Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Ecole nationale polytechnique

DER- Génie Electrique & Informatique

Département d'Electronique

Thèse

en vue de l'obtention du diplôme

de Magister en Electronique

Option

Systèmes de traitement de l'information

Présentée par

FLITTI Farid

Ingénieur d'Etat en Electronique de l'ENP.

THEME

Codeur Harmonique à Bas Débit

Soutenue le 09/06/99 devant le jury:

Président :	**A. FARAH**, Professeur à L'ENP
Rapporteur :	**D. BERKANI,** Professeur à L'ENP
Examinateurs :	**R. AKSAS**, Maître de Conférences à L'ENP
	M. GUERTI, Maître de Conférences à L'ENP
	A. BELOUCHRANI, Docteur à L'ENP

*Promotion: **98/99***

ملخص:

الهدف من الرسالة هو دراسة نوع جديد من المشفرات الجيبية ضعيفة التدفق للاشارة الصوتية. يعتمد هذا المشفر على طريقة جديدة في تمثيل إشارة الحث عند مدخل مرشح التركيب. إشارة الحث عبارة عن مزيج من مصدر عشوائي ومصدر جيبي. نسبة كل مصدر معطاة بواسطة مستوى الدورية الممثل كدالة مستمرة للتردد. الوسائط المميزة لهذا المشفر هي الوسائط LSF , التردد الاساسي, طيف إشارة الحث المواقت للتردد الاساسي, دالة الدورية و طاقة نافذة التحليل.

كلمات مفتاحية: التنبؤ الخطي, تحليل/تركيب, مشفرجيبي, تشفير الإشارة الصوتية, حث جيبي, تشفير الطيف المواقت للتردد الاساسي, تشفير الوسائط LSF .

Résumé : Le but de cette thèse est l'étude d'un nouveau type de codeur harmonique du signal vocal à faible débit. Ce codeur repose sur une nouvelle représentation de l'excitation du filtre de synthèse. L'excitation résulte du mélange d'une source harmonique et une source stochastique. Ce mélange est contrôlé par un niveau de voisement représenté comme une fonction continue de la fréquence. Les paramètres de ce codeur sont les LSF caractérisant le filtre de synthèse, le pitch, le spectre pitch synchro du résidu, l'information de voisement et l'énergie de la fenêtre.

Mots clés: Prédiction linéaire, analyse/synthèse, modèle harmonique, codage de la parole, excitation harmonique, codage du spectre pitch synchro, codage des LSF.

Abstract: The aim of this thesis is the study of a new kind of harmonic speech coder at low bit rate. This coder is based on a new representation of the synthesis filter. The excitation is the result of a mixture of a harmonic source and stochastic source. This mixture is controlled by a voisement level represented as a continuos function of frequency. The coder's parameters are the LSF , the pitch, the pitch synchronous spectrum of the residual, the voisement information and the frame energy.

Keys words: Linear prediction, analysis/synthesis, harmonic model, speech coding, sinus excitation, pitch synchronous spectrum coding, LSF coding.

A tous ceux que j'aime
et à la mémoire de ceux que j'ai perdu.

Remerciements

Ce travail a été effectué au laboratoire "Signal & Communications" du Département d'Electronique, D.E.R Génie Electrique et Informatique à l'Ecole Nationale Polytechnique (ENP).

J'adresse mes vifs remerciements à Monsieur A. FARAH, Professeur à l'ENP et Responsable du laboratoire "Techniques Digitales et Systèmes", pour l'honneur qu'il m'a fait en acceptant de présider le jury.

J'exprime ma profonde reconnaissance à Monsieur D. BERKANI, Professeur à l'ENP et Responsable du laboratoire " Signal & Communications "pour les conseils et les critiques qu'il n'a cessé de me prodiguer qui m'ont été précieux tout au cours de la réalisation de ce travail.

Je remercie Monsieur R. AKSAS, Maître de conférences à l'ENP et Responsable du laboratoire "Télécommunication", pour avoir bien voulu juger ce travail.

Je remercie M^{elle} M. GUERTI, Maître de conférences à l'ENP, pour l'honneur qu'elle m'a fait en acceptant de participer au jury de cette thèse.

Je voudrais exprimer mes remerciements les plus sincères à Monsieur A. BELOUCHRANI, Docteur à l'ENP pour avoir voulu participer au jury de cette thèse.

Mes remerciements vont aussi :

- *A mes parents, ma grand-mère et à toute ma famille.*
- *A tous ceux qui ont contribuer à ma formation.*
- *A tous les membres du laboratoire "Signal & Communications" qui, en toute circonstance, nous ont donné l'exemple de l'esprit d'équipe.*
- *A tous mes amis.*
- *Aux personnels du centre de documentation de l'ENP, et l'ensemble du corps administratif de la D.E.R Génie Electrique & Informatique.*

TABLE DES MATIERES

Introduction

Les besoins incessants d'efficacité d'utilisation des canaux de transmissions ont donné naissance à de nombreux travaux visant la réduction du débit des signaux informationnels. La parole, étant le moyen de communication le plus utilisé par l'homme, fait, et cela depuis plusieurs décennies, l'objet de nombreux travaux de compression.

En premier, on s'est intéressé au codage direct de la forme d'onde du signal vocal. Plusieurs systèmes ont été proposés tels que la modulation par impulsions codées (PCM), le système PCM différentiel (DPCM), la modulation delta, la modulation delta à pas adaptatif, le codage en sous-bandes et le codage par transformée[1,2]. A un débit de 16 kbits/s l'intelligibilité de ces codeurs reste acceptable (enveloppe spectrale correctement restituée), mais leur bruit de quantification crée une gêne trop importante pour que leur utilisation soit envisageable pour des débits plus faibles[1].

Pour atteindre des débits plus faibles, l'introduction de modèle de reproduction de la parole apporte une nette amélioration[5]. L'avantage de l'utilisation d'un modèle de la parole est que la redondance élevée de la forme d'onde est transformée en un ensemble de paramètres, à bande étroite, décrivant le modèle. Le vocodeur (voice coder) à prédiction

linéaire et le vocodeur homomorphique en sont deux exemples. Dans de tels systèmes, la parole est modélisée en sur de courts intervalles comme la réponse d'un système linéaire excité par un train d'impulsions périodiques pour les sons voisés ou par un bruit blanc pour les sons non voisés(chapitre II). Dans cette classe de vocodeurs, la parole est analysée par segment où le signal vocal est quasistationnaire. Chaque segment est pondéré par une fenêtre, comme celle de Hamming par exemple, puis les paramètres du système linéaire et de l'excitation sont déterminés. Les paramètres du système consistent en son enveloppe spectrale ou sa réponse impulsionnelle. Les paramètres de l'excitation consistent en la décision voisé/non voisé et la période de la fréquence fondamentale (pitch) pour les sons voisés. Lors de la synthèse du signal parole, les paramètres de l'analyse sont utilisés pour reconstituer le filtre linéaire et son excitation.

En plus de la bande réduite de leurs paramètres, les modèles de paroles sont utilisés pour réaliser des transformations du signal vocal à travers la modification des paramètres du modèle. Par exemple, dans le rehaussement de la parole hyperbare, on fait une correction du spectre de l'enveloppe sans modifier l'excitation[3,4].

Bien que les vocodeurs basés sur cette classe de modèles aient réussi à produire une parole intelligible, la qualité de celle-ci reste assez peu appréciable. Cette mauvaise qualité est due, d'une part, aux limitations du modèle et, d'autre part, à l'estimation imprécise des paramètres.

Une dégradation majeure présente dans les vocodeurs utilisant une simple décision voisé/non voisé est le bourdonnement détecté spécialement dans les régions contenant un mélange voisé/non voisé ou les régions voisées d'une parole bruitée. Le spectre court-term de ces régions présente des portions dominées par des harmoniques de la fondamentale et des portions dominées par le bruit. Comme la synthèse complète avec une source périodique introduit l'effet de bourdonnement et la synthèse complète avec une source de bruit introduit l'effet d'enrouement, on a pu conclure que le bourdonnement perçu dans les vocodeurs est du au remplacement du bruit dans le spectre original par une énergie périodique dans le spectre synthétisé. Cela est une conséquence directe de la décision régide voisé/non voisé prise pour tout le segment traité.

Une estimation imprécise des paramètres du modèle peut être aussi une source de dégradation de la qualité de la parole. Pour la parole dans des milieux bruités, la fréquence de ces dégradations augmente dramatiquement à cause de la difficulté supplémentaire du problème d'estimation. En conclusion, un système d'analyse/synthèse de la parole de haute qualité doit bénéficier d'un modèle amélioré et de méthodes robustes pour une estimation précise des paramètres du modèle.

De nombreux modèles à excitations mixtes ont été proposés comme solution au problème de bourdonnement des vocodeurs. Dans ces modèles, l'excitation est la somme d'harmoniques et de bruit blanc avec des spectres de formes variables ou non dans le temps.

Pour les systèmes à spectres d'excitation invariants dans le temps, le signal d'excitation est la somme d'une source périodique et une source de bruit avec une enveloppe spectrale fixe. Le rapport de mélange contrôle les amplitudes des deux sources. Les travaux de Itakura et Saito [5,6], et Krown et Goldberg[5,7] en sont deux exemples. Dans le modèle proposé par Itakura et Saito, une source de bruit blanc et une source périodique blanche (son spectre s'étend sur tout domaine du signal vocal) sont additionnées. Le rapport de mélange des deux sources est estimé à partir du pic de l'autocorrelation du signal résiduel. Les résultats n'étaient pas encourageants [5,6]. Pour le modèle de Krown et Goldberg, le rapport de mélange est estimé à partir de l'autocorrelation du résidu. La parole ainsi synthétisée est jugée légèrement étouffée et enrouée. La cause des limitations de ces modèles est que la nature des sources utilisées suppose l'invariance du spectre de l'excitation, or l'analyse des sons voisés et non voisés est en contradiction avec cette hypothèse (chapitre I).

Pour les systèmes à spectres d'excitation variables dans le temps, les spectres des deux sources d'excitations changent d'un segment d'analyse à un autre. Comme exemples, on peut citer les travaux de Fujimara [5,8], Makhoul et al.[5,9],Krown et Goldberg [5,7]et Griffin et Lim [5].

Le modèle de Fujimara est basé sur la division du spectre de l'excitation en trois sous bandes fixes. Une analyse homomorphique est effectuée à chacune d'elles a fin de faire une décision voisé/non voisé basée sur la taille des pics du cepstre pris comme mesure de périodicité.

Dans le modèle proposé par Makhoul et al., l'excitation est la somme d'une source périodique filtrée passe-bas et une source de bruit filtrée passe-haut. Les deux filtres ont la même fréquence de coupure qui est estimée en choisissant la plus haute fréquence pour laquelle le spectre est périodique. La périodicité du spectre est déterminée par l'examen des distances entres les pics successifs et décider si elle est la même, avec un certain niveau de tolérance.

Dans le second modèle d'excitation présenté par Krown et Goldberg, la source périodique est filtrée par un filtre passe-bas à gain variable puis ajoutée à elle-même. La même opération est réalisée pour la source de bruit avec filtrage passe-haut. L'excitation finale est la somme des deux sources résultantes pondérée avec le rapport de mélange

voisé/non voisé. Les gains des filtres et le rapport du mélange voisé/non voisé sont estimés à partir du signal résidu.

Dans le modèle de Griffin et Lim, appelé Multi-Band Excitation modèle (MBE), le spectre court-terme du signal vocal est modélisé comme le produit du spectre de l'excitation et de l'enveloppe spectrale du système linéaire. Le spectre de l'excitation est divisé en plusieurs sous bandes. Chacune d'elle est caractérisée par une décision voisé/non voisé. Le nombre de sous bandes peut atteindre 12.

Une autre approche consiste à modéliser directement la forme d'onde du signal vocale par un modèle sinusoïdal. Cette approche introduite par McAuley et Quierty [10,11,12,13], de type analyse et synthèse, utilise une démarche heuristique pour la détermination des amplitudes, les fréquences et les phases des sinusoïdes.

Tous les modèles cités ci-dessus sont du type analyse et synthèse. Une approche récente, utilisant le principe d'analyse par synthèse, est le modèle LP à excitation multi impulsionnelles. L'excitation du filtre LP consiste en plusieurs impulsions par période fondamentale au lieu d'une seule dans les modèles classiques. Les amplitudes et les positions des impulsions sont obtenues par la minimisation de l'erreur, pondérée par le filtre perceptuel, entre le signal original et le signal synthétisé au niveau du codeur d'où l'origine de l'appellation analyse par synthèse. Une amélioration est de concevoir un dictionnaire d'excitation du quel sera issue la série d'impulsions qui minimise l'erreur perceptuelle. Cette manière de faire est référée sous le nom de CELP (Code Exited Linear Prediction) [14].

Cette thèse a pour but d'étudier et d'évaluer un codeur basé sur un nouveau modèle de production de la parole. Ce modèle introduit par L'équipe du laboratoire Speech coding de l'université de Sherbrooke [15] utilise une excitation résultant d'un mélange entre une source harmonique (constituée par un band d'oscillateurs pilotés par le fondamental et ses harmoniques) et une source gaussienne pour solliciter le filtre de synthèse modélisant le conduit vocal. Ce mélange est fait dans des proportions contrôlées par un niveau de voisement qui est présenté comme une fonction continue de la fréquence.

Les paramètres de ce modèle sont les paramètres LSF caractérisant le filtre de synthèse, la fréquence fondamentale, le niveau de voisement, le spectre d'amplitude du signal résiduel calculé pitch synchro (i.e. à des fréquences multiples du pitch ou à l'un de ces diviseurs) et l'énergie du résidu.

Les quatre derniers paramètres sont utilisés pour régénérer l'excitation du filtre de synthèse au niveau du décodeur afin de reconstituer un signal vocal aussi prés que possible du signal original.

Le premier chapitre est consacré à l'étude du signal vocal et de ses principales caractéristiques fréquencielles et temporelles. Le modèle classique de la production de la parole y est présenté ainsi que la méthode de la prédiction linéaire permettant d'obtenir l'ensemble des coefficients représentants le filtres de synthèse.

Le second chapitre est réservé à l'introduction du modèle du codeur harmonique et la présentation des différents paramètres qui le décrivent.

Dans le troisième chapitre, l'étage de l'analyse est détaillé. On y trouvent l'ensemble des méthodes et algorithmes d'extraction des paramètres du modèle. Enfin l'algorithme d'analyse est présenté.

Dans le quatrième chapitre, réservé à l'étage de synthèse, la reconstitution du signal vocal à partir des paramètres d'analyse y est abordée et l'algorithme de synthèse est décrit.

Le cinquième chapitre est consacré aux différents essais et simulations réalisés sur ce modèle, ainsi que la simulation d'un codeur à très faible débit (2.4 Khz).

CHAPITRE I
Le Signal Vocal

I.1 Introduction

Nous avons dit dans l'introduction de ce travail, que l'utilisation de modèles de production de la parole apporte une nette amélioration par rapport au codage direct de la forme d'onde. Une bonne modélisation du signal vocal nécessite une bonne connaissance des caractéristiques de ce signal et du mécanisme de sa production. Ce chapitre est consacré à l'étude du signal vocal et de ces caractéristiques spectrales et temporelles. On y trouve aussi une description de l'appareil phonatoire humain et sa modélisation qui repose sur la prédiction linéaire.

I.2 Mécanisme de la phonation

La parole résulte de l'action volontaire et coordonnée de l'ensemble de l'appareil phonatoire (figure 1.1).

L'appareil respiratoire fournit l'énergie nécessaire lorsque l'aire est expiré à travers la trachée artère. Au sommet de celle-ci se trouve le larynx où la pression de l'air est modulée avant d'être appliquée au conduit vocal qui s'étant du pharynx jusqu'aux lèvres sur une longueur d'environ 17 cm.

Les cordes vocales sont deux lèvres symétriques placées en travers du larynx; ces lèvres peuvent fermer complètement le larynx et , en s'écartant, déterminer une ouverture triangulaire appelée glotte. L'air y passe librement pendant la respiration et la voix chuchotée, et aussi pendant la phonation des sons non voisés. Les sons voisés résultent au contraire d'une vibration périodique des cordes vocales : des impulsions périodiques de pression sont ainsi appliqués au conduit vocal.

Ce dernier constitué des trois cavités pharyngienne, buccale et nasale, peut être dans le cas statique modélisé par la succession de plusieurs tubes acoustiques de sections divers (figure 1.2). Pour obtenir le modèle électrique, il suffit de remplacer chaque tube du modèle acoustique par son circuit équivalent obtenue par l'analogie électrique acoustique. Le résultat est un filtre à plusieurs résonances et antirésonnances.

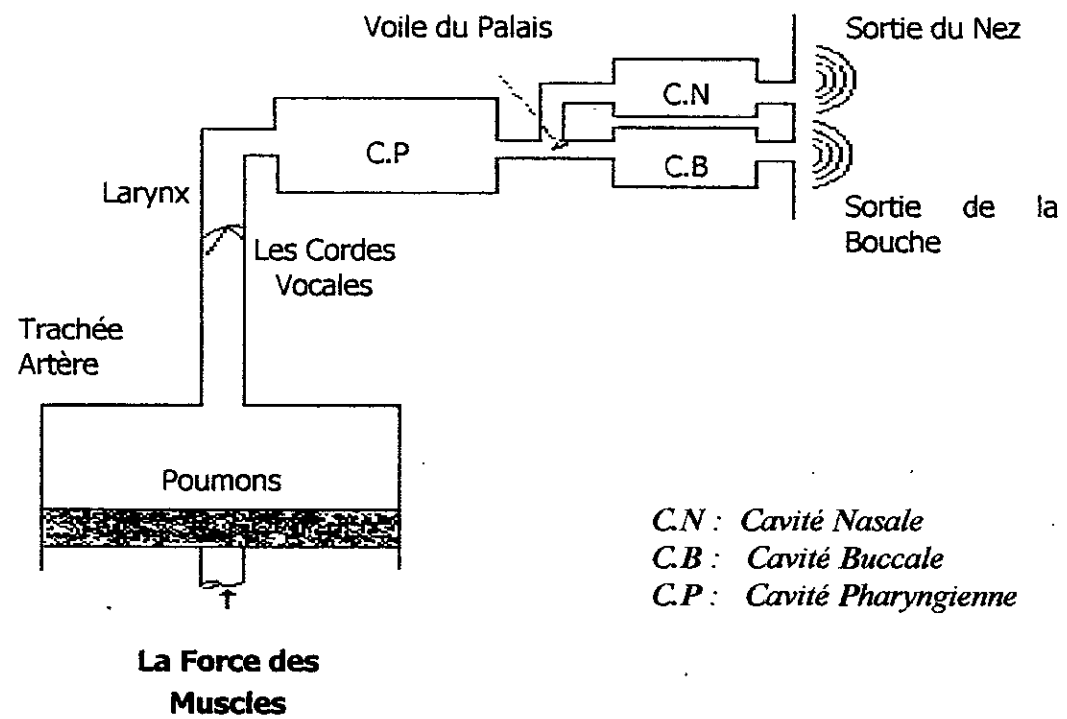


***Figure 1.1** L'appareil phonatoire humain.*

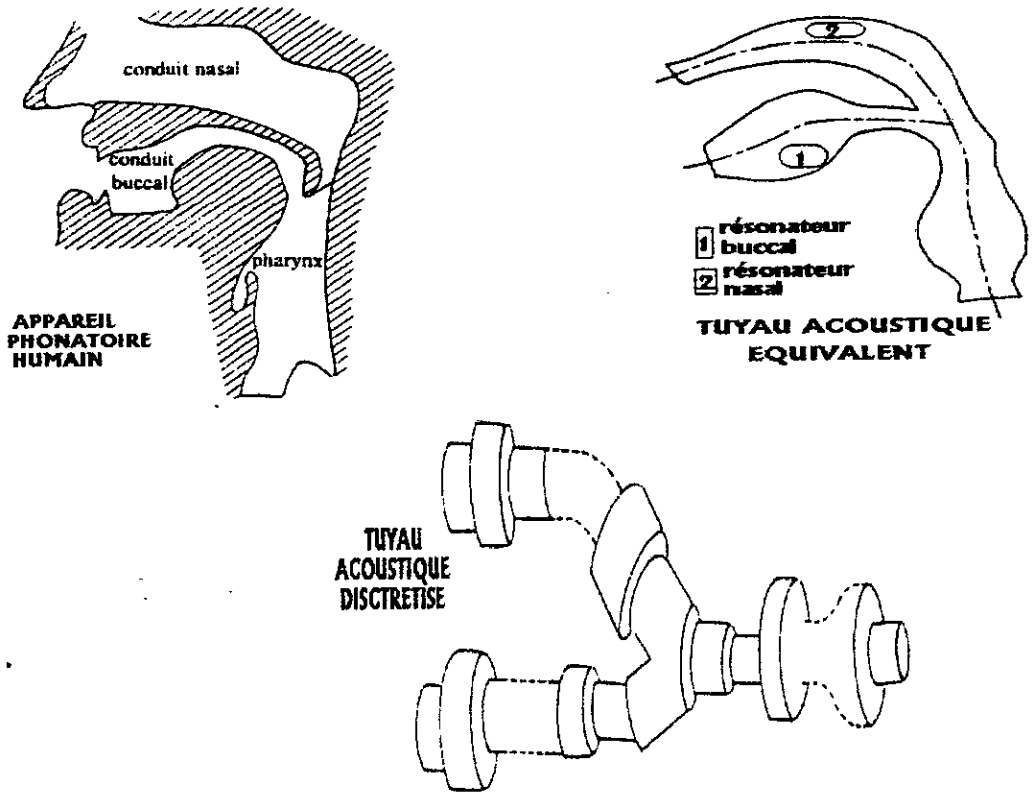


***Figure 1.2** Le conduit vocal et le tuyau acoustique équivalent.*

Les sons émis par l'être humain peuvent être classés en deux catéguories majeurs [16,17]:

- Les sons voisés: ils résultent de l'excitation du conduit vocal par des impulsions périodiques dues aux vibrations des cordes vocales avec une fréquence appelée fréquence fondamentale (pitch). Selon la longueur et la masse des cordes vocales, cette de fréquence peut varier:

de 80	à 200 Hz	pour une voix masculine.
de 150	à 450 Hz	pour une voix féminine.
de 200	à 600 Hz	pour une voix d'enfants.

La figure 1.3 représente la forme d'onde et le spectre d'un son voisé. L'aspect périodique est bien clair dans la forme d'onde, et dans le spectre on observe des raies qui correspondent aux harmoniques de la fondamentale (F_0). L'enveloppe de ces raies présente des maximums appelés formants. Ils correspondent aux fréquences propres F_i (i=1,2,3,...) du conduit vocal. Les trois premiers sont essentiels pour caractériser le spectre vocal tandis

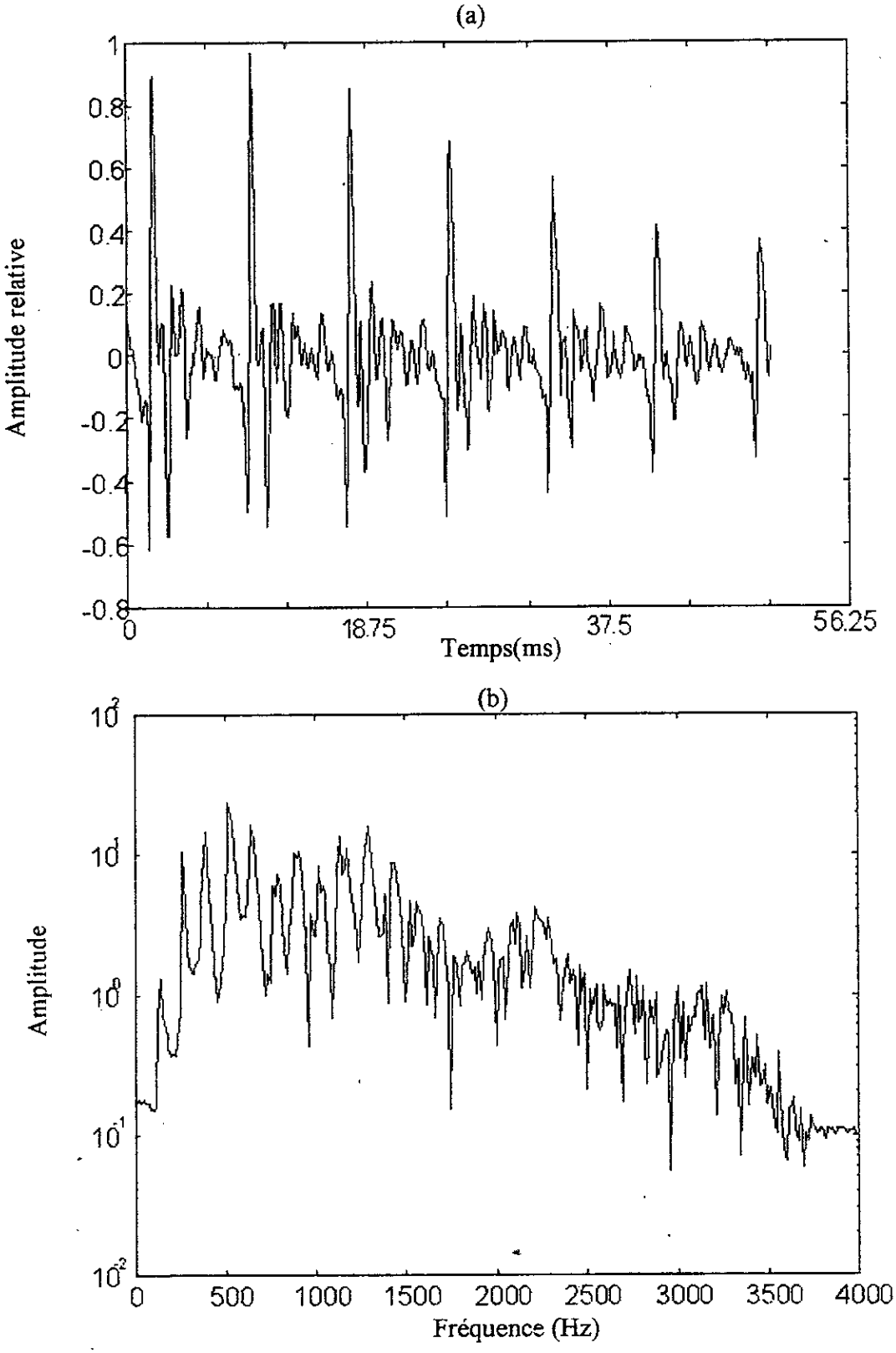


***Figure 1.3** La forme d'onde et le spectre d'un son voisé*
a)s(t): forme d'onde, b)S(f): spectre d'amplitude

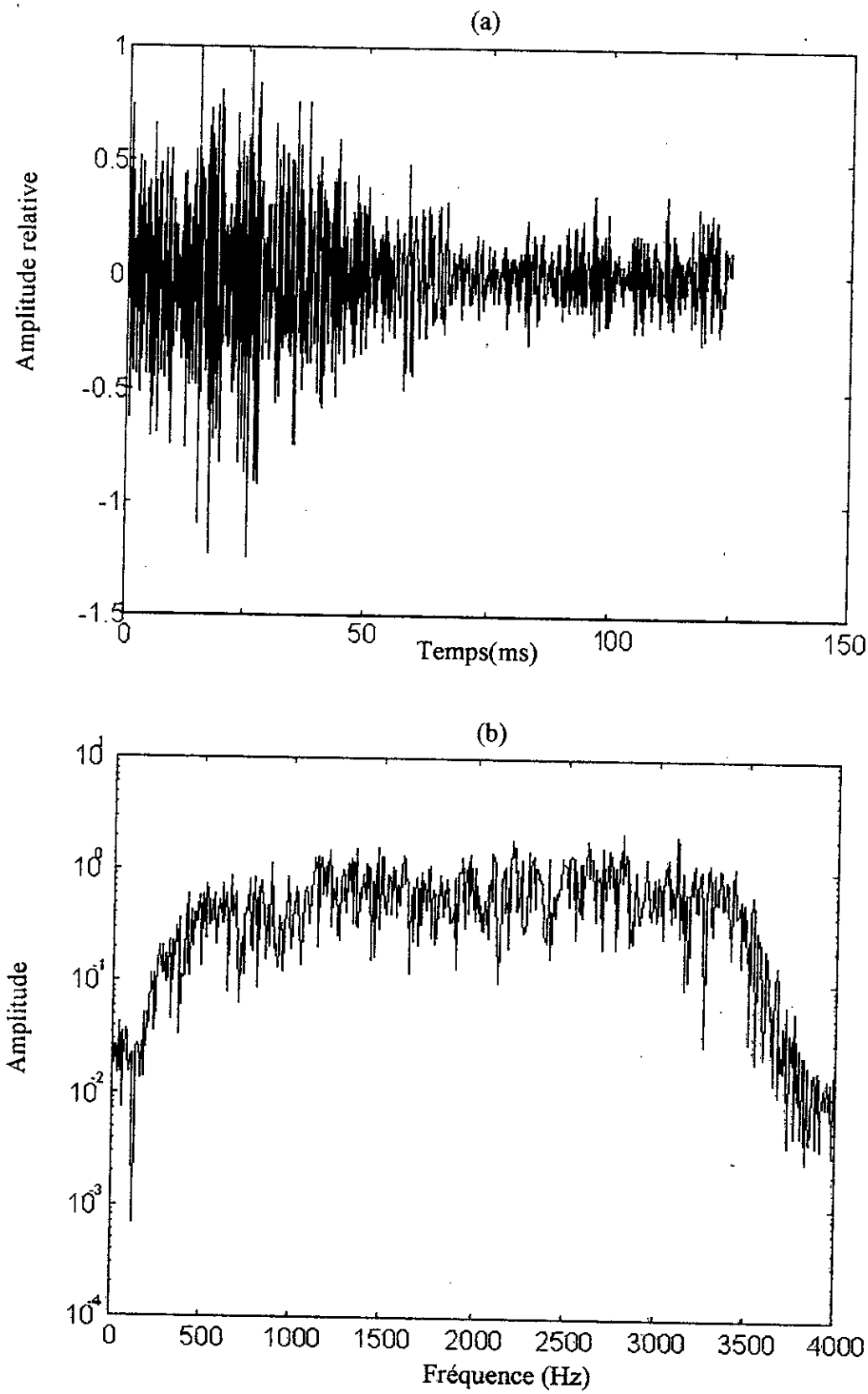


***Figure 1.4** la forme d'onde et le spectre d'un son non voisé*
a)s(t): forme d'onde, b)S(f): spectre d'amplitude

que les formants d'ordre supérieur ont une influence plus limitée pour les applications de communication.

