Université de Nancy I — U.E.R. Sciences Mathématiques

Centre de Recherche en Informatique de Nancy

ETUDE ET SIMULATION D'UNE FAMILLE DE CODEURS HYBRIDES TEMPORELS OFFRANT DES DEBITS DE 6 A 12 KBITS/S POUR DES APPLICATIONS DE QUALITE SUB-TELEPHONIQUE

THESE

Présentée et soutenue publiquement le: 12 Février 1988

pour obtention du diplôme de

DOCTEUR de l'UNIVERSITE de NANCY

en INFORMATIQUE

par

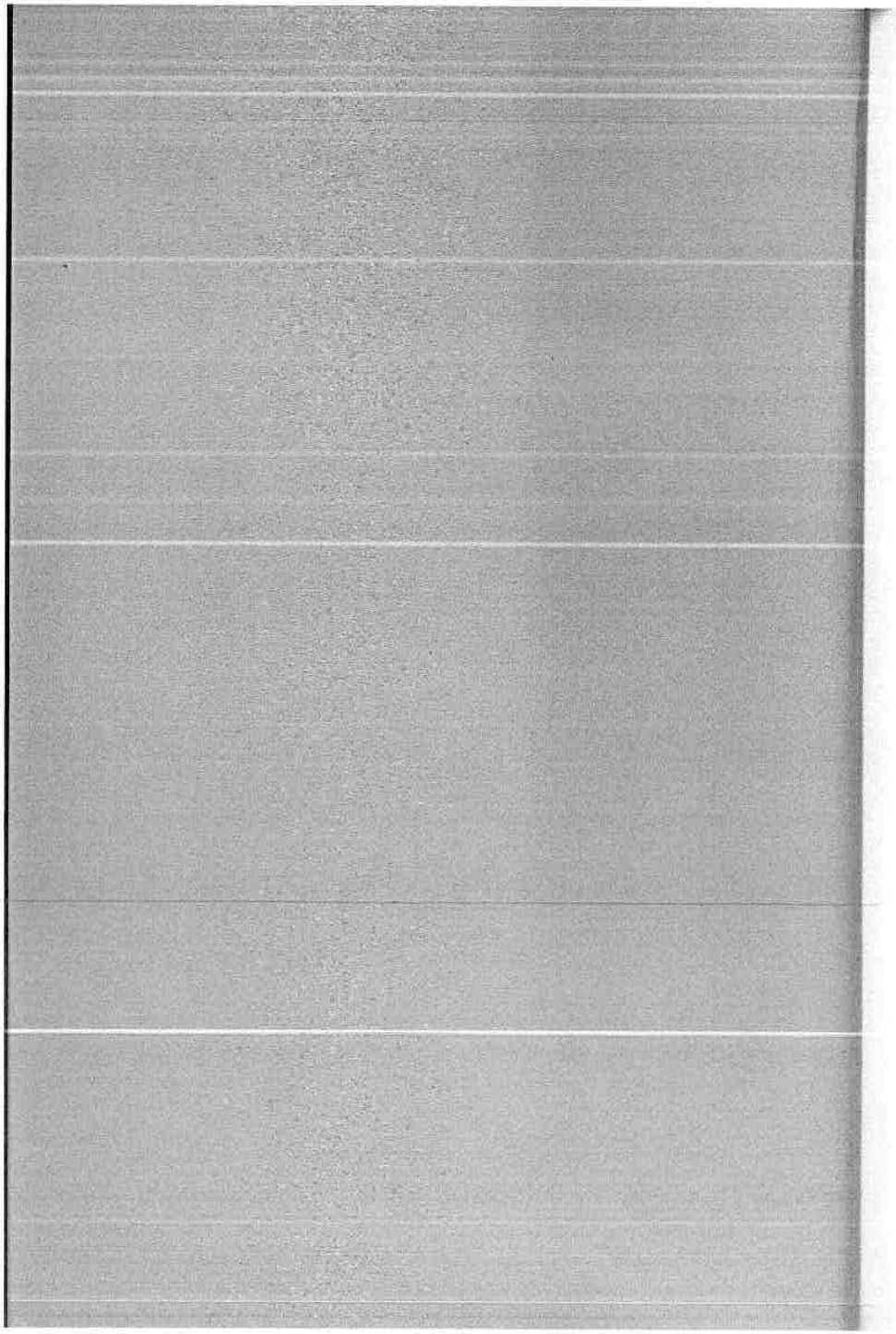
Jean-Georges FRITSCH

devant la commission d'examen

Président ..... J.P. HATON

Rapporteurs: ..... P. LESCANNE
J. MENEZ
J.M. PIERREL

Examinateurs: ..... M. FRIDISCH
M.C. HATON
N. MOREAU



Ce travail a été réalisé, dans le cadre d'une convention CIFRE, au laboratoire des Services Techniques de la Société Telic-Alcatel, en relation avec le Centre de Recherche en Informatique de Nancy.

Je remercie Monsieur J.-P. HATON, Professeur et Directeur du Laboratoire de Reconnaissance des Formes et d'Intelligence Artificielle, d'avoir accepter d'encadrer mes travaux. Il me fait l'honneur de présider ce jury et je tiens à lui exprimer ma gratitude pour la confiance qu'il m'a toujours témoignée.

Je remercie également la Direction Technique de Telic-Alcatel pour avoir mis à ma disposition les moyens nécessaires à l'aboutissement de ce travail.

Je remercie l'ANRT, pour sa contribution financière.

Que soient également remerciés ici,

Monsieur J. MENEZ, Professeur de l'université de Nice, Monsieur J.M. PIERREL Professeur à l'université de NANCY, Monsieur M. FRIDISCH Ingénieur à Telic-Alcatel, Madame M.C. HATON Maître de Conférence de l'université de NANCY, Monsieur P. LESCANNE Professeur de l'université de NANCY et Monsieur N. MOREAU, Enseignant et Chercheur de l'ENST-Paris, qui ont bien voulu soumettre ce travail à leur appréciation.

Je tiens à exprimer ma reconnaissance à Monsieur N. MOREAU, pour son altruisme et pour l'aide précieuse qu'il m'a apportée dans la rédaction de ce rapport.

Mes remerciements vont également à Monsieur H. BARRAL, Chef du Département Signaux et Systèmes de ENST à Paris, pour sa collaboration.

Je ne saurais oublier toutes les personnes du Laboratoire des Services Techniques de Telic-Alcatel et plus particulièrement Mademoiselle D. BUSSOD et Monsieur H. RINIE, pour la sympathie qu'ils m'ont témoignée.

Préface

En matière de communication, la parole occupe une place privilégiée. Depuis quelques années, les systèmes de communication peuvent bénéficier des techniques de traitement numérique du signal qui permettent de banaliser le signal de parole d'un point de vue traitement, stockage et transmission. Aussi, l'étude de procédés de codage du signal téléphonique, offrant une forte réduction du débit par rapport au **MIC 64 kbits/s**, présente un intérêt croissant face au trafic grandissant et à l'apparition de nouveaux services. D'autre part, avec le développement récent des techniques d'intégration, il devient possible d'implanter sur silicium et en temps réel des algorithmes complexes.

D'une manière générale, la numérisation présente les avantages suivants:

- une grande insensibilité au bruit; la numérisation offre une qualité de transmission pratiquement indépendante de la distance de transmission, grâce à la forme simple du signal.

- un caractère universel; contrairement à l'analogique où des systèmes de transmission spécialement adaptés à la nature des signaux sont nécessaires, en numérique les systèmes de transmission acheminent indifféremment des signaux préalablement numérisés issus de différentes sources (téléphone, vidéo, télécopie, données)

- un cryptage robuste; l'utilisation éventuelle de méthodes de cryptage numérique des données à caractère confidentiel, sont plus fiables que les systèmes analogiques.

- une réduction du coût de la liaison; le multiplexage temporel de plusieurs voies favorise l'emploi de composants à grande échelle d'intégration.

- de nouveaux services; l'utilisation des techniques de traitement numérique permet d'envisager de nouveaux services, notamment ceux liés au traitement de la parole (composeur vocal, répondeur, messagerie).

La numérisation présente néanmoins un inconvénient de poids. C'est sa grande largeur de bande passante en transmission.

Dans le cas particulier d'enregistrement et restitution de la parole pour des applications de répondeurs ou de messagerie vocale, le stockage sous forme numérique présente également des avantages par rapport au stockage sous forme analogique:

- pas d'altération au cours du temps; contrairement aux enregistrements analogiques, la qualité des enregistrements numériques n'est pas tributaire directement de la dégradation du support (bande magnétique).

- accès aléatoire possible; dans le cas d'un stockage sur disque, il est possible d'accéder directement à n'importe quel message.

- réduction du coût; l'utilisation de techniques et de circuits de codage à faible débit (8 à 14 kbits/s), permettent de réduire la taille de la mémoire, ou d'augmenter, à capacité constante, le nombre de messages. L'évolution en capacité et en prix des mémoires permettra à moyen terme d'améliorer la qualité des services en élargissant la bande téléphonique par exemle.

Ainsi cette thèse porte sur l'étude d'une famille de codeurs de la parole, dont les caractéristiques répondent aux besoins des nouveaux services tels que messagerie vocale et répondeur enregistreur solide.

Avant-Propos:

Objet de cette thèse:

Le codage de la parole est une discipline qui a subi une forte évolution depuis le début des années 1982 par avènement d'un nouveau concept de codage du signal d'excitation pour les codeurs APC (Adaptive Predictive Coding). C'est [ATAL et REMDE 1982] qui ont développé ce nouveau concept, appelé "Modélisation par Analyse-Synthèse", en proposant un codeur dit "à Excitation Multi-Impulsionnelle". Cette nouvelle technique de codqge temporelle permet de restituer un signal de parole de qualité sub-téléphonique voire téléphonique pour des débits inférieurs à 16 kbits/s. L'article introductif procure une bonne description du principe. Notre approche a été d'évaluer les performances de ce codeur. Compte-tenu des résultats attrayants que nous avons obtenus, notre effort a porté sur l'amélioration du codeur en introduisant notamment un prédicteur à long terme bouclé optimal [MOREAU, DYMARSKI et FRITSCH 1987]. De façon à réduire également la complexité de traitement plusieurs variantes ont été étudiées [BUSSOD et FRITSCH 1986]. Finalement nous avons obtenu un codeur à excitation multi-impulsionnelle qui pour un débit de 12 kbits/s restitue un signal de parole de qualité sub-téléphonique. C'est ce codeur qui nous a servi de référence en terme de qualité.

C'est en 1986 que [ATAL 1986] propose le codeur à excitation par code. Ce codeur se base toujours sur le même concept de modélisation de l'excitation par analyse-synthèse, mais dans une forme vectorielle. Ce codeur est encore appelé codeur à excitation stochastique. Les évaluations que nous avons faites, montrent que ce codeur permet de viser des débits inférieurs à 8 kbits/s, tout en procurant un signal de parole de qualité comparable à celle que procure le codeur à excitation multi-impulsionnelle à 12 kbits/s. Il présente toutefois un inconvénient majeur, compte-tenu de sa complexité qui dépasse les 350 millions de multiplications et additions par seconde. Notre étude a donc porté sur la réduction de cette complexité, en utilisant d'une part les propriétés des matrices, d'autre part en optimisant le dictionnaire d'excitation.

Au cours de cette étude, nous avons développé deux nouveaux concepts de modélisations de l'excitation, appelé "codeur optimale par code" et "codeur multi-impulsionnelle vectorielle", qui allient les avantages des deux techniques étudiées précédemment. L'évaluation de ces codeurs montre que ces techniques procurent une qualité comparable au codeur à excitation par code pour une complexité proche de celle du codeur à excitation multi-impulsionnelle.

Des procédures de codage et de quantification des paramètres sont également décrites, dont notament une technique de codage mixte des paramètres du filtre de synthèse [MOREAU, DYMARSKI, FRITSCH 1987]].

Finalement, une étude comparative des quatres codeurs, montre que ceux-ci permettent de satisfaire toutes les applications de qualité sub-téléphonique, nécessitant des débits compris entre 6 et 12 kbits/s.

Plan de la Thèse:

Le chapitre I donne une description générale du domaine du codage de la parole.

Le chapitre II présente un schéma de base efficace pour le codage de la parole. Les améliorations, apportées par le prédicteur à long terme bouclé, sont évaluées. Le principe de la procédure d'analyse synthèse incluant un filtre perceptuel est également décrit.

Le chapitre III et IV décrivent les 4 procédures de modélisation de l'excitation. Celles-ci s'intègrent dans le schéma de base proposé dans le chapitre II.

Le chapitre V est consacré aux techniques de codage et de quantification, qui ont été appliquées aux paramètres des codeurs.

Le chapitre VI présente les performances, en terme de débit, de complexité et d'effort de mémorisation, des 4 codeurs de qualité sub-téléphonique.

