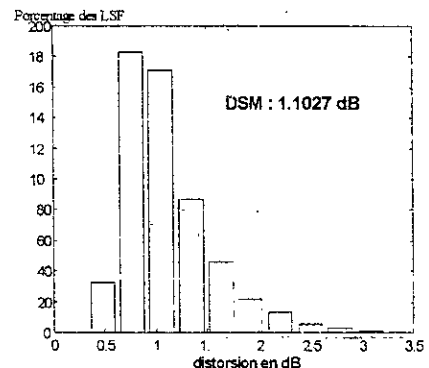


Figure 3.28

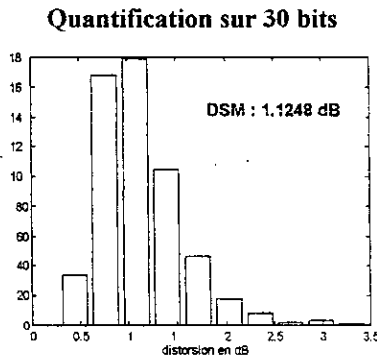


Figure 3.29

4.5.3.2-Quantification différentielle

Du fait la propriété d'ordonnement des LSF, il est concevable que les différences des fréquences consécutives des paramètres LSF possèdent un rang dynamique plus petit comparé aux LSF eux mêmes. Ce qui est confirmé par les résultats expérimentaux (figure 3.8). Ceci a motivé l'utilisation de la méthode rapportée pour la première fois par Soong et Juang dans laquelle, les paramètres ΔLSF sont quantifiés, au lieu des LSF [44].

La description de l'algorithme de codage est comme suit:

Etape1: Quantifier $LSF(i)$ par $LSFQ(i)$ et mettre $i = 1$;

Etape2 : calculer la différence entre $LSF(i+1)$ et $LSFQ(i)$ qu'on notera par :

$$\Delta LSF(i) = LSF(i+1) - LSFQ(i);$$

Etape3 : Quantifier $\Delta LSF(i)$ par $\Delta LSFQ(i)$;

Etape4: reconstruire $LSF(i+1)$ par $LSFQ(i+1) = LSFQ(i) + \Delta LSFQ(i)$;

Etape5: Si $i = m-1$ (ordre de prédiction), stop; sinon mettre $i = i + 1$ and go to 2).

Le diagramme correspondant de cette méthode est représenté sur la figure 3.30

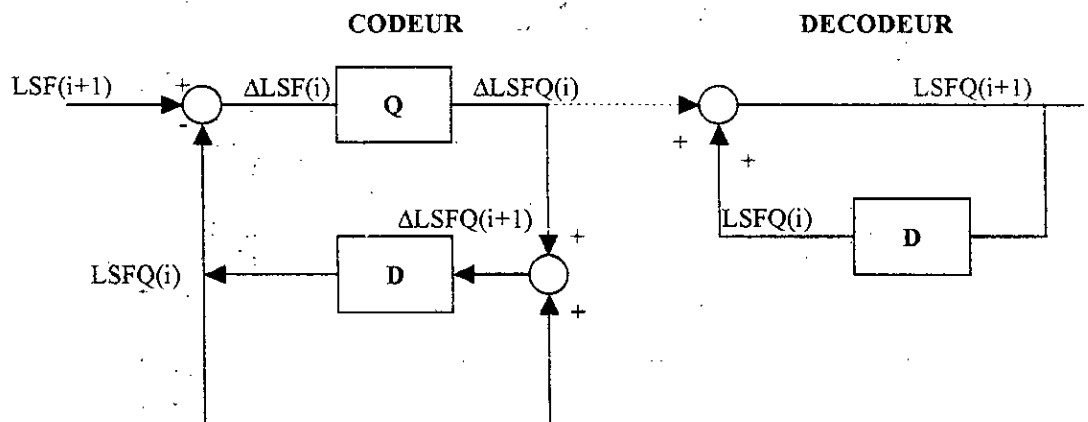


Figure 3.30: Algorithme de codage des ΔLSF .

L'allocation de bits est de 34 bits à 28 bits pour comparaison avec la quantification étudiée précédemment.

Les performances de cette méthode évaluées en terme distorsion spectrale exprimées en dB sont représentées sur les figure (3.31, 3.32, 3.33, 3.34) [61].

Q. Différentielle sur 34 bits

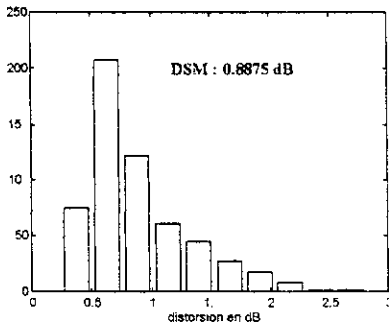


Figure 3.31

Q. Différentielle sur 32 bits

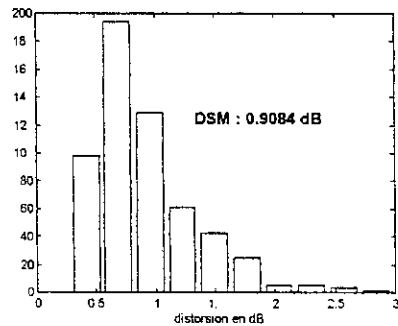


Figure 3.32

Q. Différentielle sur 30 bits

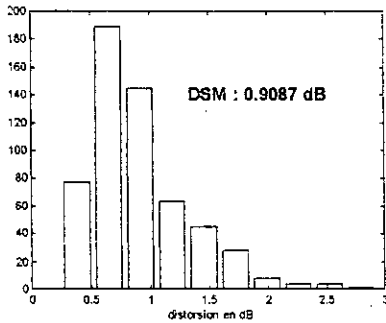


Figure 3.33

Q. Différentielle sur 28 bits

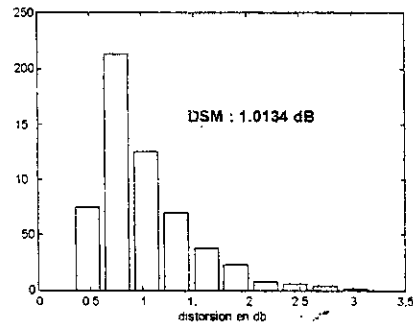


Figure 3.34

3.5.3.3- Quantification vectorielle par Split

Dans la quantification par Split le vecteur des paramètres LSF à quantifié est divisé en plusieurs parties et chaque partie est quantifiée séparément par un quantificateur vectoriel.

On sait que la quantification idéale, en terme de performances, repose sur la quantification vectorielle du bloc des 10 LSF. Cette quantification nécessite un dictionnaire bien trop puissant; en effet, si on accorde 24 bits à cette quantification, la taille du dictionnaire associée est de 2^{24} mots de code, ce qui est inconcevable aussi bien au niveau du stockage qu'au niveau de la complexité. Au terme d'une étude bibliographique sur les LSF et de l'expérience des laboratoires spécialisés dans le codage, on envisage dans un premier temps une quantification vectorielle utilisant la division des 10 LSF en sous blocs de 3-3-4 LSF, c'est à dire une quantification à l'aide d'un découpage des LSF en 3 sous-blocs (figure 3.35). Cette méthode est proposée par K.K Paliwal et B.S Atal [58]. Néanmoins, la méthode proposant la division en 4-6 LSF (nombre de sous blocs ($M=2$)) supposait une certaine indépendance entre les 4 premiers et 6 derniers LSF. D'autre part, les deux dictionnaires

accompagnant cette quantification s'étalent sur 12 bits et possèdent ainsi 2^{12} mots de code chacun, ce qui reste encore trop complexe.

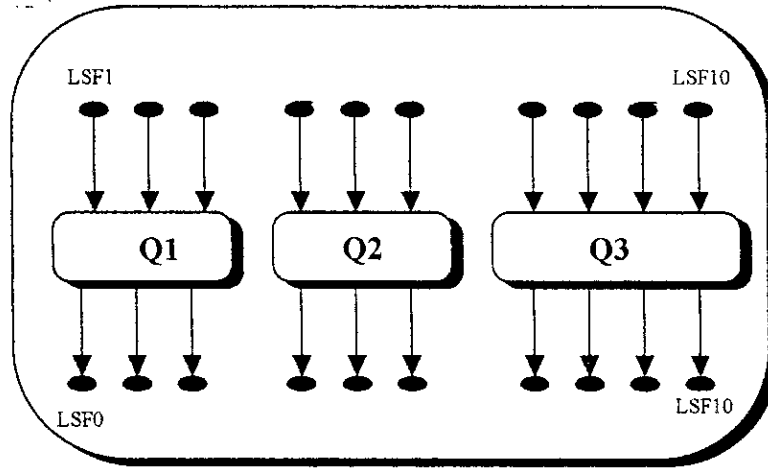


Figure 3.35: Découpage 3-3-4 LSF

Les performances de la quantification vectorielle sur 8 et 9 bits, soit 256 et 512 vecteurs par-dictionnaire, des 3 blocs des LSF sont représentées sur les figures suivantes (fig.3.36 à 3.43) avec la distance euclidienne et celle de Farvardin (fig.3.44 à 3.52) [62,63] pour une base d'apprentissage variée.