- Les sons non voisés: Ils résultent d'un écoulement turbulent de l'air à travers le conduit vocal. Il peut être assimilé à un bruit blanc filtré par la transmittance du conduit vocal. La figure 1.4 montre la forme d'onde et le spectre d'un son non voisé. La forme d'onde ne présente aucune périodicité et l'amplitude du spectre est plus importante vers les hautes fréquences.

I.3 Modèle de production du Signal Vocal

En se basant sur ce qui a été établie dans la section précédente, on peut dire que la parole résulte de l'excitation du conduit vocal par un train d'impulsions périodiques pour les sons voisés, et un bruit blanc pour les sons non voisés (figure 1.5).

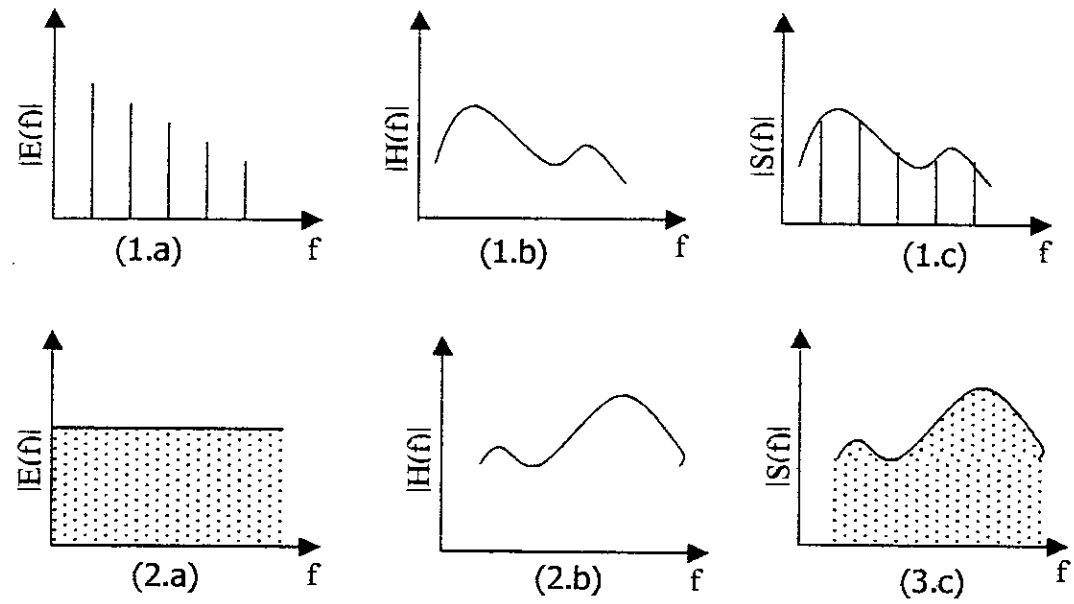


***Figure 1.5** Production du signal vocal(son voisé (1), non voisé (2))*
a)spectre de l'excitation E(f) *b)transmittance du conduit vocal* H(f)
c)spectre du signal vocal S(f)

Sur de courtes périodes, où le signal vocal peut être considérer comme quasi-stationnaire, le conduit vocal est modélisé par un filtre tout pôles variant dans le temps. Sa fonction de transfert est donnée par :

$$H(z)=\frac{G}{A(z)}=\frac{G}{1+\sum_{k=1}^{p} a_k z^{-k}} \qquad (1.1)$$

où p est l'ordre du modèle, les paramètres a_k sont appelés coefficients de prédiction et G représente le facteur de gain. H(z) est appelé filtre de synthèse et A(z) filtre inverse ou filtre d'analyse.

Dans le domaine temporel, L'expression (1.1) est exprimée par l'équation équivalente suivante :

$$s(n) = -\sum_{k=1}^{p} a_k . s(n-k) + G.u(n) \qquad (1.2)$$

qui montre qu'un échantillon s(n) peut être exprimer par une combinaison linéaire des p échantillons qui le précèdent, plus un terme dû à l'excitation. Cette représentation est justifiée par la forte corrélation qui existe entre les échantillons voisins du signal vocal. La figure 1.6 illustre le modèle classique de production de la parole.

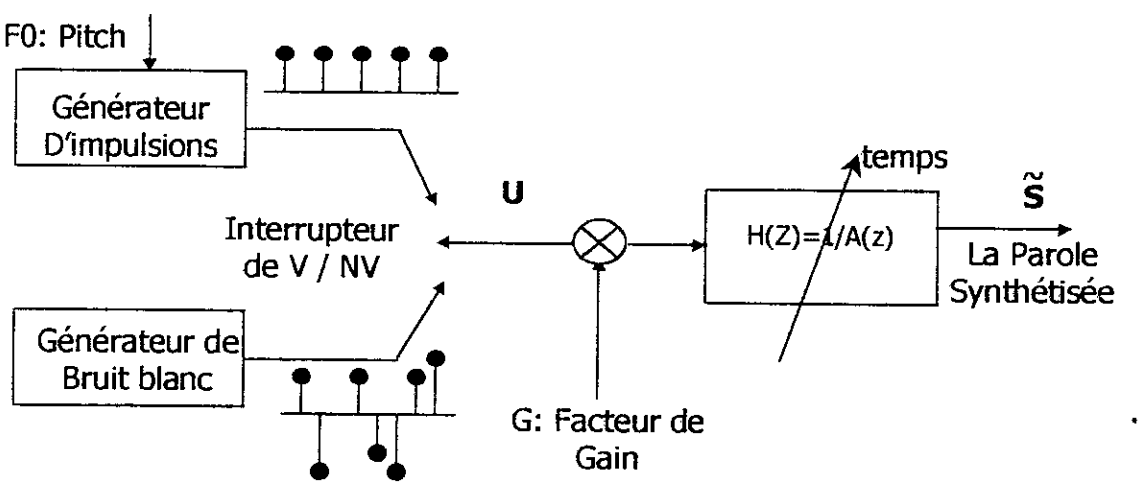


***Figure. 1.6** Modèle de production de la parole[18].*

I.4 La Prédiction Linéaire (LP)

La prédiction linéaire est l'une des méthodes les plus puissantes du traitement de la parole[16,17,18,19]. Elle est utilisée pour l'estimation des coefficients de prédiction caractérisant le filtre de synthèse H(z).

Le principe de base de cette méthode est qu'un échantillon du signal vocal peut être prédit par une combinaison linéaire d'un nombre fini d'échantillons précédents. Ainsi, un

échantillon s(n) peut être approximé par une combinaison linéaire de P échantillons précédents (p=8,...,16 échantillons) :

$$\tilde{s}(n) = -\sum_{k=1}^{p} \tilde{a}_k .s(n-k) \tag{1.3}$$

Les coefficients $\tilde{a}_k$, (k = 1,2,...,p) sont appelés paramètres LP ou coefficients de prédiction, et leur nombre p est appelé ordre de prédiction. s(n) est le n^{ieme} échantillon de la parole originale et $\tilde{s}(n)$ est la sortie du prédicteur linéaire à l'instant n. La fonction de transfert du prédicteur est donnée par :

$$P(z) = -\sum_{k=1}^{p} \tilde{a}_k .z^{-k} \tag{1.4}$$

L'erreur de prédiction *e(n)* est la différence entre l'échantillon original *s(n)* et l'échantillon prédit. Elle est donnée par :

$$e(n) = s(n) - \tilde{s}(n) = s(n) + \sum_{k=1}^{p} \tilde{a}_k .s(n-k) \tag{1.5}$$

La comparaison entre (1.2) et (1.5) montre que lorsque l'ensemble des coefficients $\tilde{a}_k$ est très proche de l'ensemble a_k ($\tilde{a}_k = a_k$, k = 1,2,...,p), l'erreur de prédiction devient égale à l'excitation à un coefficient prés *(e(n) = G.u(n))*. Ceci montre que l'erreur de prédiction e(n) contient beaucoup d'informations sur l'excitation du signal de parole. La transformée en Z de l'équation (1.5) donne :

$$E(z) = S(z).(1 + \sum_{k=1}^{p} \tilde{a}_k .z^{-k}) = S(z).(1 + \sum_{k=1}^{p} a_k .z^{-k}) = S(z).A(z) \tag{1.6}$$

S(z) est la transformée en Z de s(n). A(z) est l'inverse de H(z) dans (1.1), d'où il tire son appellation de filtre inverse. L'équation (1.6) est utilisée pour obtenir le signal erreur e(n) à partir du signal parole s(n).

A cause de la non-stationnarité du signal vocal, les coefficients de prédiction sont estimés sur des intervalles très courts (10-30 ms) où il est supposé quasi-stationnaire. Le but est de trouver l'ensemble des coefficients a_k qui minimisent l'erreur quadratique moyenne de prédiction sur tout le segment de la parole considéré. Cette erreur est définie par :

$$E = \sum_n e^2(n) = \sum_n \left[s(n) + \sum_{k=1}^{p} a_k . s(n-k) \right]^2 \qquad (1.7)$$

Les valeurs a_k qui minimisent E sont obtenues en annulant les dérivées partielles de E, par rapport à chaque coefficient prédicteur a_i (i.e: $\frac{\partial E}{\partial a_i} = 0$ pour i = 1, 2,...,p). Cette dérivée, pour chaque a_i, est donnée par :

$$\frac{\partial E}{\partial a_i} = +2.\sum_n \left\{ \left[s(n) + \sum_{k=1}^{p} a_k . s(n-k) \right] . s(n-i) \right\} = 0 \qquad (1.8)$$

Ce qui conduit au système :

$$-\sum_n s(n).s(n-i) = \sum_n \sum_{k=1}^{p} a_k . s(n-k).s(n-i) \quad ;\text{pour i} = 1,2,...,p \qquad (1.9)$$

ou :

$$-\sum_n s(n).s(n-i) = \sum_{k=1}^{p} a_k . \sum_n s(n-k).s(n-i) \quad ;\text{pour i} = 1,2,...,p \qquad (1.10)$$

En posant :

$$\phi(i,k) = \sum_n s(n-k).s(n-i) \quad ; \text{ pour } \quad i, k = 1,2,\ldots,p \qquad (1.11)$$

l'équation (1.10) devient :

$$\sum_{k=1}^{p} a_k . \phi(i,k) = -\phi(i,0) \quad ;\text{pour i} = 1,2,...,p \qquad (1.12)$$

La technique LP conduit donc à la résolution d'un système de p équations à p inconnus. En premier, on commence par le calcul des valeurs $\phi(i,k)$ (pour i = 1,2,... ,p. et k = 0,1,...,p) en utilisant l'équation (1.11) dans laquelle les limites de la sommation doivent être spécifiées. Deux méthodes, selon le choix des limites de cette sommation, sont généralement utilisées dans l'analyse LP: la méthode de l'autocorrélation et la méthode de la covariance. Nous utiliserons dans notre simulation la méthode d'autocorrélation car elle garantit la stabilité du filtre de synthèse obtenu et elle ne nécessite qu'un espace mémoire réduit [1].

I.4.1 La Méthode d'Autocorrélation

L'erreur quadratique de prédiction (1.7) est calculée sur un intervalle infini. Cependant pour des considérations pratiques, on suppose que s(n) = 0 en dehors de l'intervalle [0, N-1], N étant la durée de la fenêtre d'analyse LP. Ceci est équivalent à multiplier s(n) par une fenêtre w(n) de durée N. L'équation (1.11) devient :

$$\phi(i,k) = \sum_{n=0}^{N+p-1} s(n-i).s(n-k) \quad \text{;pour i = 1, ...,p ; k = 0, ...,p} \tag{1.13}$$

On remarque que seules les valeurs définies pour $0 \leq n \leq N + p - 1$, devront être calculés. Par un changement de variable m = n - i, l'équation (1.13) devient :

$$\phi(i,k) = \sum_{m=0}^{N-1-(i-k)} s(m).s(m+i-k) \tag{1.14}$$

Donc, $\phi(i,k)$ est l'autocorrelations à court terme de s(m) évaluée à (i - k). Par conséquent :

$$\phi(i,k) = R(i-k) \tag{1.15}$$

où :

$$R(j) = \sum_{n=0}^{N-1-j} s(n).s(n+j) = \sum_{n=j}^{N-1} s(n).s(n-j) \quad \text{;pour j = 0,1,...,p} \tag{1.16}$$

L'ensemble de p équations dans (1.12) devient :

$$\sum_{k=1}^{p} a_k.R(|i-k|) = -R(i) \quad \text{;i = 1,2 ..., p} \tag{1.17}$$

L'équation (1.17) est utilisée pour évaluer les coefficients a_k du modèle. Elle peut être représentée sous forme matricielle comme suit :

$$\begin{pmatrix} R(0) & R(1) & R(2) & \cdots & R(p-1) \\ R(1) & R(0) & R(1) & \cdots & R(p-2) \\ R(2) & R(1) & R(0) & \cdots & R(p-3) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ R(p-1) & R(p-2) & R(p-3) & \cdots & R(0) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{pmatrix} = - \begin{pmatrix} R(1) \\ R(2) \\ R(3) \\ \vdots \\ R(p) \end{pmatrix} \tag{1.18}$$

On constate que la matrice des autocorrélations est symétrique et que les éléments situés sur la diagonale sont identiques. Les propriétés de cette matricielle, appelée matrice de TOEPLITZ [1], sont exploitées pour l'élaboration d'un algorithme efficace pour la résolution de l'équation (1.17).

La procédure récursive de Wienner – Levinson - Durbin (WLD) [1,18], représente la solution la plus utilisée. Elle est donnée par :

$E(0) = R(0)$

for i= 1 à p

$$K_i = -\frac{\left[R(i) - \sum_{j=1}^{i-1} a_j^{(i-1)} . R(i-j) \right]}{E(i-1)} \tag{1.19}$$

$a_i^{(i)} = K_i$

pour j=1 à i-1

$$a_j^{(i)} = a_j^{(i-1)} - K_i . a_{i-j}^{(i-1)} \tag{1.20}$$

$$E(i) = (1 - K_i^2) . E(i-1) \tag{1.21}$$

La solution finale est :

$$a_j = a_j^{(p)} \quad \text{j = 1,2,..., p} \tag{1.22}$$

La quantité *E(i)* dans l'équation (1.21) est l'erreur de prédiction d'ordre *i*. Les quantités intermédiaires K_i sont appelées coefficients de réflexion. Ce sont les même coefficients qui apparaissent dans le modèle du tube sans pertes du conduit vocal.

Les coefficients K_i vérifient l'inégalité :

$$-1 \leq K_i \leq 1 \tag{1.23}$$

La relation (1.23) est une condition nécessaire et suffisante pour que le filtre de synthèse soit stable. La méthode d'autocorrélation garantit la stabilité de ce filtre.

L'utilisation de la méthode autocorrélation exige que le signal doit être mis à zéro en dehors de l'intervalle [0, N-1]. Cependant, l'application directe de cette opération produira une très grande augmentation de l'erreur de prédiction au début et à la fin du segment d'analyse. Ce problème est contourné par l'utilisation d'une fenêtre dont l'amplitude diminue graduellement jusqu'au zéro. La fenêtre de Hamming, généralement utilisé, est donnée par :

$$\omega(n) = 0.54 - 0.46.\cos\left(\frac{2\pi n}{(N-1)}\right) \quad n = 0,1,\ldots,N-1 \tag{1.24}$$

La longueur de la fenêtre de hamming est supérieure à la longueur du segment d'analyse de la parole. Le chauvauchement entre les fenêtres des segments adjacents produit un effet de lissage dans l'analyse LP. Il évite le changement brusque des coefficients de prédiction entre blocs voisins. Donc, au lieu du signal original s(n), l'analyse est faite sur le signal fenêtré :

$$x(n) = s(n).\omega(n) \tag{1.25}$$

I.4.2 Les Considérations sur le Choix des Conditions d'Analyse LP

Les variables dans l'analyse LP sont :

- **La méthode d'analyse** :

Bien que la méthode de covariance ne stipule aucune hypothèse sur la morphologie du signal s(n) en dehors des N points disponibles, son utilisation directe peut donner un filtre de synthèse instable. Par ailleurs, la méthode la plus utilisée est celle de l'autocorrélation, malgré l'existence d'une méthode de covariance stabilisée proposée par Atal [1].

- **Le nombre P des coefficients de prédiction** :

Pour des fins de codage, on doit utiliser un nombre minimal de paramètres pour la modélisation de l'enveloppe spectrale court - terme de la parole. Rabiner et Shafer [17] ont démontré que pour une représentation adéquate du conduit vocal, la mémoire du modèle doit être égale à deux fois le temps mis par l'onde de parole pour se propager depuis la glotte jusqu'aux lèvres. C'est à dire **(2.L)/C**, où **L** est la longueur du conduit vocal et **C** la célérité du son dans les conditions normales. Ainsi pour les valeurs de **C** = 34 (cm / ms) et **L** = 17 cm, on obtient un temps de 1 ms. Pour une fréquence d'échantillonnage de **Fs** échantillon/s, le temps de 1ms correspond à **Fs**/1000 échantillons. A la fréquence d'échantillonnage **Fs** = 8 Khz, la valeur correspondante de **p** est égale à **8** au minimum. Par ailleurs, on ajoute généralement un certain nombre de pôles pour représenter

l'influence de la source et du rayonnement. Un ordre de prédiction égal à **10** est généralement utilisé pour les applications de codage.

- **La durée de la trame d'analyse** :

La méthode de l'autocorrélation nécessite un temps d'analyse suffisant pour une bonne résolution spectrale. La pratique a montré que la fenêtre doit empiéter sur plusieurs périodes du fondamental pour les sons voisés. On utilise généralement un intervalle d'environ 20 à 25 ms. On peut dire que cet intervalle est suffisant pour maintenir une bonne qualité de la parole, bien qu'il peut introduire quelques dégradations, surtout pour les sons transitoires, qui ont des changements rapides de leurs caractéristiques spectrales.

La figure 1.7 donne une représentation complète de la prédiction linéaire sous forme d'organigramme.

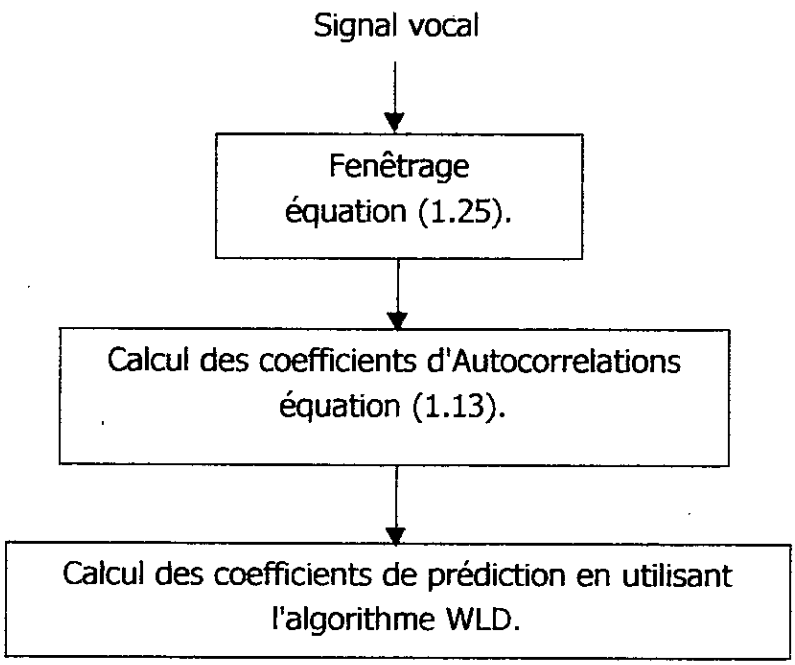


***Figure 1.7** Organigramme de la prédiction linéaire (LP)*

I.5 Conclusion

Dans ce chapitre nous avons repris les notions de base du traitement du signal vocal, développées dans une large bibliographie. Néanmoins, nous avons mis en évidence les principaux caractéristiques et outils du signal vocal qui seront appliquées par la suite dans les différents algorithmes proposés de l'analyse et la synthèse de la parole, au sein du codeur harmonique qui fera l'objet du chapitre suivant.

CHAPITRE II
Le Codeur harmonique

II.1 Introduction

Dans l'introduction, nous avons parlé des limitations des modèles classiques de production de la parole. Nous avons aussi passé en revue plusieurs modèles qui ont été proposés pour améliorer la qualité de la parole synthétisée, et palier aux problèmes de bourdonnement et de l'enrouement. Dans ce chapitre nous ferons une description du modèle du codeur harmonique basé sur la représentation harmonique de l'excitation du filtre de synthèse.

II.2 Modèle du codeur harmonique

Du fait de la quasi-stasionnarité du signal vocal s(n) sur de courtes périodes [16,17], on applique à ce dernier une fenêtre w(n) pour concentrer l'intention sur de faibles durées de 10-35 ms. Le signal fenêtré est défini par :

$$s_w(n) = s(n)\, w(n) \qquad (2\text{-}1)$$

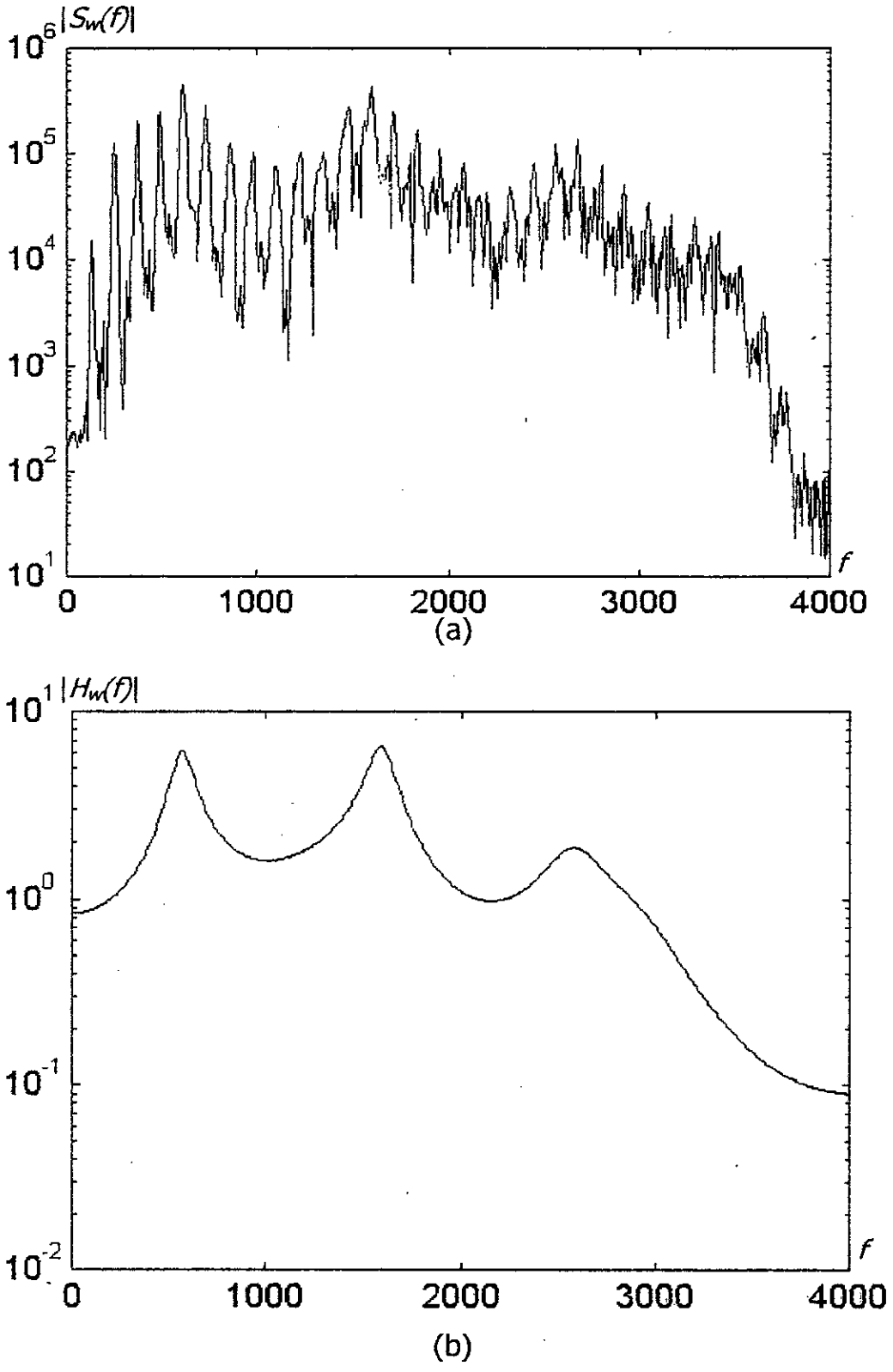


***Figure** 2.1 Le spectre du signal vocal et son enveloppe.*
a) Le spectre du signal b) L'enveloppe spectrale $|H_w(w)|$.

La fenêtre w(n) peut se déplacer sur l'axe du temps pour sélectionner le segment désiré du signal vocal. Sur ce court intervalle d'analyse la transformée de fourrier du signal fenêtré, peut être modélisé par le produit de l'enveloppe spectrale $H_w(w)$ par le spectre de l'excitation $E_w(w)$:

$$S_w(w) = H_w(w)\, E_w(w) \qquad (2\text{-}2)$$

Le spectre de l'enveloppe $H_w(w)$ est la version lissée du spectre du signal original $S_w(w)$ (figure 2.1). Cette enveloppe peut être décrite par les coefficients de prédiction linéaire (chapitre I) ,les coefficients PARCOR [1] ou les coefficients LSF(chapitre III). Le modèle du codeur harmonique utilise la représentation en LSF car ils se prêtent mieux à la quantification que les autres paramètres (section III.2.1).

La figure 2.2 illustre le principe du codeur harmonique. L'excitation résulte toujours de l'addition d'une composante harmonique et une composante stochastique (bruit blanc) sur toute la bande de fréquences du signal synthétisé(figures 2-4, 2-5, 2-6). Au lieu d'avoir une décision binaire V/NV pour la bande complète, comme pour le LPC-10 [15], ou pour chaque sous bande, comme pour le MBE [5], le codeur harmonique utilise un niveau de voisement comme fonction continue de la fréquence. Cette fonction de voisement decidera des proportions des composantes harmonique et stochastique, sommée pour former l'excitation du filtre de synthèse. Les essaies sur le codeurs harmonique ont montré qu'un petit nombre de courbes de voisement sont suffisantes. Un très bon résultat est obtenu avec huit courbes de voisement (3 bits). Par exemple, une courbe indiquant le voisement sur toute la bande, une autre indiquant le non voisement sur toute la bande, et six courbes intermediaires indiquant differents niveaux de voisement. Les courbes de voisement sont toujours des fonctions décroissantes de la fréquence.

Le modèle du codeur harmonique est basé sur une bonne représentation du spectre d'amplitude de l'excitation à l'entrée du filtre de synthèse. D'autant plus que le spectre d'amplitude du signal d'excitation est proche de celui du signal résiduel, une bonne qualité de la parole peut être assurée. On ne se préoccupe pas de reproduire les phases originales du signal résiduel. Il suffit seulement d'assurer leur continuité [15]. Ceci donne lieu à une réduction considérable du débit.

Le spectre d'amplitude de l'excitation est calculé pitch synchro, c'est à dire que les raies spectrales obtenues par la DFT sont à des fréquences multiples de la fréquence fondamentale (f_p) ou de l'un de ses diviseurs ($f_p/2$, $f_p/3$ ou $f_p/4$), selon la valeur du pitch (figure2.3). Cela est du au fait qu'une grande valeur de la fréquence fondamentale ne permettra pas une fine description du spectre d'amplitude. Il est judicieux, dans ce cas, de choisir une fréquence d'analyse du spectre égale à un sous multiple de la fondamentale augmentant ainsi la résolution de la représentation spectrale (section III.5).