SOMMAIRE

SOMMAIRE

SOMMAIRE

Principales notations du document:

Signaux:

A^i : amplitude des impulsions, ou facteur de gain

b_i : coefficients du prédicteur à long terme B(z)

c_n^i : signal d'excitation par code

C_i : autocorrélation de la réponse impulsionnelle de F(z)

e_n : signal résiduel à long terme

$\tilde{e}_n$: signal d'erreur perceptuelle

f_n : réponse impulsionnelle du filtre de pondération $F(z)= \frac{1}{A(z/\tau)}$

h_n : réponse impulsionnelle du filtre de synthèse H(z)

K_i : coefficients PARCORS du filtre de synthèse H(z)

m_i : position des impulsions

p_n : signal perceptuel

q_n : signal perceptuel synthétique

r_n : signal résiduel

R_i : autocorrélation du signal

s_n : signal original

$\tilde{s}_n$: signal synthétique

T_n : fonction de localisation de la variante 1

u_n : signal perceptuel sans prédiction à long terme

v_n : signal d'excitation multi-impulsionnel

w_n : réponse impulsionnelle du filtre perceptuel W(z)

y_n : signal d'excitation synthétique

α_n : fonction de localisation de la variante 2

Paramètres:

K : nombre d'impulsions ou de séquences d'excitation par fenêtre

-L,L : nombre de coefficients du prédicteur à long terme B(z)

P : décalage du prédicteur à long terme B(z)

MA : durée de la réponse impulsionelle du filtre de pondération F(z)

N : dimension de la fenêtre d'analyse

N' : dimension de la sous-fenêtre d'analyse

MP : ordre du filtre de synthèse H(z)

Pmax : décalage maximum du prédicteur à long terme

Pmin : décalage minimum du prédicteur à long terme

ϕ, ϕ' : fenêtre, sous-fenêtre

CHAPITRE I

VUE GENERALE SUR LE DOMAINE DU CODAGE DE LA PAROLE

I.1 LE TRAITEMENT DE LA PAROLE:

Le traitement de la parole est actuellement une discipline en pleine expansion, grâce à la compréhension croissante de la nature de la parole et de l'évolution également croissante des techniques d'intégration. Il a pour objectif le traitement, la transmission et la réception du moyen de communication qu'est la parole. Ceci n'est pas simple du fait qu'elle véhicule simultanément au moins trois types d'informations qui sont d'ordre linguistique (la langue), d'ordre socio-linguistique (l'accent spécifique a une région), d'ordre personnel (le timbre spécifique à chaque individu). Les grands axes de recherche [47,48] (la bibliographie se trouve à la fin de chaque chapitre) portent sur:

- la reconnaissance de la parole
- l'identification de locuteurs
- la synthèse de la parole
- le codage de la parole

I.1.1 LE CODAGE DE LA PAROLE:

On désigne sous le terme de codage [18] (abrégé de "codage réducteur de débit binaire") l'opération par laquelle l'information provenant d'une source de parole numérisée est transformée en une autre de moindre débit telle que par une opération inverse, appelée décodage, la parole d'origine plus ou moins approximée peut être retrouvée. Ce concept de codage recouvre donc le tandem codeur-décodeur. Lorsque on parlera de la qualité de la parole à débit réduit ou codée, il faudra comprendre après décodage.

Potentiellement le codage de la parole fait appel à la reconnaissance de la parole dans le sens où les limites intrinsèques du nombre d'informations nécessaires sont dictées par le nombre de phonèmes de la langue et par la perception de la parole. Les phonèmes sont, rappellons-le, les éléments phonétiques de base d'une langue et leur nombre est généralement inférieur à une cinquantaine (pour le français 36). Ainsi, la théorie de l'information situe la limite fondamentale du débit aux environs de 2000 informations par seconde. Comme le laisse prévoir ce débit informationnel théorique, la parole est très redondante sur le plan linguistique et le signal qui la représente l'est tout autant. Ce sont les redondances liées au signal que l'opération de codage de la parole tente de réduire.

Ces redondances sont de deux types, à long terme et à court terme. Elles sont liées respectivement à la périodicité de l'excitation et aux résonances du conduit vocal. C'est ainsi que les premiers travaux sur le codage s'appelaient "réduction de redondance", laissant supposer que la réduction du débit portait exclusivement sur les redondances. Ceci n'est généralement pas le cas, étant donné que la réduction du débit supprime également de l'information utile. Cette perte d'information utile provoque une dégradation du signal codé.

La conception d'un codeur requiert un nécessaire compromis entre trois paramètres antagonistes qui sont, la qualité de la parole codée, le débit et la complexité du codeur [19]. Comme nous le verrons plus loin, il existe de nombreuses techniques de codage dont les performances ne sont acceptables que pour un débit particulier. Ceci est lié aux hypothèses sous-jacentes à la technique considérée. En prenant comme référence la qualité, il est possible de classer les codeurs en trois grandes classes [18,34], comme le visualise la figure 1.1. Ces trois classes sont:

- les codeurs dits de "qualité téléphonique", dont la dégradation introduite par le codeur est inaudible, mais dont la facteur de compression peut difficilement dépasser deux.

- les codeurs dits de qualité "sub-téléphonique", dont le signal décodé est très intelligible, tout en présentant des distortions audibles. Ils procurent un facteur de compression compris entre trois et dix, pour une complexité dix à dix milles fois supérieure à celle des codeurs de qualité téléphonique.

- les codeurs dits de "synthèse", dont la qualité est insuffisante pour des applications téléphoniques grand public, mais adaptés aux applications militaires, par exemple. Ils procurent un facteur de compression compris entre quinze et trente. La complexité de tels codeurs n'excède pas cent fois celle des codeurs de qualité téléphonique, compte-tenu justement des hypothèses simplistes qui les caractérisent.

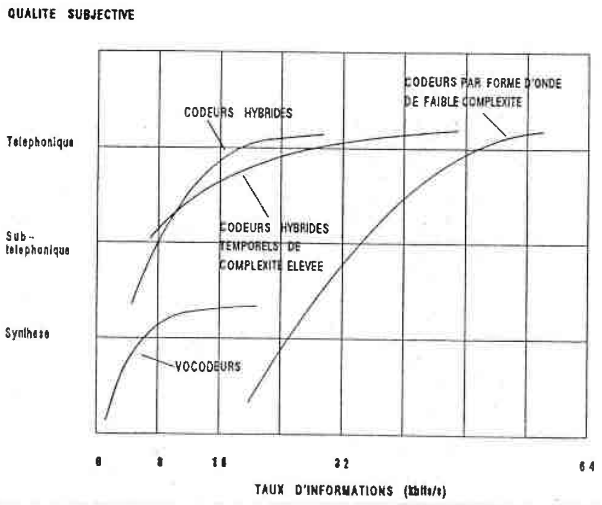


Figure 1.1: Performances des différentes classes de codeurs

Il est à noter que dans le cadre d'une application en transmission, deux paramètres supplémentaires interviennent. Il s'agit du délai de restitution et de la robustesse aux erreurs de transmission, mais ils ne sont pas déterminants, compte-tenu des applications de messageries et répondeurs que nous envisageons. La famille de codeurs que nous avons étudiée entre dans la deuxième classe car elle offre le meilleur rapport qualité-débit. En effet de telles applications nécessitent une restitution de bonne qualité mais qui n'a pas besoin d'être téléphonique, étant donné que les messages codés ne subissent qu'une ou deux opérations de codage-décodage en cascade. D'autre part, un facteur de compression de 6 à 10 est très satisfaisant car il faut garder à l'esprit, que le prix des mémoires chutant, il n'est pas judicieux d'exagérer la réduction du débit au risque d'une complexité démesurée. Notre choix s'est porté sur une classe de codeurs qui met en oeuvre le principe

d'une modélisation de l'excitation par déconvolution itérée, qui fut proposée pour la première fois en 1982 par ATAL et REMDE [3]. On trouve dans cette classe, le codeur dit "à excitation multi-impusionnelle" et le codeur "à excitation par code". C'est en 1986 que ATAL [7] a proposé ce codeur qui correspond à la formulation vectorielle du codeur à excitation multi-impulsionnelle. Ces codeurs offrent respectivement des facteurs de compression de l'ordre de 5 et de 10. L'application de telles techniques de codage au signal MIC permet d'encoder le signal de parole à environ 12 et 6 kbits/s. Afin de pouvoir couvrir les débits intermédiaires, nos recherches ont également porté sur l'étude de nouvelles procédures de modélisation d'une excitation mixte, qui ont abouti au codeur "à excitation optimale par code" d'une part, au codeur à excitation multi-impulsionnelle vectorielle d'autre part. Les différents codeurs de cette classe ne diffèrent que dans le choix de la procédure de modélisation de l'excitation. Les performances des techniques d'intégration s'accroissant, il est possible de faire évoluer le codeur à excitation multi-impulsionnelle, qui est le moins complexe, vers un codeur à excitation multi-impulsionnelle vectorielle, puis vers un codeur à excitation par code, qui est le plus complexe.

I.1.2 PROPRIETES DE LA PAROLE:

La communication vocale se décompose fonctionnellement en deux processus [17,48]:

- la production d'un message oral par un locuteur

- la perception de ce message par un auditeur.

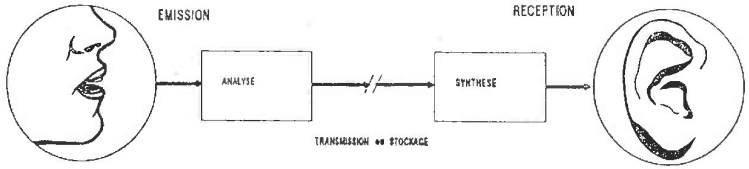


Figure 1.2: Modèle de la communication vocale

Des propriétés liées à ces deux processus ont été mises en évidence puis appliquées au traitement de la parole et plus particulièrement au codage de la parole. Voyons ces propriétés.

I.1.2.1 PROPRIETES LIEES A LA PRODUCTION:

Le mecanisme de production de la parole, proposé par Flanagan [18] (fig. 1.3), se décompose en une excitation et un filtrage. L'excitation est produite par la vibration des cordes vocales ou par des turbulences liées à des constrictions du conduit vocal. Ces deux modes d'excitation peuvent être simultanés. Cette excitation est modifiée spectralement par le conduit vocal (le pharynx suivi des cavités bucales et des lèvres) qui se comporte comme un filtre. Quelques fois, le conduit nasal est couplé en parallèle avec le conduit vocal.

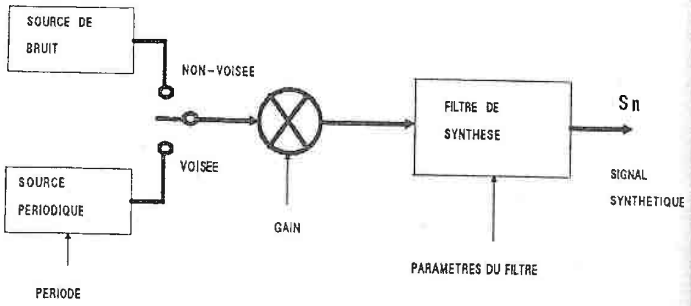


Figure 1.3: Modèle de production de la parole

Le comportement de l'appareil phonatoire humain est non-linéaire compte-tenu des couplages entre l'excitation et le conduit vocal, des déformations plastiques et des pertes par absorption. La prise en compte de tous ces facteurs dans la modélisation du système, aboutit à un modèle dit de connaissance, adapté à la compréhension du processus de production de la parole. Mais il n'est pas adapté au codage de la parole vu sa lourdeur et sa complexité. Un modèle dit de représentation, qui se contente de décrire le comportement externe du système, a pu être obtenu grâce à des simplifications telles que linéarisation et découplage.

Classification des sons:

Selon le mode d'excitation, les sons liés à la parole se classent en trois catégories (fig. 1.4):

- les sons voisés qui sont produits par la vibration quasi-périodique des cordes vocales à la fréquence du fondamental (le "pitch").

- les sons fricatifs qui sont produits par du bruit dus à des constrictions du conduit vocal (s,ch,f). La combinaison simultanée de constrictions et de la vibration des cordes vocales génère des sons fricatifs voisés (z,j,v).

- les sons plosifs qui sont le résultat d'une occlusion du conduit vocal créant une surpression derrière l'occlusion suivi d'un relâchement brutal. Les sons plosifs peuvent être voisés (b,d,g) ou non-voisés (p,t,c).

Les caractéristiques spectrales de la parole sont déterminées par la réponse spectrale du conduit vocal et permettent de limiter son spectre à 8 kHz. Toutefois, une réduction plus importante de la bande passante, dans les hautes fréquences à 3400 kHz, est possible pour les systèmes téléphoniques sans dégradation significative. Ceci permet d'échantillonner le signal à 8 kHz. L'enveloppe spectrale est déterminée par la réponse spectrale du conduit vocal et ses résonances, appelées formants, y apparaissent sous forme de pics.

Les caractéristiques temporelles de la parole sur une durée longue (1 seconde et plus), comme le présente la figure 1.4a, mettent en évidence sa non-stationnarité qui est due à son évolution en énergie, en mode d'excitation, et en contenu spectral. Toutefois sur des durées plus courtes

(10 à 20 ms), la parole est quasi-stationnaire. D'autres propriétés liées à la forme d'onde peuvent être mises en évidence par des critères statistiques tels que la distribution en amplitude ou l'autocorrélation des échantillons. Une distribution gaussienne approxime de manière satisfaisante la distribution à court terme de l'amplitude des échantillons. L'autocorrélation à court terme confirme que la corrélation des sons voisés est élevée (fig 1.4c et fig 1.4d), tandis que pour les sons non-voisés celle-ci est faible (fig 1.4.e). On peut ajouter à cela que pour les sons voisés, il existe également une forte corrélation à long terme, dont la période est synchrone avec la période du fondamental (fig 1.4b).