Les performances obtenues avec la distance euclidienne pour une base d'apprentissage non variée sont représentées sur les (fig.3.52 à 3.63) et avec la distance de Farvardin (fig.3.64 à 3.75) [64,65,66].

Quantification avec la base d'apprentissage variée
Distance Euclidienne

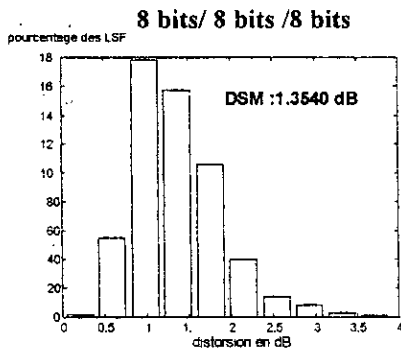


Figure 3.36

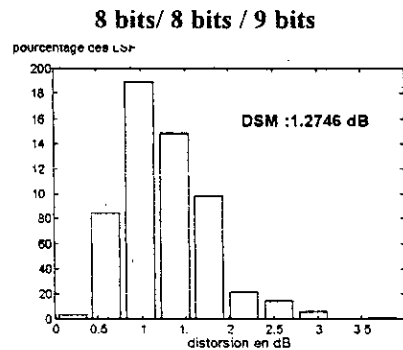


Figure 3.37

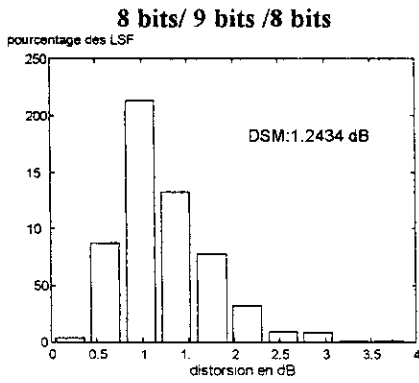


Figure 3.38

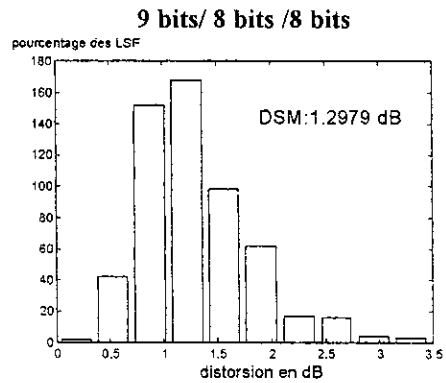


Figure 3.39

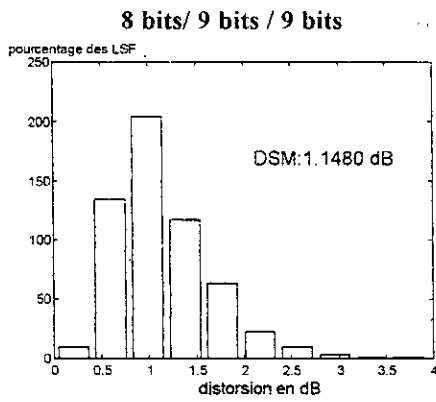


Figure 3.40

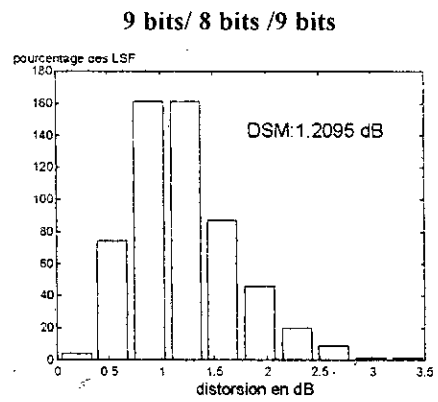


Figure 3.41

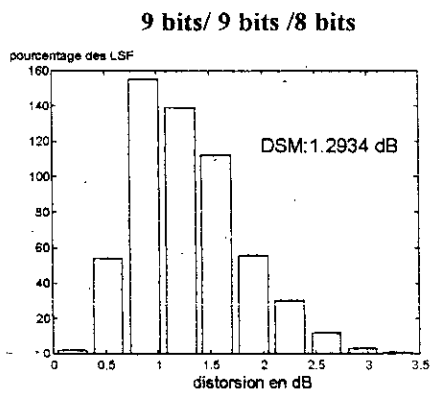


Figure 3.42

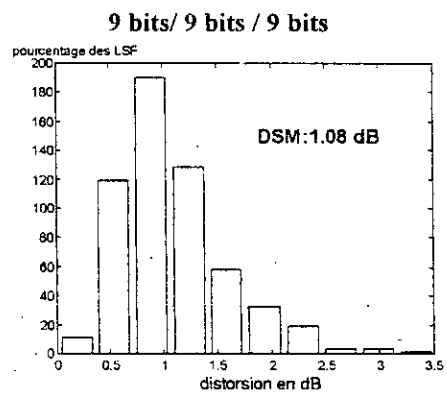


Figure 3.43

Quantification avec la base d'apprentissage variée
Distance de Farvardin

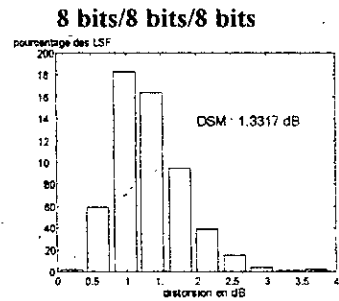


Figure 3.44

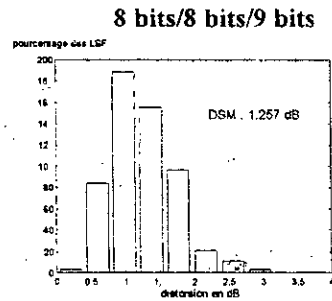


Figure 3.45

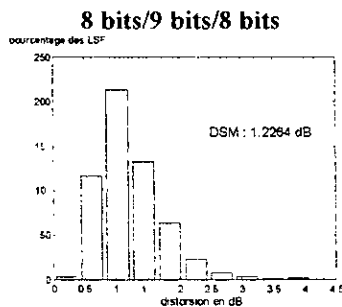


Figure 3.46

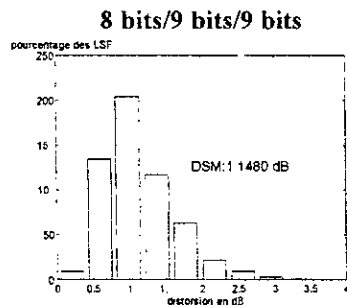


Figure 3.47

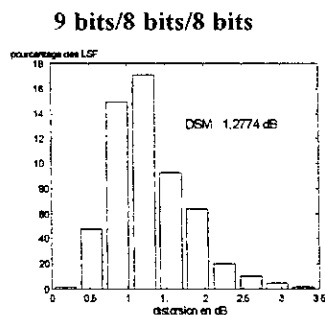


Figure 3.48

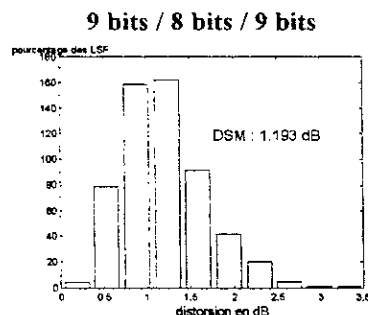


Figure 3.49

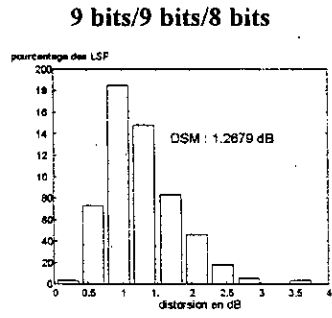


Figure 3.50

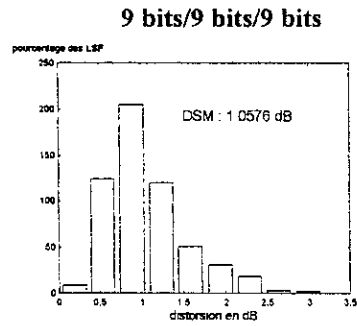


Figure 3.51

**Quantification avec la base d'apprentissage non variée
Distance Euclidienne**

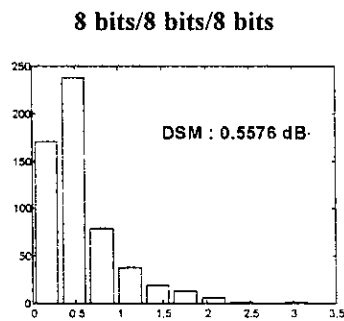


Figure 3.52

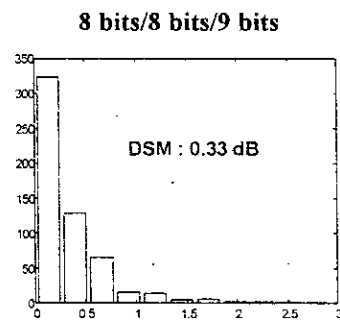


Figure 3.53

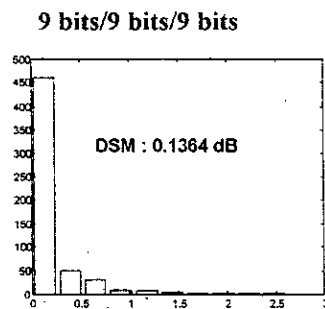


Figure 3.54

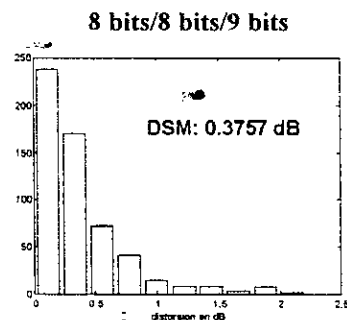