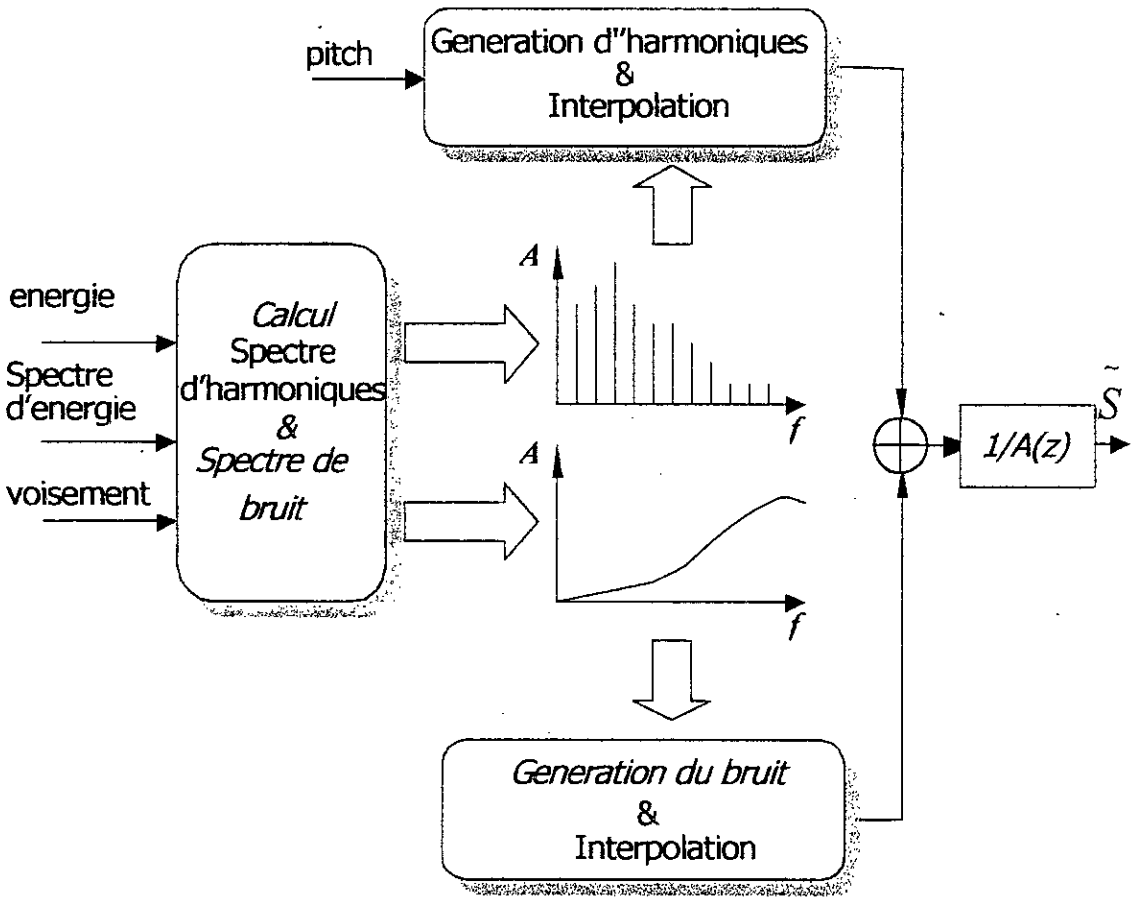


Figure 2.2 Principe du codeur harmonique

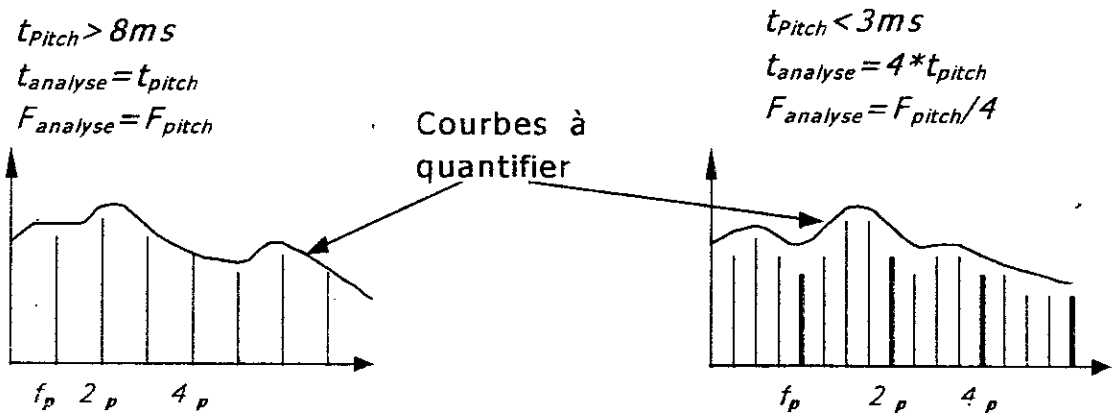


Figure 2.3 Variation de la fréquence d'analyse en fonction du pitch.

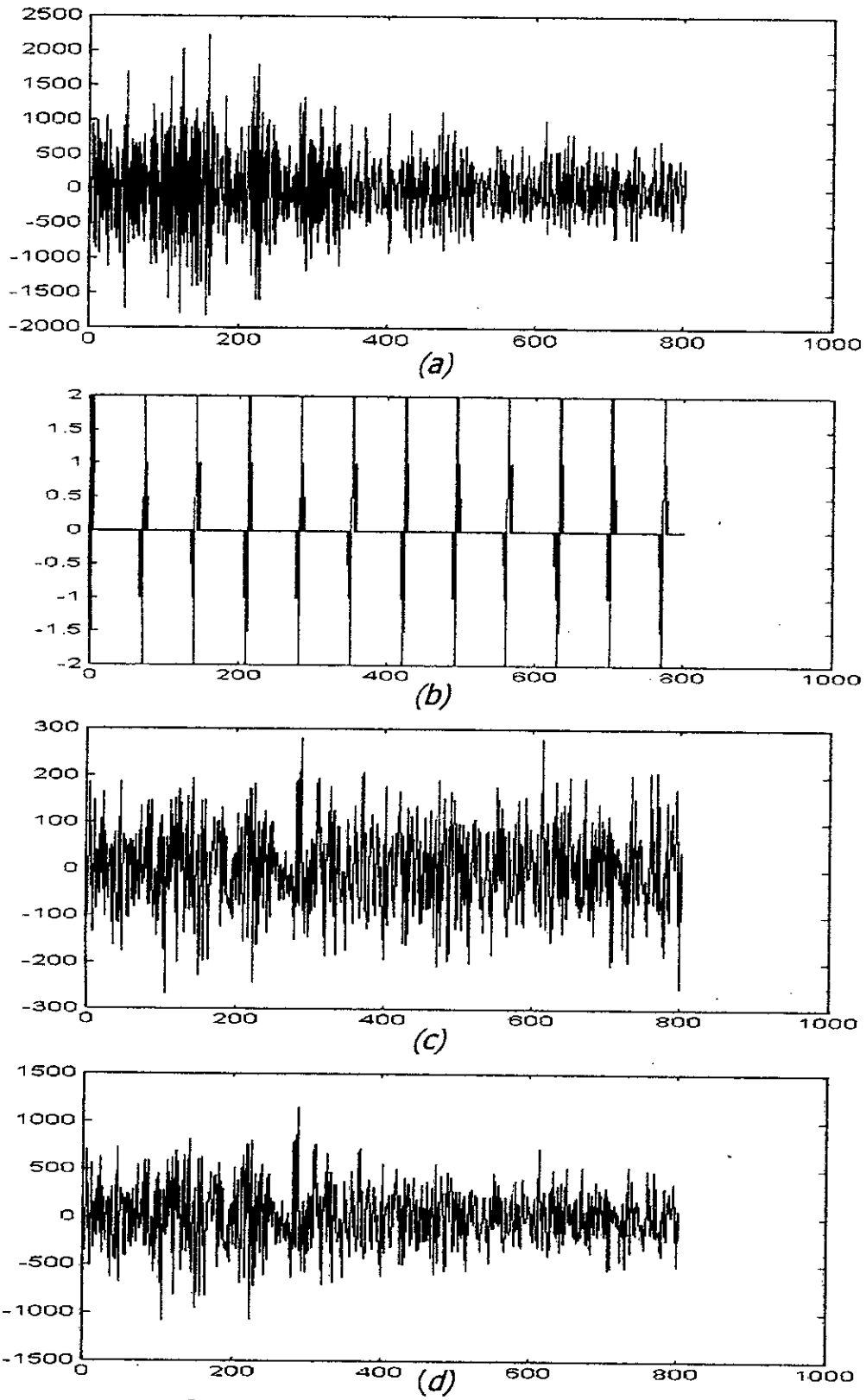
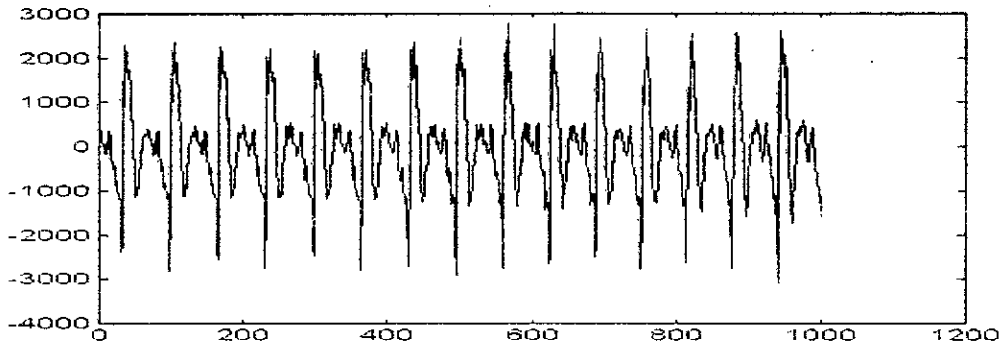
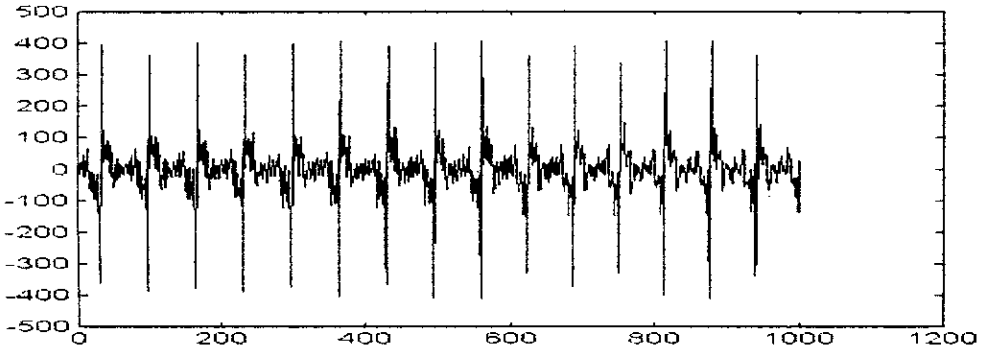


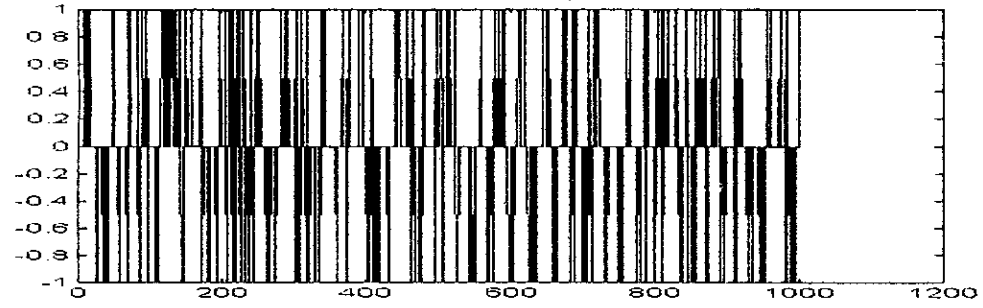
Figure 2.4 *L'excitation pour un signal non voisé :*
a) signal original b) excitation harmonique
c) excitation gaussienne d) excitation totale.



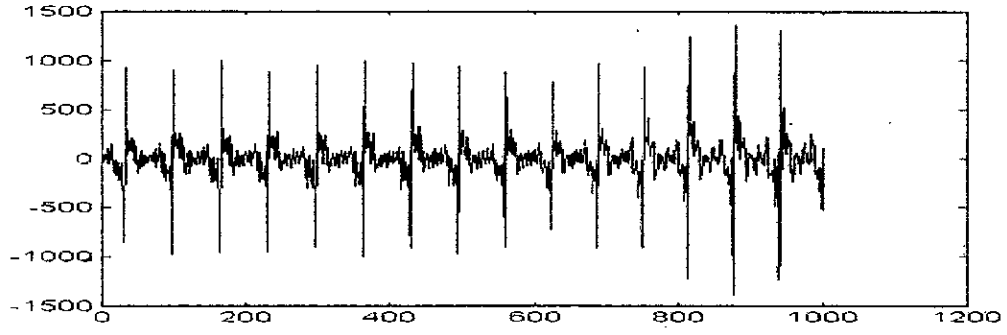
(a)



(b)



(c)



(d)

***Figure 2.5** L'excitation pour un signal voisé :*
a) signal original b) excitation harmonique
c) excitation gaussienne d) excitation totale.

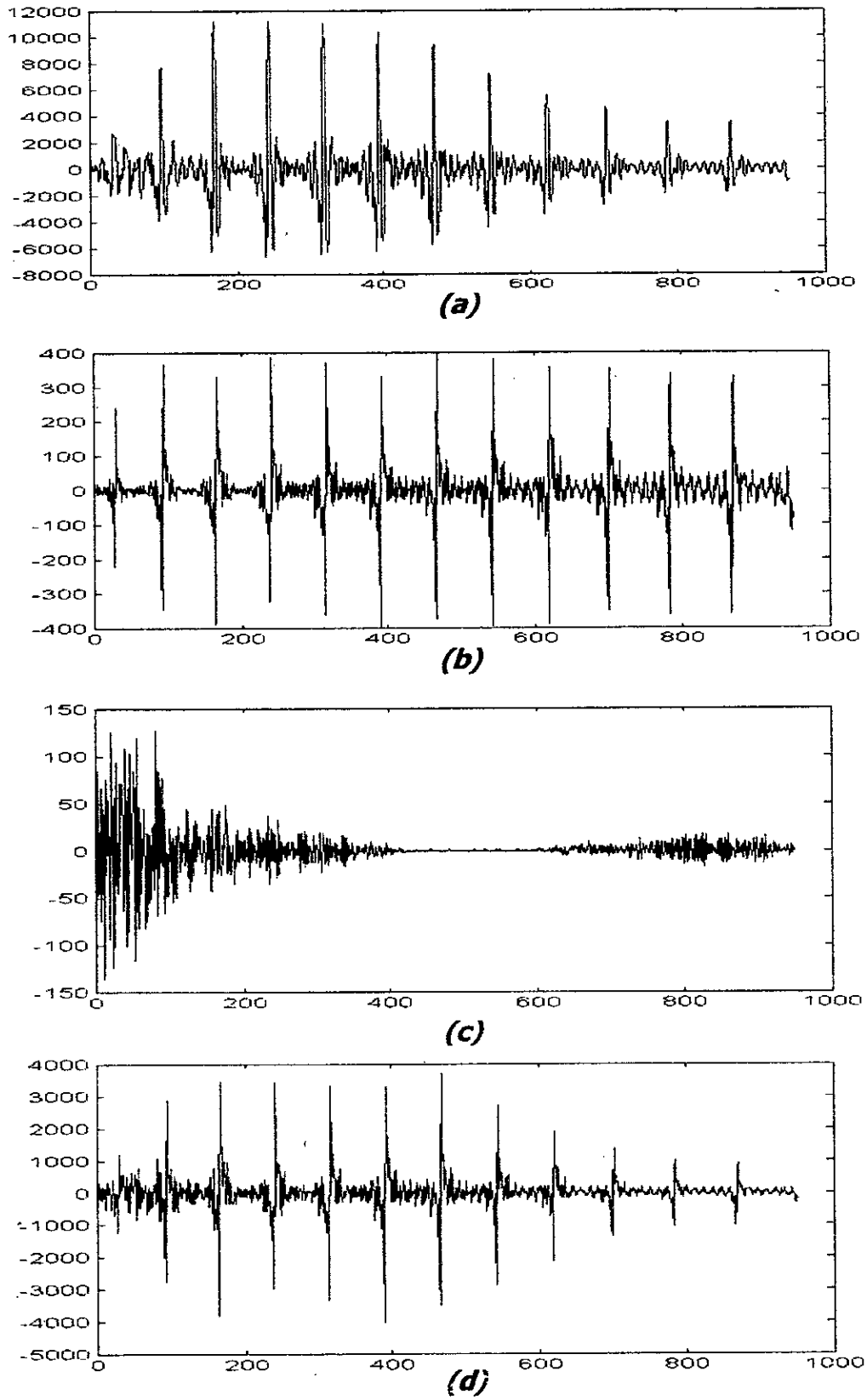


***Figure 2.6** L'excitation pour un mélange v/nv:*
a) signal original b) excitation harmonique
c) excitation gaussienne d) excitation totale.

Il est clair que le pitch conditionne complètement le nombre de raies présentes dans le spectre. Faire une quantification vectorielle directe de ses raies est donc impossible. La solution est de ne conserver qu'un nombre fixe de raies par conditionnement du spectre pitch synchro. En effet l'analyse pitch synchro du spectre engendre un nombre de raies variable à cause du changement de la fréquence d'analyse (f_a) avec la variation du pitch. Le spectre à nombre fixe de raies, adapté à la quantification vectorielle, est, cependant, décrit avec un pas fréquenciel fixe (f_f). La methode la plus simple pour la détermination d'une raie du spectre fixe est de faire une interpolation lineaire des raies les plus proches d'elle dans le spectre pitch synchro (figure 2.7). Une autre manière de réaliser cette operation est décrite à la section III.5.

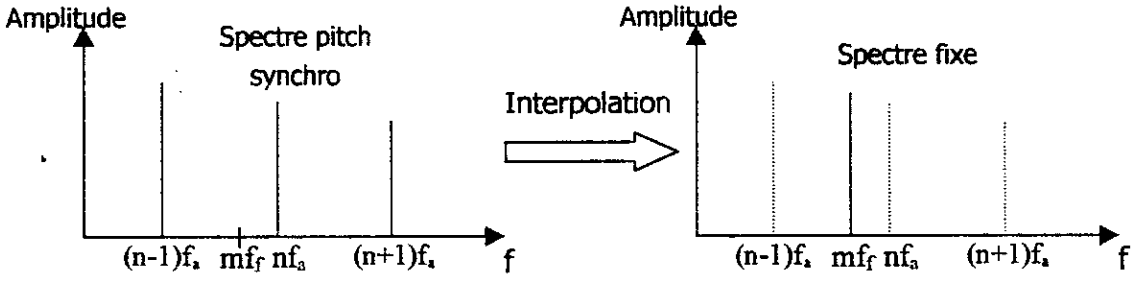


***Figure 2.7** Conditionnement du spectre pitch synchro par interpolation.*

Au niveau du décodeur le spectre pitch synchro sera régénéré à partir du spectre à pas fixe en utilisant l'information sur le pitch. Chaque raie de ce spectre pilote, alors un de l'ensemble des oscillateurs formant le générateur de la composante harmonique de l'excitation.

La composante stochastique, quand à elle, est obtenue par un bruit blanc dont le spectre est conditioné en utilisant l'information de voisement (chapitre IV).

Notons, en fin, que les les amplitudes des raies des spectres des deux composantes ne sont pas utilisée directement, mais elles font l'objet d'une interpolation linéaire toute au long de la trame d'analyse pour assurer une bonne continuité dans les régions de transitions entre les fenêtres.

Dans l'étage d'analyse ,pour chaque segment, les paramètres suivants sont déterminés:

- Les paramètres LSF, caractérisant la fonction de transfert du conduit vocal.
- Le pitch du segment, dont la précision est fixée au 1 /4 d'échontillon.

- Le niveau de voisement, représenté par une fonction continue de la fréquence.
- Le spectre d'amplitude du signal résiduel.
- L'énergie du résidu pour le segment.

Ces paramètres sont ensuite quantifiés avant d'être transmis. Au niveau du recepteur, ils sont utilisés pour reconstituer un signal aussi près que possible de l'original selon le schéma de la figure 2.2. Au chapitre III, nous allons présenter les differents algorithmes d'éxtraction de ces paramètres d'analyses.

CHAPITRE III

Analyse du Signal Vocal

III.1 Introduction

Dans le chapitre II, le modèle du codeur harmonique a été introduit. Les paramètres de ce modèle sont le filtre de synthèse représenté par les coefficients LSF, la fréquence de la fondamentale, le niveau de voisement, le spectre de l'excitation et l'énergie du résidu. Ces paramètres doivent être déterminés pour chaque segment d'analyse. La qualité de la parole synthétisée est d'autant plus bonne que l'estimation de ces paramètres est meilleure.

Dans ce chapitre, nous présenterons les différentes méthodes utilisées pour l'extraction des paramètres du codeur harmonique.

III.2 Le filtre de synthèse

III.2 Le filtre de synthèse

L'utilisation directe des paramètres a_i pour la représentation du filtre LPC n'est pas recommandée. Cela est du au fait que la quantification de ces paramètres peut introduire des instabilités du filtre de synthèse [1]. Différentes représentations, telles que les coefficients de réflexion, les paramètres LAR (Log Area Ratio), les coefficients ceptraux et les paramètres LSF (Line spectral frequencies), ont été proposées pour contourner ce problème. Le modèle du codeur harmonique utilise la représentation en LSF qui est la plus efficace à cause de leurs caractéristiques intéressantes qui vont être citées à la section suivante.

III.2.1Calcul des LSF[1]

Le filtre inverse A(z) associé au filtre de synthèse H(z) satisfait la relation de récurrence suivante[1] :

$$A_n(z) = A_{n-1}(z) - K_n z^{-n} A_{n-1}(z^{-1}) \quad ;n = 1,\ldots, p \tag{3.1}$$

avec $A_0(z)=1$, et K_n est le $n^{ième}$ coefficient de réflexion. Ces coefficients appelés aussi coefficients PARCOR (Partial Corrélation), sont interprétés comme les limites du modèle du tube acoustique. En étendant l'ordre du filtre à n=p+1, l'équation (3.1) devient :

$$A_{p+1}(z) = A_p(z) - K_{p+1} z^{-(p+1)} A_p(z^{-1}) \tag{3.2}$$

On considère les deux conditions limites artificielles $K_{p+1} =1$ et $K_{p+1} =-1$, qui correspondent, respectivement, à l'ouverture complète et la fermeture complète au niveau de la glottes dans le modèle du tube acoustique. Sous ces conditions, nous obtiendrons les deux polynômes suivants :

$$\begin{aligned} F_1(z) &= A_p(z) + z^{-(p+1)} A_p(z^{-1}) \\ &= 1+ a_1 z^{-1}+ a_2 z^{-2} +\ldots+ a_2 z^{-(p-1)} + a_1 z^{-p}+ z^{-(p+1)} \end{aligned} \tag{3.3}$$

pour $K_{p+1} =-1$, et:

$$\begin{aligned} F_2(z) &= A_p(z) - z^{-(p+1)} A_p(z^{-1}) \\ &= 1+ a_1 z^{-1}+ a_2 z^{-2} +\ldots- a_2 z^{-(p-1)} - a_1 z^{-p} - z^{-(p+1)} \end{aligned} \tag{3.3}$$

pour $K_{p+1} =1$.

Les polynômes $F_1(z)$ et $F_2(z)$ possèdent les propriétés importantes suivantes :

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

2- les racines de $F_1(z)$ et $F_2(z)$ alternent sur le cercle unité.

3- A(z) est de phase minimum (la stabilité de H(z) est facilement préservée après la quantification des racines de $F_1(z)$ et $F_2(z)$).

Comme les racines de $F_1(z)$ et $F_2(z)$ sont sur le cercle unité, elles sont données par $e^{i2\pi fi}$, et il est facile de montrer que pour le cas d'un ordre de prédiction p pair, $F_1(z)$ et $F_2(z)$ sont données par :

$$F_1(z) = (1 + z^{-1}) \prod_{i=2,4,..,p} (1 - 2\cos(2\pi f_i)\, z^{-1} + z^{-2}) \qquad (3.5)$$

$$F_2(z) = (1 - z^{-1}) \prod_{i=1,3,..,p-1} (1 - 2\cos(2\pi f_i)\, z^{-1} + z^{-2}) \qquad (3.6)$$

Les fréquences f_i, qui correspondent aux racines des polynômes $F_1(z)$ et $F_2(z)$, dans les équations (3.5) et (3.6) sont normalisées par la fréquence d'échantillonnage f_s. Les paramètres f_i, i =1,...,m, sont appelés les paires de raies spectrales LSP (line spectrum pairs) ou LSF(line spectrum frequencies). Il est important de noter que f_0=0 et f_{p+1}=0.5 sont des racines fixes correspondant à z=1 et z=-1 respectivement. Ils sont donc exclus de l'ensemble des paramètres LSF nécessaires pour caractériser le filtre de synthèse. Les LSF peuvent être interpréter comme les fréquences de résonances du conduit vocal sous les deux conditions limites au niveau de la glotte (ouverture complète ou fermeture complète). Une deuxième propriété des paramètres LSF peut être donnée sous la forme suivante :

$$f_0 < f_1 < f_2 < \ldots < f_{p-1} < f_p < f_{p+1} \qquad (3.7)$$

avec f_0=0 et f_{p+1}=0.5. Cette relation est connue sous le nom de propriété d'ordonnancement des LSF. Aussi bien que est cette propriété est conservée la stabilité de H(z) est assurée.

Les paramètres LSF se prêtent mieux à la quantification que les autres représentations du filtre LPC à cause des propriétés suivantes :

1-les LSF ont de bonnes propriétés statistiques, et la stabilité du filtre de synthèse est assurée par la préservation de la propriété d'ordonnancement. En plus, cette propriété permet la détection des erreurs de transmissions des LSF sans introduire de redondance.

2- il y a une relation évidente entre les LSF et le spectre du filtre LPC. Une concentration des LSF dans une certaine bande de fréquences correspond approximativement à une résonance dans cette bande.

3-Les LSF entre deux fenêtres d'analyse adjacentes sont fortement corrélés.

III.2.2 Algorithme de Calcul des LSF[18]

Les deux polynômes $F_1(z)$ et $F_2(z)$, qui sont symétrique et antisymétrique respectivement, ont pour racines fixes $z=1$ et $z=-1$. Pour un ordre de prédiction pair, ces deux racines peuvent être enlever par division polynomiale. Ceci donne :

$$G_1(z)= F_1(z)/(1+z^{-1}) \quad \text{et} \quad G_2(z)= F_2(z)/(1-z^{-1}) \qquad (3.8)$$

Les résultants $G_1(z)$ et $G_2(z)$ sont des polynômes symétriques pour un ordre p pair. Comme les racines n'apparaissent que par pairs complexes conjuguées, il est possible de ne déterminer que ceux situées sur la moitié supérieure du cercle unité. Les racines en question sont e^{jw_i} pour i=1,2,...,p. Les LSF sont les positions angulaires des racines, $0<w_i<0.5$. En faisant le changement de variable suivant :

$$M=p/2 \qquad (3\text{-}9)$$

on peut exprimer sous forme explicite la symétrie des polynômes $G_i(z)$ comme suit:

$$G_i(z)=1+g_i(1)\, z^{-1}+\ldots+g_i(1)\, z^{-(2M-1)}+ z^{-2M} \quad ; i=1,2 \qquad (3.10)$$

En enlevant le terme de la phase linéaire, nous aurons deux développement en séries de cosinus à phase nulle :

$$G_i(e^{jw}) = e^{-jwM}\, G_i'(w) \qquad (3.11)$$

avec

$$G_i'(w) = .2\ \cos Mw + 2g_i(1)\cos(M-1)w + \ldots +2\, g_i(M-1)\cos w + g_i(M) \qquad (3.12)$$

En faisant le changement de variable x=cosw, on obtient :

$$\text{Cos } mw = T_m(x) \qquad (3.13)$$

où $T_m(x)$ est un polynôme de chebychev ($1^{ère}$ espèce)d'ordre m. Les polynômes de chebychev vérifient la récurrence :

$$T_k(x) = 2 \times T_{k-1}(x) - T_{k-2}(x) \qquad (3.14)$$

avec les polynômes d'ordre zéro et un donnés par : $T_0(x)=1$ et $T_1(x)=x$. L'équation (3.12), en utilisant le développement en polynômes de chebychev, devient :

$$G_i'(w) = 2\,T_M(x) + 2g_i(1)\,T_{M-1}(x) + \ldots + 2\,g_i(M-1)\,T_1(x) + g_i(M) \qquad (3.15)$$

Lorsque les racines x_i de $G_1'(x)$ et $G_2'(x)$ sont déterminées, les LSF correspondants sont donnés par $w_i=\arccos(x_i)$. Le changement de variable $x=\cos w$ transforme la moitié supérieure du cercle unité, dans le plan z, en l'intervalle [-1,+1], sur la droite des x (figure 3.1). Ainsi tout les x_i sont situées entre –1 et 1, avec la racine correspondante à la plus faible fréquence située plus près de 1.

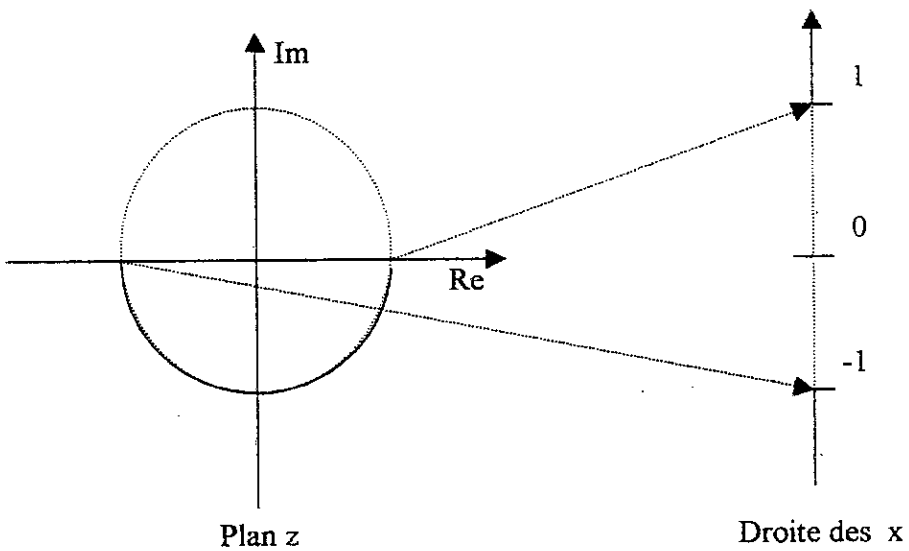


Figure 3.1 *Correspondance entre le plan z et la droite des x pour le changement de variable x=cosw*

Soit la série de polynômes de chebychev :

$$Y(x) = \sum_{k=0}^{N-1} c_k\,T_k(x) \qquad (3.16)$$

Considérons la récurrence suivante :

$$b_k(x) = 2.x\,b_{k+1}(x) - b_{k+2}(x) + c_k \qquad (3.17)$$

avec les conditions initiales $b_N(x) = b_{N+1}(x) = 0$. Cette récurrence est utilisée pour calculer $b_0(x)$ et $b_2(x)$. Après, Y(x) peut être exprimée en termes de $b_0(x)$ et $b_2(x)$ par la relation:

$$Y(x) = \sum_{k=0}^{N-1} [\, b_k(x) - 2.x\, b_{k+1}(x) + b_{k+2}(x)]\, T_k(x)$$

$$= (b_0(x) + b_2(x) + c_0)/2 \qquad (3.18)$$

L'avantage de cette formulation est que les erreurs d'évaluation de $b_0(x)$ et $b_2(x)$ tendent à s'annuler. En se rappelant de la propriété d'entrelacement et sachant que la plus proche racine de x=1 correspond au polynôme $G_1(x)$, la procédure d'extraction des x_i peut être abordée :

$x_0 = x_1 = 1$, i=1

Pour j = 1 à P

1- Evaluer $G_i'(x_0)$ utilisant 3.17 et 3.18

2- $x_1 = x_1 + \Delta x$

3- Evaluer $G_i'(x_1)$ utilisant 3.17 et 3.18

4- Si $G_i'(x_0) \,.\, G_i'(x_1) > 0$ aller à 2

5- Recherche de la solution précise x_j par divisions successives de $[\, x_0 , x_1 \,]$

6- Si i = 1 alors i=2 sinon i = 1

7- $x_0 = x_1$

8- aller à 2

fin .