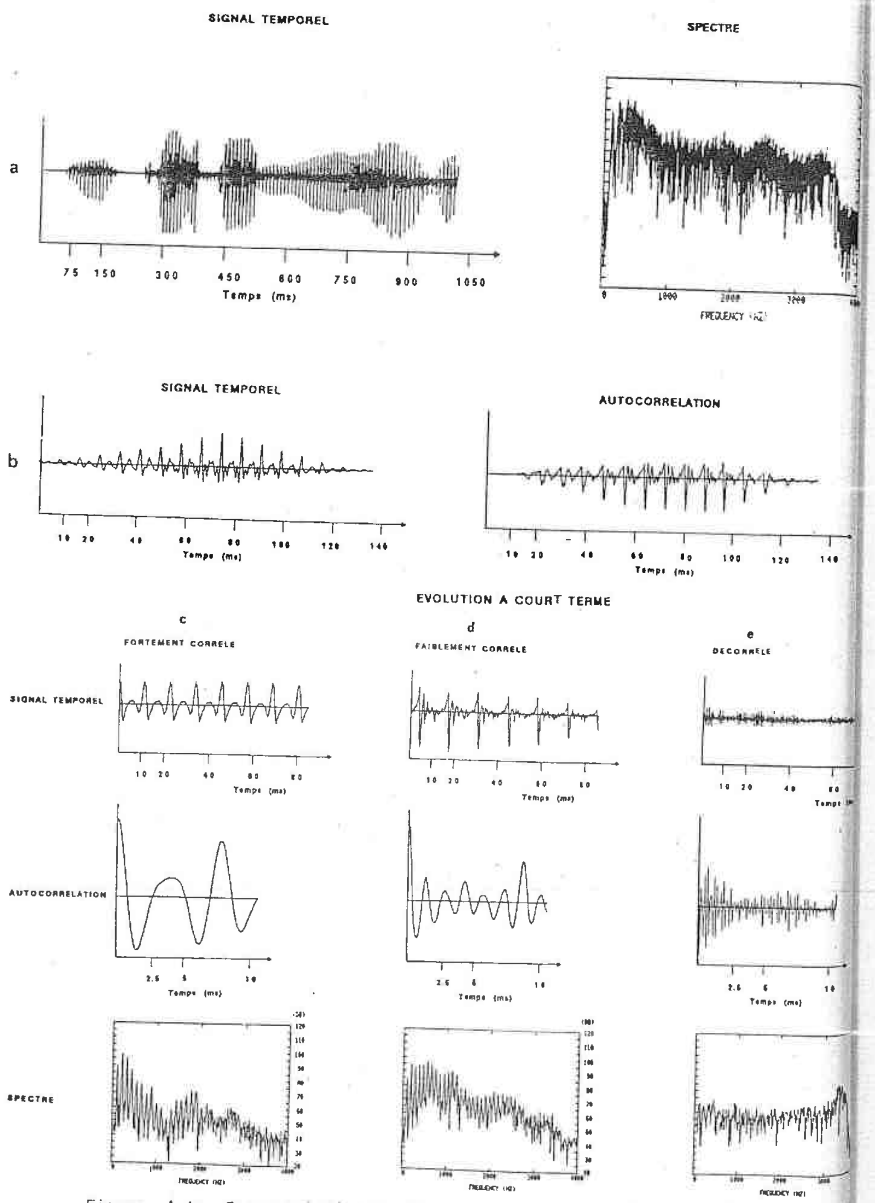


Figure 1.4: Caractéristiques temporelles et spectrales du signal de parole a) évolution à long terme; b) évolution à court terme

I.1.2.2 PROPRIETES LIES A LA PERCEPTION:

L'audition possède des propriétés symétriques à la production de la parole. A la lente évolution du spectre des sons correspondent des processus d'intégration de l'ordre de quelques millisecondes. De même on peut associer à l'enveloppe spectrale à court terme du conduit vocal, un analyseur de spectre à bande relativement large (100 à 150 Hz allant des basses fréquences vers les hautes fréquences). L'audition a une sensibilité logarithmique, comme d'ailleurs tous les organes des sens, ainsi qu'une insensibilité au déphasage. L'intégration des signaux par le système auditif s'accompagne d'un phénomène de masquage des signaux de faible amplitude par rapport aux signaux de forte amplitude. Il apparaît dès que deux signaux, dont les composantes spectrales sont proches, sont présents simultanément ou à des instants voisins de quelques millisecondes. Grâce à cette propriété, le bruit de quantification est partiellement voire totalement masqué par la parole. Il a été appliqué pour la première fois au codage du signal de parole par Atal [4], ce dernier ayant défini une structure légèrement différente permettant d'obtenir un effet similaire.

En conclusion, le mécanisme de production de la parole est modélisé par un système faiblement stationnaire découplé de l'excitation. Celle-ci est constituée soit de trains d'impulsions à la période du fondamental, soit de bruit aléatoire. D'autre part, la forte redondance présente dans le signal de parole est mise à profit par les différents algorithmes de codage pour réduire le débit.

I.2 CLASSIFICATION DES CODEURS DE LA PAROLE:

Nous avons vu précédemment qu'il est possible de classer les codeurs en fonction de leurs performances. Une autre classification est possible suivant les techniques (en terme de propriétés et de modélisation) utilisées pour encoder efficacement la parole. On peut distinguer quatre catégories [18,19,25] qui sont proposées dans la figure ci-dessous.

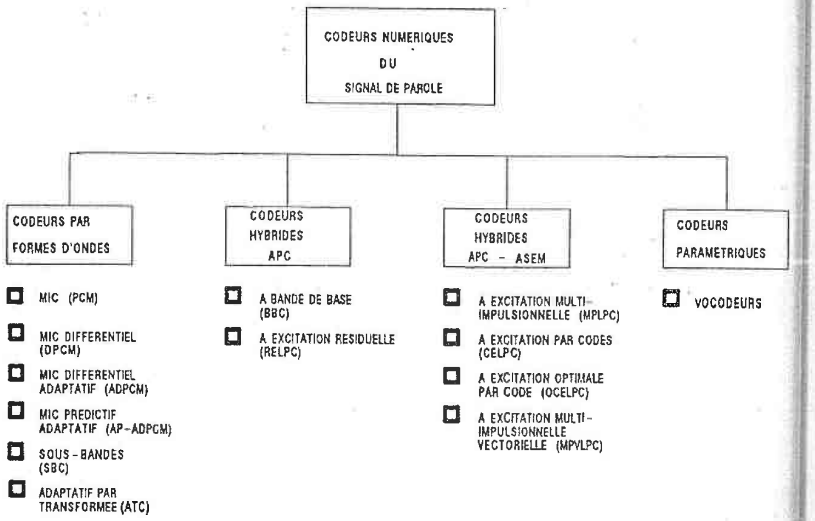


Figure 1.5: Classification des techniques de codage de la parole en fonction de leurs performances

Comme on peut le constater, il existe une forte corrélation entre la catégorie du codeur et ses performances. Les techniques de codage à haut débit sont pratiquement toujours temporelles parce qu'elles cherchent avant tout à préserver la forme d'onde du signal. A l'autre bout de l'échelle, correspondant aux faibles débits, se trouvent les vocodeurs dont les performances sont insuffisantes pour des applications téléphoniques. La plage médium située entre 8 et 16 kbits/s est couverte par les codeurs hybrides.

Une description plus détaillée de ces différentes catégories, dans le cadre d'applications téléphoniques, est donnée ci-après.

I.2.1 CODEURS PAR FORME D'ONDE:

Le codage par forme d'onde n'utilise pas la connaissance à priori du processus de production de la parole. En principe, ces types de codeurs sont adaptés à une grande variété de signaux. Ils présentent l'avantage d'une grande robustesse aux variations entre locuteurs, mais également aux environnements bruités. Ces codeurs opèrent soit dans le domaine temporel soit dans le domaine spectral.

I.2.1.1 CODEURS TEMPORELS:

CODEUR MIC (en anglais PCM):

Une forme simple de codage temporel est le MIC [31] (Modulation par Impulsion Codée), où le signal analogique est échantillonné à la fréquence W et quantifié en $N = 2^R$ niveaux. Le débit associé à ce codage est de W.R bits/s, et la bande passante du signal échantillonné est limitée à W/2 Hz (Shannon). Ainsi la loi de codage MIC à 64 kbits/s, normalisée par le CCITT (Comité Consultatif International des Téléphones et Télégraphes), fournit pour la transmission un code sur 8 bits, correspondant à une quantification non-uniforme sur 256 niveaux du signal de parole, filtré dans la bande 300 à 3400 Hz et échantillonné à 8 kHz. Cette quantification non-uniforme est pseudo-logarithmique (fig. 1.6b). Elle utilise l'effet de masquage en codant avec plus de précision les échantillons de faible amplitude. On obtient ainsi une erreur de quantification proportionnelle au niveau du signal, d'où un rapport signal à bruit constant.

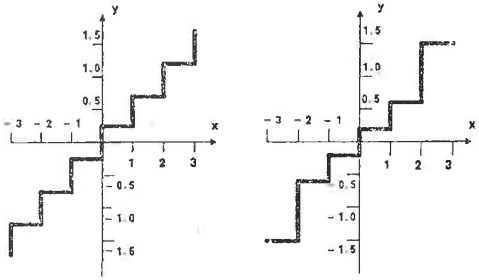


Figure 1.6: Loi de quantification; a) uniforme; b) non-uniforme pseudo logarithmique

CODEUR MIC DIFFERENTIEL (en anglais DPCM):

Le codage temporel peut être optimisé en exploitant les propriétés spécifiques du signal de parole comme la stationnarité. En effet, la corrélation d'ordre 1 à long terme du signal de parole est élévée. Par conséquent, la variance de la différence entre deux échantillons successifs est plus faible que la variance des échantillons. Le principe du codage différentiel [34] consiste donc à coder la différence entre échantillons e_n et non plus les échantillons s_n. Ceci permet de réduire le pas de quantification à débit égal. Le bruit de quantification sera d'autant plus faible que e_n est petit c'est à dire que la différence entre échantillons successifs est faible. On montre que la variance est minimale en codant la différence pondérée qui s'exprime:

$$e_n = s_n - a.s_{n-1} \qquad (1.1)$$

Le coefficient a est ajusté sur le spectre moyen expérimental à long terme de la parole au sens des moindres carrés. Un tel type de codage porte le nom de codage prédictif, car la quantité $a.s_{n-1}$ correspond à la composante prédictible de l'échantillon s_n. Dans ce cas, le codage est

appliqué à l'erreur de prédiction e_n. Ce concept peut être étendu en considérant les termes de corrélation supérieurs. L'erreur de prédiction e_n devient alors:

$$e_n = s_n - \sum_{i=1}^{i=P} a_i . s_{n-i} \qquad \text{où P est l'ordre du prédicteur} \qquad (1.2)$$

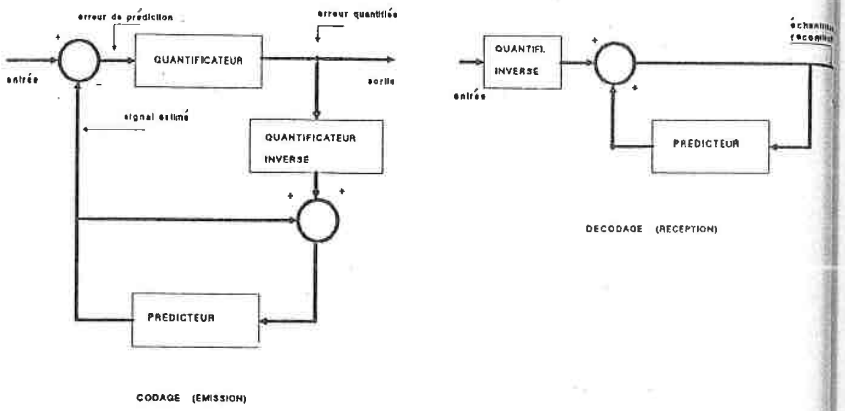


Figure 1.7: Principe d'un MIC différentiel

En pratique, les codeurs MICD sont mis en oeuvre avec un prédicteur d'ordre 2 ou 3 et fournissent une parole de qualité satisfaisante pour des débits supérieurs à 40 kbits/s.

CODEUR MIC DIFFERENTIEL A QUANTIFICATION ADAPTATIVE (en anglais ADPCM):

Compte-tenu de la dynamique du signal de parole, le signal de différence e_n à l'entrée du quantificateur peut varier dans des proportions considérables. Or le codeur décrit précédemment est constitué d'un quantificateur qui reste fixe au cours du temps. Par conséquent, celui-ci n'utilise à un instant donné qu'une faible partie de sa dynamique. D'où l'intérêt d'un quantificateur, dont la loi caractéristique évolue au cours du temps en s'adaptant au niveau moyen du signal. Classiquement le quantificateur fournit 16 niveaux de quantification encodés sur 4 bits comme le visualise la figure 1.8. La connaissance de cette évolution du quantificateur nécessaire au décodage, peut être transmise séparément. Mais plus généralement elle est retrouvée implicitement à partir des échantillons reçus.

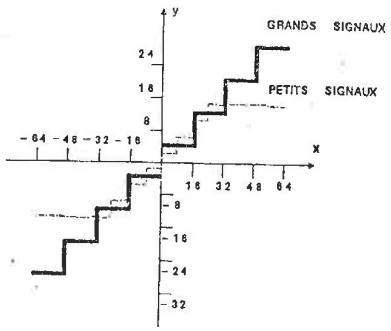


Figure 1.8: Principe du MIC différentiel adaptatif

De tels quantificateurs associés au codeur différentiel décrit précédemment permettent de restituer une parole de qualité téléphonique pour des débits voisins de 32 kbits/s [13,29].

CODEUR MIC PREDICTIF ADAPTATIF (en anglais AP-ADPCM):

Les codeurs décrits jusque là, mettent en oeuvre un prédicteur fixe, qui est représentatif du spectre moyen à long terme de la parole. Ils sont par conséquent mal adaptés aux caractéristiques locales du signal. C'est pourquoi, il est profitable d'adapter les coefficients du prédicteur au rythme de l'échantillon de façon à suivre l'évolution des caractéristiques spectrales à court terme du signal. Les techniques d'adaptation du prédicteur mettent en oeuvre des procédures telles que celle du gradient [47].

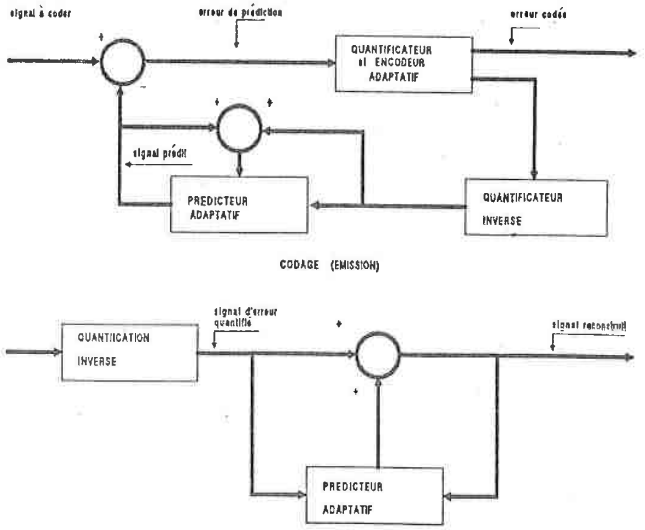


Figure 1.9: Principe du MIC prédictif adaptatif

L'algorithme retenu par le CCITT [CCITT 1983], pour la réalisation de CODEC (codeur-décodeur) à 32 kbits/s entre dans cette catégorie. Il permet de résoudre la quasi-totalité des problèmes de réduction de débit sur le réseau téléphonique, ainsi que ceux de transmission terrestre ou par satellite, compte-tenu de l'augmentation de trafic qu'ils offrent par rapport au codage MIC à 64 kbits/s. Ce codage procure une qualité voisine de celle du MIC. En utilisant des jeux de paramètres particuliers, il est également adapté à des services annexes tels que la télécopie et la télégraphie par exemple.