Figure 3.55

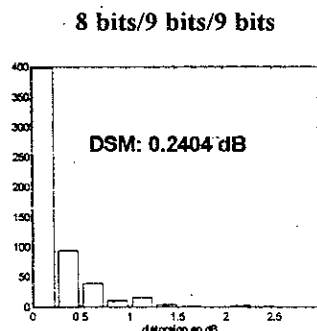


Figure 3.56

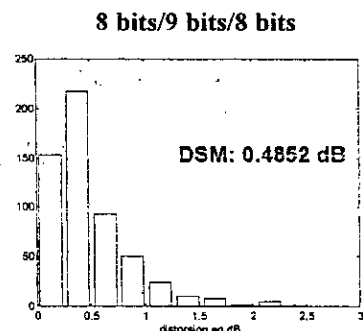


Figure 3.57

9 bits/9 bits/8 bits

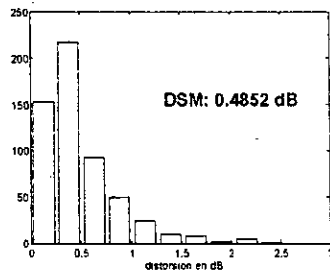


Figure 3.58

9 bits/8 bits/8 bits

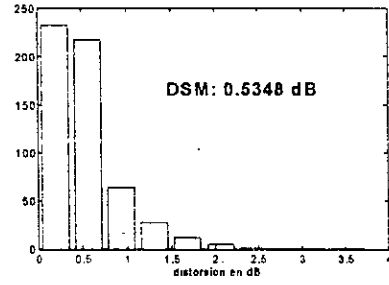


Figure 3.59

6 bits/7 bits/6 bits

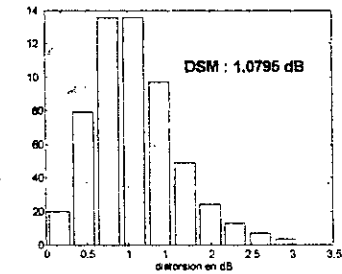


Figure 3.60

6 bits/6 bits/7 bits

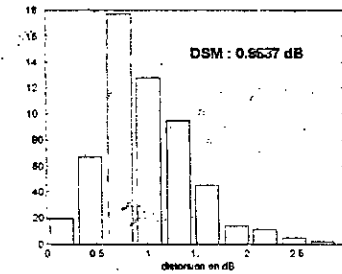


Figure 3.61

6 bits/7 bits/7 bits

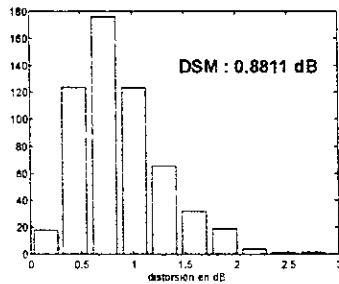


Figure 3.62

6 bits/6 bits/6 bits

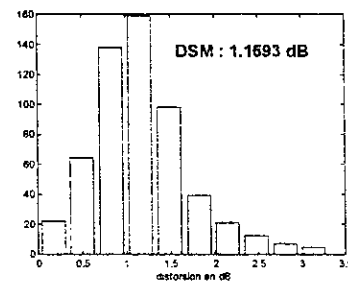


Figure 3.63

Quantification avec la base d'apprentissage non variée
Distance Farvardin

8 bits/8 bits/8 bits

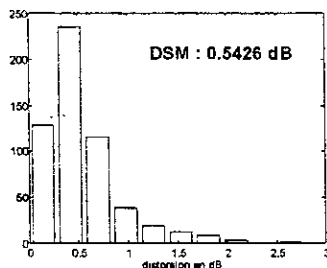


Figure 3.64

8 bits/8 bits/9 bits

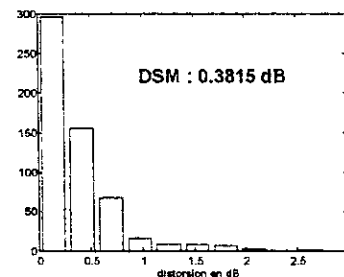


Figure 3.65

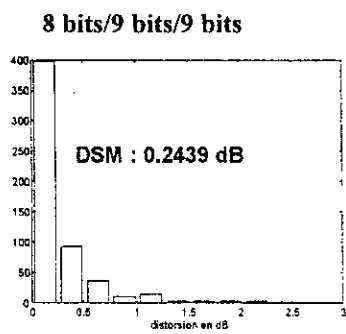


Figure 3.66

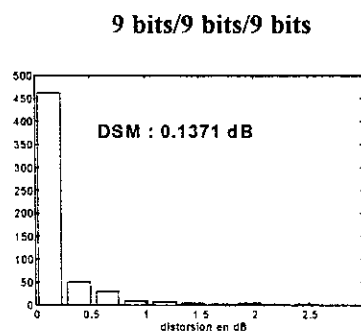


Figure 3.67

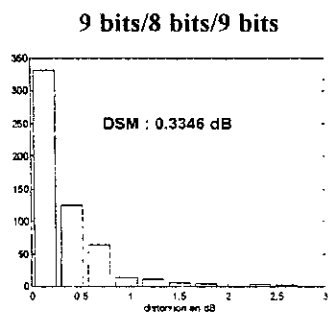


Figure 3.68

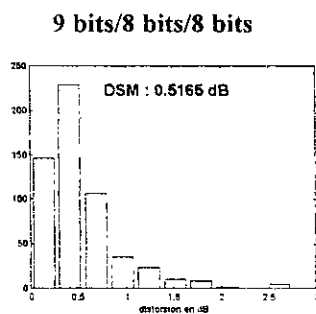


Figure 3.69

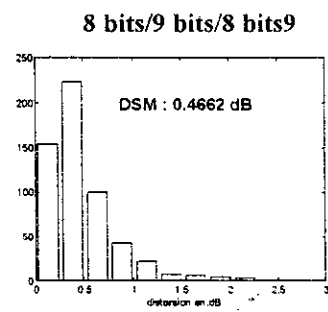


Figure 3.70

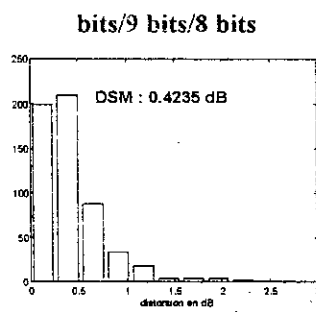


Figure 3.71

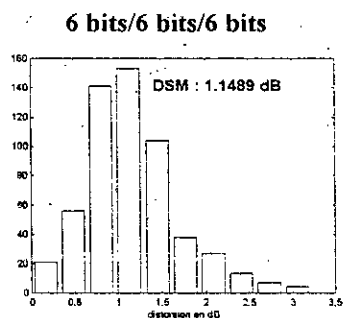


Figure 3.72

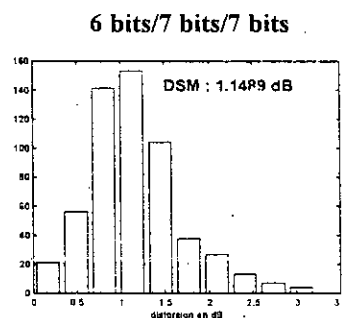


Figure 3.73

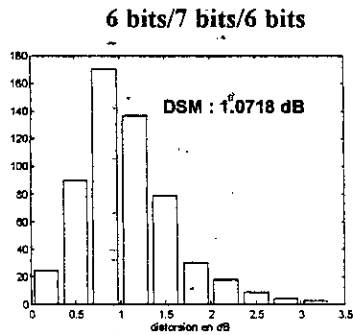


Figure 3.74

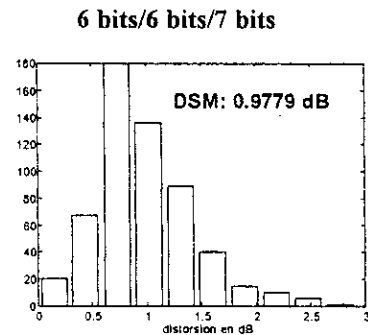


Figure 3.75

La figure 3.76 illustre le découpage 5-5 LSF. On utilise cette division pour une comparaison avec la structure 3-3-4 LSF.