Enfin, il faut noter que le pas Δx doit être assez faible pour pouvoir détecter toutes les racines, mais non pas trop afin d'accélérer l'algorithme de résolution.

III.3 Détermination du pitch

La méthode fait appel au calcul de la fonction d'autocorrelation normalisée. Le pitch correspond à l'intervalle séparant le premier maximum de l'autocorrélation différent du zéro de l'origine. Cependant le calcul direct de cette fonction à partir du signal vocal présente des inconvénients. Le signal de parole est une convolution entre le signal quasi-periodique d'excitation et la réponse impulsionnelle du conduit vocal. Or, les formants du conduit vocal ont des largeurs de bandes assez étroites pour produire plusieurs oscillations d'amplitudes dans la fonction d'autocorrelation. Par ailleurs, il peut exister des

interférences entre ces oscillations et la composante représentant la période de la fondamentale. L'estimation du pitch peut être erronée en raison de telles interférences qui sont dues essentiellement au premier formant. Par conséquent, il faut éliminer la contribution de la réponse du conduit vocal du signal de la parole avant de calculer l'autocorrélation. Le moyen le plus simple pour y parvenir est l'extraction du signal erreur (résidu) par filtrage inverse du signal vocal. Le signal erreur possède un spectre plus aplati que celui du signal vocal. Il a été constaté lors des différents essaies que le signal erreur obtenu par un filtre inverse pondéré $A(\gamma_p z)$ $(0.9<\gamma_p<1)$ donne de meilleurs résultats. Ceci est due à la bonne redistribution de la puissance du bruit. Ainsi, on réduit la densité spectrale du bruit dans les zones de fréquence où le niveau du signal est faible et on l'augmentant dans les zones des formants où le signal est fort et peut masquer efficacement le bruit. Le signal erreur ainsi obtenu est appelé signal erreur perceptuel. Un filtrage passe-bas est, ensuite, appliqué à ce signal. Ceci permet, en quelque sorte, d'augmenter la corrélation. En effet, dans le cas où le pitch ne coïnciderait pas avec une valeur entière de l'échantillon, il se peut que la corrélation soit plus forte à un multiple du pitch, grâce, en grande partie, à la contribution des hautes fréquences. Le filtrage passe-bas évite ce problème en éliminant les fréquences élevées.

La figure (3.2) illustre l'algorithme de détection du pitch. Ce dernier repose sur le test de l'énergie de la trame pour diriger la recherche du pitch. Si la trame est de faible énergie, ce qui correspond le plus probablement à une zone de silence, le pitch de la trame précédente est conservé. Sinon, on teste s'il y a une grande variation de l'énergie en comparant l'énergie de la trame avec celle de la trame précédente. Si tel est le cas, nous sommes en présence d'une transitoire. La détection du maximum est faite autour du pitch de la trame précédente. Pour des petites variations de l'énergie, une recherche complète du pitch est entreprise, entre les deux valeurs extrêmes qu'il peut prendre. Si la valeur du pic trouvé dépasse un certain seuil, une recherche autour des sous-multiples du pitch est entamée. Cette recherche touche tout les sous-multiples situés en dessus de la valeur minimum que peut prendre le pitch. Elle vise la résolution du problème de détection d'un pic correspondant à une harmonique et non pas au pitch lui-même. Dans le cas où la recherche complète fournie un pic inférieur au seuil, la recherche autour des sous multiples n'est pas faite et la recherche entamée est dirigée vers le pitch de la trame précédente.

Enfin le pic du pitch obtenu après différents tests est comparé à un seuil. S'il le dépasse, une interpolation est faite pour augmenter la précision du pitch, sinon, le pitch initial est conserver sans aucune interpolation.

L'exigence que le pic du pitch trouvé dépasse un certain seuil découle de ce qui suit. Le signal résiduel (signal d'excitation du filtre de synthèse) pour les sons non voisés

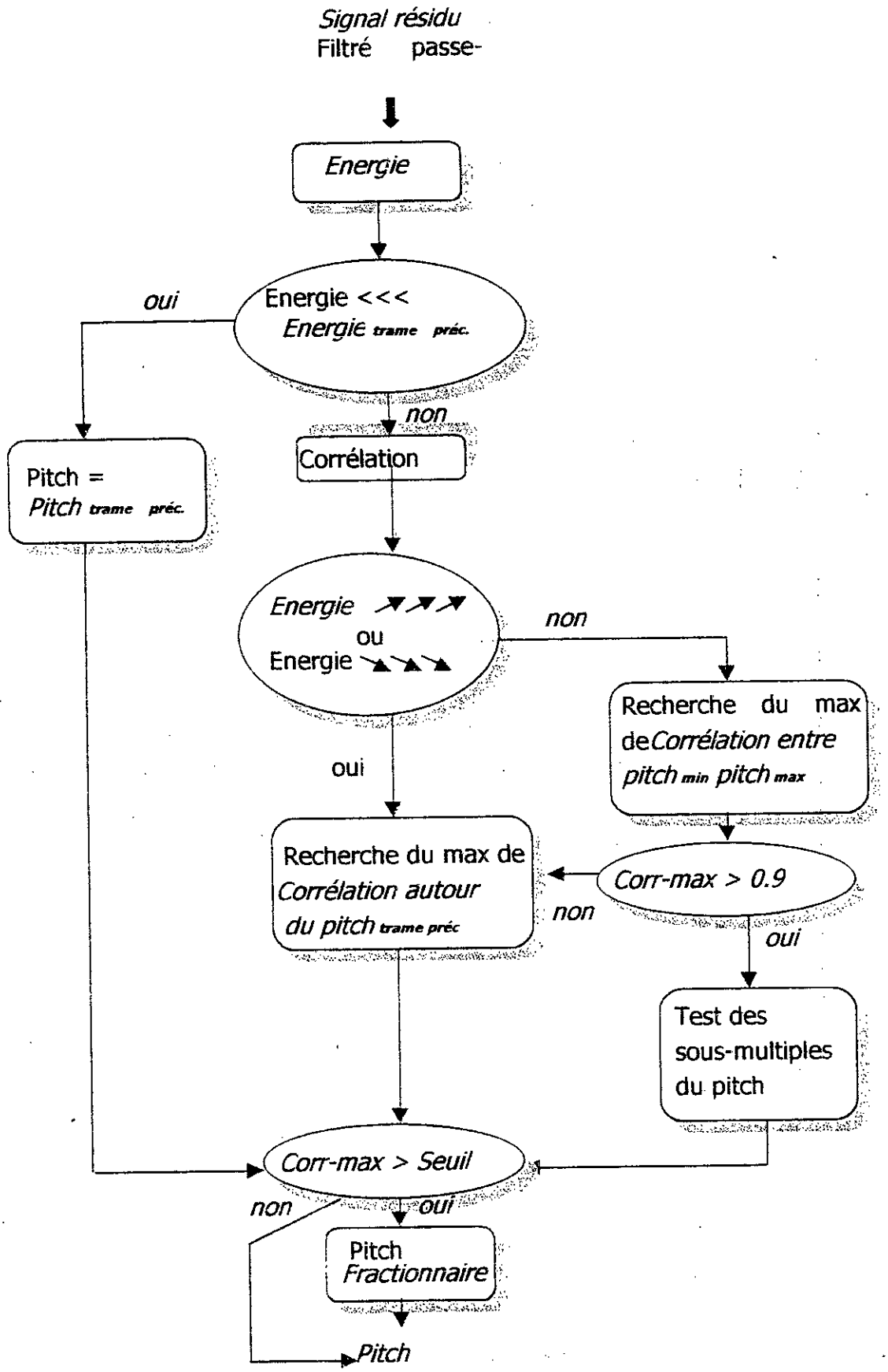


Figure 3.2 Algorithme de determination du pitch.

est assimilé à un bruit blanc gaussien de valeur moyenne nulle. Ainsi, quand la séquence du signal analysé devient grande, la fonction d'autocorrélation Rn tend vers zéro pour n différent de zéro. Par ailleurs, pour une longueur L finie de la séquence du signal analysé, on peut déterminer un seuil tel que, avec une certaine probabilité et pour un certain intervalle de confiance β, aucun échantillon de la fonction d'autocorrélation d'un bruit blanc (excepté celui de l'origine) ne dépasse la valeur de seuil.

Le terme $r=Rn/R_0$ a une densité de probabilité assez complexe, mais on montre dans la référence [16] qu'à l'aide d'un changement de variable :

$$z = \log [(1+r)/(1-r)]/n \tag{3.19}$$

r se transforme en un processus, z, de distribution pratiquement gaussiènne, de valeur moyenne nulle et son écart type vaut :

$$\sigma_z=(L - z)^{1/2} \tag{3.20}$$

Pour un intervalle de confiance égal à la probabilité que z ne dépasse pas une valeur fixe α_z, on a :

$$Pr [z \leq \alpha_z] = 1/2 + \varphi(\alpha_z /\sigma_z) = 0.01 \beta \tag{3.21}$$

où β est estimé en % et où :

$$\varphi(k) = (1/ 2\pi)^{1/2} \int_0^k e^{-u^2/2} du \tag{3.22}$$

Connaissant σ_z et β, on peut déterminer α_z à partir d'une table de la fonction φ(k). Puis, en substituant $z=\alpha_z$ dans la relation (3.19), nous pourrons obtenir α_r satisfaisant $Pr[r\leq\alpha_r] = 0.01 \beta$, à partir de :

$$\alpha_r = (e^{n\alpha z} - 1)/ (e^{n\alpha z} + 1) \tag{3.23}$$

La figure 3.3 donne la variation de α_r en fonction de L pour différents intervalles de confiance β.

Pour des pics dépassant le seuil, la trame est considérée comme voisé. Elle est, alors trop dépendante du pitch. C'est pourquoi l'algorithme effectue dans un tel cas une recherche minutieuse du pitch en testant les sous multiples ou en interpolant la valeur trouvée. Par contre, pour des pics inférieurs au seuil, la trame est considérée comme non voisé, donc faiblement dépendante du pitch, d'où une valeur grossière de ce dernier est suffisante.

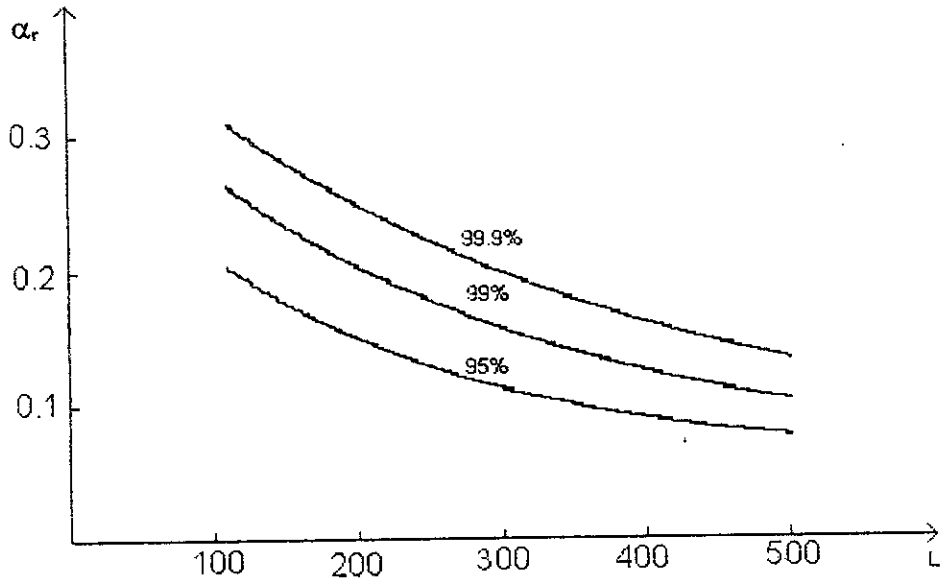


Figure 3.3 Variations de α_r en fonction de L pour différents intervalles de confiances[16].

III.4 Le niveau de voisement

La formulation du niveau de voisement repose sur la détection des pics de la fonction de l'autocorrélation normalisée du signal résiduel. Un signal résiduel filtré passe-bas, le même que celui utilisé pour la détection du pitch, est utilisé pour repérer un maximum de la fonction d'autocorrélation normalisée. Ce maximum est associé à la fréquence 500Hz qui n'est pas nécessairement sa position exacte dans le spectre, mais juste une approximation approchée. Le choix de la valeur 500 Hz peut être justifié par le fait que c'est autour de cette fréquence que ce situe, généralement, le premier formant. Un autre signal résiduel, non filtré passe-bas, est utilisé pour détecter un autre maximum de la fonction d'autocorrélation normalisés correspondant à des fréquences élevées. Ce pic est associé à la fréquence 2000 Hz autour de laquelle ce situe, généralement, le deuxième formant. La droite reliant ces deux pics est prise comme une fonction V(f), continue de la fréquence, représentant la proportion du voisé présent dans le signal résiduel. L'erreur due à l'inexactitude dans la localisation des pics n'aura que peut d'influence sur les résultats de la synthèse car le niveau de voisement n'est pas utilisé pour donner les valeurs exactes des deux composantes de l'excitation, mais seulement un niveau relatif de chacune d'elles. le choux des fréquences 0.5 khz et 2khz a été ensuite valider par les différents essaies réalisés. Nous avons utilisé plusieurs combinaisons d'emplacement fréquenciels de ces deux pics et nous avons constater que même si l'on s'écartait de ces positions (500 ± 50 hz et 2 ± 0.5 khz) la qualité de la parole synthétisé reste assez bonne.

La figure (3.4) illustre ce principe. A une fréquence f du spectre, le niveau relatif Nh(f) de la composante harmonique est donné par le rapport :

$$Nh(f) = V^2(f)/[V^2(f) + (1 - V(f))^2] \quad (3.24)$$

où V(f) est la valeur de la fonction de voisement à la fréquence f. le terme $V^2(f)$ représente le niveau d'amplitude de la composante harmonique à la fréquence f, et le terme $(1-V^2(f))$celui de la composante gaussienne en cette même fréquence. Cette expression du niveau d'amplitude de la gaussienne est due au fait que le maximum que peut atteindre un pic est 1. Pour les trames voisées, où les pics de l'autocorrélation atteignent des valeurs très proches de 1, le niveau de la gaussiènne est presque nul. Le niveau relatif de la composante gaussiènne Ng(f) en une fréquence f est donné par le rapport :

$$Ng(f) = (1 - V(f))^2/[V^2(f) + (1 - V(f))^2] \quad (3.25)$$

Ces niveaux relatifs, Nh(f) et Ng(f), sont utilisés pour pondérer les deux composantes, harmonique et gaussienne, à la fréquence f.

Cette méthode heuristique, inspirée d'une bonne connaissance des caractéristiques du signal vocal, a donné des résultats satisfaisant lors des différents essais (figures 2.4, 2.5, 2.6).

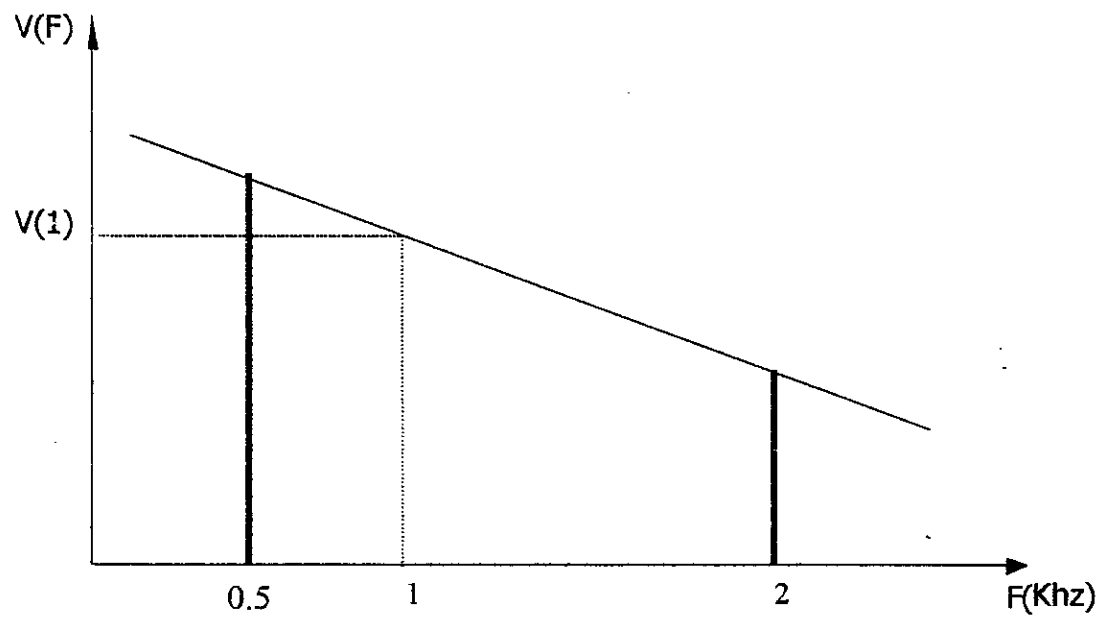


***Figure 3.4** Courbe de voisement caractérisée par les deux pics à 500hz et 2khz.*

III.5 Le spectre de l'excitation

Le codeur harmonique est basé sur une représentation efficace du spectre d'amplitude du signal d'excitation à l'entrée du filtre de synthèse. Aussi bien que ce spectre

est proche du spectre du signal résiduel, une bonne qualité de la parole synthétisée peut être assurée. Les phases du signal original ne sont pas déterminées lors de l'analyse [15]. Il suffit seulement d'assurer leur continuité lors de la synthèse. Ceci donne lieu à une réduction considérable du débit.

Le spectre d'amplitude du signal d'excitation est déterminé par le calcul de la DFT du signal résiduel à des fréquences multiples du pitch ou à l'un de ses diviseurs (figure 2.3). Cela est possible après une bonne détection du pitch, comme nous l'avons vu à la section III.3. La période du pitch conditionne totalement l'analyse du spectre. Ainsi pour des fréquences du pitch assez faibles, le spectre est décrit avec une bonne précision. Mais dés que la fréquence fondamentale devient grande, son utilisation directe pour l'analyse du spectre ne permet plus une représentation fidèle de ce dernier. Une fréquence d'analyse égale à un diviseur de la fréquence fondamentale est alors utilisée. L'algorithme d'analyse utilise une fréquence d'analyse donnée par la formule suivante :

$$f_a = \begin{cases} f_p & f_p < 125Hz \\ f_p/2 & 125Hz \leq f_p < 200Hz \\ f_p/3 & 200Hz \leq f_p < 330Hz \\ f_p/4 & 330Hz \leq f_p \end{cases} \qquad (3.26)$$

où f_p et t_p représentent la fréquence et la période du pitch respectivement. Le choix de ces fréquences peut être justifié par le désire de garder une fréquence d'analyse inférieur ou assez proche du pas fréquenciel fixe (100Hz) utilisé pour décrire le spectre à nombre de raies fixe adapté pour la quantification vectorielle.

La DFT pitch synchro donne, alors, un nombre de raies variable avec le pitch, et comme le modèle du pitch utilise une quantification vectorielle du spectre, il est donc impératif de ne conserver qu'un nombre fixe de raies. Une première méthode consiste à interpoler linéairement l'amplitude des raies variables du spectre pour aboutir à un nombre fixe de raies (tous les 100 Hz). Une seconde méthode, qui a conduit à de meilleurs résultats, est décrite ci-dessous [15] :

Dans le cas où le pitch engendrerait un nombre de raies inférieur au nombre fixe nécessaire (fondamentale>100 Hz), on choisit à chaque multiple de 100 Hz la raie la plus proche provenant du spectre variable. Le nombre de raies final étant plus grand, il est claire que la redondance permet une reconstitution transparente du spectre analysé.

Dans le cas où le nombre de raies serait supérieur, on moyenne les amplitudes du

groupe de raies le plus proche du multiple de 100 Hz, afin d'y placer une raie unique.

La figure (3.5) illustre le principe de cette transformation. Les raies sont groupées en 03 blocs de 12 raies pour permettre leur quantification. On aura ainsi 36 raies distincte l'une de l'autre d'un pas fixe égal à 100 Hz pour décrire une bande de 3600 Hz qui correspond à la bande couverte par l'analyse. Il vrai que la bande du signal original s'étant au delà de 3600Hz (signal échantillonné à 8KHz), mais en communication on se contente d'une bande assez réduite (3400 Hz en téléphonie par exemple) à cause de la concentration de la plus grande partie de l'énergie du signal de parole dans cette partie.

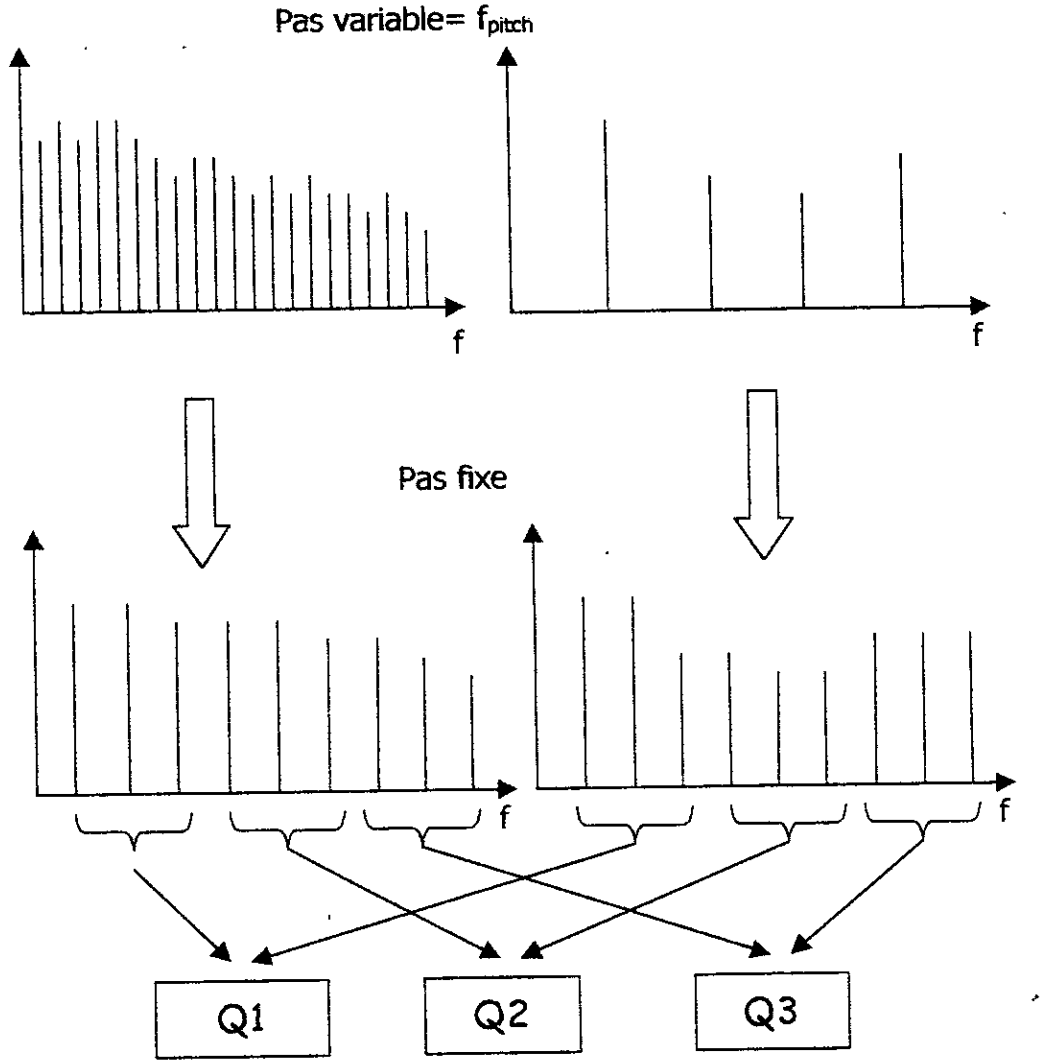


***Figure 3.5** Transformation du spectre à nombre de raies variable en spectre à nombre fixe.*

III.6 L'énergie du signal résiduel

Cette énergie est calculée pitch synchro, c'est à dire à des fréquences multiples du pitch ou à l'un de ses diviseurs. L'échantillonnage du spectre, avec la fréquence d'analyse f_a, engendre une périodicité dans le domaine temporel. La période temporelle est égale à

$t_a=1/f_a$. Au lieu de calculer l'énergie du signal, on calcule sa puissance qui représente, en quelque sorte, l'énergie par unité de temps. Ce calcul est fait dans le domaine temporel sur un intervalle égal à t_a. Cette puissance est donnée par :

$$\text{Ener} = \left(\sum_{i=0}^{N_a-1} x^2(i)\right)/N_a \qquad (3.27)$$

avec $N_a = t_a/t_{ech}$, $1/t_{ech}$ étant la fréquence d'échantillonnage du signal vocal.

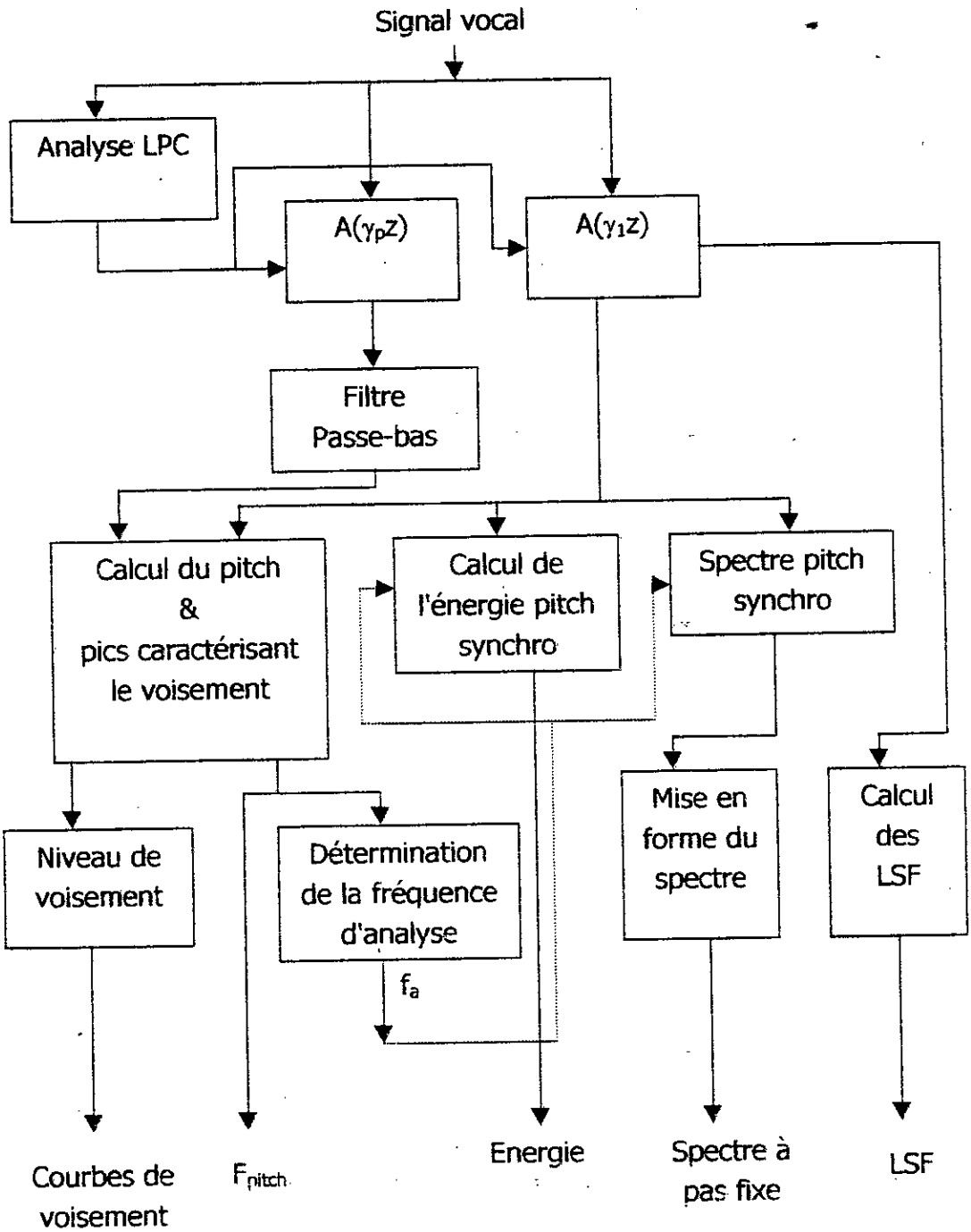


Figure 3.6 Schéma global de l'étage d'analyse

III.7 Présentation générale de l'étage d'analyse

La figure (3.6) illustre le schéma global de l'étage d'analyse. En premier, on procède à une analyse LP du signal vocal afin d'extraire les paramètres a_i représentant le filtre de synthèse. Ces paramètres sont, ensuite, pondérés pour obtenir deux filtres inverses pondérés. Le premier, très proche du filtre inverse original, est obtenu avec une pondération γ_1=0.99. Ce filtre est utilisé pour calculer le signal résiduel. Le deuxième filtre, obtenu avec une pondération γ_p=0.96, est utilisé pour obtenir le signal erreur perceptuelle qui va être utilisé pour le calcul du pitch. La pondération est faite dans le but de redistribuer de la puissance du bruit. Ainsi, on réduit la densité spectrale du bruit dans les zones de fréquence où le niveau du signal est faible et on l'augmentant dans les zones des formants où le signal est fort et peut masquer efficacement le bruit Le bloc de calcul du pitch donne aussi les deux pics de la fonction d'autocorrélation caractérisant le niveau de voisement. A partir du pitch, on détermine la fréquence d'analyse avec laquelle on calcule l'énergie et le spectre pitch synchro. Le spectre est ensuite mis en forme pour ne garder qu'un nombre fixe de raies séparées d'un pas fréquenciel fixe d'environ 100 Hz.