I.2.1.2 CODEURS SPECTRAUX:

Les codeurs spectraux [40], contrairement aux codeurs temporels qui considèrent le signal de parole comme une entité spectrale unique, subdivisent le signal de parole en un certain nombre de composantes dans le domaine spectral. Le principal avantage de cette approche est, qu'elle permet de répartir de manière dynamique les bits alloués au codage de chaque composante en fonction de leur contenu. Ensuite, le codage de chaque composante fait appel à des techniques temporelles qui sont largement inspirées des techniques décrites précédemment. Il existe principalement deux codeurs que nous allons décrire maintenant.

CODEUR EN SOUS-BANDES:

Cette technique [12] sépare le contenu spectral du signal de parole en sous- bandes contiguës (2 à 8 typiquement) à l'aide de filtres miroirs en quadrature. Ensuite la sortie de chaque filtre est ramenée dans la bande de base puis sous échantillonnée et encodée. Une procédure adaptative ajuste dynamiquement le nombre de bits à allouer aux différentes sous-bandes en favorisant les plus énergétiques. Une autre approche met en oeuvre, pour chaque sous-bande, un codage du type ADPCM voire même AP-ADPCM. A la réception, les sous-bandes sont restituées par les décodeurs correspondant. La bande spectrale totale est reconstruite à l'aide des filtres miroirs en quadrature homologues à ceux de la séparation. Ce mode de filtrage limite fortement (à la reconstruction du signal) les bruits de quantification et évite surtout son extension aux sous-bandes voisines où le signal peut être faible.

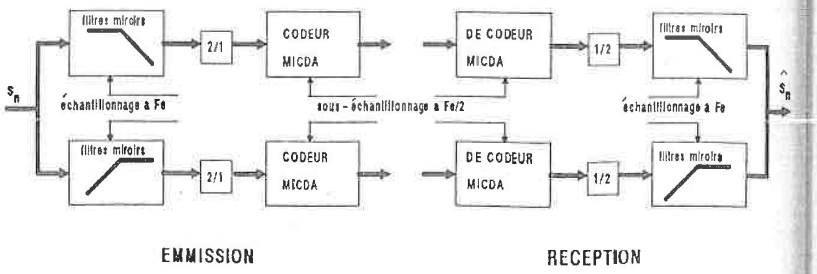


Figure 1.10: Principe du codeur en sous-bandes avec filtres miroirs

Des tests d'écoutes [20] montrent que ce codeur avec codage AP-ADPCM des sous-bandes est performant. L'accroissement de complexité qu'introduit le prédicteur adaptatif est tout à fait justifié, compte-tenu de l'amélioration qu'il apporte. Pour des débits inférieurs à 32 kbits/s, cette technique de codage offre une qualité supérieure au codage temporel AP-ADPCM, l'écart allant en s'accentuant avec la diminution du débit.

CODEUR ADAPTATIF PAR TRANSFORMEE:

Les performances du codage en sous-bandes peuvent être améliorées en augmentant le nombre de sous-bandes. On arrive ainsi à des codeurs dits par transformée [19], pour lesquels le banc de filtres est remplacé par une transformation unitaire temps fréquence. Ce type de codage a été appliqué pour la première fois au signal de parole par Campanella [11], puis proposé dans sa forme adaptative par Zelinski [43]. En principe la transformée de Karuhen-Loeven [42] assure le meilleur gain moyen. Toutefois des considérations tant liées à la perception qu'à la méthode de calcul, conduisent à utiliser en pratique la transformée discrète en cosinus [43]. Cette transformation permet d'obtenir un gain moyen de prédiction très proche du gain optimal moyen assuré par la transformation de Karuhen-Loeven. La figure ci-dessous représente le schéma de principe de ce codeur.

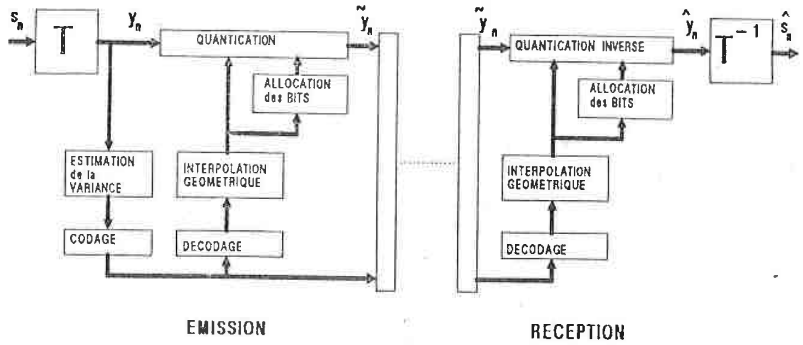


Figure 1.11: Principe du codeur par transformée

Le signal de parole est analysé par blocs de N échantillons (s_n n=0 à N-1). Chaque bloc de signal s_n est transformé en un bloc d'échantillons (y_n n=0 à N-1), qui sont quantifiés de façon indépendante et transmis en même temps que les paramètres du quantificateur adaptatif. A la réception, la quantification inverse est effectuée, puis la transformation inverse restitue les blocs d'échantillons de parole décodés s_n.

Ce codeur permet d'obtenir un signal de parole décodé de qualité téléphonique aux environs de 20 kbits/s. Au débit de 16 kbits/s, le signal décodé présente une qualité sub-téléphonique. Quant à 9,6 kbits/s, les performances se détériorent très rapidement car le signal est dégradé par un effet de "cliquetis" propre à cette méthode.

I.2.2 VOCODEURS:

Nous venons de voir que les codeurs par forme d'onde se comportent de façon indifférente (ou presque) vis à vis du spectre du signal. Le spectre à long terme du signal de parole est plus riche en basses fréquences car il est représentatif de l'excitation voisée (mode d'excitation le plus fréquent). Le spectre à court terme est riche en hautes fréquences car il est représentatif du filtrage introduit par le conduit vocal. Ces propriétés spectrales peuvent être mises à profit en quantifiant non-plus le signal, mais des paramètres représentatifs du spectre du signal. Le principe [6] du vocodeur, qui utilise ces propriétés, s'explique aisément en considérant la partie synthèse représentée par la figure 1.12. Celle-ci est constituée d'une excitation et d'un filtre adaptatif, dont les paramètres réunis sous forme de trames binaires sont renouvelés au rythme de 40 à 100 fois par seconde.

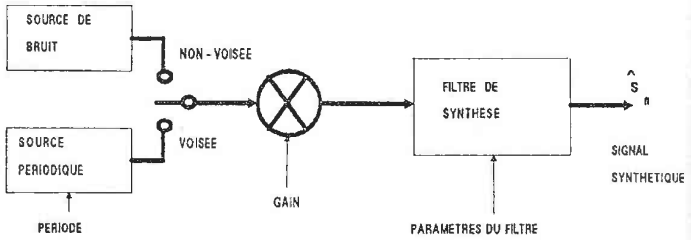


Figure 1.12: Partie synthèse du vocodeur

La partie analyse (Fig. 1.13) consiste donc d'une part à modéliser puis encoder les paramètres du filtre adaptatif, d'autre part à estimer et encoder l'excitation qui peut être soit voisée soit bruitée. Dans le cas d'une excitation voisée les paramètres à coder sont la période du fondamental et son énergie. Dans le cas d'une excitation non-voisée l'énergie est suffisante.

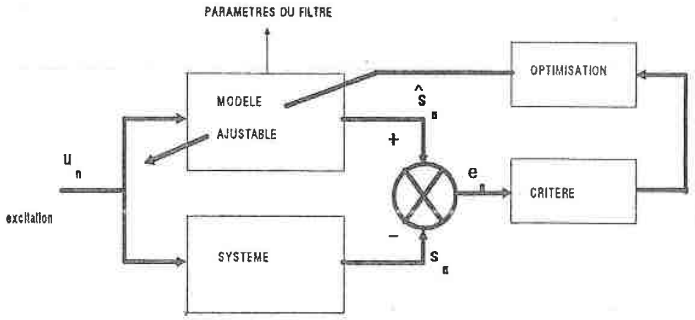


Figure 1.13: Partie analyse du vocodeur

Le vocodeur ne permet pas de restituer de la parole de bonne qualité, compte-tenu des hypothèses simplificatrices qui sont appliquées au signal d'excitation. En effet, l'augmentation du débit au-dessus de 5 à 6 kbits/s

n'augmente guère la qualité subjective de la parole décodée. En réalité le problème de la détection du voisement et du calcul de la périodicité du fondamental est très complexe car cette dernière évolue. Ainsi, l'emploi de vocodeurs reste-t-il limité à des applications particulières telles que certaines transmissions militaires.

I.2.3 CODEURS HYBRIDES:

Les systèmes de codage temporel, même s'ils sont relativement sophistiqués, présentent des performances insuffisantes pour des débits inférieurs à 24 kbits/s. Nous venons de voir également que l'utilisation du vocodeur ne permet pas d'obtenir un signal de qualité sub-téléphonique. En réalité, il existe un troisième mode d'excitation qui est la combinaison des deux modes, voisé et non-voisé. C'est le cas des fricatives voisées (z,j,v). Les transitions entre phonèmes, qui sont entachées de phénomènes de coarticulation, donnent lieu également à des modes d'excitation mixte. Ainsi les recherches se sont-elles orientées vers des systèmes d'architecture hybride qui utilisent de manière complémentaire les avantages du codage par forme d'onde (évitant la mesure du fondamental) et ceux des vocodeurs (assurant un codage efficace de l'enveloppe spectrale). Cette classe de codeur [18,19] permet de restituer un signal de parole dont les caractéristiques temporelles et spectrales sont conservées, ouvrant ainsi la voie à un grand nombre d'applications nécessitant une qualité sub-téléphonique.

L'enveloppe spectrale du signal est codée efficacement grâce à des techniques empruntées aux vocodeurs, telles que la prédiction linéaire ou les bancs de filtres, qui procurent un débit compris entre 1 et 3 kbits/s.

Quant à la modélisation de l'excitation, elle est réalisée avantageusement par le codage temporel d'un signal représentatif de la source ou des caractéristiques temporelles de celle-ci. Selon la technique de codage d'une part et les caractéristiques temporelles et spectrales du signal d'excitation d'autre part, le débit alloué à celle-ci varie entre 3 et 12 kbits/s. Ce débit est évidemment très nettement supérieur aux 400 à 800 bits/s nécessaires au codage de l'excitation d'un vocodeur.

En pratique, la combinaison des différentes techniques temporelles et spectrales peut s'effectuer de nombreuses façons, donnant lieu à de nombreux codeurs hybrides. Ainsi, nous nous limiterons aux codeurs hybrides temporels qui sont basés sur le principe du codage par prédiction linéaire adaptative (en anglais Adaptive Predictive Coding).

CODEUR HYBRIDE A PREDICTION ADAPTATIVE:

Le codage prédictif adaptatif [5] est basé sur la modélisation du conduit vocal et du signal d'excitation glottal à l'aide des techniques de prédiction linéaire. La modélisation du conduit vocal réduit les redondances à court terme contenues dans l'enveloppe spectrale. La modélisation du signal d'excitation a pour rôle de réduire la redondance liée au caractère périodique de ce signal. Le codeur comprend alors deux prédicteurs appellés respectivement prédicteur à court terme et prédicteur à long terme.

Plusieurs schémas de codeurs prédictifs sont possibles:

- le codeur prédictif direct
- le codeur prédictif bouclé

Ils sont équivalents en l'absence de quantificateur et pour une précision de calcul infinie. En revanche, lorsqu'on ajoute un quantificateur, le prédicteur bouclé permet d'atteindre en théorie, pour des hypothèses de stationnarité et de distribution gaussienne du signal, le gain de codage maximum [19]. Ceci n'est pas possible pour le prédicteur direct.

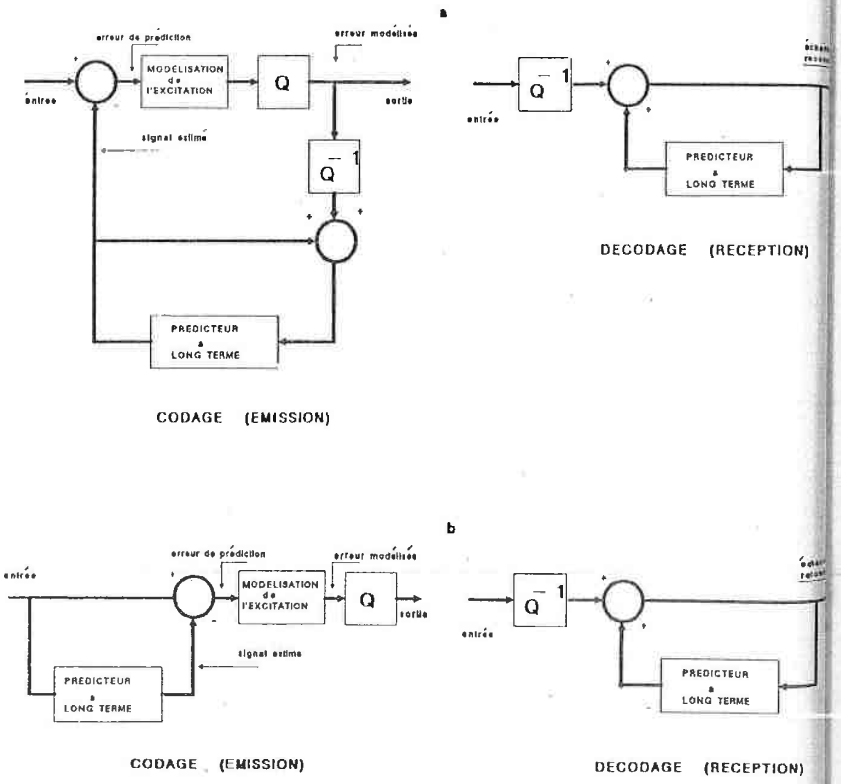


Figure 1.14: a) prédicteur bouclé; b) prédicteur direct

VOCODEUR A BANDE DE BASE:

Ces systèmes fonctionnent selon le principe de base schématisé sur la figure 1.15. La bande haute du spectre, qui est comprise entre 800 et 3400 Hz, est traitée suivant une technique vocodeur. La bande des basses fréquences appelée bande de base est encodée par des techniques de codage par forme d'onde. Le débit nécessaire à l'encodage de la bande de base est bien inferieur à celui requis par les codeurs par forme d'onde classique opérant directement sur le signal. A la synthèse, le signal d'excitation est reconstruit à partir d'une déformation non-linéaire de la bande de base décodée.