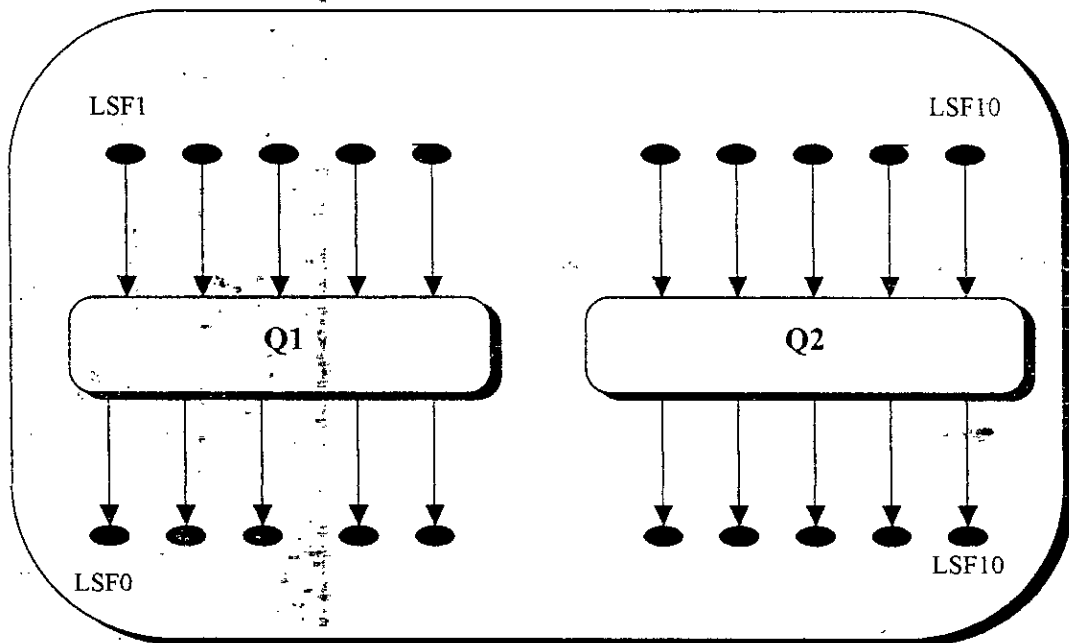


Figure 3.76: Découpage 5-5 LSF

les résultats obtenus pour le découpage 5-5 LSF sont représentés sur les figures 3.51 et 3.52 avec la distance de Euclidienne et celle de Farvardin.

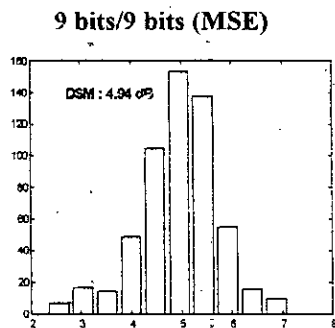


Figure 3.77

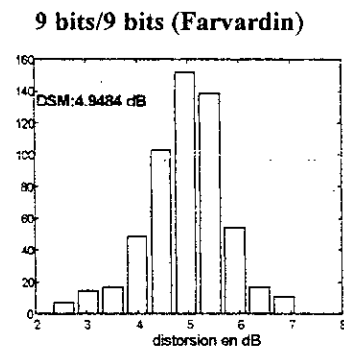


Figure 3.78

4.5.3.4-Quantification vectorielle à étages

La quantification à étages est une autre forme de quantification vectorielle, qui permet de réduire la complexité au prix d'une dégradation en performance.

Dans la quantification à étages, le vecteur des 10 LSF est quantifié par le premier étage, les étages successifs permettent de quantifier les erreurs. La version quantifiée du vecteur des 10 LSF est obtenue par la sommation des sorties des différents étages. (figure 3.79)

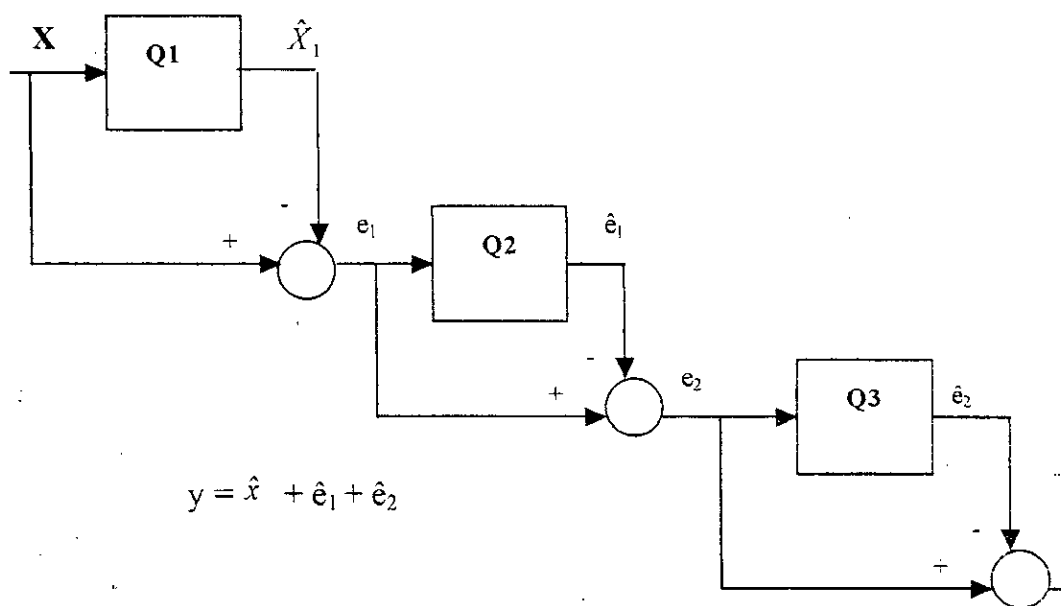


Figure 3.79: Quantificateur vectoriel à 3 étages. La version quantifiée Y du vecteur X est égale à la somme des trois sorties des quantificateurs (Q1, Q2, Q3).

Pour une comparaison des performances entre la quantification vectorielle par Split à 24, 25, 26 et 27 bits, la quantification scalaire et différentielle, étudiés précédemment, nous avons les structures de quantificateurs vectoriels suivants:

- Un quantificateur vectoriel à 4 étages avec 6 bits pour chaque étage .
- Un quantificateur à 4 étages : 7 bits pour 3 premiers étages et (3, 4, 5 et 6 bits) pour le dernier étage.
- Un quantificateur à 4 étages : 8 bits pour 3 premiers étages et (2, 3 et 1 bit) pour le dernier.
- Un quantificateur à 3 étages : 8 bits pour chaque étage.
- Un quantificateur à 3 étages : 7 bits pour le premier et second étages et (3,4) pour le dernier étage.

- Un quantificateur à 3 étages : 6 bits pour le premier et second étages et (3, 4) pour le dernier étage.

Les performances obtenues sont représentées sur les figures suivantes: (fig.3.80 à fig.3.92)

**Quantification (MSVQ) avec la base d'apprentissage non variée
Distance Euclidienne**

**3 étages à 24 bits
(8 bits/ 8 bits / 8 bits)**

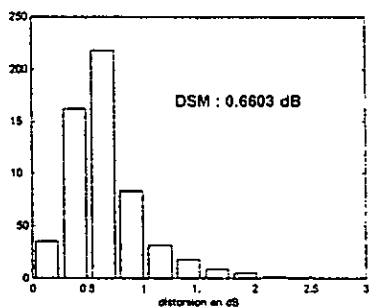


Figure 3.80

**4 étages à 24bits
(6 bits / 6 bits/ 6 bits / 6 bits)**

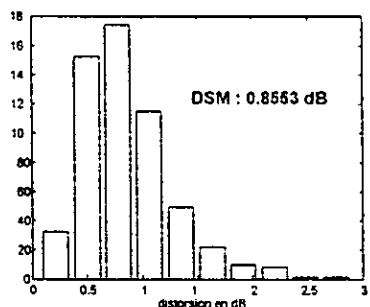


Figure 3.81

**3 étages à 24bits
(7 bits / 7 bits/ 7 bits/ 3 bits)**

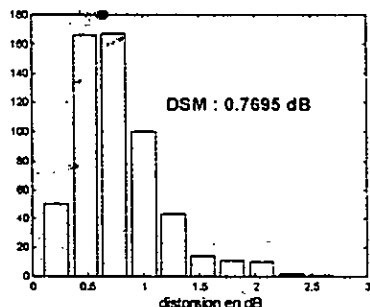


Figure 3.82

**3 étages à 25bits
(7 bits / 7 bits/ 7 bits/ 4 bits)**

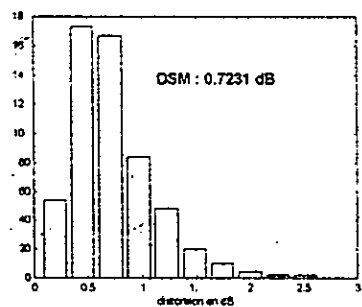


Figure 3.83

**3 étages à 26bits
(7 bits / 7 bits/ 7 bits/ 5 bits)**

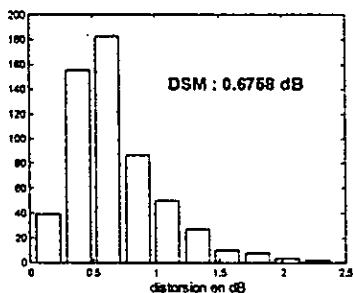


Figure 3.84

**3 étages à 27bits
(7 bits / 7 bits/ 7 bits/ 6 bits)**

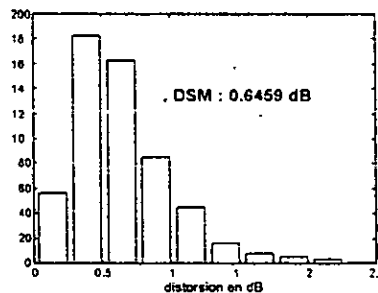


Figure 3.85

4 étages à 27bits
(8 bits / 8 bits / 8 bits / 3 bits)