La sortie de cet étage donne les cinq paramètres caractéristiques du modèle du codeur harmonique. Ces paramètres sont quantifiés et transmis.

La reconstitution du signal originale à partir de ces paramètres de synthèse fera l'objet du prochain chapitre.

CHAPITRE IV
Reconstruction du signal vocal

IV.1 Introduction

Dans les deux chapitres précédents, les paramètres du codeur harmonique ont été décrits ainsi que les différents algorithmes qui permettent leur extraction. Dans ce chapitre, nous allons présenter les méthodes utilisées pour la reconstitution du signal vocal à partir de ces paramètres. Nous présenterons en premier l'algorithme de transformation inverse des paramètres LSF en coefficients de prédiction. Nous décrirons ensuite les générateurs des composantes harmoniques et stochastique de l'excitation. Enfin, la représentation générale de cet étage est abordée.

IV.2 Calcul des coefficients de prédiction

La conversion des paramètres LSF en coefficients de prédiction est moins complexe que le calcul des LSF à partir des coefficients a_i. Le polynôme A(z) est obtenu à partir des polynômes $F_1(z)$ et $F_2(z)$ (section 3.2.1), en utilisant la relation [1,20] :

$$A(z) = (F_1(z) + F_2(z))/2 \qquad (4.1).$$

La détermination des polynômes $F_1(z)$ et $F_2(z)$ à partir des paramètres LSF est réalisée en utilisant les équations (3.5) et (3.6). Cela revient à une série de multiplications de polynômes du type $(1 - 2 \cos w_i\ z^{-1} + z^{-2})$ pour obtenir les polynômes $G_1(z)$ et $G_2(z)$ (équation 3.8), puis les multiplier par (1+z) et (1-z), respectivement, pour obtenir $F_1(z)$ et $F_2(z)$.

Ayant à calculer le polynôme $G(Z) = \prod_{i=1}^{L}(1 - 2 \cos w_i\ z^{-1} + z^{-2})$, l'algorithme suivant peut être utilisé pour obtenir les coefficients g(n) de G(z) à partir des w_i [15] :

```
g(0) = 1
b = - 2 cos w1
g(1) = b
Pour i = 2,...,L
        b = -2 cos wi
        g(i) = b g(i-1) + 2 g(i-2)
        Pour j = i - 1,...,1
                g(j) = g(j) + b g(j-1) + g(i-2)
        fin
        g(1) = g(1) + b
fin .
```

Pour calculer le polynôme $G(z) = G(z)\ (1 - z)$, l'algorithme suivant est utilisé :

```
Pour i = 2,...,L
        g(i) = g(i) - g(i-1)
fin.
```

Pour la multiplication de G(z) par (1+z), le moins est remplacé par un plus dans l'algorithme.

IV.3 Générateur de la composante harmonique

Avant de calculer la composante harmonique, les raies du spectre analysé pitch synchro sont régénérées en utilisant le spectre à nombre fixe de raies reçu de l'étage d'analyse. Ces raies sont ensuite modulées par le niveau de voisement Nh(m), calculé à chaque harmonique m en utilisant l'équation 3.24. La composante harmonique est calculée dans le domaine temporel comme la somme des sorties de plusieurs oscillateurs sinusoïdaux. Chaque oscillateur est piloté par une harmonique de la fondamentale. A chaque instant t, entre deux instants d'analyse t_{n-1} et t_n (figure 4.1), la composante harmonique est donnée par :

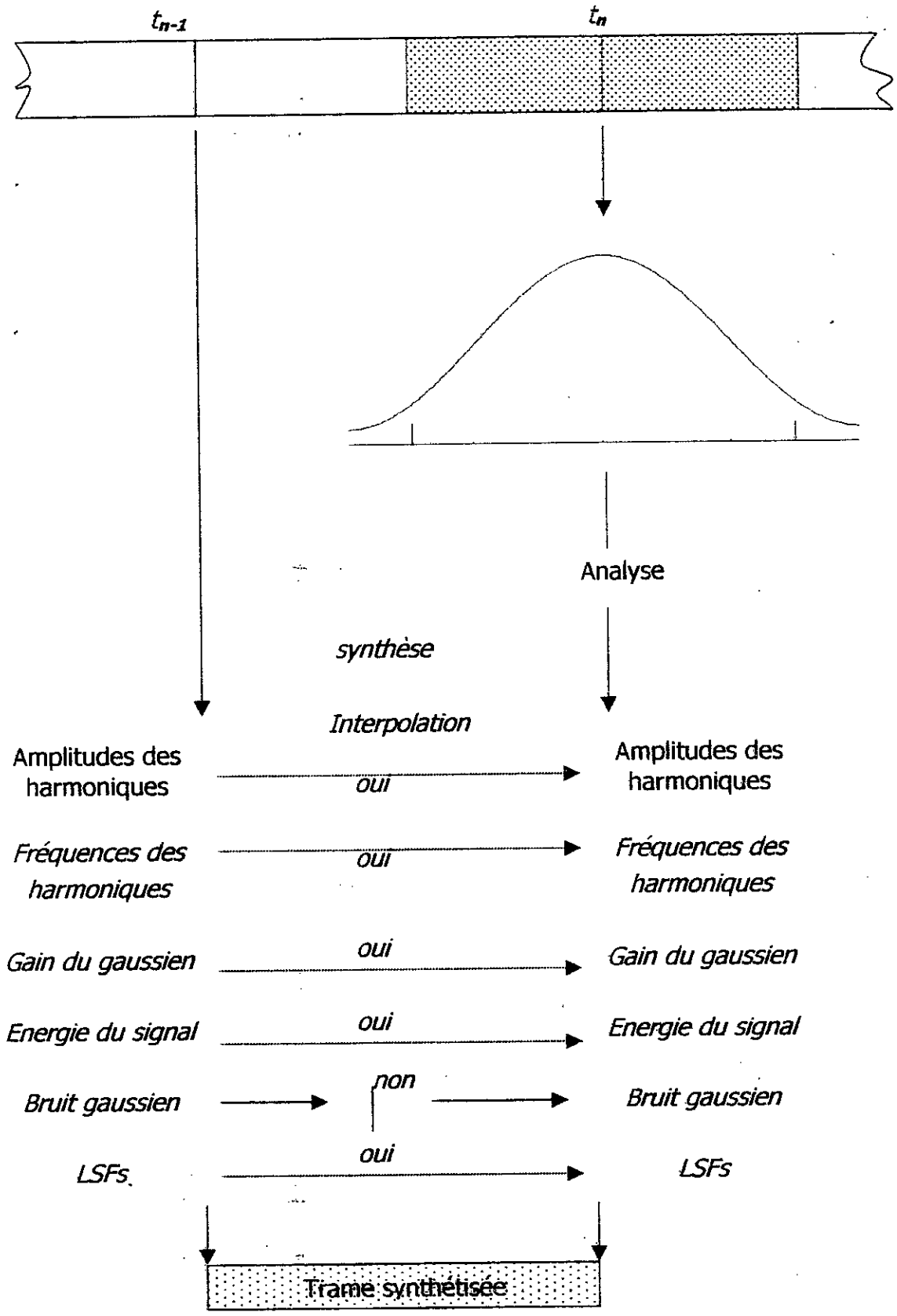


***Figure 4.1** Illustration de l'enchaînement des trames analysée et synthétisée.*

$$S_h(t) = G(t) \sum_{m}^{L} A_m(t) \cos(\theta_m(t)) \quad (4.2)$$

où G(t), $A_m(t)$ et $\theta_m(t)$ sont, respectivement, l'énergie, l'amplitude de l'harmonique m et sa phase à l'instant t.

Les amplitudes $A_m(t)$ sont calculées par interpolation linéaire des amplitudes de l'harmonique m entre les deux instants d'analyse t_{n-1} et t_n. La différence du pitch entre les deux instants engendre un nombre de raies différent aux instants t_{n-1} et t_n. L, le nombre de raies d'analyse, est choisi égal au maximum de raies entre les deux instants et les raies absentes d'un côté ou de l'autre sont ajoutées en considérant leurs amplitudes nulle. Ainsi pour le cas où le nombre de raies à l'instant t_{n-1} serait le pus grand, certaines harmoniques, au début de la trame de synthèse, verront leurs amplitudes diminuer jusqu'à atteindre zéro à la fin de la trame. Dans le cas contraire, certaines harmoniques démarreront avec des amplitudes nulles qui vont augmenter tout au long de la trame jusqu'à atteindre leurs valeurs finales à l'instant t_n. Nous aurons ainsi, la mort de certaines harmoniques, pour le premier cas, et la naissance d'autre pour le deuxième. Cette manière de faire le lien entre les spectres aux instants t_n et t_{n-1} est beaucoup plus simple que celle proposée par MacAuley et Quaterie dans [10,11].

L'amplitude à l'instant t est donnée par :

$$A_m(t) = A_m(t_{n-1}) + \{ [A_m(t_n) + A_m(t_{n-1})] [(t-t_{n-1})/(t_n-t_{n-1})] \} \quad (4.3)$$

où $A_m(t)$ est l'amplitude de l'harmonique m à l'instant t..

la phase $\theta_m(t)$ est déterminée par la phase initiale μ_m et la pulsation $w_m(t)$ de l'harmonique m à l'instant t. Elle est donnée par :

$$\theta_m(t) = \int_0^t w_m(x)\, dx + \mu_m \quad (4.4)$$

La fréquence $w_m(t)$ de l'harmonique m à l'instant t résulte de l'interpolation linéaire des fréquences de cette même harmonique entre les instants t_{n-1} et t_n. Elle est donnée par :

$$w_m(t) = m\, w_0(t_{n-1}) + m \{ [w_0(t_n) - w_0(t_{n-1})] [((t-t_{n-1})/(t_n-t_{n-1}))] \} \quad (4.5)$$

où $w_0(t)$ est la fréquence fondamentale à l'instant t. la phase initiale μ_m est choisie égale à la phase de l'harmonique m à l'instant t_{n-1}. Ceci permet d'assurer la cohérence des phases aux jonctions des trames.

L'énergie G(t) est obtenue par interpolation linéaire de l'énergie entre les deux instants t_{n-1} et t_n :

$$G(t) = G(t_{n-1}) + \{ [G(t_n) + G(t_{n-1})] [(t-t_{n-1})/(t_n-t_{n-1})] \} \quad (4.6)$$

La figure 4.1 illustre le générateur de l'excitation harmonique. Les différentes interpolations y sont clairement indiquées. L'interpolation des différents paramètres évite les transitions rapides entre trames adjacentes, en introduisant un effet de lissage de ces paramètres. On obtient ainsi de lentes variations assurant la continuité du signal synthétisé.

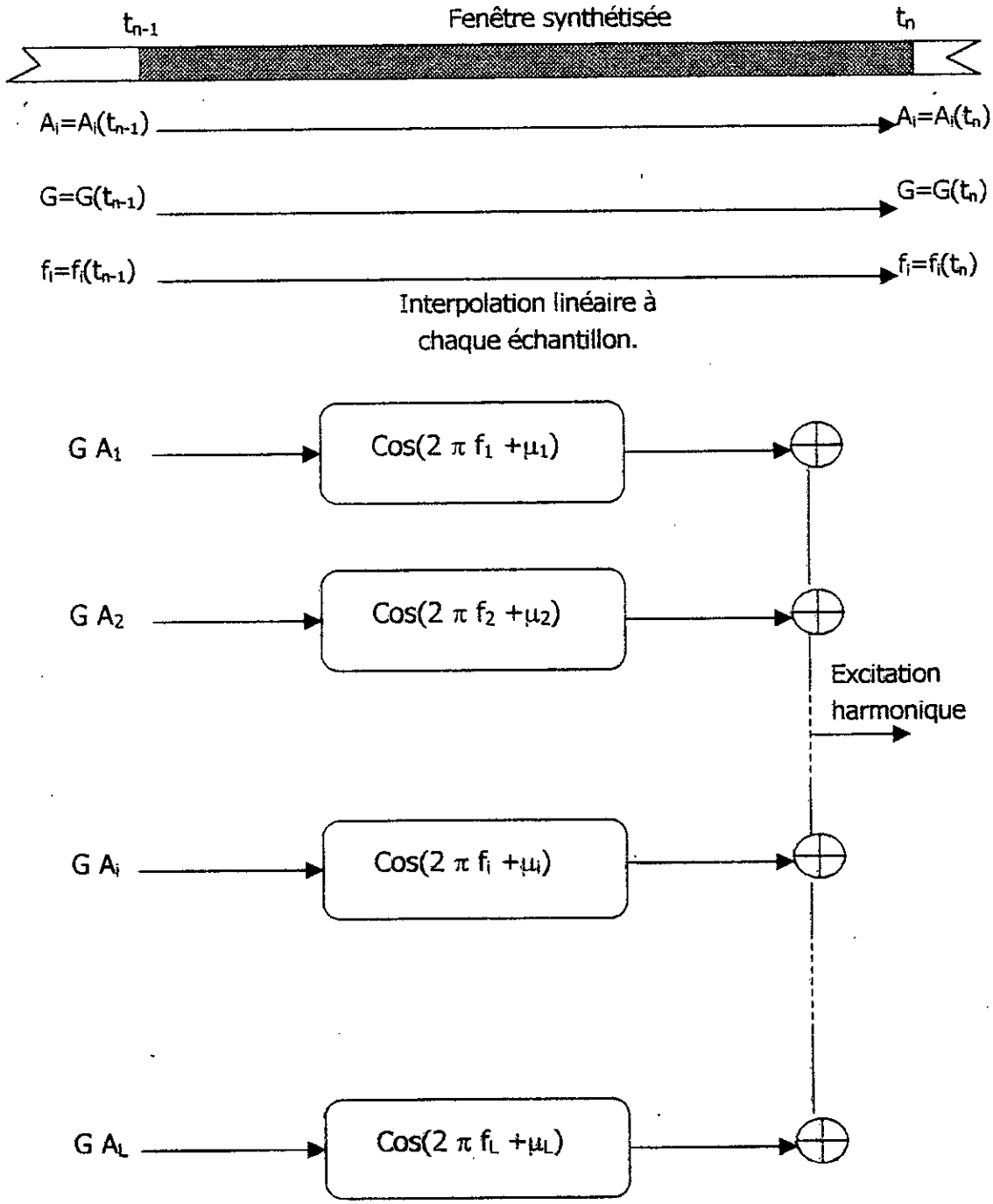


Figure 4.2 Générateur harmonique

IV.4 Générateur de la composante stochastique

Une première méthode, utilisée pour la génération de la composante stochastique, consiste à créer un bruit gaussien, effectuer une DFT, conditionner le spectre(en le modulant par le niveau relatif de la composante stochastique Ng(m) calculé à chaque raie par la relation 3.25), puis revenir dans le domaine temporel afin d'ajouter le résultat à la composante harmonique pour obtenir l'excitation totale. Néanmoins, cette méthode est quelque peu complexe et coûteuse en temps de calcul (DFT, DFT^{-1}) (figure 4.3).

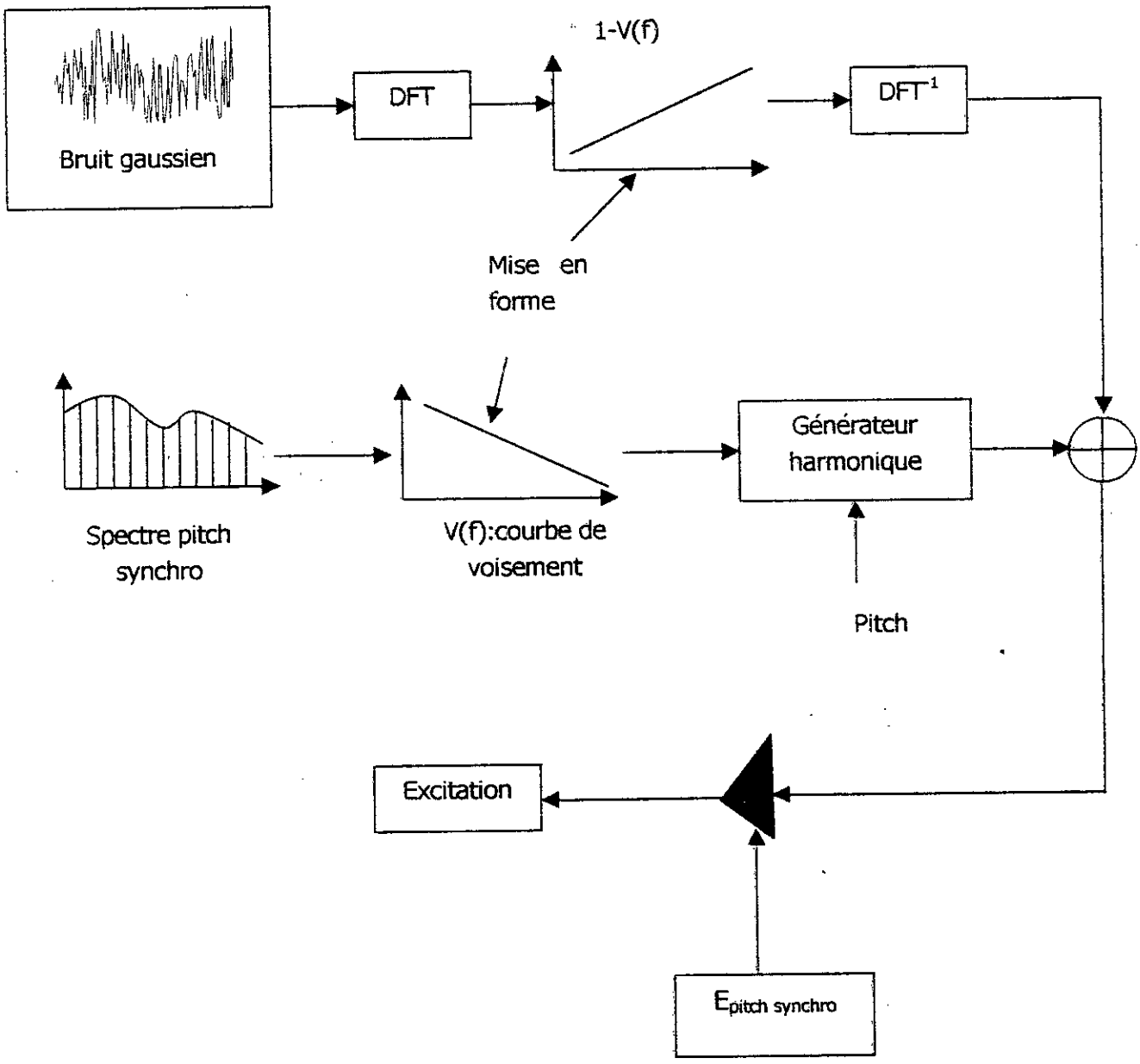


Figure 4.3 Première méthode de génération de la composante stochastique.

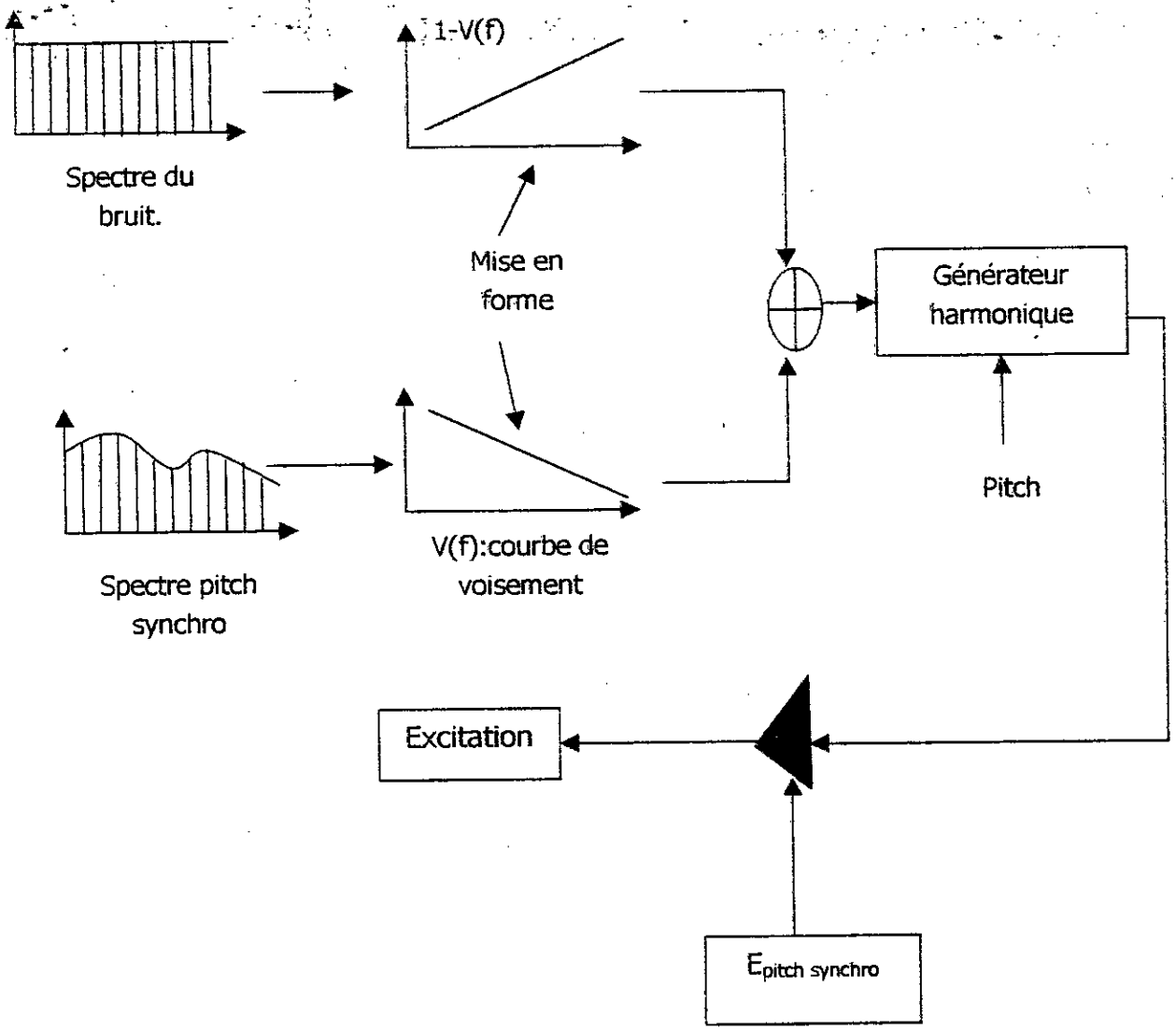


***Figure 4.4** Méthode directe de génération de l'excitation.*

Une deuxième méthode consiste à ajouter la contribution de la composante stochastique directement à l'amplitude de l'harmonique, et l'excitation totale est obtenue en une étape au lieu de deux(figure 4.4). L'amplitude de l'excitation, e(n), est donnée par [15]:

$$E(n) = \sum_{m}^{L} G(t)\,[A_m(t) + N_m(t)]\cos(\theta_m(t)) \qquad (4.7)$$

où G(t) est l'énergie interpolée à l'instant t (équation 4.6), $A_m(t)$ est l'amplitude de l'harmonique m calculée avec 4.3, et $N_m(t)$ est l'amplitude de la raie correspondante (à l'harmonique m) du spectre de bruit obtenu par interpolation linéaire entre les instants t_{n-1} et t_n. L'amplitude de la raie du spectre de bruit à l'instant t_n est obtenue par multiplication de l'amplitude de cette raie par le niveau relatif de bruit en cette fréquence calculé avec 3.25.

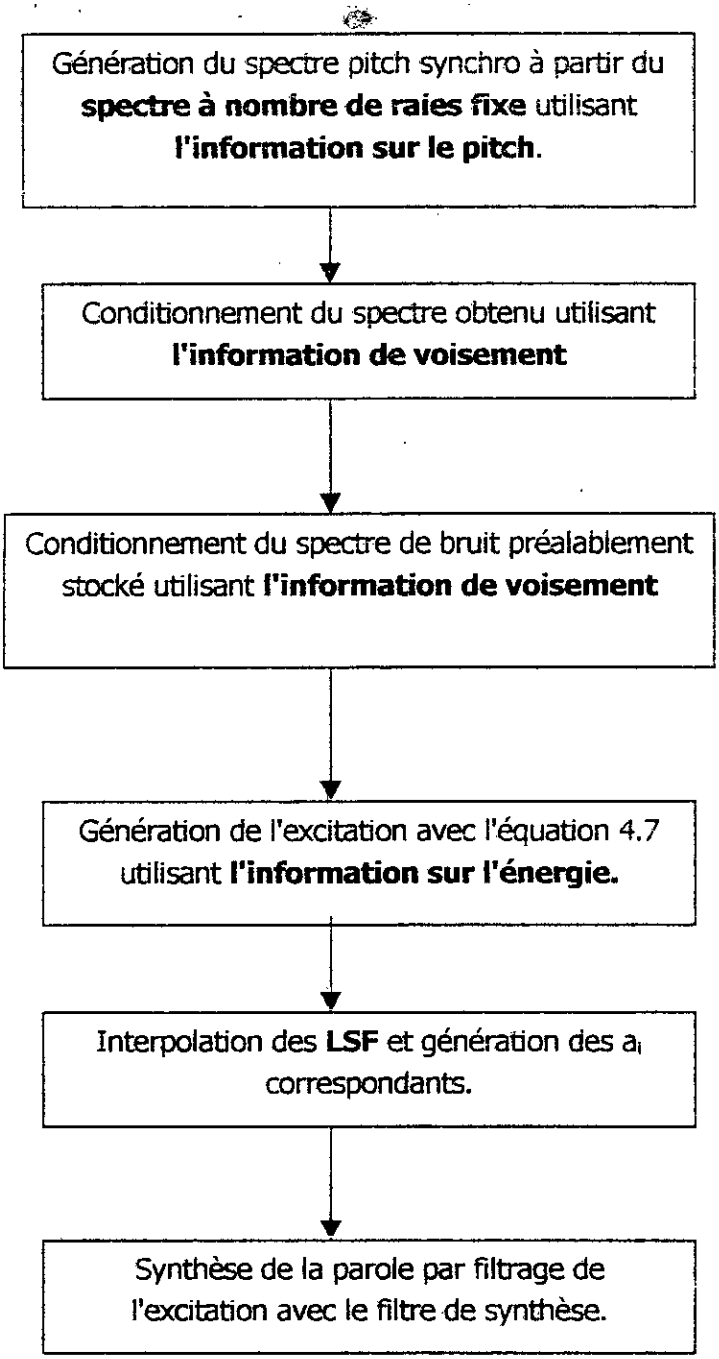


***Figure 4.5** Algorithme de reconstruction.*

Le spectre de bruit est en réalité préalablement stocké dans une table pour éviter de le régénérer à chaque fenêtre. Au sein de chaque trame, son spectre est conditionné par l'information de voisement.

IV.5 Interpolation des LSF

Les paramètres LSF ne sont pas utilisés directement lors de la synthèse pour la totalité de la trame, car cela peut entraîner des transitions désagréables au niveau des limites des fenêtres à cause d'une variation brusque de ces paramètres d'une fenêtre à l'autre. Pour éviter ce problème et avoir une lente variation entre trames, ces paramètres sont interpolés quatre fois au cours de la même fenêtre. Chaque fenêtre est divisée en quatre sous fenêtres. Pour chacune d'elle un ensemble des paramètres LSF est déterminé pour donner le filtre de synthèse de cette sous fenêtre. Ainsi, on démarre d'un ensemble plus proche des LSF à l'instant t_{n-1}, pour la première sous fenêtre, pour arriver à un ensemble plus proche des LSF à l'instant t_n, pour la quatrième sous fenêtre. Le filtre de synthèse caractérisant chaque sous fenêtre est obtenu par la conversion de l'ensemble des LSF de cette sous fenêtre nen paramètres a_i. La figure 4.1 illustre les différentes interpolations réalisées au sein du synthétiseur.

IV.6 Organisation générale de l'étage de reconstruction

La figure 4.5 illustre Les différentes opérations réalisées dans cet étage. On commence par la génération du spectre pitch synchro à partir du spectre à nombre de raies fixe. Le spectre obtenu est, ensuite, conditionner en utilisant l'information de voisement. Le spectre du bruit, préalablement stocké, fait, lui aussi, l'objet d'un conditionnement en utilisant l'information de voisement. Ceci étant fait, l'excitation peut être générée en utilisant l'équation 4.7. Pour chaque sous fenêtre (section IV.5), les LSF sont déterminés par interpolation linéaire de leurs valeures entre la fenêtre précédente et la fenêtre actuelle. Les a_i correspondants sont alors déterminés. La parole synthétisée, pour cette sous fenêtre, est obtenue en filtrant l'excitation par le filtre de synthèse de cette sous fenêtre. Cette opération est répétée pour les quatre sous fenêtres.

CHAPITRE V

Simulation et résultats

V.1 Introduction

Dans les chapitres précédents, le modèle du codeur harmonique a été introduit, les étages d'analyse et de reconstruction ont été décrits. Dans ce chapitre, nous allons utiliser ce modèle pour la simulation d'un vocodeur à faible débit (2400 bits/s). Mais avant d'arriver à ce point, nous allons discuter l'influence des différentes conditions d'analyse sur la qualité de la parole synthétisée. Ensuite, Nous présenterons la quantification des différents paramètres d'analyse. En fin, nous discuterons les résultats obtenus.