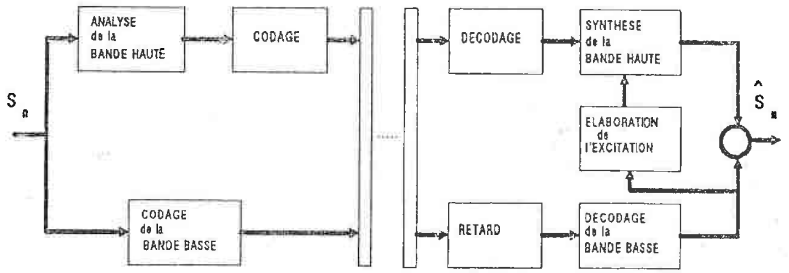


Figure 1.15: Principe du vocodeur à bande de base

Les performances qu'offre cette famille de codeurs sont satisfaisantes pour des débits de l'ordre de 16 kbits/s. De plus, elles sont sensiblement meilleures lorsque la partie vocodeur est estimée par prédicion linéaire plutôt qu'à partir d'un banc de filtres [32].

CODEUR A EXCITATION PAR LE RESIDUEL:

Ce codeur, connu sous le nom RELPC [24] (en anglais Residual Excited Linear Predictif Coder), opère une déconvolution du signal de parole à travers un filtre évolutif blanchissant, dont la fonction de transfert est l'inverse de celle du filtre de synthèse. Ce filtre est estimé à partir du signal de parole. Le signal issu de ce filtrage, dont la corrélation ou les redondances à court terme ont été éliminées, est appelé signal résiduel. Il correspond également à l'erreur de prédiction. Il est clair qu'en utilisant ce signal comme entrée du filtre de synthèse, on reconstruit exactement le signal original. Dans ce type de codeur, qui est la généralisation à un ordre élévé du codeur AP-ADPCM, un gros effort lors du codage instantané du signal résiduel est indispensable. Aussi, selon la complexité du codeur, les performances obtenues varient dans des proportions importantes.

Ce codeur offre une qualité quasi-téléphonique pour des débits avoisinant les 12 kbits/s. Toutefois, il présente un inconvénient majeur qui est sa grande sensibilité aux locuteurs.

I.2.4 CODEURS HYBRIDES A PREDICTION ADAPTATIVE ET A MODELISATION PAR ANALYSE-SYNTHESE DE L'EXCITATION[3,25]:

Contrairement aux codeurs RELPC, les codeurs à modélisation de l'excitation par analyse/synthèse mettent en oeuvre de manière générale des méthodes globales qui supposent disponibles tous les échantillons sur un intervalle donné. De telles méthodes ont l'inconvénient d'introduire des retards à la synthèse. Les modélisations du filtre de synthèse et de son excitation s'effectue bloc par bloc. Le signal d'excitation optimal est obtenu en minimisant un critère d'erreur perceptuel fondé sur les propriétés du système auditif humain. Cette méthode de modélisation de l'excitation a été introduite par ATAL et REMDE [3]. Le principe dont le bloc diagramme général est présenté ci-dessous peut se formuler de la manière suivante.

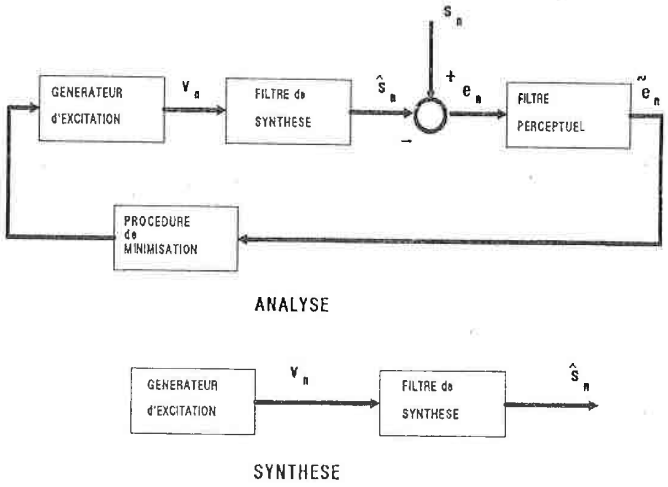


Figure 1.16: Principe des codeurs à modélisation de l'excitation par analyse-synthèse

Connaissant le filtre de synthèse (à court terme et à long terme éventuellement), il faut trouver la séquence d'excitation optimale qui minimise un critère d'erreur sur un intervalle donné. Plusieurs alternatives sont possibles selon que le signal d'excitation est modélisé par:

- un train d'impulsions, donnant lieu à la modélisation multi-impulsionnelle
- des signaux plus élaborés choisis dans un dictionnaire de références, donnant lieu à la modélisation par code.
- un signal qui est combinaison des deux modélisations proposées, donnant lieu à une modélisation mixte qui peut être optimale.

CODEUR A EXCITATION MULTI-IMPULSIONNELLE:

L'hypothèse retenue est d'exciter le filtre de synthèse LPC par une séquence d'impulsions (typiquement 800 à 1200 impulsions par seconde) de telle sorte que, le signal synthétique soit très proche du signal original au sens du critère retenu. Ces impulsions sont caractérisées par leur position et leur amplitude respectives. Toutefois, la recherche optimale de la position et de l'amplitude des impulsions se heurte à une grande complexité, compte-tenu de la non-linéarité du processus de modélisation.

Aussi, des méthodes sous-optimales [10], ont été proposées. Les paramètres transmis sont d'une part les coefficients du filtre de synthèse, d'autre part les positions et amplitudes de l'excitation multi-impulsionnelle. Il est à noter, que ce codeur nécessite un codage-décodage sans erreur des positions des impulsions, sous peine d'une forte dégradation. Aussi, le débit correspondant à l'excitation représente environ les 2/3 voire les 3/4 du débit global.

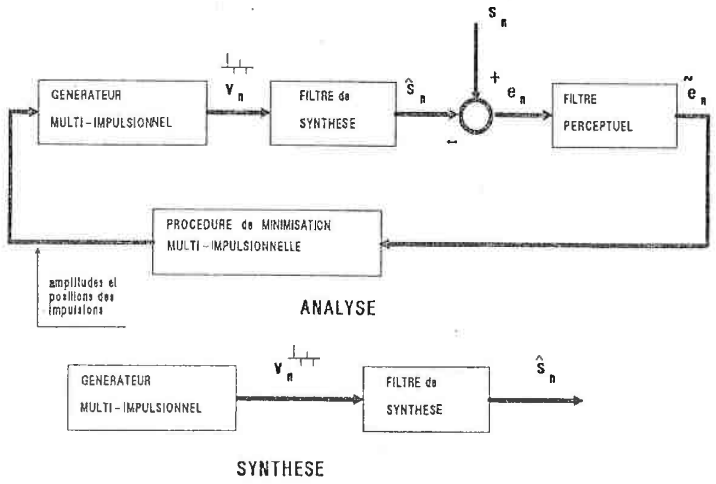


Figure 1.17: Codeur à excitation multi-impulsionnelle

Selon le nombre d'impulsions par unité de temps et les techniques de codages des paramètres retenus, le codeur à excitation multi-impulsionnelle offre une qualité sub-téléphonique pour un débit compris entre 10 et 14 kbits/s. L'algorithme choisi par le CCITT, dans le cadre de la normalisation du 16 kbits/s pour les radio-téléphones, entre dans la catégorie des codeurs à excitation multi-impulsionnelle.

CODEUR A EXCITATION PAR CODE:

Ce codeur [7] permet de réduire le débit nécessaire à l'excitation. L'entrée du filtre de synthèse, comme le présente la figure 1.18, est constituée non plus d'impulsions mais de séquences d'excitations plus

complexes. Celles-ci sont choisies dans un dictionnaire identique à l'analyse comme à la synthèse. Comme pour les méthodes de quantification vectorielle le débit associé à l'excitation se réduit à la transmission de l'index des séquences d'excitation et du facteur de gain associé.

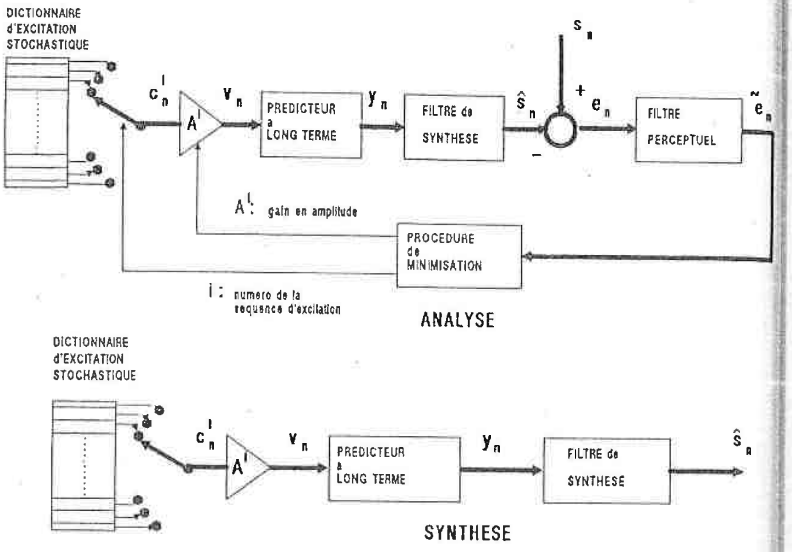


Figure 1.18: Codeur à excitation par code

Selon la façon dont le dictionnaire à été constitué, celui-ci contient entre 512 et 1024 séquences de 40 échantillons. Il suffit donc respectivement de 9 ou 10 bits pour encoder vectoriellement les séquences d'excitation, car celles-ci se juxtaposant, leurs positions sont implicites. Il faut y ajouter 2 à 4 bits pour encoder le gain moyen de chaque séquence, portant ainsi le débit global de l'excitation à environ 2,5 kbits/s (pour une fréquence d'échantillonnage de 8 kHz). Ce codeur permet de limiter le débit global à queqlques 6 kbits/s tout en procurant une qualité comparable à celle que fournit le codeur multi-impulsionnel pour un débit de l'ordre de 12 kbits/s. La complexité de cette technique, qui s'élève à 250 millions de multiplications et accumulations par seconde pour un dictionnaire de 1024 références, est hors de portée des calculateurs. Des simplifications dans les calculs matriciels proposées par TRANCOSO [38] permettent de réduire à environ 25 millions de multiplications et accumulations par seconde.

On peut constater également que cette complexité est directement proportionnelle au nombre de vecteurs du dictionnaire ainsi qu'à la dimension des vecteurs. Aussi, il nous a paru important de porter un effort sur la constitution et le stockage du dictionnaire dans le but de réduire sa taille. Son extraction de manière déterministe permet de limiter sa taille à environ 512 vecteurs, réduisant également la complexité à environ 12 millions de multiplications et accumulations par seconde. Nous avons développé une formulation de la procédure de modélisation par code, qui prend en compte directement le signal résiduel, ce qui entraine certaines simplifications, notamment

l'élimination du filtrage perceptuel. Cette approche procure également, à paramètres identiques, des performances sensiblement meilleures, car la décomposition de la formulation matricielle est bien symétrique. Moyennant toutes ces simplifications, cette technique de codage n'est pas encore à la portée d'un circuit spécialisé unique.

Il est à noter, que ce codeur, restitue difficilement les segments du signal de parole qui sont de nature impulsionnels. Ceci est lié, de manière intrinsèque, au critère des moindres carrés, dont les propriétés de moyennage nivellent le signal.

CODEUR A EXCITATION OPTIMALE PAR CODE:

Ce codeur met en oeuvre une nouvelle procédure de modélisation de l'excitation que nous avons développée. Elle est la généralisation de la procédure d'analyse par synthèse décrite précédemment [7]. Cette approche modélise l'entrée du filtre de synthèse par des séquences d'excitation, dont l'amplitude mais également la position sont déterminées. Un dictionnaire de taille très réduite (5 à 10 références) suffit, car l'information de phase est encodée séparement comme pour le codeur multi-impulsionnel.

Ce codeur permet de réduire d'un facteur 1.5 le nombre d'excitations par unité de temps nécessaires à un codeur à excitation multi-impulsionnelle. La réduction de débit qu'il procure n'est pas significative car les positions doivent être transmises explicitement comme pour un codeur à excitation multi-impulsionnelle. Toutefois, cette approche est intéressante d'un point de vue méthodologique, car elle réalise une reconnaissance de forme du signal. Elle peut être mise en application en reconnaissance de la parole par exemple.

CODEUR A EXCITATION MULTI-IMPULSIONNELLE VECTORIELLE:

Ce codeur met également en oeuvre une nouvelle procédure de modélisation de l'excitation qui combine les modes d'excitation multi-impulsionnelle et par code. Il modélise le signal d'excitation en deux temps.

Premier temps, les positions des impulsions sont déterminées par une modélisation multi-impulsionnelle [3].