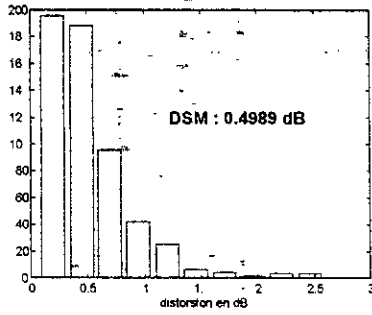


Figure 3.86

2 étages à 12 bits
(6 bits / 6 bits)

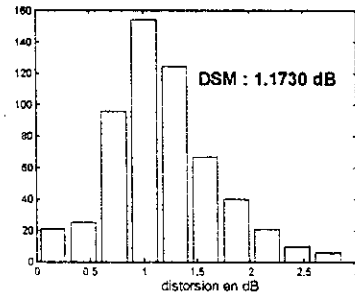


Figure 3.87

3 étages (6 bits / 6 bits / 3 bits)

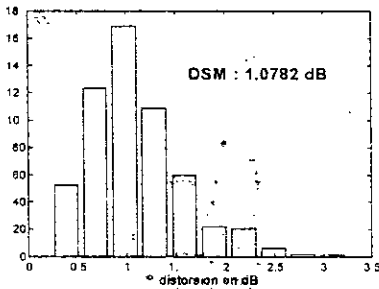


Figure 3.88

3 étages (6 bits / 6 bits / 4 bits)

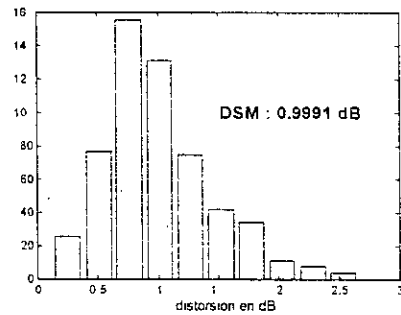


Figure 3.89

3 étages (6 bits / 6 bits / 5 bits)

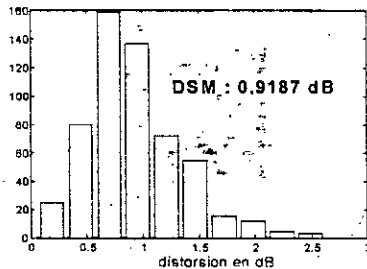


Figure 3.90

3 étages (6 bits / 6 bits / 6 bits)

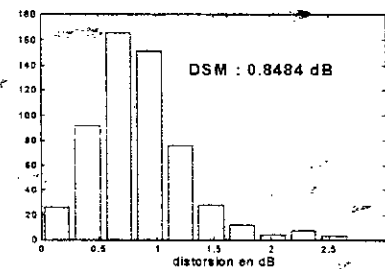


Figure 3.91

3 étages à 17 bits
(7 bits / 7 bits / 3 bits)

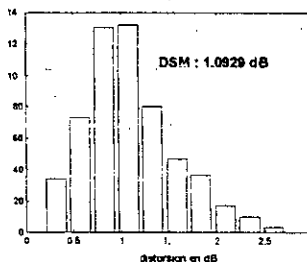


Figure 3.92

Conclusion

Dans ce chapitre, nous avons présenté les performances des différentes variantes de quantification des paramètres LSF, pour cela nous avons réalisé les étapes suivantes:

- dans un premier lieu, nous avons présenté les différentes étapes suivies pour l'extraction des paramètres LSF par la méthode autorégressive LPC, ainsi que le choix des conditions d'analyse. Nous avons présenté également leurs propriétés déterministes et statistiques.
- dans une seconde étape, nous avons étudié, la qualité, l'efficacité et la convergence des dictionnaires composés des vecteurs des LSF, par l'algorithme itératif LBG. Ces dictionnaires vont servir à la quantification de s LSF.
- dans une troisième étape, nous avons étudié quatre méthodes de quantification. La quantification directe, la quantification des Δ LSF, la quantification vectorielle par split et à étage. Nous avons utilisé deux mesures de distances, la distance euclidienne et une distance pondérée appelée distance de Farvardin.
Les performances obtenues sont représentées sous forme d'histogrammes en terme de distorsion spectrale.

Les commentaires sur les différents résultats obtenus seront présentés au chapitre suivant.

Chapitre 4

Interprétation des résultats

Introduction

Dans ce chapitre, on reprend les résultats déjà présentés pour les interpréter et les comparer mutuellement. Cela permettrait au lecteur de tirer un meilleur profit en faisant un bon usage des résultats.

Dans un premier lieu, on parlera de la représentation des LSF dans le plan complexe.

En deuxième lieu, on discutera *l'efficacité, la qualité et la convergence* des différents vecteurs des dictionnaires des LSF utilisés.

Dans une troisième étape, une comparaison des performances obtenues est faite pour les différentes méthodes de quantification étudiées.

4.1-Les paramètres LSF

Nous avons fait dans le chapitre précédent l'extraction des paramètres LSF par la méthode autorégressive LPC. Les résultats obtenus et qui sont représentés sur des cercles unité confirment bien la théorie. En d'autres termes, les LSF sont bien alternés sur le cercle unité.

Cette propriété a été vérifiée aussi bien pour les vecteurs des LSF avant et après quantification.

4.2-Les dictionnaires utilisés

Les dictionnaires utilisés pour la quantification des LSF ont été générés par l'algorithme itératif LBG, comme nous l'avons vu au chapitre précédent.

Une étude a été réalisée sur des dictionnaires des LSF de dimensions allant de 3 à 10 pour des débits de 1 à 10 bits pour chaque dimension.

Les figures (3.9 à 3.12) concernant les entropies, représentent *l'efficacité* des dictionnaires des LSF.

Chapitre 4

On constate que lorsque la taille des dictionnaires augmente, l'entropie augmente également.

Les figures (3.13 à 3.16) concernant la variation du RSB pour différents débits mesurent la *qualité* de codage.

On constate que, les dictionnaires de grande taille, donnent les meilleurs résultats c'est à dire un rapport signal à bruit de plus grande valeur. Cette considération permet de donner un avantage certain à la quantification vectorielle sur la quantification scalaire.

Les figures (3.17 à 3.20) concernant la variation en % du RSB pour différents débits, représentent la convergence. Cette dernière est bien vérifiée pour tous les vecteurs utilisés.

En conclusion, les dictionnaires qui utilisent plus de bits; occupent plus d'espace, donnent des résultats meilleurs.

4.3-Quantification directe de chaque LSF

L'étude la quantification directe sur 34 bits/ trame a donné un bon résultat. C'est à dire une distorsion spectrale moyenne de presque 1 dB.

Les résultats obtenus pour les débits 32 et 30 restent comparables à la littérature concernant le codeur CELP de "federal standard US" à 4.8 kb/s de 1.45 dB.

En conclusion, les résultats obtenus par cette méthode de quantification sont très satisfaisants. Une distorsion spectrale moyenne de 1 dB pour une bonne reproduction du signal parole est atteinte à 34 bits / trame.

4.4-Quantification différentielle

Nous souhaitons maintenant comparer les résultats obtenus par la quantification directe à ceux de la quantification des Δ LSF. Ces résultats sont repris sur le tableau 4.1.

On constate d'après ce tableau, que ces résultats sont satisfaisants, comme l'indique la théorie c'est à dire la quantification des Δ LSF donne de meilleures performances.

La quantification des Δ LSF permet d'atteindre une distorsion spectrale moyenne de presque 1 dB à 28 bits/ trame. Ceci permet un gain de 6 bits par rapport à la quantification scalaire.

En conclusion, il reste plus intéressant de quantifier les Δ LSF plutôt que les LSF.

Débit Bits utilisés par trame	Paramètres utilisés	Distorsion Spectrale moyenne en (dB)
34	LSF	1.05
34	Δ LSF	0.8875
32	LSF	1.1027
32	Δ LSF	0.9084
30	LSF	1.1248
30	Δ LSF	0.9087
28	Δ LSF	1.0134

Tableau 4.1: Performances exprimées en terme de distorsion spectrale moyenne (DSM) du quantificateur scalaire utilisant les représentations LSF et Δ LSF avec la distance euclidienne.

4.5- Performances de la quantification vectorielle par split

Dans ce paragraphe, on va s'intéresser à l'interprétation des résultats obtenus par la quantification vectorielle par split et la comparaison de ces derniers à ceux obtenus par la quantification directe et différentielle déjà présentées.

La comparaison des résultats obtenus à ceux de la littérature, montre que les performances réalisées par ce type de quantificateur sont satisfaisantes.