I.2 choix des conditions d'analyse

V.2.1 Acquisition du signal vocal

Le signal vocal utilisé pour la simulation est limité à une fréquence inférieur à 4KHz. La fréquence d'échantillonnage est de 8000 Hz. Chaque échantillon est codé sur 16 bits.

V.2.2 Ordre de prédiction

Le choix de l'ordre de prédiction P, résulte d'un compromis : il doit être suffisamment élevé pour reproduire correctement la structure formantique du signal vocal. Inversement, l'ordre doit être le plus faible possible pour réduire le débit. Pour une fréquence d'échantillonnage de 8 Khz, l'ordre de prédiction P doit être supérieur à 8 (section I.4.2). Le calcul des coefficients de prédiction s'effectuera en utilisant l'algorithme de **Winer-Livenson-Durbin** (section I.4.1). Ces coefficients seront ensuite convertis en paramètres LSF pour leurs meilleures propriétés de quantification (section III.2.1).

Pour montrer l'influence de l'ordre de prédiction sur la qualité de la parole synthétisée, nous avons simulé l'analyse/synthèse du signal vocal, selon le modèle du codeur harmonique, en variant l'ordre de prédiction avec tous les autres paramètres d'analyse fixes.

La figure 5.1 montre une tranche voisée du signal original comparée avec les tranches correspondantes du signal synthétisé, pour différentes valeurs de l'ordre de prédiction. Les spectres de toutes les tranches y sont montrés. Les figures 5.2 et 5.3 font la même comparaison pour une tranche non voisée et la totalité du signal, respectivement. Bien que l'enveloppe spectrale parait bien constitué pour les trois cas, on remarque que celle-ci est légèrement plus élevée vers les hautes fréquence pour l'ordre 8 et cela pour les trois types de signaux choisis. La différence entre les enveloppes d'ordres 10 et 12 est presque nulle. Cela est bien illustré par la figure 5.4 qui donne le gain de prédiction (rapport signal sur l'erreur de prédiction) en fonction de l'ordre [1]. A partir de l'ordre 10, la variation du gain de prédiction est presque nulle. L'ordre 10 est donc le meilleur compromis entre une reproduction correcte de la structure formantique du signal vocal et la minimisation du débit.

V.2.3 Longueur de la fenêtre d'analyse

La durée des tranches d'analyse dépend de la méthode de résolution choisie (du système donné pour l'équation 1.12) et des conditions dans lesquelles elle est appliquée. La méthode de l'autocorrélation que nous avons utilisé (l'algorithme de **Winer-Livenson-Durbin**), nécessite l'utilisation d'une fonction «fenêtre» et impose un temps d'analyse suffisant pour assurer une bonne résolution spectrale. Nous avons utilisé plusieurs longueurs de tranches d'analyse : de 10 ms et 30 ms (80 et 240 échantillons respectivement).

La figure 5.5 illustre une tranche voisé du signal original comparée avec les tranches synthétisées correspondantes, pour différentes longueurs de la fenêtre d'analyse. la figure 5.6 montre le résiduel d'une tranche voisé et l'excitation correspondante pour différentes longueurs de la fenêtre d'analyse. les figures 5.7 et 5.8 font de même pour une tranche non voisé.

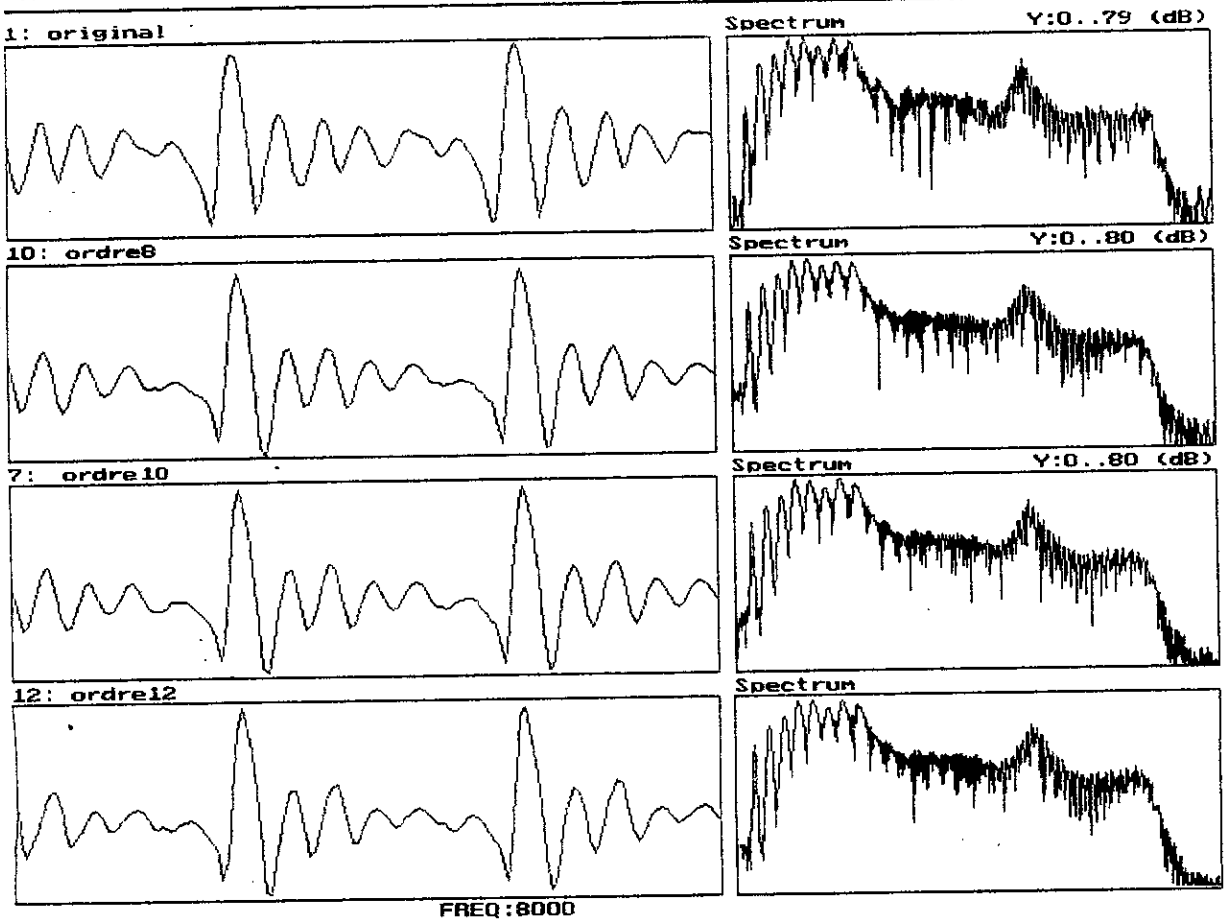


Figure 5.1 Une tranche voisé du signal original et la même tranche du signal synthétisée pour différents ordres du filtre LPC avec leurs spectres.

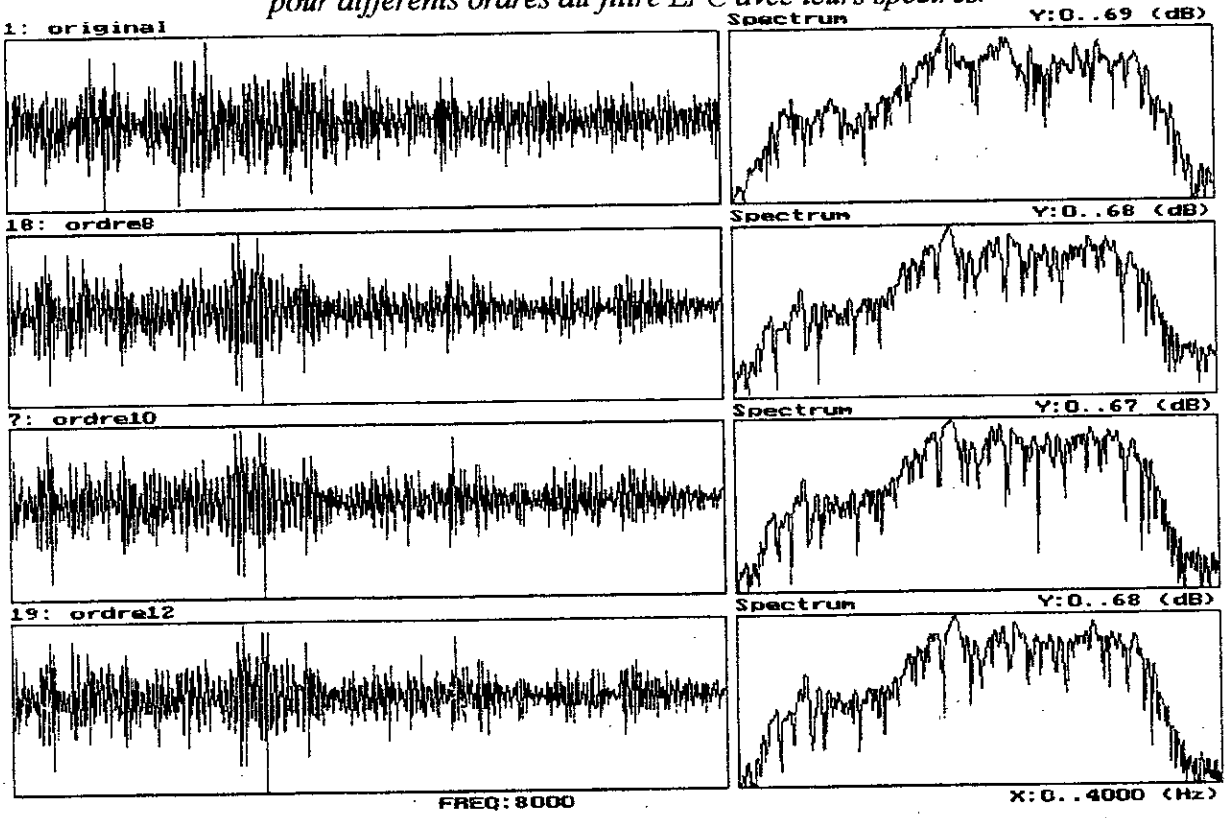


Figure 5.2 Une tranche non voisé du signal original et la même tranche du signal synthétisée pour différents ordres du filtre LPC avec leurs spectres.

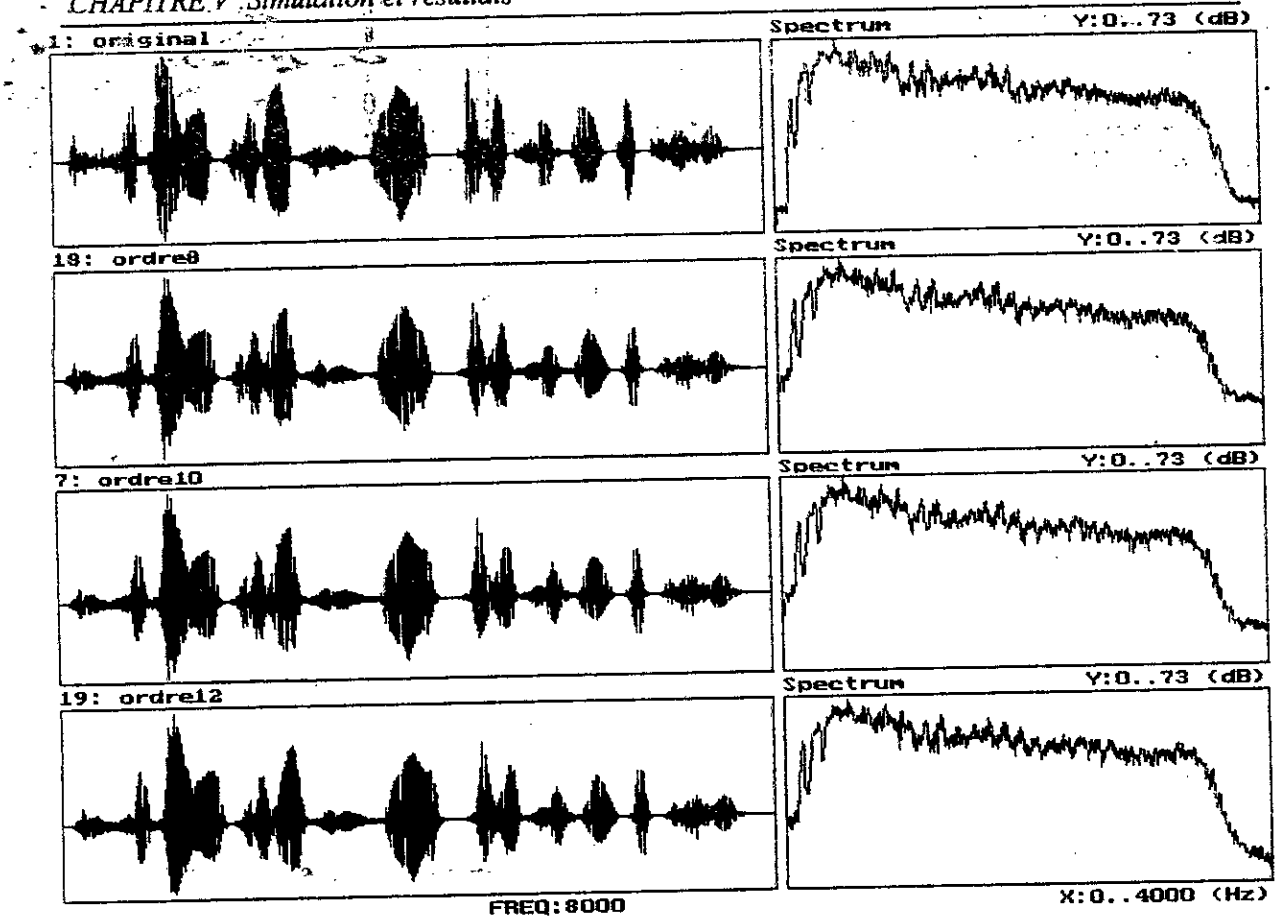


Figure 5.3 *Le signal original et le signal synthétisée pour différents ordres du filtre LPC avec leurs spectres.*

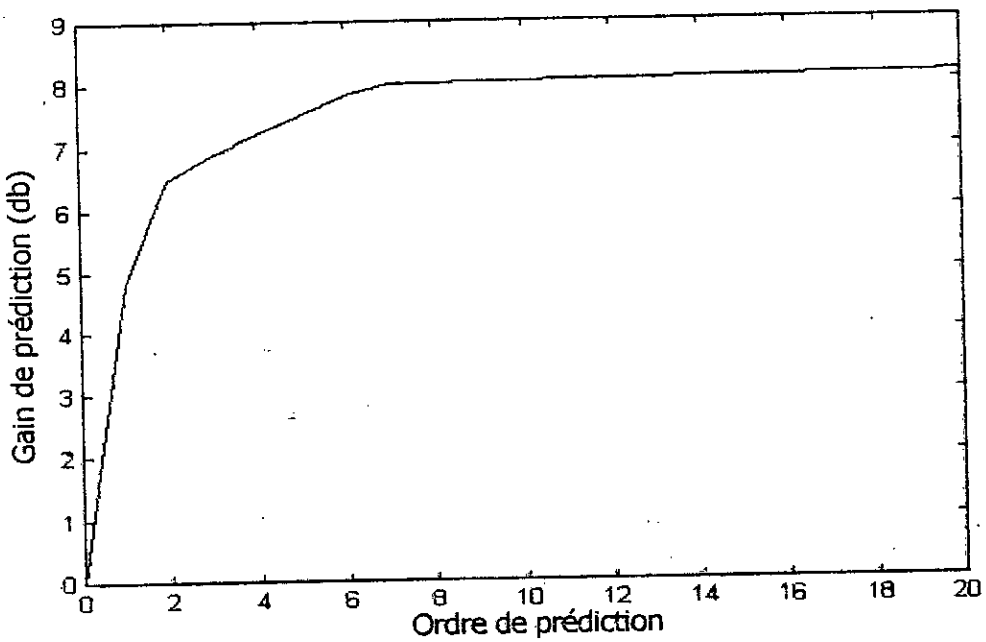


Figure 5.4 *Le gain de prédiction en fonction de l'ordre de prédiction .*

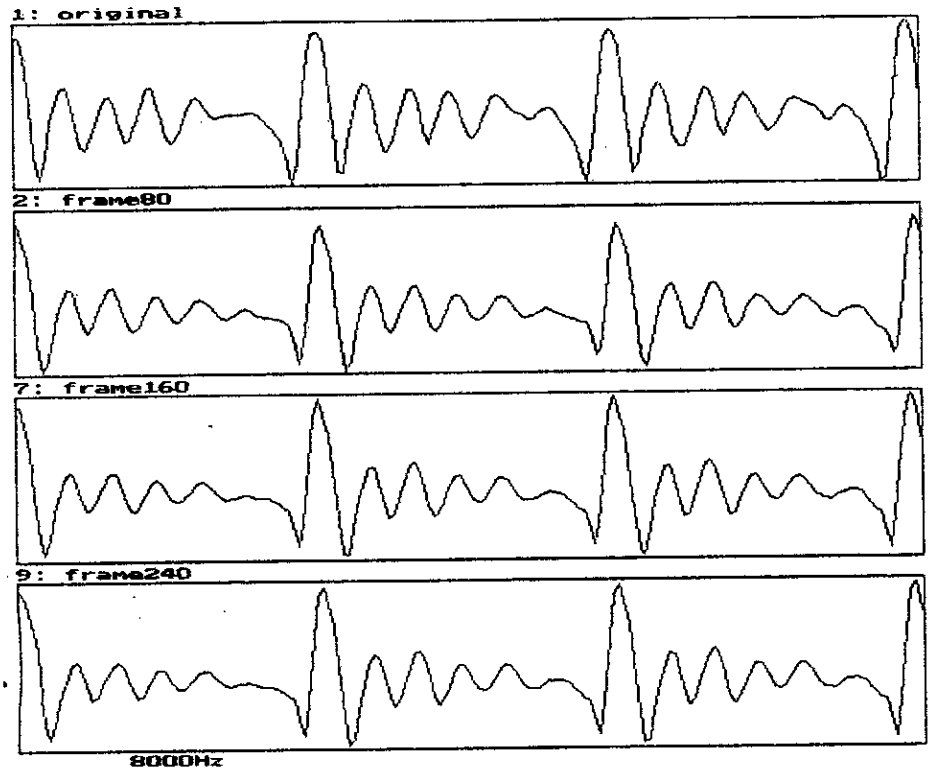


Figure 5.5 Une tranche voisé du signal original et la même tranche du signal synthétisé pour différentes longueurs des frames d'analyse.

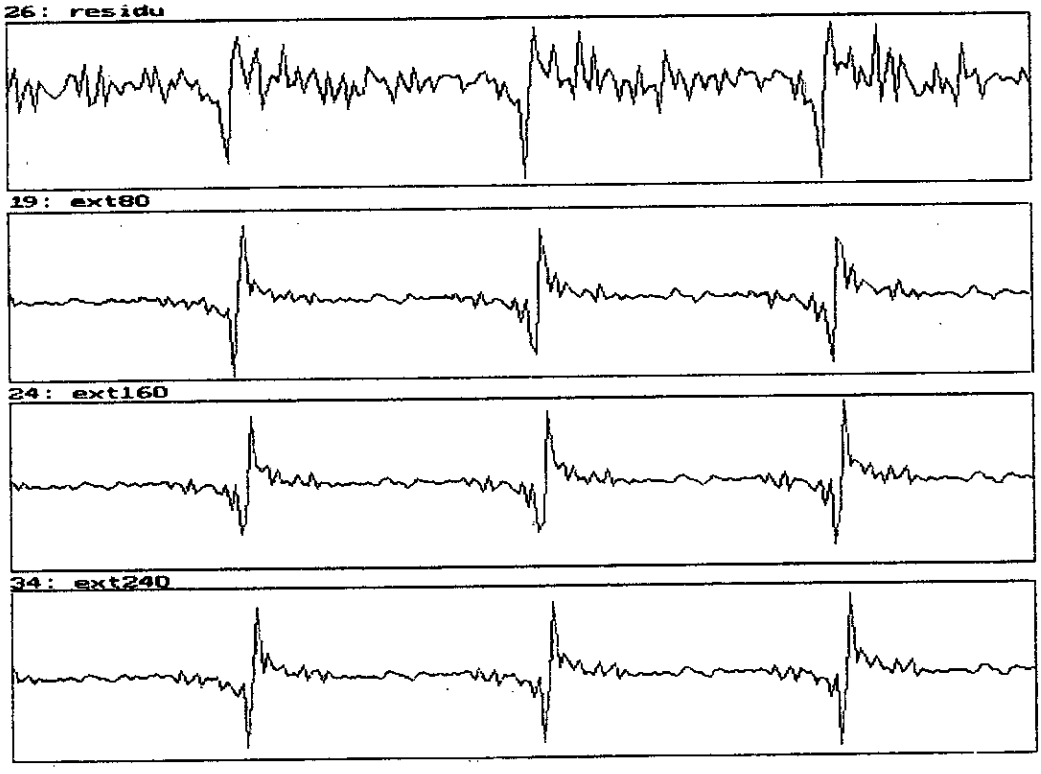


Figure 5.6 Le résiduel correspondant à une tranche voisé ainsi que les excitations des tranches synthétisées correspondantes pour différentes longueurs des frames.

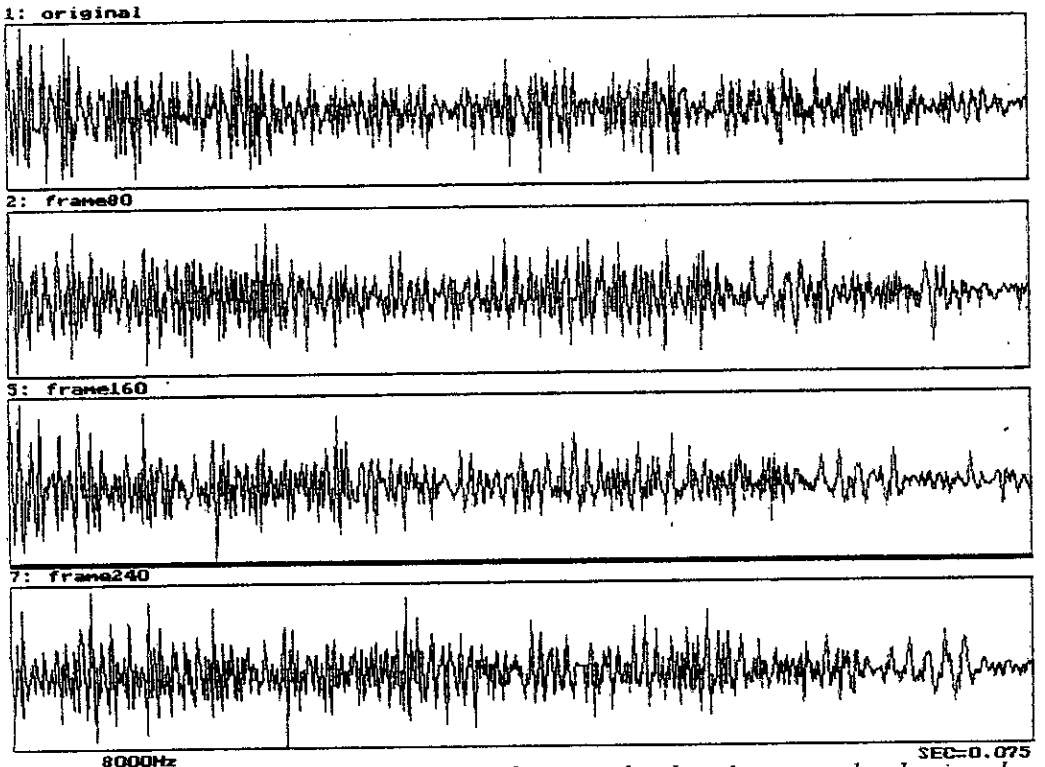


Figure 5.7 Une tranche NV du signal original et la même tranche du signal synthétisé pour différentes longueurs des trames d'analyse.

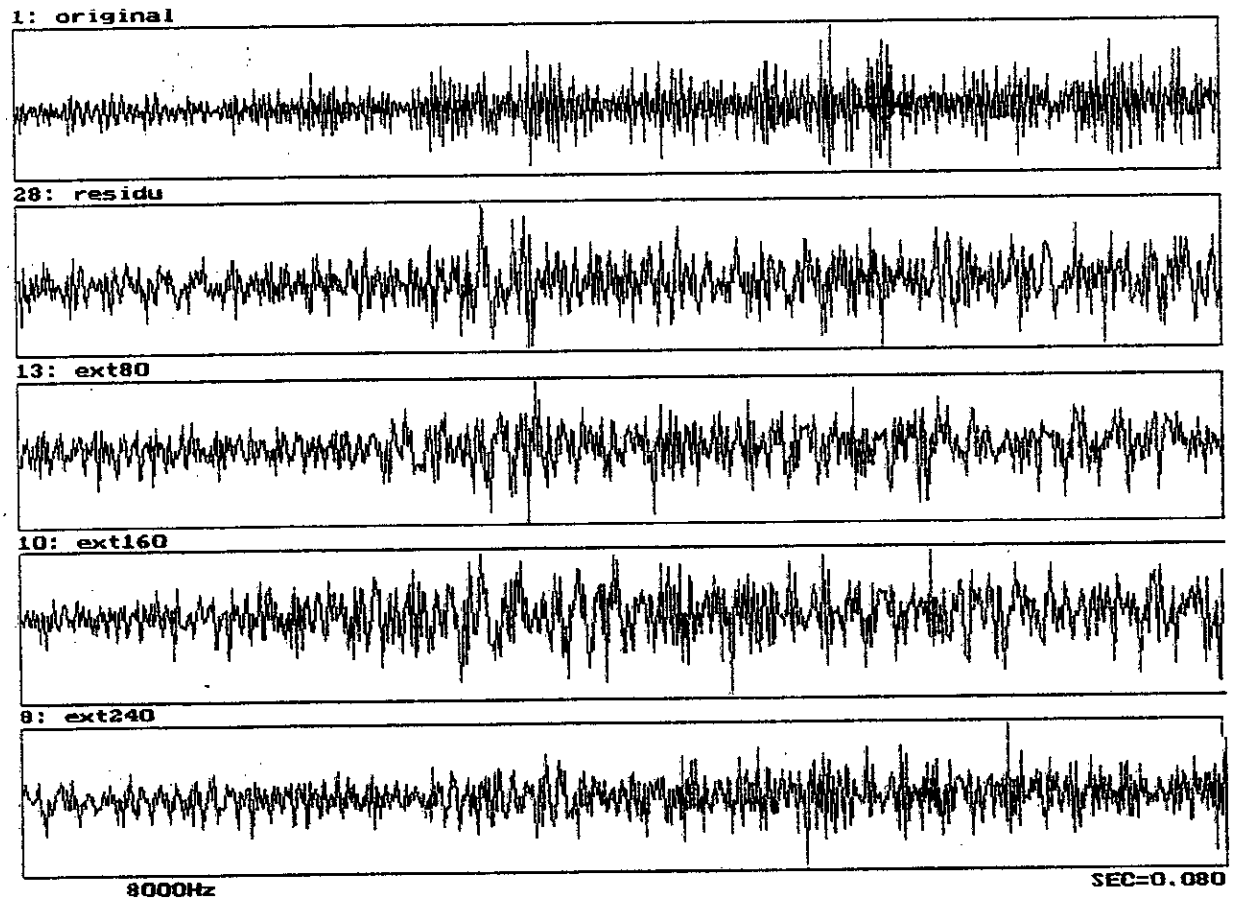


Figure 5.8 Une tranche NV et le résiduel correspondant ainsi que les excitations des tranches synthétisées correspondantes pour différentes longueurs des frames.

Lors de l'écoute, nous avons constaté que la parole synthétisée correspondante à une longueur de la fenêtre d'analyse variant entre 140 et 200 échantillons était de meilleur qualité que celle correspondante à une longueur de fenêtre superieur. On choisira pour notre analyse des tranches de longueurs allant de 20 à 25 ms selon le débit voulu.

Notons enfin que la fenêtre de hamming utilisée déborde de la tranche analysée de 7.5 ms de chaque côté. On aura donc un recouvrement de 60 échantillons (figure 5.1).

V.2.4 Nombre dè raies du spectre du résidu

Nous avons dit que le spectre du résidu est calculé pitch synchro, c'est à dire à des fréquences multiples du pitch ou à l'un de ses diviseurs (section III.5). nous avons également signalé que le spectre ainsi décrit ne peut être quantifié directement avec un quantificateur vectoriel. C'est pourquoi le spectre pitch synchro est transformé en un spectre à pas fréquenciel fixe en utilisant une interpolation ou une décimation. Un petit pas est théoriquement le mieux, mais cela risque d'augmenter le nombre de bits nécessaires à la représentation du spectre. Un compromis entre la minimisation du pas et l'augmentation du débit s'impose. En variant le pas fréquenciel de 25 Hz à 100 Hz (figures 5.9,10,11,12), nous avons constater que l'utilisation d'un pas égal à 100 Hz est le meilleur choix. Ce qui donne 36 raies pour la représentation du spectre.

En conclusion, les conditions d'analyse retenues sont un ordre de prédiction égale à 10, une longueur de fenêtre d'analyse de 20 à 25 ms et un pas fréquenciel de 100 Hz pour le spectre à nombre de raies fixe. La figure 5.13 montre une tranche voisé du signal original et la tranche synthétisée correspondante, ainsi que leurs spectres, pour les conditions d'analyse citées ci-dessus. La figure 5.14 fait de même pour une tranche non voisé. Les figures 5.15 et 5.16 donnent les spectrogrammes courts termes de deux fichiers original ainsi que les fichiers synthétisés correspondants.

V.3 Simulation du codeur harmonique à faible débit

Après avoir obtenu un signal synthétisé satisfaisant, sans introduire les différentes quantifications des paramètres d'analyse, nous allons, dans ce qui suit, aborder la quantification des paramètres d'analyse afin d'aboutir à un codeur à faible débit (2.4 kHz).