Deuxième temps, les amplitudes de ces impulsions sont encodées par une procédure de quantification vectorielle qui prend en compte la fonction de filtre perceptuel. Cette procédure de quantification vectorielle n'est autre que la procédure de modélisation par code [7], qui dans ce cas ne s'applique pas à tous les échantillions de l'intervalle à modéliser, mais uniquement aux quelques impulsions dont les positions ont été déterminées précédemment. Les positions des impulsions sont transmises séparemment.

Ce codeur a l'avantage d'être nettement moins complexe qu'un codeur à excitation par code tout en procurant un débit inférieur à 8 kbits/s. En effet la puissance de traitement ainsi que la dimension du dictionnaire d'excitation sont nettement réduites. En fonction du nombre d'impulsions et taille du dictionnaire, il est possible de couvrir de manière continue les débits intermédiaires (7 à 10 kbits) qui ne sont pas couverts par le codeur à excitation par code et le codeur à excitation multi-impulsionnelle.

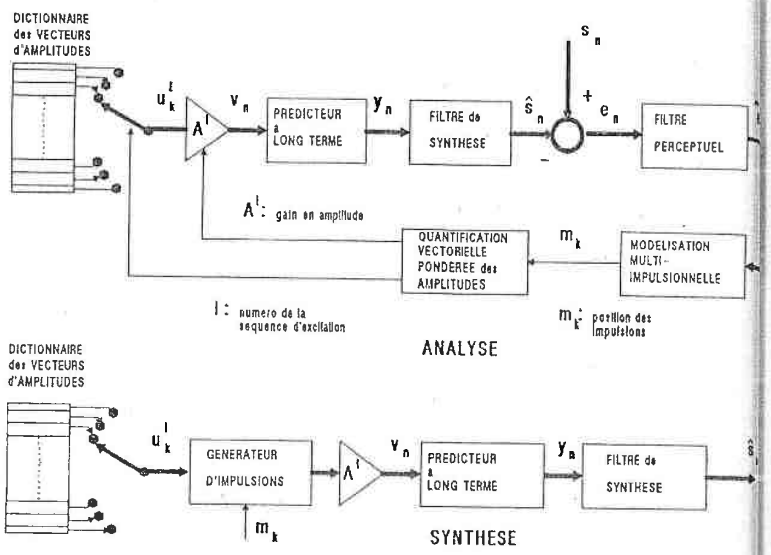


Figure 1.19: Principe du codeur à excitation multi-impulsionnelle vectorielle

1.3 CODAGE ET QUANTIFICATION DES PARAMETRES:

La transmission comme le stockage de la parole codée à débit fini nécessitent que les paramètres transmis soient encodés de façon à les représenter par un nombre limité de bits. L'opération de quantification consiste donc à passer d'un espace de représentation continu en un espace de représentation discret d'une ou plusieurs variables donnant lieu respectivement à des techniques de quantification scalaire et vectorielle.

1.3.1 QUANTIFICATION SCALAIRE:

Pour une seule variable, le processus de quantification est bien connu. Il fait correspondre à la variable d'entrée x une valeur de l'espace de représentation discret y^i la plus proche. Un quantificateur fournit deux informations qui sont:

- la valeur quantifiée de x qui est y
- la valeur codée de x qui est i

Parfois la valeur quantifiée n'est pas utile. Aussi, dans le cas du codage scalaire, on peut tolérer une certaine ambiguité sémantique entre "quantification" et "codage". La figure ci-dessous donne 2 diagrammes équivalents qui mettent en relief la distinction.

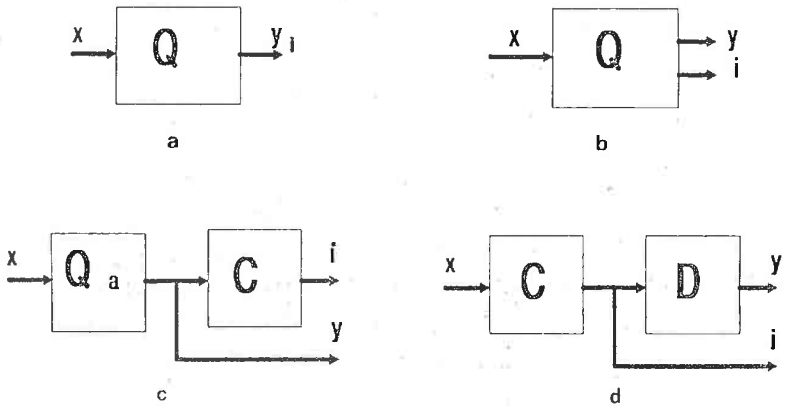


Figure 1.20: Bloc-diagrammes des diffétentes formes de quantificateurs scalaires, a) valeur quantifiée uniquement, b) deux informations distinctes, c) d'abord la valeur quantifiée puis la valeur codée, d) valeur codée uniquement

De manière générale, deux transformations, compensant la sensibilité non-uniforme et la distribution non-uniforme [33] des paramètres à coder, permettent d'améliorer les performances du codage scalaire.

Le codage scalaire est simple à mettre en oeuvre. Il utilise fréquemment des tables de codages dans lesquelles sont figées les valeurs de l'axe de représentation discret. L'effort de mémorisation engendré par les tables de codage scalaire ne dépassent guère quelques kbits. La technique de codage scalaire calculée que nous proposons, permet de réduire (suivant le type de

paramètre) la taille de cette table dans un rapport deux à quatre.

Toutefois, quelles que soient les performances du quantificateur scalaire utilisé, le débit minimum que permet d'atteindre le codage scalaire, est de 1 bit par paramètre.

I.3.2 QUANTIFICATION VECTORIELLE:

Contrairement à la quantification scalaire, qui encode individuellement les paramètres, la quantification vectorielle [1,22], appellée également quantification multi-dimensionnelle, traite les paramètres ou échantillons par blocs ordonnés. Le codage vectoriel est défini par un ensemble de représentants, appelé "dictionnaire" à partir duquel il est possible de reproduire l'espace considéré. D'autre part une mesure de distorsion ou métrique permet d'apprécier de manière quantitative le meilleur représentant.

De manière générale, la métrique [20] retenue est une fonction réelle positive, qui garantit que la distance d'un point à lui-même est toujours inférieure à celle de ce point à tout autre point. Elle est définie par exemple par:

$$d(x,y^i) = \| x - y^i \|^L \qquad (1.3)$$

Dans le cas particulier où L=2, $d(x,y^i)$ est le carré de la distance euclidienne. Elle correspond, dans le cas de vecteurs de signal par exemple, à la puissance de l'erreur de quantification. Pour tout vecteur à encoder, on cherche le représentant du dictionnaire le plus proche au sens de la métrique retenue. Le candidat retenu et son index définissent respectivement, le vecteur quantifié et son code binaire.

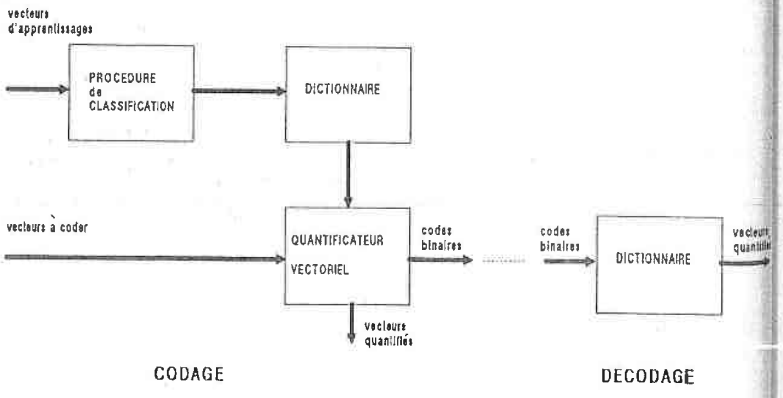


Figure 1.21: Principe de la quantification vectorielle

Ce codage est d'autant plus efficace que la dimension des blocs est grande. Il permet, contrairement au codage scalaire, d'atteindre des débits inférieurs à un bit par paramètre. Toutefois, compte-tenu de l'effort de mémorisation et de la complexité du traitement qu'engendre ce codage, il ne peut être appliqué qu'à des vecteurs de dimension relativement réduite. Lorsque le parcours du

dictionnaire, lors de la recheche du meilleur candidat est intégral, la complexité du codage croît proportionnellement avec la dimension du vecteur. A cela s'ajoute que l'augmentation de la dimension des vecteurs accroît également le nombre de vecteurs du dictionnaire. Les codeurs vectoriels structurés, hiérarchisés [21] ou multi-niveaux [26] sont des exemples de codeurs sous-optimaux qui permettent de réduire soit la complexité soit la taille du dictionnaire soit les deux.

Une forme particulière de codage vectoriel, apte à encoder les signaux temporels, est le codage vectoriel sphérique [2]. Il consiste à quantifier séparemment la norme et l'orientation du vecteur. Ceci est avantageux à plus d'un titre:

- la taille du dictionnaire est réduite, car la norme n'est plus prise en compte dans le vecteur à quantifier, diminuant ainsi la dimension du vecteur de un.

- dans le cas du signal de parole, la norme évolue lentement par rapport à la durée que représente le vecteur. Elle peut être encodée efficacemment par un codage du type prédictif.

Cette technique de codage comporte encore une grande part d'empirisme particulièrement dans le choix et la détermination du dictionnaire. Deux approches sont envisageables:

- l'approche statistique

- l'approche algébrique

L'APPROCHE STATISTIQUE:

Elle [36,45] est recommandée d'une part pour des signaux dont la distribution des orientations est non-uniforme et d'autre part pour des vecteurs de paramètres relativement corrélés qui dans leur espace de représentation révèlent des concentrations localisées. L'extraction du dictionnaire à l'aide d'un algorithme d'apprentissage, permet de réduire sa taille en mettant à profit cette non-uniformité ou corrélation. C'est le cas, comme le présente la figure ci-dessous, des coefficients du filtre de synthèse dans le modèle à prédiction linéaire.

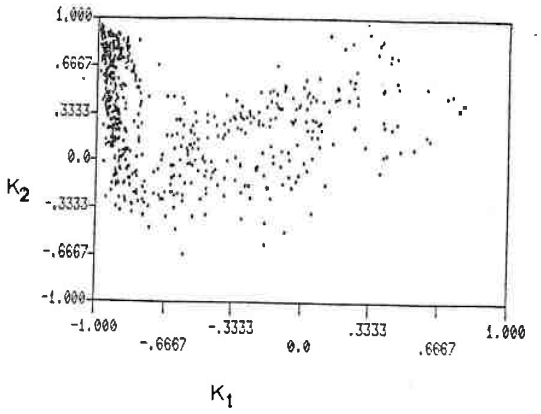


Figure 1.22: Projection des deux premiers coefficients parcors K_1 et K_2 du filtre de synthèse sur le plan principal

L'exctraction statistique du dictionnaire fait appel à des algorithmes de classifications qui peuvent être classés en deux catégories, à savoir:

- les méthodes de classification globale

- les méthodes de classification en ligne

METHODES DE CLASSIFICATION GLOBALE:

Parmi ces méthodes de classification globales, on peut citer les méthodes des nuées dynamiques, qui sont appliquées dans le domaine de la reconnaissance de forme et d'analyse de données. L'inconvénient majeur de ces méthodes est qu'elles nécessitent la présence permanente des données jusqu'à leur convergence. Il en résulte que, compte-tenu de la complexité de traitement, ces méthodes sont difficilement applicables à des ensembles de données de grande taille nécessaires au codage de la parole.

METHODES DE CLASSIFICATION EN LIGNE:

Les méthodes de classification en ligne et plus particulièrement la méthode de classification à seuil [44], ne nécessitent pas la présence des éléments déjà affectés. Par conséquent, ces méthodes sont adaptées à traiter des bases de données de grande taille. La création du dictionnaire est séquentielle. Le premier élément de la base de donnée, devient le représentant de la première classe. Ensuite, les éléments suivants ne créent des classes supplémentaires que si la distance à la classe la plus proche est supérieure à la valeur d'un seuil fixe S. En revanche, si la distance est inférieure au seuil, le représentant de la classe la plus proche est actualisé. L'organigramme de la figure ci-dessous illustre le principe de la methode de classification par seuil.

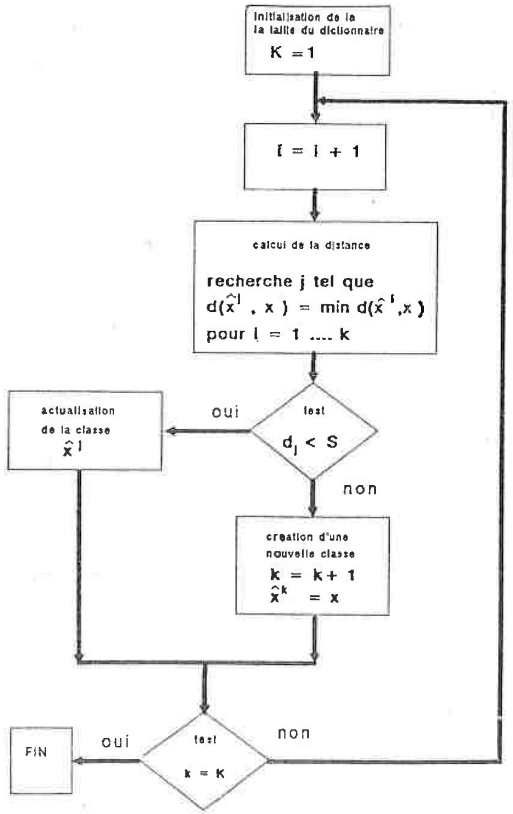


Figure 1.23: Organigramme de l'algorithme de classification à seuil

La valeur à donner au seuil de classification est importante, car elle détermine la taille finales du dictionnaire. La méthode à suivre, consiste à faire varier le seuil S par petits pas. Une évaluation des performances des dictionnaires successifs permet de retenir le

dictionnaire optimal.

Cette approche présente, actuellement, un inconvénient majeur. En effet, l'effort de mémorisation qu'engendre le stockage du dictionnaire est exhorbitant (plusieurs centaines de kbits). Néanmoins pour les nouvelles technologies cette contrainte est fortement atténuée.