Les performances obtenues avec une base d'apprentissage variée et qui sont représentées sur les figures (3.36 à 3.51) pour le découpage 3 / 3 / 4 LSF sont reprises sur les tableaux 4.2 et 4.3. Ces résultats sont encourageants, en ce sens que la distorsion spectrale moyenne atteinte à 27 bits, est comparable à celle obtenue par la quantification scalaire sur 34 bits. On remarque d'autre part, sur ces tableaux qu'il est plus intéressant, en terme de distorsion spectrale moyenne, d'attribuer des bits supplémentaires au second et troisième quantificateur.

- en ce qui concerne le quantificateur Q3, la dimension des vecteurs de 3 au lieu de 4 peut expliquer cette observation.
- pour Q2, on constate que la dispersion des vecteurs rend moins performant le dictionnaire; ce qui s'explique par la variété de possibilités existant au niveau des 2nd formants.

Débit Bits utilisés	Allocation de bits de chaque sous vecteur			Distorsion Spectrale moyenne en dB
27	9	9	9	1.0800
26	9	9	8	1.2934
26	9	8	9	1.2095
26	8	9	9	1.1480
25	8	8	9	1.2746
25	8	9	8	1.2434
25	9	8	8	1.2979
24	8	8	8	1.3540

Tableau 4.2: Performances en terme de distorsion spectrale moyenne(DSM) pour un quantificateur par un split de 3-3-4 LSF en fonction du débit pour une distance euclidienne. (base d'apprentissage variée)

Débit Bits utilisés par trame	Allocation de bits à chaque sous vecteur			Distorsion Spectrale moyenne en dB
27	9	9	9	1.0576
26	9	9	8	1.9980
26	9	8	9	1.1930
26	8	9	9	1.1480
25	8	8	9	1.2570
25	8	9	8	1.2264
25	9	8	8	1.2774
24	8	8	8	1.3317

Tableau 4.3 :Performances en terme de distorsion spectrale moyenne(DSM) pour un quantificateur par un split de 3-3-4 LSF en fonction du débit pour une distance de Farvardin. (base d'apprentissage variée)

Débit Bits utilisés par trame	Allocation de bits de chaque sous vecteur			Distorsion Spectrale moyenne en dB
27	9	9	9	0.1371
26	9	9	8	0.4235
26	9	8	9	0.3346
26	8	9	9	0.2439
25	8	8	9	0.3815
25	8	9	8	0.4662
25	9	8	8	0.5165
24	8	8	8	0.5426
20	6	7	7	0.8619
19	6	7	6	1.0718
19	6	6	7	0.9779
18	6	6	6	1.1489

Tableau 4.5: Performances en terme de distorsion spectrale moyenne (DSM) pour un quantificateur par un split de 3-3-4 LSF en fonction du débit pour une distance de Farvardin (base d'apprentissage non variée)

On remarque d'après les tableaux 4.4 et 4.5 que, pour un débit de 19 bits / trame, on a une distorsion spectrale moyenne est presque de 1 dB. Donc les performances réalisées avec une base d'apprentissage non variée sont meilleures que celles réalisées avec une base d'apprentissage variée dans la mesure où on a un gain de 8 bits / trame en comparaison avec les tableaux 2 et 3.

Il est donc plus intéressant d'utiliser une base d'apprentissage non variée pour avoir de meilleures performances avec le découpage 3-3-4 LSF.

Comme pour précédemment, la mesure de distance de Farvardin permet d'avoir de meilleures performances par rapport à la mesure de distance euclidienne.

En comparaison avec la quantification directe sur 34 bits on a un gain de 15 bits et avec la quantification des Δ LSF un gain de 9 bits.

En conclusion le découpage 3-3-4 LSF reste le plus simple et le facile à mettre en oeuvre et donne les meilleures performances avec la base d'apprentissage non variée.

Débit Bits utilisés par trame	Allocation de bits de chaque sous vecteur			Distorsion Spectrale moyenne en dB
27	9	9	9	0.1371
26	9	9	8	0.4235
26	9	8	9	0.3346
26	8	9	9	0.2439
25	8	8	9	0.3815
25	8	9	8	0.4662
25	9	8	8	0.5165
24	8	8	8	0.5426
20	6	7	7	0.8619
19	6	7	6	1.0718
19	6	6	7	0.9779
18	6	6	6	1.1489

Tableau 4.5: Performances en terme de distorsion spectrale moyenne (DSM) pour un quantificateur par un split de 3-3-4 LSF en fonction du débit pour une distance de Farvardin (base d'apprentissage non variée)

On remarque d'après les tableaux 4.4 et 4.5 que, pour un débit de 19 bits / trame, on a une distorsion spectrale moyenne est presque de 1 dB. Donc les performances réalisées avec une base d'apprentissage non variée sont meilleures que celles réalisées avec une base d'apprentissage variée dans la mesure où on a un gain de 8 bits / trame en comparaison avec les tableaux 2 et 3.

Il est donc plus intéressant d'utiliser une base d'apprentissage non variée pour avoir de meilleures performances avec le découpage 3-3-4 LSF.

Comme pour précédemment, la mesure de distance de Farvardin permet d'avoir de meilleures performances par rapport à la mesure de distance euclidienne.

En comparaison avec la quantification directe sur 34 bits on a un gain de 15 bits et avec la quantification des Δ LSF un gain de 9 bits.

En conclusion le découpage 3-3-4 LSF reste le plus simple et le facile à mettre en oeuvre et donne les meilleures performances avec la base d'apprentissage non variée.

4.6- Performances de la quantification vectorielle à étages

Les résultats obtenus par la quantification vectorielle à étages sont repris sur le tableau 4.6.

Débit Bits utilisés par trame	Nombre d'étages	Allocation de bits pour chaque étage				Distorsion Spectrale moyenne (dB)
27	4	8	8	8	3	0.4989
27	4	7	7	7	6	0.6459
26	4	8	8	8	2	
26	4	7	7	7	5	0.6758
25	4	8	8	8	1	
25	4	7	7	7	4	0.7231
24	3	8	8	8	-	0.6603
24	4	7	7	7	3	0.7695
24	4	6	6	6	6	0.8553
18	3	6	6	6	-	0.8484
17	3	6	6	5	-	0.9187
17	3	7	7	3	-	1.0929
16	3	6	6	4	-	0.9991
15	3	6	6	3	-	1.0782

Tableau 4.6: Performances en terme de distorsion spectrale moyenne (DSM) pour un quantificateur à étages en fonction du débit pour une distance euclidienne.

D'après le tableau 6, on constate que l'obtention d'une distorsion spectrale moyenne de presque 1 dB est réalisée par les deux quantificateurs à trois étages, le premier à 17 bits (7 bits, 7 bits, 3 bits) et le second à 15 bits (6 bits, 6 bits, 3 bits).

On constate donc, qu'on a un gain considérable en bits qui est de 17 pour le premier et de 19 bits pour le second par rapport à celui de la quantification directe des LSF.

La comparaison des résultats du tableau 6 avec ceux des tableaux 2 et 4 pour le quantificateur par split de 3- 3 - 4 LSF permet de sauver 2 bits par rapport au quantificateur à étages de (7 bits, 7 bits, 3 bits) et 4 bits pour le quantificateur de (6 bits, 6 bits, 3 bits).

Conclusion

Nous avons comparé les performances de quatre méthodes de quantification des paramètres LSF, la quantification directe, la quantification des Δ LSF, la quantification vectorielle par split et à étages.

Les performances en terme de distorsion spectrale sont déterminées pour un fichier de parole composé de 565 trames.

Basé sur les résultats obtenus, on constate qu'ils sont très satisfaisants. La quantification des Δ LSF permet un gain de 6 bits comparée à la quantification scalaire, par contre le quantificateur par un split de 3 un gain de 7 bits avec la base d'apprentissage variée un gain de 15 bits avec une base d'apprentissage non variée.

La comparaison des résultats obtenus avec le quantificateur à étages a permis de constater qu'on obtient un gain de 19 bits pour le quantificateur à 3 étages (7 bits, 7 bits, 3 bits) et 17 bits pour le quantificateur à 3 étages (6 bits, 6 bits, 3 bits).

Donc la structure à 3 étages permet de sauver 2 à 4 bits par rapport au quantificateur par split de 3-3-4 LSF. Aussi, il faut noter que la complexité est inférieure comparée à split VQ.

En conclusion les quantificateurs à étages présentés dans cette étude permettent d'obtenir de meilleures performances.

Conclusion

L'objectif de notre travail concerne la quantification des LSF. Nous avons été amenés à faire l'extraction des paramètres LSF à l'aide de la méthode autorégressive LPC. Les propriétés des LSF obtenus sont conformes à celles annoncées par la théorie. En effet, les LSF sont entrelacés et occupent bien les deux premiers quadrants du plan complexe Z .

Les propriétés statistiques des LSF ont été étudiées et les résultats montrent bien que la réduction de la gamme de variabilité est obtenue par l'utilisation des différences des LSF successifs. Cette propriété servira à l'étude de la quantification différentielle des LSF.

Pour réaliser la quantification, nous avons testé l'efficacité et la convergence des dictionnaires composés des vecteurs des LSF par l'algorithme itératif LBG. On conclut que les vecteurs de grande dimension permettent d'avoir de meilleurs résultats en terme de rapport signal à bruit. Ce qui donne un avantage certain à la quantification vectorielle par rapport à la quantification directe.