V.3.1 Quantification des LSF

Pour les paramètres LSF, le codeur harmonique utilise la quantification vectorielle par split. Les dix paramètres LSF sont divisés en trois vecteurs de dimensions 3, 3 et 4 respectivement. Chaque vecteur est quantifié séparément par un quantificateur vectoriel sur

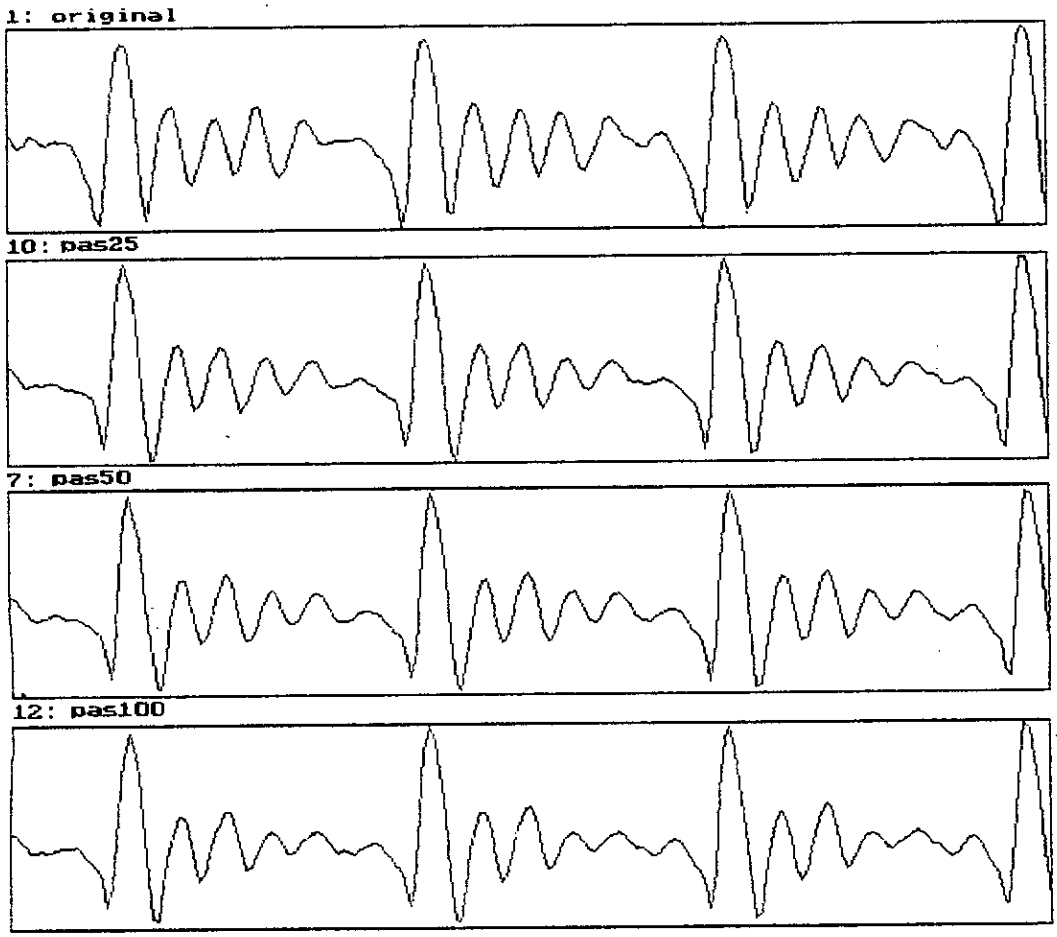


Figure 5.9 *Une tranche voisé du signal original et la même tranche du signal synthétisé pour différents pas du spectre d'excitation.*

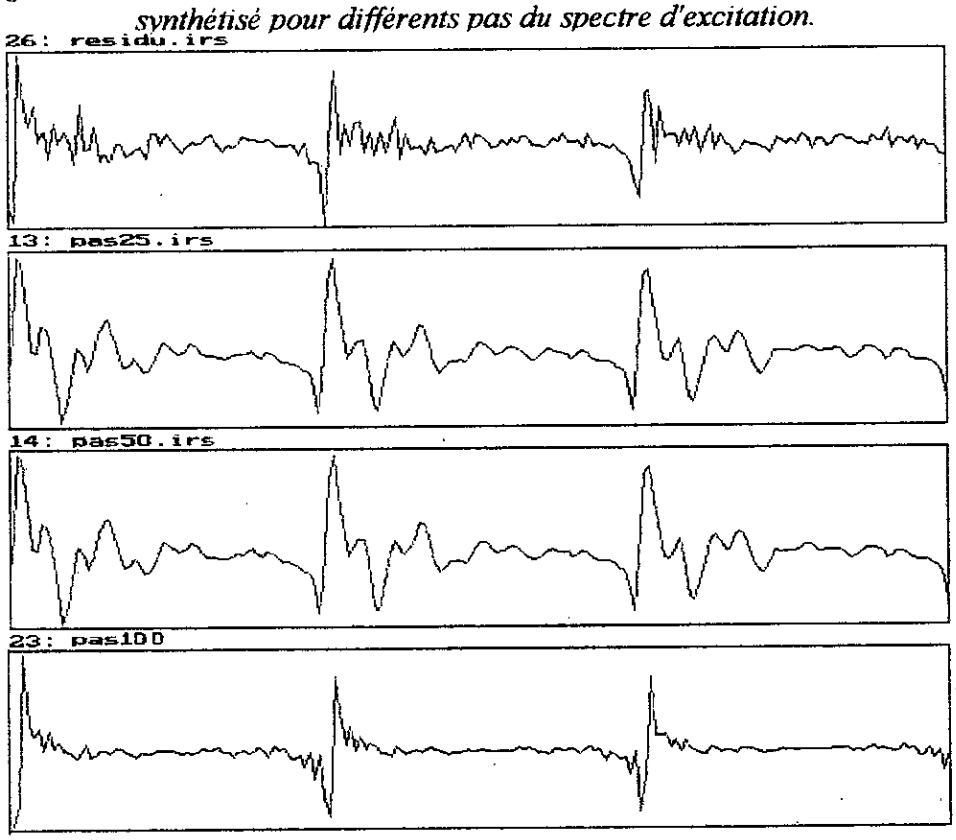


Figure 5.10 *Le résiduel correspondant à une tranche voisé et ainsi que les excitations des tranches synthétisées correspondantes pour différents pas du spectre*

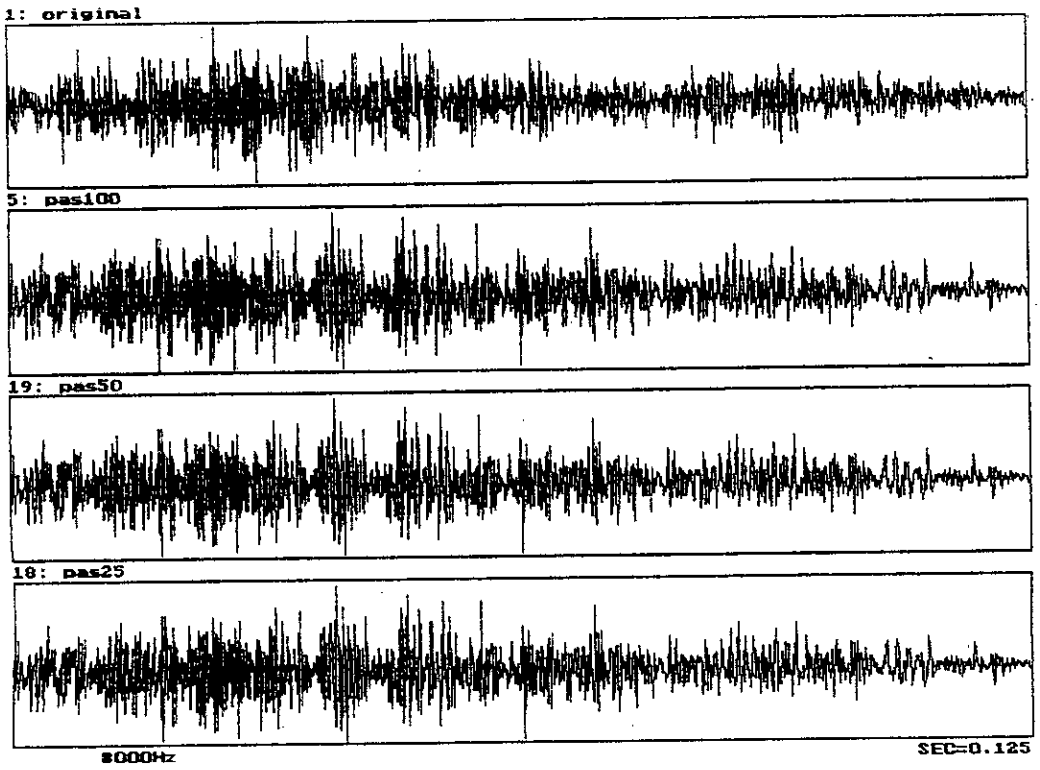


Figure 5.11 Une tranche NV du signal original et la même tranche du signal synthétisé pour différents pas du spectre d'excitation.

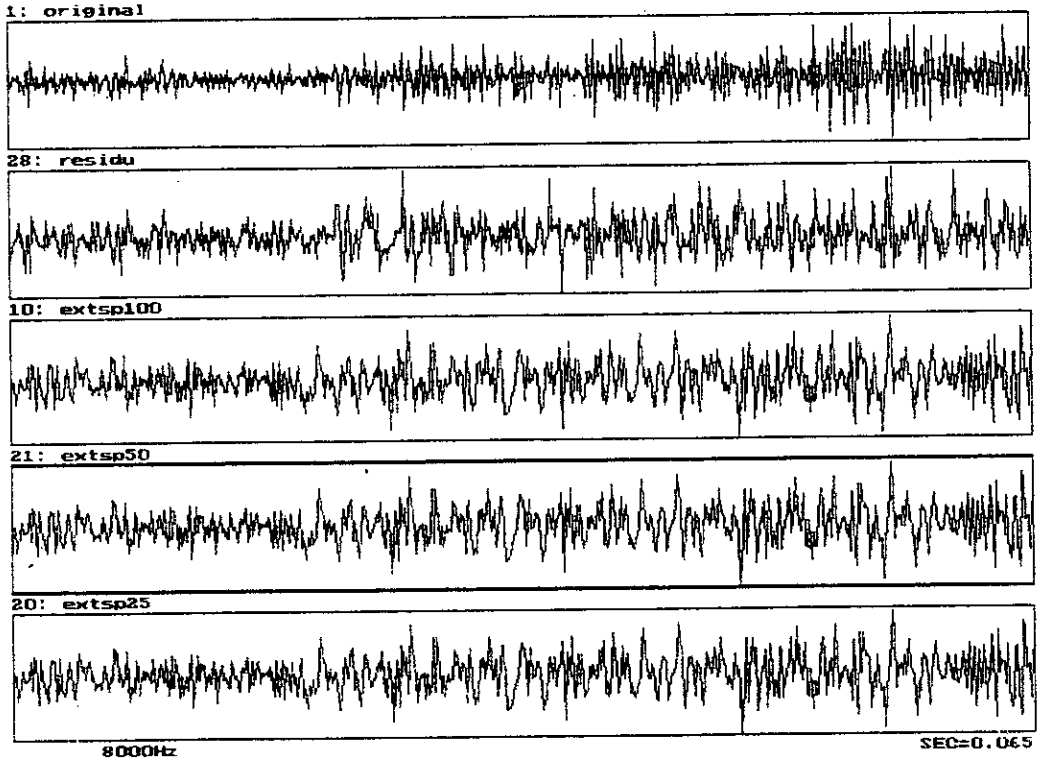


Figure 5.12 Une tranche NV et le résiduel correspondant ainsi que l'excitation des tranches correspondantes synthétisées pour différents pas du spectre d'excitation.

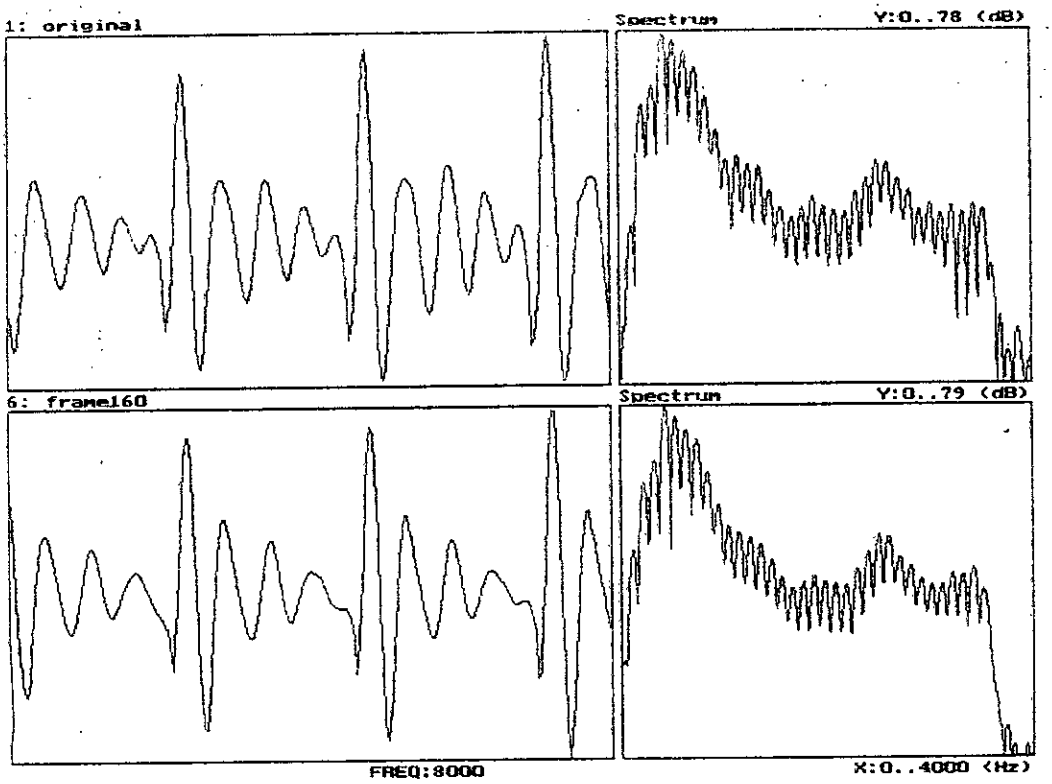


Figure 5.13 *Une tranche voisé du signal original et la même tranche du signal synthétisé (frame=160), ainsi que leurs spectres.*

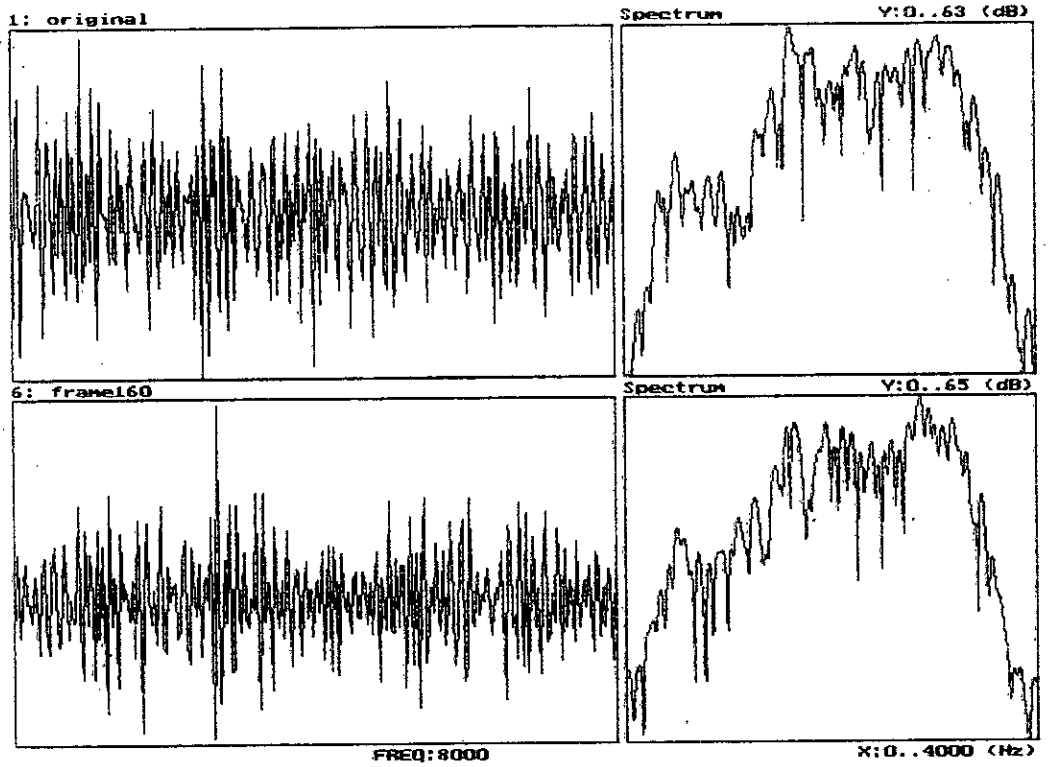


Figure 5.14 *Une tranche non voisé du signal original et la même tranche du signal synthétisé (frame=160), ainsi que leurs spectres.*

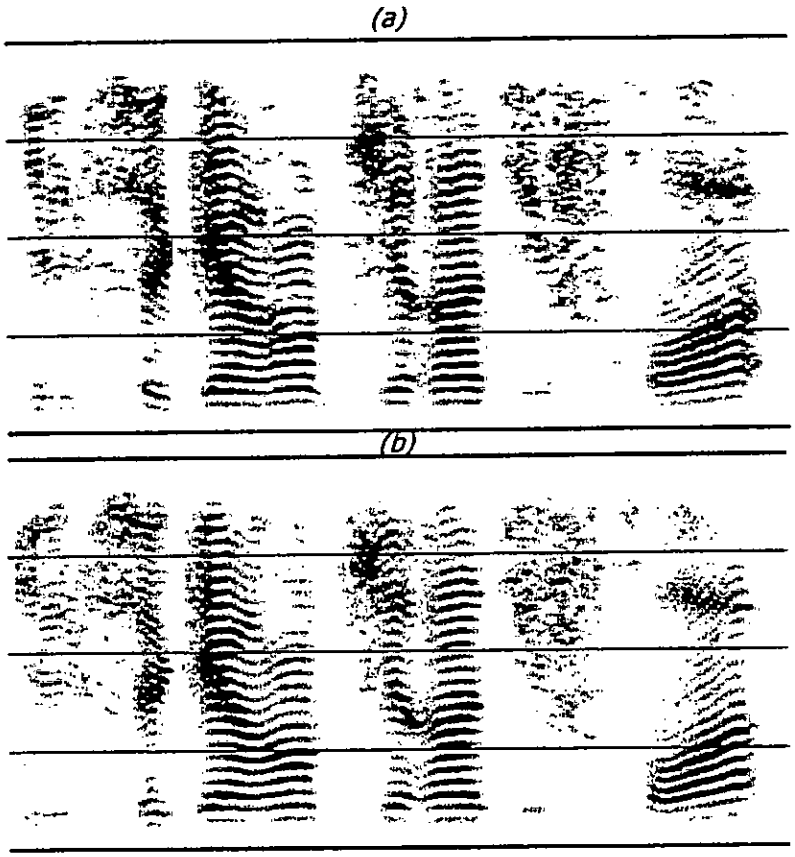


Figure 5.15 Spectrogramme du signal original (a) et du signal synthétisé(b).

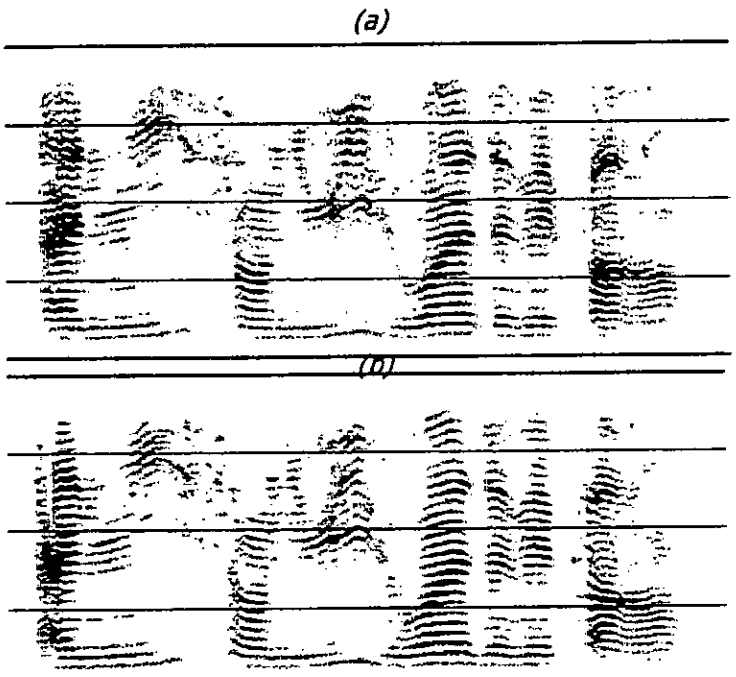


Figure 5.16 Spectrogramme du signal original (a) et du signal synthétisé(b).

8 bits [20,21,22,23,24,25] (figure 5.17). Elle évite le problème d'avoir un dictionnaire de taille trop grande. En effet, la quantification des LSF en un seul bloc de dimension 10 avec les 24 bits réservés à la quantification de ces paramètres dans le modèle du codeur harmonique, engendre un dictionnaire de taille 2^{24}. Ceci entraîne un temps trop long de recherche du meilleur représentant et un espace de stockage important. Le dictionnaire utilisé est du type statistique. Il a été obtenu par l'utilisation de l'algorithme LBG [26,27,28] (voir Annexe).

- La distance utilisée lors de la quantification vectorielle est donnée par :

$$d(LSF,LSFQ) = \sum_{i=1}^{M} [w_i(LSF_i - LSFQ_i)]^2 \qquad (5.1)$$

où w_i est le poid assigné au $i^{ième}$ LSF, M la dimension des vecteurs LSF et LSFQ (LSF quantifiées).

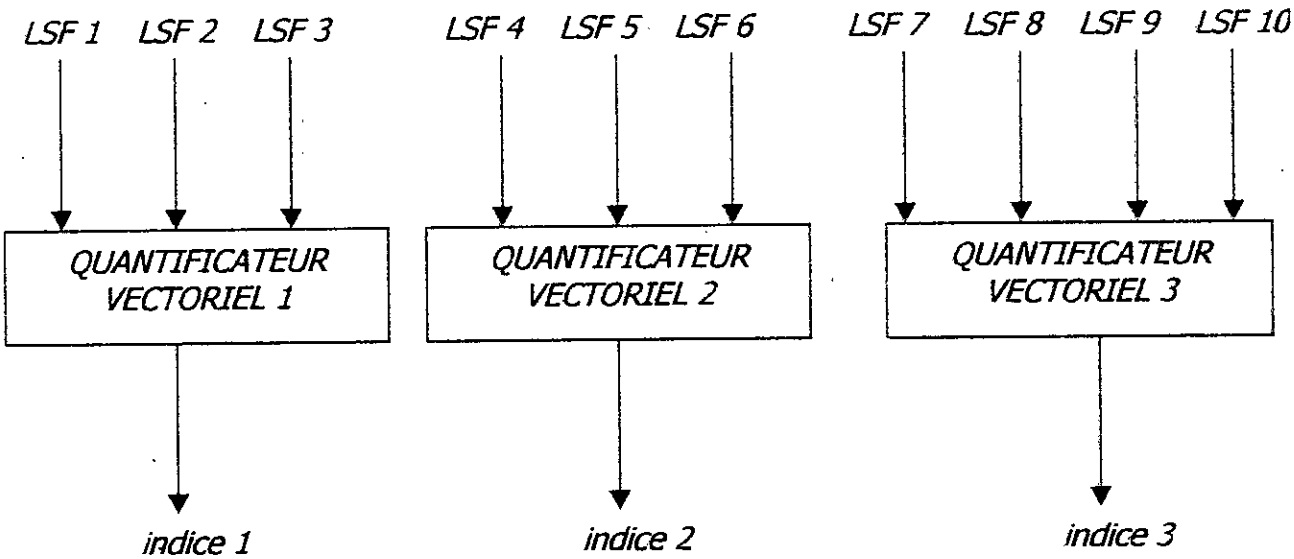


Figure 5.17 Quantification vectorielle des LSF

Selon le choix de w_i, nous aurons trois distances différentes [21] :

- Distance euclidienne : $w_i = 1$. (5.2)
- Distance de Farvardin : $w_i = 1/[LSF(i+1) - LSF(i)] + 1/[LSF(i) - LSF(i-1)]$. (5.3)
- Distance de Atal : $w_i = [DSP(\text{fréquence du LSF})]^{0.15}$. (5.4)

La distance euclidienne est la plus évidante. Elle est facile à mettre en œuvre. Son inconvénient est qu'elle attribue la même importance à tous les LSF quelles que soient

leurs dispositions. Or, en réalité ce n'est pas toujours vrai. Nous avons vu à la section III.2.1 que la concentration de certains LSF dans une bande de fréquence correspond approximativement à l'existence d'un formant dans cette région. Il est alors légitime d'accorder une plus grande importance aux LSF de cette bande. C'est ce qui a poussé les chercheurs dans ce domaine à introduire de nouvelles distances pour obtenir une meilleure pondération des LSF.

La distance de Farvardin lie chaque LSF à son voisinage. Le poids d'un LSF est d'autant plus grand que ce dernier est proche de ces voisins.

La distance de Atal utilise la densité spectrale du filtre LPC, évaluée à la fréquence correspondante à chaque LSF. Cette distance est trop complexe à cause de la nécessité d'une DFT après l'analyse LPC.

L'algorithme de quantification du codeur harmonique utilise la distance de Farvardin à cause de sa simplicité et sa bonne pondération des LSF selon leur disposition.

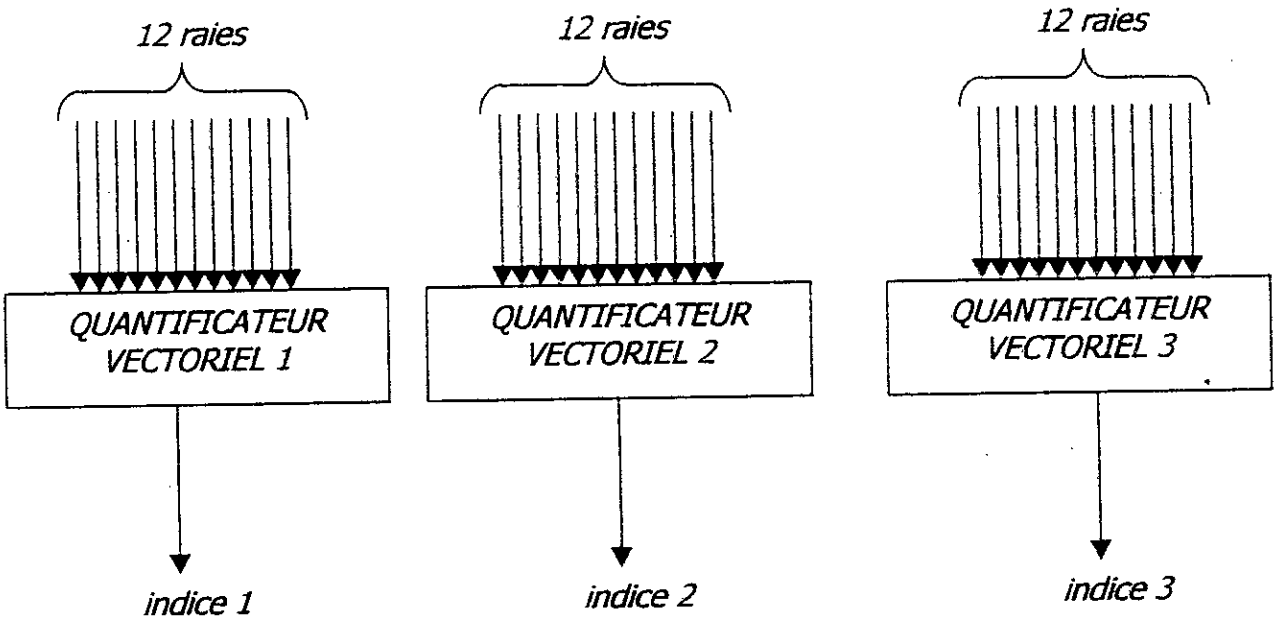


***Figure 5.18** Quantification vectorielle du spectre*

V.3.2 Quantification du spectre

Le spectre du signal vocal est analysé jusqu'à 3600 Hz. Le pas fixe utilisé, après obtention du spectre variable analysé pitch synchro, est égal à 100 Hz. Ceci donne 36 raies à quantifier. La quantification vectorielle par split est encore utilisée pour le spectre. Les 36 raies sont divisées en trois vecteurs de 12 raies chacun. Chaque vecteur est quantifié vectoriellement avec 7 bits.

L'algorithme de quantification choisi dans le dictionnaire le meilleur représentant au sens d'Euclide du vecteur résultant de l'analyse (qui minimise la distance euclidienne).

L'algorithme LBG est utilisé pour l'élaboration des dictionnaires à partir de trois bases d'apprentissage obtenues après plusieurs exécution de l'algorithme d'analyse.

La figure (5.18) illustre l'opération de quantification du spectre.

V.3.3 Quantification de la fonction de voisement

Les deux pics caractérisant la fonction de voisement ne sont pas quantifiés directement. La fonction de voisement est représentée comme une droite caractérisée par l'équation :

$$V(f) = \text{pente} \cdot f + c^{ste} \tag{5.5}$$

Où pente et c^{ste} sont les nouveaux paramètres caractérisants la fonction de voisement Ils sont obtenus à partir des deux pics P_1 et P_2 à 500 Hz et 2 KHz, respectivement; comme suit :

$$\text{pente} = (P_2 - P_1)/(2000-500) \tag{5.6}$$
$$c^{ste} = P_1 - \text{pente} \cdot 500 \tag{5.7}$$

Ces deux paramètres sont quantifiés vectoriellement en utilisant un dictionnaire statistique de taille égale à 8 vecteurs, obtenu par l'algorithme LBG. 03 bits sont donc suffisant pour caractériser la fonction de voisement.