L'APPROCHE ALGEBRIQUE:

Elle fait usage des propriétés de certains réseaux réguliers [37]. La quantification algébrique est appropriée aux signaux dont la distribution de l'orientation est approximativement uniforme sur une hypersphère unité. Elle a été appliquée au codage du résiduel par ADOUL [2]. Cette approche, qui pour l'instant n'est pas beaucoup utilisée, est pleine d'avenir.

I.4 COMPLEXITE DES ALGORITHMES DE CODAGE:

Lorsqu'on souhaite implanter un algorithme, on se trouve confronté au compromis à gérer entre calculateur banalisé et performances. Dans le domaine du traitement du signal, ce compromis disparaît, compte-tenu des contraintes liées d'une part au temps réel, d'autre part au nombre d'opérations à réaliser par unité de temps. En effet, des traitements réalisant des millions de multiplications et accumulations par secondes ne peuvent être implantés que sur des processeurs ou plus généralement des architectures spécialisés. Dans le domaine du codage, cette complexité devient [19] considérable lorsque l'on souhaite réduire fortement le débit. La figure ci-dessous, qui présente l'évolution de cette complexité en fonction du débit pour une qualité comparable, montre que diviser le débit par deux multiplie la complexité du codeur par un facteur dix.

TYPE de CODEUR	DEBIT en Kbits/s	COMPLEXITE en 10^6 +,* / s
PCM	64	0.01
AP-ADPCM	24	0.1
SBC ADAPTATIF	16	1
MPLPC	12	2
CELPC	6	25
VOCODEUR LPC	2	1

Figure 1.24: évolution de la complexité des codeurs en fonction du débit à qualité comparable

Comme nous l'avons déjà signalé, la nature des signaux de test ainsi que leurs conditions d'enregistrement sont de toute première importance. Les signaux employés sont des suites de phrases phonétiquement équilibrées [23,30], enregistrées à travers un combiné téléphonique. Elles sont prononcées par différents locuteurs (masculin, féminin, enfant).

CRITERE DE FIDELITE A LA VOIX:

Ce critère purement subjectif, sert à évaluer si le locuteur peut encore être identifié à l'aide de sa voix après codage. Ce critère n'est pas utile pour les codeurs de qualité téléphonique et sub-téléphonique.

Les mesures subjectives sont applicables à tous les codeurs quelques soient leur débits, mais leur mise en oeuvre est longue. De ce fait, celles-ci sont souvent remplacées par des mesures physiques, qui ne nécessitent pas l'intervention humaine.

I.5.2 EVALUATION OBJECTIVE:

On distingue principalement 2 indices objectifs qui sont:

- l'indice de qualité temporel
- l'indice de qualité spectral

Ces deux indices sont complémentaires. Le premier est le reflet de la puissance du bruit entre le signal original et le signal synthétique. Le deuxième évalue la distribution fréquentielle du bruit dans l'enveloppe spectrale du signal synthétique.

INDICES DE QUALITE TEMPORELS:

Actuellement, les indices temporels les plus fréquemment utilisés sont le rapport signal à bruit global et le rapport signal à bruit segmental. Ils s'expriment en dB.

Le rapport signal à bruit global est défini à partir de la puissance de l'erreur entre le signal synthétique $\hat{s}_n$ et le signal original s_n:

$$RSB\text{-}glob = 10 \log \frac{\sum_n \hat{s}_n^2}{\sum_n (\hat{s}_n - s_n)^2} \qquad (1.6)$$

La sommation est effectuée sur l'ensemble des échantillons d'une phrase.

Le choix d'une méthode de codage particulière sera non-seulement fonction de la qualité désirée, mais également du compromis obtenu entre l'économie réalisée en transmission ou stockage et le coût de conception du codeur.

I.5 EVALUATION DE LA QUALITE:

La qualité [8,27] de la parole codée peut être évaluée, soit par des mesures subjectives faisant appel à des essais d'écoute, soit par des mesures objectives réalisées entre le signal original et le signal synthétique. Quelque soit la méthode d'évaluation retenue, la nature des signaux de test et les conditions d'enregistrement sont de première importance.

I.5.1 EVALUATION SUBJECTIVE:

La qualité subjective s'exprime suivant plusieurs critères qui sont:

- l'intelligibilité
- l'agrément à l'écoute
- la fidélité à la voix du locuteur.

CRITERE D'INTELLIGIBILITE:

Il est évalué à l'aide de logatomes. Les logatomes sont des mots sans signification, de structure consonne-voyelle-consonne. Le pourcentage de reconnaissance aux logatomes caractérise de manière stable l'intelligibilité. Il est à noter que ces tests ne sont utiles que pour des codeurs dont le débit est inférieur à 16 kbits/s.

CRITERE D'AGREMENT A L'ECOUTE:

Il met en oeuvre des méthodes d'isopréférence ou de préférence relative. Le bruit multiplicatif est fréquemment utilisé comme mesure de dégradation. Ces évaluations consistent à comparer le signal original dégradé par le bruit multiplicatif (pour différentes valeurs de k) avec le signal décodé. L'expression du signal dégradé par le bruit multiplicatif est:

$$s\_dég_n = s_n + k.b_n.s_n \qquad \text{où } s_n \text{ est le signal original, } b_n \text{ est un bruit aléatoire.} \qquad (1.4)$$

le rapport signal à bruit en dB est:

$$R\_dég = -20 \log k \qquad (1.5)$$

Lorsque le signal original dégradé par le bruit multiplicatif et le signal décodé sont jugés proche en qualité, la dégradation introduite par le codeur est alors équivalente à celle qu'apporte le bruit multiplicatif. La valeur de la dégradation due au codeur est proche de R_dég = -20 log k. Notons qu'il existe d'autres méthodes qui utilisent comme étalon le signal MIC sur 8, 7 ou 6 bits.

Le rapport signal à bruit segmental est défini comme la moyenne de rapports signal sur bruit:

$$RSB\text{-}seg = \frac{1}{N} \sum_{i=0}^{N-1} 10 \log \frac{\sum_{n=M.i}^{M.i+M-1} s_n^2}{\sum_{n=M.i}^{M.i+M-1} (s_n - \hat{s}_n)^2} \qquad (1.7)$$

La longueur du segment de sommation M est de 16 ms soit 128 échantillons pour une fréquence d'échantillonnage de 8 kHz. Elle est en rapport avec les caractéristiques de stationnarité du signal de parole. N correspond au nombre de segments constituant la phrase.

INDICES DE QUALITE SPECTRALE:

Si les indices temporels sont reconnus et peu coûteux en temps de calcul, il n'en est pas de même pour les indices spectraux. Une manière intuitive de comparer deux spectres est d'utiliser la distance spectrale. Elle est définie par:

$$D(w) = \frac{S(e^{jw})}{S'(e^{jw})} \qquad (1.8)$$

où $S(e^{jw})$ et $S'(e^{jw})$ représentent respectivement les spectres de puissance d'un segment de signal original et synthétique.

Pour p=1, D(w) correspond à l'erreur moyenne entre les deux spectres.

Pour p=2, D(w) correspond à l'erreur quadratique moyenne.

En conclusion, à la question " quelle méthode employer pour évaluer la qualité de la parole codée ? " il n'est pas simple de répondre.

Il semble que les méthodes subjectives donnent les résultats les plus stables et les plus fiables quels que soient le débit et la méthode de codage. Elles sont retenues pour les évaluations finales des codeurs à moyen et bas débits car la corrélation entre le signal original et le signal synthétique est réduite.

Quant au problème du passage des critères objectifs aux critères subjectifs, il n'a pas encore été résolu de façon satisfaisante. Aucun indice objectif à l'heure actuelle, ne peut prédire les performances d'un ensemble de codeurs basés sur des principes différents. Toutefois, pour une même famille de codeurs, les indices objectifs de rapport signal à bruit varient de manière monotone avec la qualité subjective. Ceci nous a conduit à retenir les indices objectifs de rapport signal à bruit segmental et global comme critères d'évaluation et d'optimisation au cours de l'étude.

I.6 CONCLUSION:

Dans ce chapitre nous avons tenté de présenter les nombreux aspects que couvrent le codage de la parole.

Pour les applications téléphoniques envisagées, seuls les codeurs hybrides répondent à la double exigence de qualité quasi téléphonique pour un débit aux environs de 10 kbits/s. En effet, les vocodeurs à détection du fondamental ne peuvent être retenus à cause de la modélisation excessive de la source. D'autre part, les codeurs par forme d'ondes, dont la qualité du signal décodé est certe excellente, ne permettent pas de réduire le débit à moins de 16kbits/s.

Les codeurs hybrides mettent en oeuvre les techniques de codage les plus sophistiquées. Leur complexité est importante. Notre choix s'est porté plus particulièrement sur la famille de codeurs basés sur la prédiction linéaire et la modélisation temporelle de l'excitation par des procédures d'analyse-synthèse. Les codeurs les plus connus de cette famille sont, le codeur à excitation multi-impulsionnelle, et codeur à excitation par code. Notre étude a porté sur le développement de deux autres codeurs qui sont: le codeur à excitation optimale par code et le codeur à excitation multi-impulsionnelle vectorielle. On peut noter que ce dernier permet de couvrir efficacement, pour une qualité sub-téléphonique, des débits compris entre 7 et 10 kbits/s.

Les codeurs de cette famille, qui ne diffèrent que par la procédure de modélisation de l'excitation, couvrent l'ensemble des applications nécessitant une qualité sub-téléphonique. Le coût opératoire qu'ils engendrent s'échelonne entre 3 et 17 millions de multiplications et additions par seconde. Ceci n'est pas sans incidences sur la réalisation de ces codeurs.

Il existe un tronc commun aux différents codeurs à prédiction adaptative avec modélisation de l'excitation par analyse synthèse. Ceci est décrit dans le chapitre II.

Bibliographie:

Publications:

[1] "Vector Quantization of Speech and Speech Waveforms"
ABUT H., GRAY R.M., REBOLLEDO G.
IEEE Proc. Int. Conf. on ASSP, 1982, pages 423-435

[2] "La Quantification Vectorielle des Signaux: Approche Algébrique"
ADOUL J.P.
Ann. Télécommun., 1986, pages 158-177

[3] "A New Model of LPC excitation for Producing Natural Sounding Speech at Low Bit Rates"
ATAL B.S., REMDE J.R.
IEEE Proc. Int. Conf. on ASSP, 1982, pages 614-617

[4] "Predictive Coding of Speech Signals and Subjective Error Criteria"
ATAL B.S., SCHROEDER M. R.
IEEE Proc. Int. Conf. on ASSP, 1978, pages 573-576

[5] "Adaptive Predictive Coding of Speech Signals"
ATAL B.S., SCHROEDER M. R.
Bell. Syst. Tech., 1970, pages 1973-1986

[6] "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave"
ATAL B.S., HANAUER S.L.
Journ. Acoust. Soc. Ameri.,1971, pages 201-212

[7] "High-Quality Speech at Low bit Rates Multi-pulse and Stochastically Excited Linear Predictive Coders"
ATAL B.S.
IEEE Proc. Int. Conf. on ASSP, 1986, pages 1681-1684

[8] "An Analysis of Objectively Computable Measures for Speech quality Testing"
BARNWELL T.P., QUACKENBUSH S.R.
IEEE Proc. Int. Conf. on ASSP, 1982, pages 996-999

[9] "A Robust Adaptative Transform Coder for 9.6 kbps Speech Transmission"
BERGERON R.E., GOLDBERG A., KWON S., MILLER M.
IEEE Proc. Int. Conf. on ASSP, 1980, pages 344-347

[10] "Efficient Computation and Encoding of the Multi-Pulse Excitation for LPC"
BEROUTI M., GARTEN H., KABAL P., MERMELSTEIN P.
IEEE Proc. Int. Conf. on ASSP, 1984, pages 1011-1014

[11] "A Comparison of Orthogonal Transformation for Digital Speech Processing"
CAMPARELLA S., ROBINSON G.S.
IEEE Trans. on Commun. Tech., 1971, pages 1045-1049

[12] "Digital Coding of Speech in Subbands"
CROCHIERE R.E., WEBBER S. A., FLANAGAN J.L.
Bell. Syst. Tech., 1976, pages 1069-1085

[13] "A Digital Signal Processor Implementation of the CCITT 32 kbits/s ADPCM Algorithm"
CHARBONNIER A., MAITRE X., PETIT J.P.
ICC-85, Chicago, June 1985

[14] "20 Listes de Dix Phrases Phonétiquement Equilibrées"
COMBESCURE P.
CMC/TSS-CNET LA/A-22301 Lannion

[15] "Codage Numérique des Signaux"
COMBESCURE P., MATHIEU M.
Echo des Recherches, 3 ème trimestre1985, Pages 13-22

[16] "Rate-Distorsion Theory and Application"
DANISSON D.
Proc. IEEE 1972, pages 800-808

[17] "Acoustic Theorie of Speech Production"
FANT G.
Mouton-The Hague, 1970, Paris

[18] "Speech Coding"
FLANAGAN J.L., SCHROEDER M.R., ATAL B.S., CROCHIERE R.E., JAYANT N.S., TRIBOLET J.M.
IEEE Trans. on Comm., april 1979, pages 711-737

[19] "Techniques de Codage de la Parole à Débit Moyen (5 à 16 kbits/s)"
GALAND Cl., ESTEBAN D., MENEZ J.
l'Onde Electronique, septembre 1981, pages 41-53

[20] "Distance Measure for Speech Processing"
GRAY A.H., MARKEL J.D.
IEEE Trans. on ASSP, 1976, pages 380-391

[21] "Full Search and Tree Searched Vector Quantization of Speech Waveforms"
GRAY R.M.
IEEE Proc. Int. Conf. on ASSP, 1982, pages 593-596

[22] "Vector Quantization"
GRAY R.M.
IEEE Proc. on Comm., 1984, pages 4-29

[23] "Etude Statistique des Phonèmes et Diphonèmes dans le Français Parlé"
HATON J.P., LAMOTTE M.
Revue d'Acoustique, 1971, pages 258-262