La convergence a été assurée pour tous les vecteurs et les différentes dimensions. Les résultats obtenus sont conformes à ceux annoncés par les auteurs de l'algorithme LBG. C'est à dire une convergence obtenue après six itérations.

Pour la quantification des paramètres LSF, nous avons étudié la quantification directe, la quantification différentielle, la quantification vectorielle par split et à étages pour réduire les 34 bits attribués aux LSF dans le codeur CELP "federal standard US à 4.8 kbits/s.

Les performances des différentes méthodes sont mutuellement comparées. Les résultats obtenus sont très satisfaisants.

Nous avons utilisé une distorsion spectrale exprimée en dB pour permettre une comparaison objective des performances des différentes méthodes de quantification et de les comparer à celles de la littérature. Une distorsion spectrale moyenne de 1 dB représente un niveau communément accepté pour une bonne reproduction de la parole reconstruite.

La quantification directe des LSF à 34 bits/ trame a permis d'avoir une distorsion spectrale de presque 1 dB.

L'utilisation de la quantification des Δ LSF a permis d'obtenir une distorsion spectrale de 1 dB à 28 bits/trame. En comparaison avec la quantification directe ceci permet une réduction de 6 bits. ce qui confirme bien la théorie.

La quantification vectorielle par un Split de 3-3-4 LSF testée sur une base d'apprentissage variée et non variée nous permet un gain en bits de 7 à 15 respectivement. Ce type de quantificateur permet un gain en bits de 1 à 9 par rapport au quantificateur différentiel. Ce type de quantificateur est simple et facile à mettre en oeuvre. Nous avons comparé deux

Conclusion

mesures de distances, la distance euclidienne et la distance de Farvardin. Les performances obtenues avec la distance de Farvardin sont meilleures que celles obtenues par la distance euclidienne en terme de distorsion spectrale moyenne.

La comparaison des résultats précédents avec le quantificateur vectoriel à étage, nous permet un gain considérable en bits. Ce type de quantificateur permet de sauver 17 bits avec la structure à 3 étages (7, 7, 3) et 19 bits avec la structure à 3 étages (6, 6, 3), par rapport à la quantification directe et 11 à 13 bits par rapport à la quantification des Δ LSF. Le gain en bits par rapport à la split VQ est de 10 à 12 bits pour une base d'apprentissage variée et de 2 à 4 bits avec une base d'apprentissage non variée. Basé sur ces résultats, la structure à étage reste la plus performante.

L'indisponibilité d'une carte DSP autour d'un microprocesseur de type C30 ou C31, ne nous a pas permis de faire quelques évaluations subjectives.

Par ailleurs, cette étude, loin d'être exhaustive, se base sur l'approche statistique de quantification vectorielle. L'approche algébrique ou hybride pourrait être une autre orientation future de ce thème.

Bibliographie

- [1] A. Gersho, "Principles of quantization," IEEE Trans. on circuits and syst., vol. CAS 25, pp 427-436, July 1978.
- [2] T. Berger, "Rate distortion theory, a mathematical basis for a data compression," Englewood Cliffs, N.J. ; Prentice-Hall, 1971.
- [3] R.M. Gray, "Source coding theory," Kluwer, Academic Press, 1990.
- [4] N. S. Jayant and P. Noll, "Digital coding of waveforms," Prentice Hall, 1984.
- [5] R. E. Blahut, "Principles and practice of information theory," Addison Wesley, 1987.
- [6] J. P. Adoul, "La quantification vectorielle des signaux : approche algébrique," Annales des télécommunications vol 41, 1986.
- [7] J.P. Adoul, "Speech-coding algorithms and vector quantization, in Advanced digital communications," Ed. K. Feher, 1987.
- [8] S.P. Lloyd, "Least squares quantization in PCM," IEEE Trans. on information theory, vol.IT-28, pp 129-137, March 1987.
- [9] J. Max, "Quantizing for minimum distortion," IRE Trans. on Information theory, vol. IT-6, pp7-12, Mar. 1960.
- [10] M.D. Peaz, T.H. Glisson, "Minimum mean squared error quantization in speech PCM and DPCM systems," IEEE Trans. on Communications, pp 225-230, April 1972.
- [11] W.A. Pearlman and G.H. Senge, "Optimal quantization of the Rayleigh probability distribution," IEEE Trans. on Communications, vol. COM-2, pp 101-112, Jan. 1979.
- [12] A. Gersho, "Asymptotically optimal block quantization," IEEE Trans. on information theory, vol.IT -25, pp 373-380, July 1979.
- [13] R.M. GRAY, A.H. Gray Jr, "Asymptotically optimal quantizers," IEEE Trans. on information theory, vol.IT -23, pp 143-154, July 1977.
- [14] Y. Yamada, S. Tazaki, R.M. Gray, "Asymptotic performance of block quantizers with difference distortion measures," IEEE Trans. on Information theory, vol IT 26 pp 6-14, January 1980.
- [15] Z. Ben Néticha, "Nouvelles techniques de quantification vectorielle algébrique par réseaux réguliers, codes et treillis," Ph. D dissertation, Sherbrooke, 1990.
- [16] R.M. Gray, "Vector Quantization," IEEE ASSP Magazine, vol., pp 4-29, April 1984.

Bibliographie

- [17] Y. Linde, A. Buzo, and R. M. Gray, " An Algorithm for Vector Quantization Design," IEEE Trans. on Communications, vol. COM-28, No. 1, January 1980, pp. 84-95.
- [18] J. Mac Queen, "Some methods for classification and analysis of multivariate observation," The 5th. Berkley Symp. Mathematics, statistics and probability, 1967.
- [19] J. P. Adoul & al., " Spectral distance measure applied to the optimum design of DPCM with L predictors," ICASSP 80 pp 512-515, April 1980.
- [20] J. P. Adoul, " Backward adaptive reencoding : a technique for reducing the bit rate of m-Law PCM transmissions," IEEE Trans on Communications vol COM- 30, pp 581-592, April 1982.
- [21] A. Buzo, A. H. Gray, Gray Jr R.M and J.D Markel "Speech coding based upon vector quantization ," IEEE Trans Acoust Speech and signal Process ASSP628 no 5 pp 562-574 Oct 1980.
- [22] D. L. Gray and D. H. Lee. " On the performance of tree-structured vector quantization," Proc, Int Conf ASSP (Toronto) May 1991, vol 4 pp 2277-2280.
- [23] B.H Juang and A.H Gray, Jr., " Multiple stage vector quantization for speech coding, " Proc. Intl. Conf. on Acoust., Speech, and Signal Process., pp 597-600, Paris, 1982.
- [24] B. Bhattacharya, W. P. LeBlanc, S. A. Mahmoud, and V. Cuperman, "tree.Searched Multistage vector quantization of LPC Parameters For 4 kbits/s Speech Coding," ICASSP, pp. 105-108, May 1992.
- [25] N. Phando, N. Farvardin, and T. Moriya, "Combined Source Channel Coding of LSP Parameters Using Multistage Vector Quantization, " IEEE Workshop on Speech Coding for Telecommunications, pp 36-38, 1991.
- [26] W. P. LeBlanc, V. Cuperman, B. Bhattacharya and S. A. Mahmoud, "Efficient Search and Design Procedures for Robust Multistage VQ of LPC Parameters For 4 kbits/s Speech Coding," IEEE Trans on ASSP, vol. 1, NO, 4, October 1993, pp 373-384.
- [27] A. Gersho and V. Cuperman, " Vector Quantization : A Pattern-Maching Technique for Speech Coding ", IEEE Communication Mag. vol. 21, pp. 15-21, Dec. 1983.
- [28] N. Sloane , " Les empilements de spheres pour la science, " Mars 1984, pp 44-55.
- [29] J. Leech, N; Sloane " Sphere packing and error correcting codes", Canad J. Math vol 23 pp 718-745 1871.
- [30] J. H Conway, N. Sloane" Fast quantizing and decoding algorithms for lattice quantizers ", vol 28 pp 227-232, 1982.
- [31] J. P. N. Schalkwijk, " An algorithm for source coding ", IEEE Trans. On Information theory , vol. 18, pp. Mai 1982.