V.3.4 Quantification de l'énergie

Pour l'énergie, le codeur harmonique utilise une quantification scalaire. Après avoir obtenu un fichier des énergies à partir de l'algorithme d'analyse, l'histogramme de ces valeurs est dressé (figure 5.19). La distribution des énergies est à peu prés gaussienne centré prés du zéro. Cette information nous a permis de distribuer les 05 bits réservés à la quantification de l'énergie de manière à ce que les zones avec les plus grandes probabilités d'apparition aient le plus grand nombre de bits. Nous avons ainsi obtenu une quantification scalaire non-uniforme de l'énergie (figure 5.20). La distance utilisée est la distance euclidienne.

V.3.5 Quantification du pitch

La quantification du pitch, de même que celle de l'énergie, est scalaire. Comme nous avons fait pour l'énergie, nous avons pu détecter la plage de variation du pitch en stockant les différentes valeurs obtenues lors de l'exécution de l'algorithme d'analyse dans un fichier. Cela nous a permis de constater qu'un dictionnaire uniforme était suffisant pour couvrir toute la plage avec une bonne résolution.

Le pitch est donc quantifié avec un quantificateur scalaire uniforme sur 7 bits.

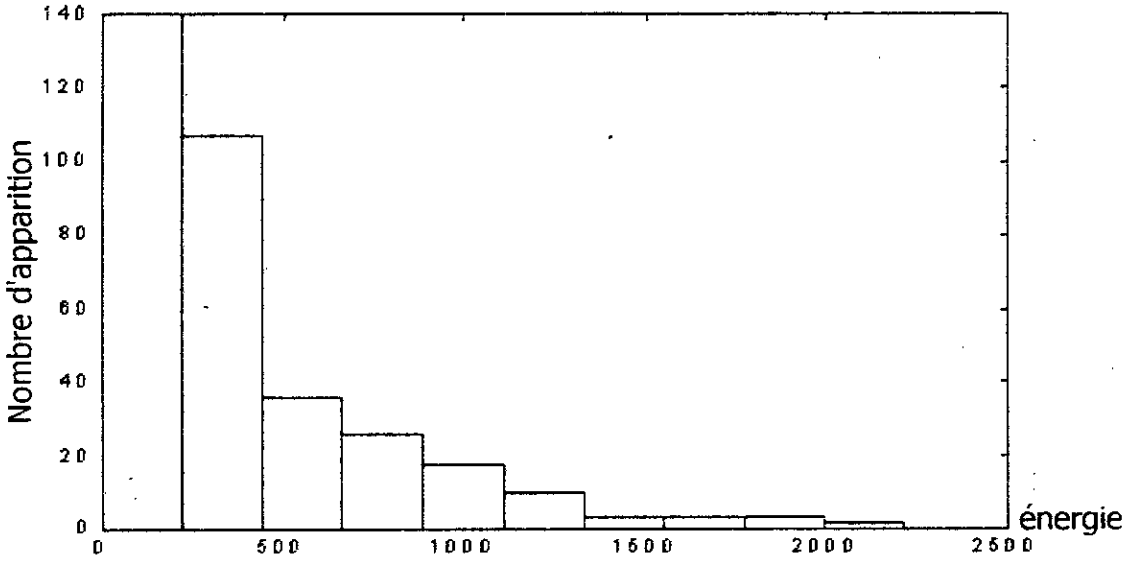


***Figure 5.19** Histogramme de l'énergie.*

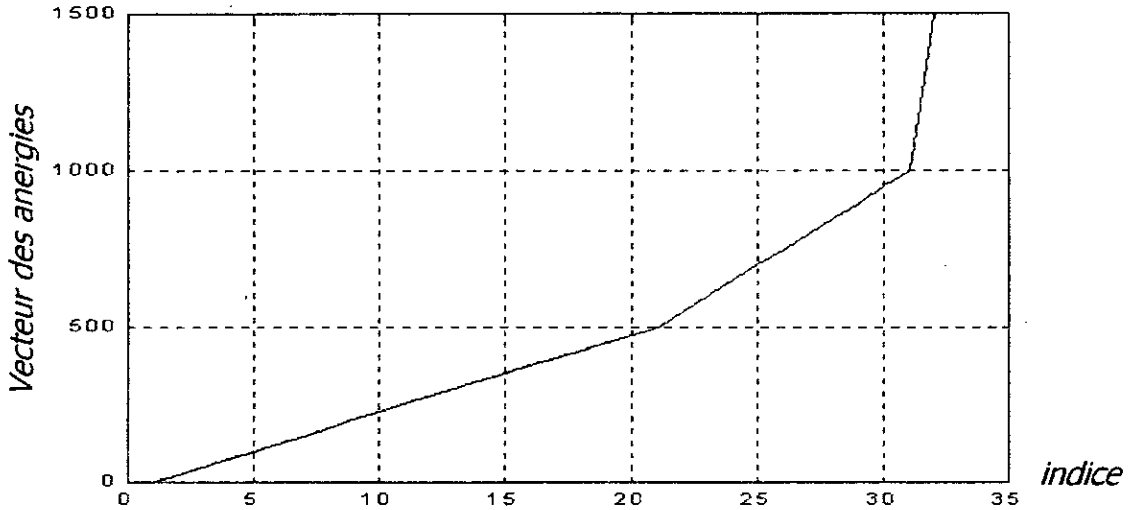


***Figure 5.20** Quantification non-uniforme de l'énergie.*

V.3.6 Allocation de bits

Le tableau 5.1 résume les différentes conditions d'analyse choisies et le tableau 5.2 donne le nombre de bits alloué à chaque paramètre d'analyse.

Paramètre	Valeur
Longueur de la fenêtre d'analyse	200 échantillons (25 ms)
Ordre de prédiction	10
Pas fréquenciel fixe	100 Hz
Nombre de raies du spectre fixe	36
Débordement de la fenêtre de Hamming de chaque côté de la fenêtre d'analyse	60 échantillons (7.5 ms)

***Tableau 5.1** Résumé des conditions d'analyse.*

Paramètre	Nombre de De bits/25ms	Débit bit/s
10 coefficient LSF	8+8+8	960
Spectre du résidu	7+7+7	840
Pitch	7	280
Fonction de voisement	3	120
Energie de la fenêtre	5	200
total	60	2400

***Tableau 5.2** Allocation de bits pour les différents paramètres d'analyse.*

Les figures 5.21 et 5.22 présentent deux tranches de la parole synthétisée (voisé et non voisé respectivement) comparées avec les tranches originales correspondantes avec leurs spectres. Les figures 5.23 et 5.24 donnent les spectrogrammes large bande de deux fichiers originaux ainsi que les fichiers synthétisés correspondants.

V.4 Evaluation de la parole synthétisée

Pour évaluer la qualité de la parole reconstruite, les mesures provenant de tests d'écoute sont les plus significatives, surtout pour le codage à faible débit. Les tests objectifs qui reposent sur la mesure du RSB ne sont utilisés généralement que pour les débits moyens ou plus.

Pour notre codeur, nous avons utilisé une mesure subjective appelée ACR (Absolute Category Rate). Elle consiste à demander aux auditeurs d'exprimer une opinion sur la qualité des séquences synthétisées (en comparaison avec une communication téléphonique) en choisisant une appréciation parmi les cinqs suivantes:

Excellent – Bon – Passable – Médiocre – Mauvais.

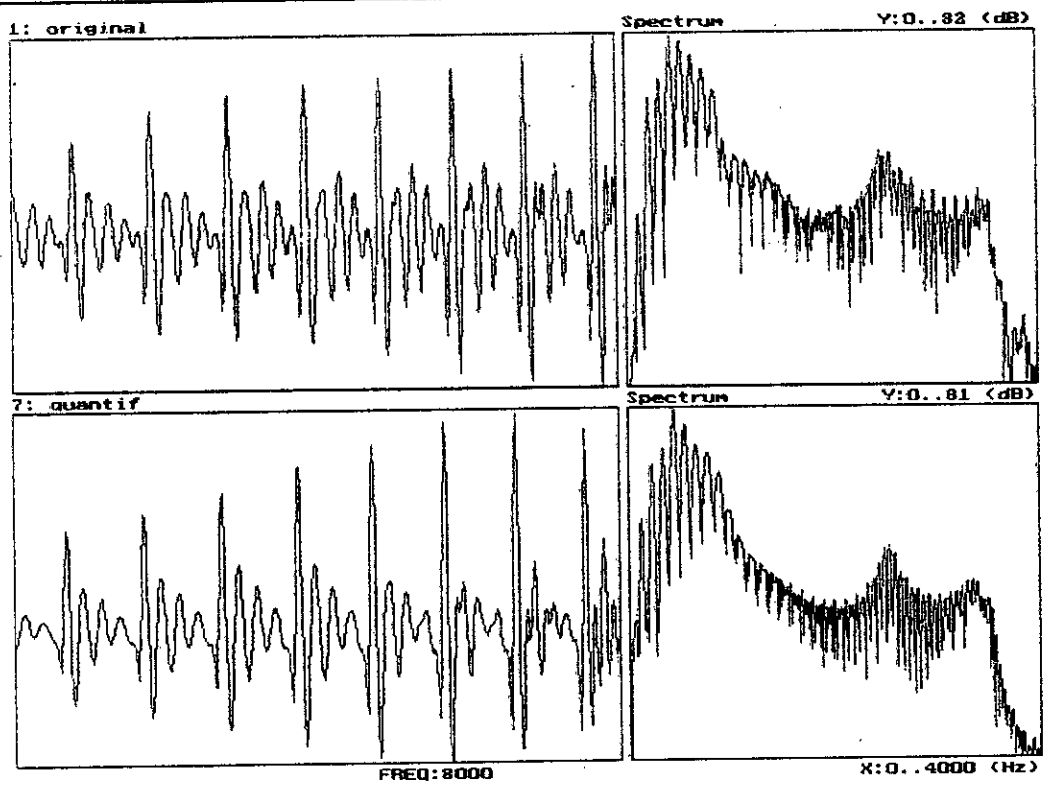


Figure 5.21 Une tranche voisé du signal original et la même tranche du signal quantifié (2400bits/s), ainsi que leurs spectres.

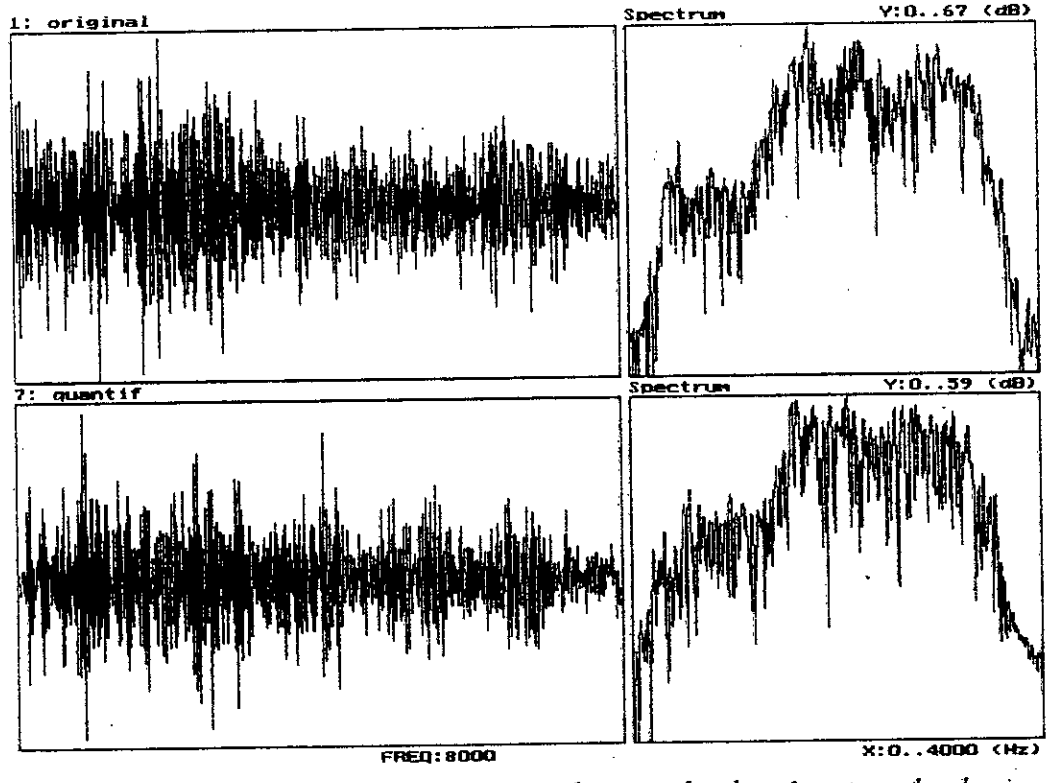


Figure 5.22 Une tranche non voisé du signal original et la même tranche du signal quantifié (2400bits/s), ainsi que leurs spectres.

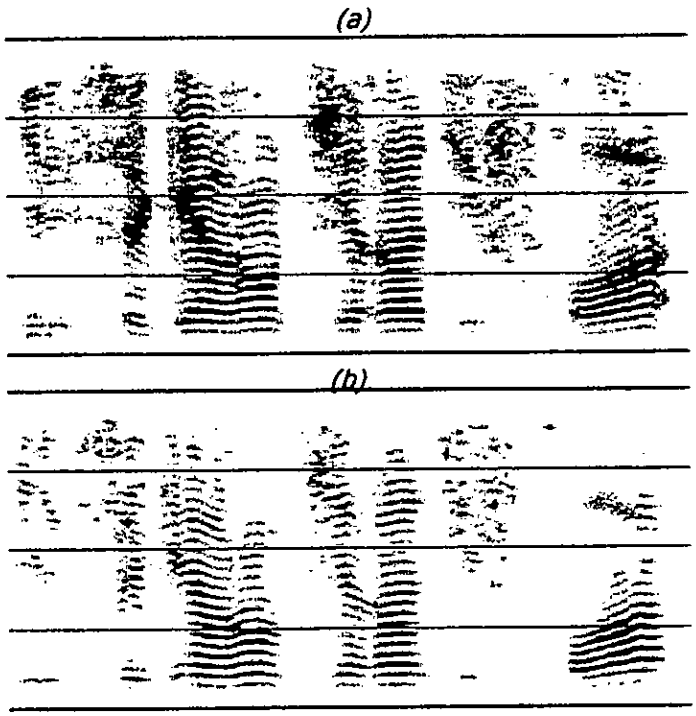


Figure 5.23 Spectrogramme du signal original (a) et du signal codéé(b).

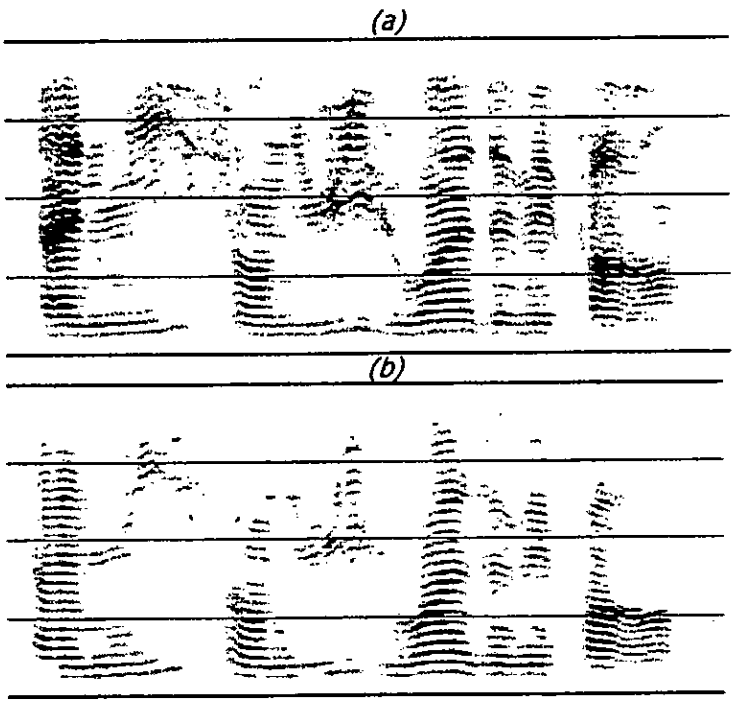


Figure 5.24 Spectrogramme du signal original (a) et du signal codé(b).

Pour exploiter les réponses, on attribue les notes 5,4,3,2 et 1, respectivement, à ces appréciations et on fait le calcul du score d'opinion moyen (MOS).

L'expérience a été réalisée avec quinze auditeurs. Nous avons obtenu un MOS de 3.6 pour la parole synthétisée sans quantification et un MOS de 3 pour la parole codée à 2400 bits/s. Ce qui situe le codeur harmonique dans une assez bonne catégorie.

Conclusion

Dans ce travail, nous avons étudié un nouveau codeur du signal vocal se basant sur le modèle harmonique de l'excitation. Ce codeur introduit par L'équipe du laboratoire Speech coding de l'université de Sherbrooke (Canada) [15], repose sur une nouvelle représentation de l'excitation du filtre de synthèse. Elle résulte toujours d'un mélange d'une source harmonique et une source stochastique (bruit blanc).

Avant d'entamer ce codeur, nous avons commencé par la présentation des caractéristiques fréquencielles et temporelle du signal vocal. Ceci nous a permis d'introduire le modèle classique de reproduction de la parole. Nous nous somme basé sur la modélisation AR du conduit vocal ainsi que la méthode de la prédiction linéaire permettant d'obtenir les paramètres a_i de ce modèle.

Ensuite, nous avons présenté le modèle du codeur ainsi que les différents paramètres qui le caractérisent. L'extraction de ces paramètres a été, ensuite, abordée. Nous avons exposé les différents algorithmes permettant l'extraction des paramètres LSF, du pitch, du spectre du résidu, de la fonction de voisement et de l'énergie de la trame. Nous nous sommes intéressé par la suite à la synthèse de la parole a partir des paramètres d'analyse. Nous avons présenté les différentes méthodes de génération de l'excitation avec toutes les interpolations effectuées sur les paramètres d'analyse.

Enfin, nous avons entamé la simulation du codeur harmonique. Nous avons commencé par la discussion de l'influence des différentes conditions d'analyse sur la qualité de la parole synthétisée. Ensuite, Nous avons présenté la quantification des différents paramètres du codeur. En fin, nous avons simulé le codeur harmonique à faible débit, 2400 bits/s.

Les résultats de l'écoute ont montré que le signal vocal synthétisée, sans quantification, est assez proche du signal original. L'introduction des quantificateurs des paramètres du modèle, en vue de l'obtention d'un codeur à faible débit (2400 bits/s), a remarquablement dégradé l'écoute, mais la qualité reste satisfaisante. La mesure subjective, appelée ACR (Absolute Category Rate), qui a été utilisée pour évalué ce codeur a donnée un score d'opinion moyen (MOS) égal à 3.6 pour la parole synthétisée sans quantification et un MOS de 3 pour la parole codée. Ce qui situe le codeur harmonique dans une assez bonne catégorie.

Plusieurs modifications peuvent être apportées au codeur harmonique que nous avons étudié. La fonction de voisement, qui semble imprécise, peut être modélisée autrement pour mieux conditionner le spectre de bruit. L'extraction des phases du spectre pitch synchro peut donner lieu à une amélioration de la parole synthétisée. Enfin, La quantification des différents paramètres du codeur peut être aussi sujette à une optimisation.

Bibliographie

[1] R.Steele, R.A.Salami, " Speech Coding," Chapitre 3, *in Mobile Radio Communications*,–Pentech Press.1992

[2] M.Xie et D.Berkani, "Amelioration des Performances des Codeurs de Parole," *AJOT 97* vol.-12, No.1, PP.109-115.

[3] D.Berkani, H.Hasseinem et J.P.Adoul "Intelligibility Enhancement of Diver's speech," *Bienval Symposium on Communication*, IEEE com. Canada, Kingston, 1994.

[4] D.Berkani, H.Hasseinem et J.P.Adoul "Single DSP System for High Enhancement of Diver's speech," *IEICE Transaction on information theory.*, No.10, October 1998, Japan.

[5] D.W.Griffin,"Multi-Band Excitation Vocoder,"*Ph.D.Disertation*,M.I.T.,Cambridge ,MA., 1987.

[6] F.Itakura and S.Saito,"Analysis Synthesis telephony Based upon The Maximum Likelihood Methode," *Report of 6th int. Cong. August*, Tokyo, Japan, Paper C-5-5, pp.C17-20, 1968.

[7] S.Y.Kwon and A.J.Goldberg, "An Enhanced LPC Vocoder with No Voiced/Unvoiced Switch," *IEEE Trans. ASSP*, vol.ASSP-32, No.4, PP.851-858, August 1984.

[8] O.Fujimara, "An Approximation to Voice Aperiodicity," *IEEE Trans. Audio and Electroacoust.*, PP.68-72, March 1968.

[9] J.Makhoul R.Wiswannathan, R.Schwartz and W.F.Huggins, "A Mixed-Source Excitation Model for Speech Compression and Synthesis," *IEEE int. Conf. On ASSP*, pp.163-166, April 1978.

[10] R.J.McAulay and T.F.Quatieri, "Speech Analysis-Synthesis based on Sinusoidal Representation," *IEEE Trans. ASSP*, vol.ASSP-34, No.4, PP.744-754, August 1986.

[11] R.J.McAulay and T.F.Quatieri, "Speech Transformations based on Sinusoidal Representation," *IEEE Trans. ASSP*, vol.ASSP-34, No.6, PP.1449-1464, December 1986.

[12] R.J.McAulay and T.F.Quatieri, "Low Rate Speech Coding Based on a Sinusoidal Model," Chapitre 1.6, pp. 165-208 in *Advences in Speech Signal Processing*, S.Furui and M. Sondhi (Eds.), Marcel Dekker, New York,1992.

[13] R.J.McAulay and T.F.Quatieri, "Sin-wave Amplitude Coding at Low Data Rates," Chapitre 1.6, pp. 165-208 in *Advences in Speech Coding*, B.S.Atal, V.Cuperman and A. Gersho (Eds.), Klewer Academic, Boston, MA, pp. 203-213,1991.

[14] B.Atal and R.Schroeder, "Code Excited Linear Prediction (CELP) : High Quality Speech at Very Low Bit Rates," *in proc .ICASSP*, March 1986.

[15] C.Laflamme et al, "Le Codeur Harmonique," *Rapport interne*, Université de sherbrooke.

[16] J.L.Flanagan, *Speech Analysis, Synthesis, and Perception*, Second Edition. New York Springer-Verlag, 1972.

[17] L.R.Rabiner and R.W.Schafer, *Digital Processing of Speech Signals*, Englewood Cliffs, New Jersey : Prentice-Hall, 1978.

[18] J.Makhoul "Linear Prediction: A Tutorial Review," *Proceeding of the IEEE*, VOL.63, NO.4, April 1975.

[19] A.H.Gray, Jr.and and D.Y.Wong. "The Burg Algorithm for LPC Speech Analysis/Synthesis," *IEEE Trans. On ASSP, vol.28, No6*, PP.609-615, Dec 1980.

[20] P.Kabal and R.P.Ramachanndran, "The computation of Line Spectral Frequencies Using Chebeychev Polynomials," *IEEE Trans. ASSP*, vol.ASSP-34, No.6, PP.1419-1425, December 1986.

[21] K.K.Paliwal and B.S.Atal, "Efficient Vector Quantization of LPC Parameters at 24 Bits/Frame," *in ICASSP 1991*, PP.661-664.

[22] F.Merazka, "Quantification des Paramètres LSFs," *Thèse de Magister, E.N.P.*, Juin 1991.

[23] F.Merazka and D.Berkani, "LSP Vector Quantization," *IASTED, International conf.; Signal Processing and Communication*, Canaria Islands, Spain, February 1998.

[24] F.Merazka and D.Berkani, "Vector Quantization of LSP Parameters by Split," *30th IEEE Southeastern Synposium on System Theory*, (SSST'98) WVU Morgantown, West Virginia, pp. 334-337, Mars 1998.

[25] F.Merazka and D.Berkani, "Efficient Vector Quantization of LSP Parameters," *IEEE, SMC Multiconference IMACS*, CESA'98.

[26] F.Merazka and D.Berkani, "Vector Quantization of LSP Parameters at Low Bit Rates," *International conf. CCM'98, AMSE*, Lyon, France, Julay 1998.

[27] Y.Lind, A.Buzo and R.M.Gray, "An Algorithm for Vector Quantizer Design," *IEEE Trans. On Communications*, vol.COM-25, No.1, PP.84-95, January 1980.

[28] D.Berkani, G.Turgeon, A.Chekima et B.Derras, "Utilisation de l'Algorithme LBG et du Treillis en Quantification Vectorielle," *AJOT 91*,No. 8.

[29] A.Gersho and R.M.Gray , *Vector Quantization and Data Compression* , Klewer Academic Publishers 1997.

ANNEXE

L'algorithme LBG

Soit un nombre N de vecteurs dans un espace à k dimensions formés à partir de kN échantillons. Nous disposons au départ d'un alphabet vectoriel **Y'** contenant m vecteurs types **y'** disposés d'une façon quelconque dans l'espace. Notre but est d'obtenir un alphabet vectoriel **Y** qui permettra de coder les échantillons de **X** avec une erreur moindre que celle que nous aurions obtenu en utilisant l'alphabet **Y'**. L'algorithme **LBG**, du nom de ses auteur Lind ,Buzo et Gray [27], emploi un procédé itératif pour trouver **Y** à partir de **Y'**.

Lors du codage, chacun des vecteurs **x** sera associé à l'un des vecteurs **y** la plu petite erreur **d(x,y)** possible. Cette erreur est obtenue par :

$$d(x, y) = \sum_{i=0}^{k-1} (x_i - y_i)^2 \qquad (A.1)$$

La relation (A.1) exprime la distance euclidienne entre **x** et **y**.

Chacun des vecteurs **y** servira donc à coder ou à représenter avec plus ou moins de précision un certain nombre de vecteurs de l'ensemble **x**. La distance euclidienne **d(x,y)** reste le critère de fidélité de cette représentation.

Pour obtenir l'alphabet vectoriel **Y** à partir de **Y'**, nous commençons par associer chacun des vecteurs **x** à l'un des vecteurs **y'** qui est le moins éloigné (au sens de la distance euclidienne). La figure (A.1) montre cette association.

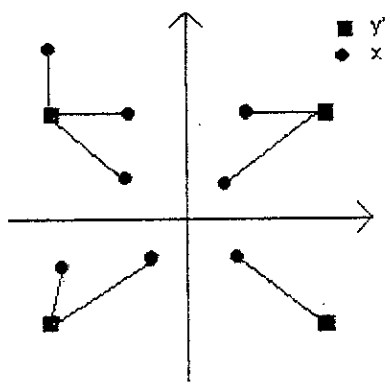


***Figure A.1** chacun des vecteurs x est associé au vecteur y' le plus proche (au sens de la distance euclidienne.*

La deuxième étape consiste à trouver le centre de chacune des classes ainsi constituées et a y placer un vecteur **y** comme le montre la figure A.2.

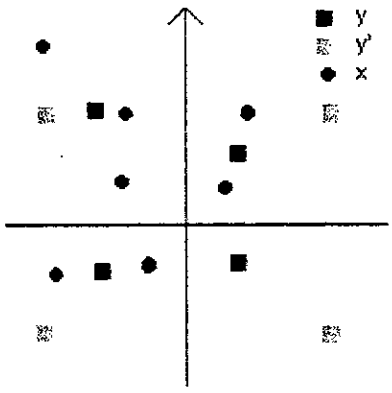


***Figure A.2** on positionne les vecteurs de l'alphabet résultant Y au centre des classes des vecteurs **x**.*

Comme les vecteurs **y** changent de position, il est probable que plusieurs vecteurs **x** changent d'allégeance. Nous recommencerons alors le processus de l'étape 1 jusqu'à ce qu'au moment où une certaine stabilité sera atteinte dans la formation des classes. On peut schématiser cette opération par la figure A.3.

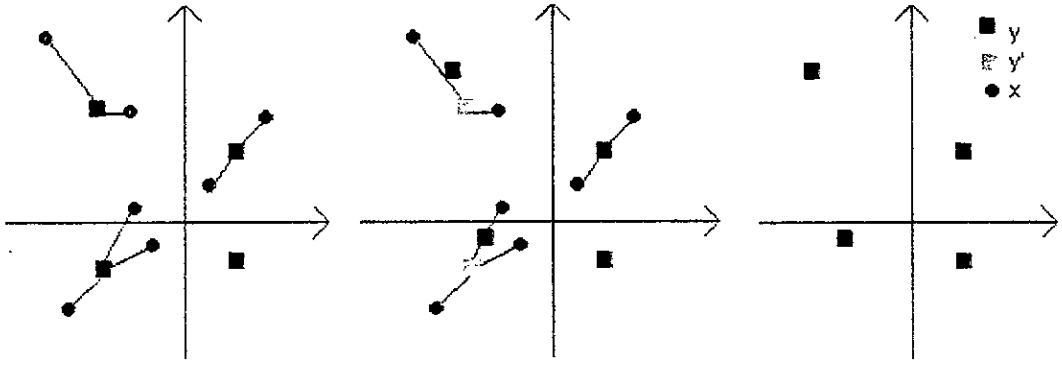


***Figure A.3** succession des assignations **x**, **y** et des changements d'allégeances des vecteurs **x**.*

L'algorithme **LBG** peut être résumé comme suit :

1. Associer chacun des vecteurs **x** à un des vecteurs **y** qui le plus proche ; la distance d(**x**,**y**) est minimale.
2. Calculer le centre de gravité de chacune des classes ainsi formées.
3. Placer chacun des vecteurs **y** au centre de gravité de sa classe.
4. Recommencer l'étape 1 si l'erreur de quantification est supérieure à un seuil prédéterminé.