Bibliographie

- [32] P. E. Papamichlis, "Practical Approaches to Speech Coding," Prentice-Hall, Englewood Cliffs, N. J. 1987.
- [33] F. Itakura and S. Saito, "Analysis synthesis telephony 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.
- [34] 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.
- [35] 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.
- [36] L.R. Rabiner and W. Shafer, " Digital processing of speech signals", Prentice Hall Int., 1978.
- [37] J.D. Markel and A.H; Gray, Jr, " linear prediction of speech ", Springer-Verlag, New York, 1976.
- [38] A. Jennings, " Matrix computation for engineers and scientists ", Wiley and Sons Ltd., 1977.
- [39] B. S. Atal and M. R. Schroeder, " Predictive Coding of Speech Signals and Subjective Error Criteria ," IEEE Trans. on Acoustic, Speech, and Signal Processing, vol. ASSP-27, No. 3, June 1979.
- [40] M. Xie et D.Berkani. " Amélioration des performances des codeurs de parole," Submit. to Al. Journal of Technology, N. 12, ENP 1997.
- [41] Viswanathan and J. Makhoul. "Quantization properties of transmission parameters in linear predictive systems." *IEEE Trans Acoustic Speech Signal Processing* , vol. ASSP-23 pp 309- 321. June 1975.
- [42] A. Gray and J. Markel, " Quantization and bit allocation in speech processing," *IEEE Trans Acoust.,Speech Processing*. Vol. ASSP-24 Dec.1976.
- [43] N. Sugamura and F. Itakura, "Speech data compression by LSP Speech Analysis-Synthesis Technique," IECE. Trans . , vol J64-A, No. 8, pp 599-605, Aug 1981.
- [44] F. Soong and B. H. Juang, "Line spectrum pairs and speech data compression." *In Proc. ICASSP-San Diego, CA, Mar 84* pp. 1101- 1104.1984.
- [45] G. S. Kang and L. J. Fransen, "Application of line spectrum pairs to low bit rate speech encoders," in *Proc, IEEE Int. Conf. Acoust. Speech Signal. Processing*. Tampa, FL pp 731- 734. Apr. 1985

- [46] G. S. Kang and L. J. Fransen, "Low bit rate speech encoders based on line spectrum frequencies (LSFs)," Naval Res. Lab. Rep. 8857. Nov. 1985
- [47] F. Itakura, "Line spectrum representation of linear predictor coefficients of speech signals," *J. Acoustic. Soc Amer.*, vol 57, S36 (A), 1975.
- [48] H. Wakita, "Linear prediction voice synthesizers: Line spectrum pairs (LSP) is the newest of several techniques," *Speech Technol.*, Fall 1981.
- [49] F. Itakura and N. Sagamura. "LSP synthesizer, its principles and implementation." *Tech Rep S79.46. speech Study Group. Acoust. Soc Japan.* Nov 1979.
- [50] N. Sagamura. and F. Itakura "Speech analysis and synthesis methods developed at ECL. in NTT- From LPC to LSP," *Speech Techno.*, vol. 1. pp 17-22. Fall 1981.
- [51] J. Makhoul. "linear Prediction : A tutorial review, " *Proc. IEEE.* vol.63 pp 124-143. Apr. 1975.
- [52] F. Merazka, D. Berkani, "Extraction des paramètres LSF par prédiction Linéaire," COMAEI, Tlemcen, Algeria. Dec. 1996.
- [53] F. Merazka, D. Berkani, " Statistiques des paramètres LSP," Submit. to Al. Journal of Technology, N.12, ENP. 1997
- [54]. P. Kabal and R. Ramachandran, "The computation of Line Spectral Frequencies using Chebychev polynomials, " *IEEE Trans, on Acoustic, Speech Signal Processing*, vol.34, pp.1419-1426, Dec. 1986
- [55] J. P. Campbell, Jr., V.C. Welch and T.E. Tremain, "An expandable error-protected 4800 bps CELP coder (U.S. federal standard 4800 bps voice coder)," *Proc. IEEE Int Conf. on Acoustic, Speech Signal Processing*, Glasgow, Scotland, pp. 735-738, May 1989.
- [56] R. Laroia. N. Phamdo. and N. Farvardin. "Robust and efficient quantization of speech LSP parameters using structured vector quantizers. " *in Proc. ICASSP.* 1991. pp. 641-644.
- [57] K. K. Paliwal and B. Atal. " Efficient Vector Quantization of LPC Parameters at 24 bits/frame," *IEEE Int Conf. on Acoust. Speech and Signal Processing*, vol.1 pp, 3-14, Jan. 1993.
- [58] K. K. Paliwal and B. Atal. " Efficient Vector Quantization of LPC Parameters at 24 bits/frame," *IEEE Ttrans. Speech Audio Processing* pp 661-664 Mar. 1991.
- [59] N. Kitawaki. K. Itoh. and K. Kehiki. "Speech quality measurement methods for synthesized speech." *Review of ECL. NTT Japon.* vol 29. no 9-11. Sept.- Dec. 1981.
- [60] B. S. Atal, R. V Cox and P. Kroon, "Spectral quantization and interpolation for CELP coders," *Proc. IEEE Int Conf. on Acoust. Speech and Signal Processing* Glasgow, Scotland, pp. 69-72, May 1989.

Bibliographie

- [61] F.Merazka, D. Berkani, " La Comparaison d'Algorithmes de Quantification des paramètres LSF," Submit. to Syria Damas, Novembre 1997.
- [62] F.Merazka, D. Berkani, "La Quantification des Paramètres LSF," Inter. Convention on Engineering and Electronics, Boumerdes, Algerie. Dec. 1996.
- [63] F.Merazka, D. Berkani, " Quantification vectorielle des paramètres LSF," Submit. to 17th. JTE'A Novembre 1997, Nabeul - Tunisie
- [64] F.Merazka, D. Berkani, " Utilisation de la " Split VQ " en Quantification Vectorielle des Paramètres LSP," Submit. To IEEEA Octobre 1997, Batna Algerie
- [65] F.Merazka, D. Berkani, "LSP Vector quantization," Submit. To IASTED, Las Palmas, Spain Fevrier 1998.

Annexe A

Algorithme de Wiener-Livinson-Durbin

Dans ce paragraphe, nous allons calculer les solutions de système d'équations donnant les coefficients a_k $k = 1$ à m ; du filtre prédicteur en fonction des coefficients d'autocorrélation r_j $j = 1$ à m . Cet algorithme fonctionne par récurrence. On suppose le problème résolu à l'ordre m et l'on calcule la solution à l'ordre $m+1$ à l'aide de celle à l'ordre m qui est:

$$\sum_{k=1}^m a_{m,k} r_{j-k} = -r_j \quad j = 1 \text{ à } m \quad (\text{A.1})$$

où les $a_{k,m}$ sont les coefficients du filtre prédicteur à l'ordre m . On se propose donc de calculer les mêmes coefficients à l'ordre $m+1$ qui seront solution du système suivant:

$$\sum_{k=1}^{m+1} a_{m+1,k} r_{j-k} = -r_j \quad j = 1 \text{ à } m+1 \quad (\text{A.2})$$

Ce système d'équations peut être divisé en un système d'ordre m et une équation.

$$\sum_{k=1}^m a_{m+1,k} r_{j-k} + a_{m+1,m+1} r_{j-m-1} = -r_j \quad j = 1 \text{ à } m \quad (\text{A.3})$$

$$\sum_{k=1}^m a_{m+1,k} r_{m+1-k} + a_{m+1,m+1} r_0 = -r_{m+1} \quad (\text{A.4})$$

En substituant les valeurs de r_j dans le système (A.3), par celles obtenues par le système (A.1), la solution du problème à l'ordre m , nous obtenons:

$$\sum_{k=1}^m (a_{m+1,k} - a_{m,k}) r_{j-k} + a_{m+1,m+1} r_{j-m-1} = 0 \quad j = 1 \text{ à } m \quad (\text{A.5})$$

D'autre part, en remplaçant l'indice k par $m+1-k$, il est possible de récrire le système (A.1) sous la forme suivante:

$$\sum_{k=1}^m a_{m,m+1-k} r_{j-m+1+k} = -r_j \quad j = 1 \text{ à } m \quad (\text{A.6})$$

Or r_{j-m-1} est en fait r_{m-1-j} car la fonction d'autocorrélation du signal est paire, ce qui permet de substituer les seconds membres du système (A.5) par les valeurs de r_j données par (A.6). Cette opération se fait en remplaçant l'indice j par l'indice $m+1-j$ dans le système (A.6) avant la substitution. Le système (A.3) devient finalement:

$$\sum_{k=1}^m (a_{m+1,k} - a_{m,k}) r_{j-k} + a_{m+1,m+1} - \sum_{k=1}^m a_{m,m+1-k} r_{j-k} = 0 \quad (\text{A.7})$$

De ces équations l'on peut directement déduire les relations qui lient les $a_{m-1,k}$ avec les $a_{m,k}$

$$a_{m+1,k} = a_{m,k} + a_{m+1,m+1} a_{m,m+1-k} \quad k = 1 \text{ à } m \quad (\text{A.8})$$

A cette étape $a_{m-1,m+1}$ reste inconnu, mais il est possible de substituer dans (A.4) les valeurs de $a_{m-1,k}$ $k = 1$ à m par celles obtenues de (A.8). Il résulte une équation qui permet de trouver les $a_{m+1,m+1}$ en fonction de la solution à l'ordre m .

$$\sum_{k=1}^m (a_{m,k} + a_{m+1,m+1} a_{m,m+1-k}) r_{m+1-k} + a_{m+1,m+1} r_0 = -r_{m+1} \quad (\text{A.9})$$

ce qui donne

$$a_{m+1,m+1} = - \frac{\sum_{k=1}^m a_{m,k} r_{m+1-k} + r_{m+1}}{r_0 + \sum_{k=1}^m a_{m,m+1-k} r_{m+1-k}} \quad (\text{A.10})$$

Les équations (A.10) et (A.8) permettent donc de calculer les $a_{m,k}$ en fonction des $a_{m,k}$ et des r_k . Pour $m=1$ le système admet pour solution:

$$a_{1,1} = - \frac{r_1}{r_0} \quad (\text{A.11})$$