[24] "RELP-Vocoding with Uniform and Non-Uniform Down-Sampling"
HEDELIN P.
IEEE Proc. Int. Conf. on ASSP, 1983, pages 1320-1323

[25] "Coding Speech at Low Bit Rates"
N. S. JAYANT
IEEE Spectrum, Août 1986, pages 58, 63

[26] "Multiple Stage Vector Quantization for Speech Coding"
JUANG B.H., GRAY A.H.
IEEE Proc. Int. Conf. on ASSP, 1982, pages 597-600

[27] "Comparison of Objective Speech Quality Measures for Voiceband Codecs"
KITAWAKI N., ITOH K., HONDA M., KAKEHI K.
IEEE Proc. Int. Conf. on ASSP, 1982, pages 1000-1003

[28] "An overview of Digital Techniques for Processing Speech Signals"
KUNT M., HUGLI H.
N Advance Studies Institute (Bonas,Gers, France), 1984

[29] "Codage Différentiel de la Parole: Algorithme de Prédiction Adaptative et Performances"
LEGUYADER A., GILLOIRE A.
Ann. Télécom., 1983, pages 381-398

[30] "Phrases Françaises Phonétiquement Equilibrées"
LENNIG M., MERMELSTEIN P.
JEP , 1980, pages 210-211

[31] "Least Squares quantization in PCM"
LLOYD S.P.
IEEE Trans. on Inf. Theory, 1957, pages 129-137

[32] "Linear Prediction: A Tutorial Review"
MAKHOUL J.
IEEE Trans. on ASSP, 1975, pages 561-580

[33] "Implementation and Comparison of Two Transformed Coefficient Scalar Quantization Methods"
MARKEL J.D., GRAY A.H.
IEEE Trans. on ASSP, 1980, pages 575-583

[34] "Digital Speech Technoly - Telecommunication Applications"
MARSH D.J.
Speech Tec.,mars 1987, pages 76-79

[35] "Minimum Mean Squared Error Quantization in Speech PCM and DPCM Systems"
PAEZ M.D., GLISSON T.H.
IEEE Trans. on Commun., 1972, pages 225-230

[36] "Product Code Vector Quantizers for Waveform and Voice Coding"
SABIN M.J., GRAY R.M.
IEEE Trans. on ASSP, 1984, pages 474-488

[37] "Tables of the Spheres Packings and Spherical Codes"
SLOANE N.J.A.
IEEE Trans. on Inf. Theory, 1981, pages 327-338

[38] "Efficient Procedures for Finding the Optimal Innovation in Stochastic Coders"
TRANCOSO I.M., ATAL B.S.
IEEE Proc. Int. Conf. on ASSP, 1986, pages 2375-2378

[39] "Analysis/Synthesis Frame Work for Transform Coding of Speech"
TRIBOLET J.M., CROCHIERE R.E.
IEEE Proc. Int. Conf. on ASSP, 1979, pages 81-84

[40] "Frecency Domain Coding of Speech"
TRIBOLET J.M., CROCHIERE R.E.
IEEE Trans. on ASSP, 1979, pages 336-339

[41] "A Comparison of the Performances of Four Low-Bit-Rate Speech Waveform Coder"
TRIBOLET J.M., NOLL P., Mc DIRMOTT B.J., CROCHIERE R.E.
Bell Syst. Tech. J., 1979, pages 699-712

[42] "Adaptative Transform Coding of Speech Signals"
ZELINSKI R., NOLL P.
IEEE Trans. on ASSP-25, pages 299-309

[43] "Approches to Adaptative Transform Speech Coding at Low Bit Rates"
ZELINSKI R., NOLL P.
IEEE Trans. on ASSP, 1979, pages 89-95

Thèses:

[44] "Transmission de la Parole a Faible Débit par Vocodeur à Classification"
DABOUZ M.
1984, ENST Paris

[45] "Time-Domain Coding of (Near) Toll Quality Speech at Rates Below 16 kb/s"
KROON P.
1985, Delft university

[46] "Conception d'un Vocodeur à Excitation Vocale à 9600 bits/seconde"
MOURIKIS C.
1979, ENST Paris

Ouvrages:

[47] "Computer Speech Processing"
FALLSIDE F., WOODS W.A.

[48] "Speech Analysis, Synthesis and Perception"
FLANAGAN J.L.
Springer-Verlag, New-York,1976

[49] "Adaptive Signal Processing"
WIDROW B., STEARNS S.D.

CHAPITRE II

STRUCTURE EFFICACE D'UNE FAMILLE DE DE CODEURS HYBRIDES TEMPORELS

CHAPITRE II

II.1 INTRODUCTION:

Dans le chapitre précédent, nous avons décrit différents codeurs, en particulier les codeurs hybrides qui offrent une qualité sub-téléphonique pour un débit compris entre 6 à 12 kbits/s.

A l'analyse, une structure intéressante de tels codeurs est fournie par la combinaison des méthodes APC (Adaptative Prédictive Coding), avec les méthodes de minimisation globales par analyse/synthèse [3]. Ils encodent le signal de parole en 2 étapes successives:

1: Un filtre linéaire évolutif dans le temps estime l'enveloppe spectrale du signal de parole. Ce filtre peut être un prédicteur à court terme ou la combinaison d'un prédicteur à court terme et à long terme.

2: La séquence d'excitation optimale pour ce filtre est déterminée de manière à minimiser le critère d'erreur retenu. Ce processus de modélisation est amélioré en incorporant une fonction de pondération. Elle a pour rôle de distribuer de manière non-uniforme le spectre du signal d'erreur.

A la synthèse, la procédure de décodage est beaucoup plus simple. La séquence d'excitation excite le filtre de synthèse qui reconstruit les échantillons du signal de parole. On peut noter que le traitement à la synthèse est complètement indépendant de la procédure d'analyse.

Les performances d'un tel codeur sont fonction:

- du critère de minimisation retenu
- des caractéristiques des filtres de modélisation
- de la fonction de pondération
- de la procédure de recherche de l'excitation
- des hypothèses sur la nature du signal d'excitation

Dans ce chapitre, nous nous proposons de présenter en détail les trois premiers points. La nature de l'excitation et la procédure de recherche définissent les différents codeurs de la famille: le codeur à excitation multi-impulsionnelle, le codeur à excitation par code, le codeur à excitation optimale par code et le codeur à excitation multi-impulsionnelle vectorielle. Ceux-ci seront décrits en détail dans les chapitres III, IV respectivement.

II.2 DEFINITION DU CRITERE DE MINIMISATION:

De manière générale, plusieurs critères de minimisation peuvent être incorporés dans les procédures de minimisation, moyennant une pondération adéquate. En pratique, on se limite à un critère unique de façon à réduire la complexité du traitement.

La grande majorité des techniques d'estimation ou de modélisation adaptatives choisissent un critère des moindres carrés. Il se définit comme la somme quadratique du critère d'erreur. Ce critère, que l'on appelle également norme L2 est intéressant à plus d'un titre.

- Il se prête bien aux calculs, car les paramètres inconnus sont solutions d'un système linéaire. Ce dernier est obtenu par dérivation du critère par rapport aux paramètres inconnus.

- Si les hypothèses statistiques de l'erreur à minimiser sont gaussiennes, l'estimation au sens des moindres carrés correspond également au maximum de vraissemblance.

- Il s'interprète directement comme la puissance du signal d'erreur à minimiser.

On peut noter cependant, que plusieurs auteurs [7,21] ont également appliqué le critère de moindre valeur absolue, appelée norme L1, à l'estimation du filtre et à la modélisation de l'excitation car ses propriétés sont adaptées aux bruits impulsifs. Ce critère conduit à l'utilisation d'algorithmes de simplex peu adaptés à l'intégration.

II.3 STRUCTURES DES FILTRES:

Le modèle de représentation général de tels prédicteurs est donné par le modèle ARMA évolutif. La relation entre l'entrée et la sortie du filtre est définie par l'équation récurrente:

$$s_n = G\Big(e_n + \sum_{j=1}^{MP'} d_j . e_{n-j} \Big) - \sum_{i=1}^{MP} a_i . s_{n-i} \qquad (2.1)$$

ou, en prenant la transformée en z

$$\frac{S(z)}{E(z)} = G \, \frac{1 + \sum_{j=1}^{MP'} d_j . z^{-j}}{1 + \sum_{i=1}^{MP} a_i . z^{-i}} \qquad (2.2)$$

Ce modèle a été étudié par GRENIER [10,11] mais son estimation est complexe. Aussi se limite-t-on de manière générale à la modélisation de filtres tout pôle dits filtres AR, sachant que tout filtre ARMA d'ordre fini peut être approximé par un filtre AR d'ordre plus élévé. Il est défini par la relation récurrente:

$$s_n = G\, e_n - \sum_{i=1}^{MP} a_i . s_{n-i} \qquad (2.3)$$

ou, en prenant la transformée en z

$$\frac{S(z)}{E(z)} = H(z) = \frac{1}{A(z)} = \frac{G}{1 + \sum_{i=1}^{MP} a_i . z^{-i}} \qquad (2.4)$$

La réalisation d'un filtre tout zéro A(z) ou tout pôle 1/A(z) peut être envisagée dans sa forme directe proposée par OPPENHEIM et SCHAFER [28] ou dans sa forme en treillis proposée par GRAY et MARKEL [9,27] et MAKHOUL [19]. La forme directe correspond à la transposition des équations différentielles décrivant la fonction de transfert. En revanche, le filtre en treillis à deux multiplieurs est la représentation canonique de la fonction de transfert. Il permet de réaliser la fonction de transfert des filtre AR et MA par la mise en cascade de cellules à deux entrées et deux sorties. En notant $E_p^+(t)$ et $E_p^-(t)$ les erreurs "progressives" et "rétrogrades" et K_{p+1} le coefficient PARCOR (coefficient de corrélation partielle) de la cellule p, les relations:

$$\left.\begin{aligned} E^+_{p+1}(n) &= E^+_p(n) + K_{p+1}\, E^-_p(n-1) \\ E^-_{p+1}(n) &= E^-_p(n-1) + K_{p+1}\, E^+_p(n) \end{aligned}\right| \quad \text{pour } p = 0 \text{ à } MP-1 \qquad (2.5)$$

$$E^+_0(n) = E^-_0(n) = s_n; \quad E^+_{MP}(n) = r_n$$

permettent de réaliser le filtre d'analyse A(z) et les équations:

$$\left.\begin{aligned} E^+_{p-1}(n) &= E^+_p(n) - K_p\, E^-_{p-1}(n-1) \\ E^-_p(n) &= E^-_{p-1}(n-1) + K_p\, E^+_{p-1}(n) \end{aligned}\right| \quad \text{pour } p = MP-1 \text{ à } 0 \qquad (2.6)$$

$$E^+_{MP}(n) = r_n; \quad E^+_0(n) = E^-_0(n-1) = s_n$$

le filtre de synthèse 1/A(z).

Les filtres dans leur forme directe présentent plusieurs inconvénients, en particulier leur grande sensibilité à la précision des calculs et à la quantification qui peut provoquer l'instabilité du filtre. En revanche, les filtres en treillis se prêtent bien aux traitements en précision finie. De plus, leurs coefficients permettent d'utiliser des techniques de quantification très efficaces qui ne mettent pas en question la stabilité du filtre. Le filtrage d'un échantillon à travers un filtre en treillis nécessite le double de multiplications et d'additions qu'un filtre direct.

La figure ci-dessous représente la forme directe et la forme en treillis des filtres tout pôle et tout zéro.

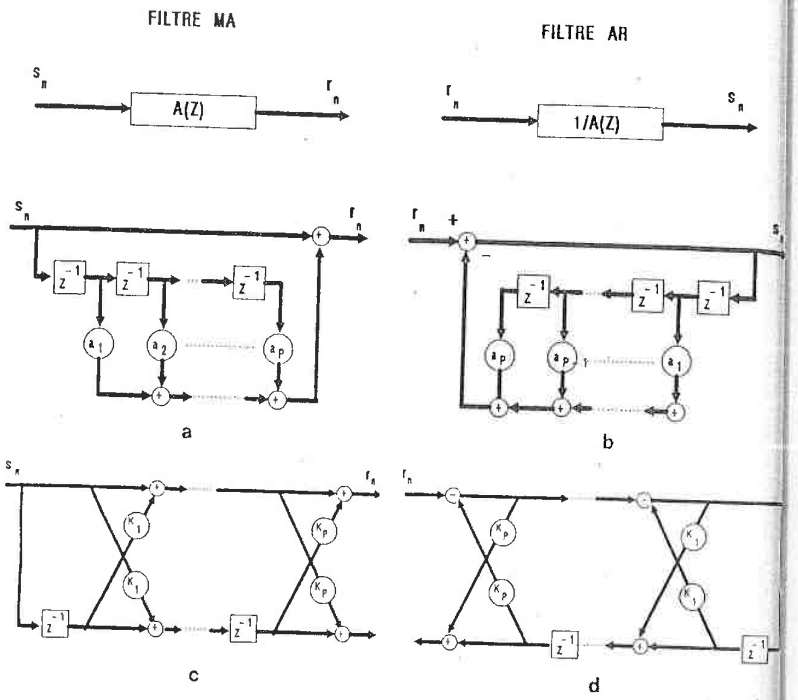


Figure 2.1: Structure des filtres a) Forme directe du filtre tout zéro; b)Forme directe du filtre tout pôle; c) forme en treillis du filtre tout zéro; d) forme en treillis du filtre tout pôle.

On constate également que le comportement de ces deux filtres n'est pas identique même si on fait abstraction des points cités précédemment. Dans des applications d'analyse-synthèse de la parole, où les coefficients des filtres sont renouvelés à intervalle régulier, la dégradation introduite par le filtrage en treillis est sensiblement supérieure à celle introduite par le filtrage direct. Néanmoins, cette dégradation (fig. 2.2) s'estompe dès que la précision des calculs n'est plus illimitée. Notre choix se porte finalement sur la forme en treillis car comme nous le verrons, une procédure efficace permet d'extraire les coefficients de ce type de filtre.

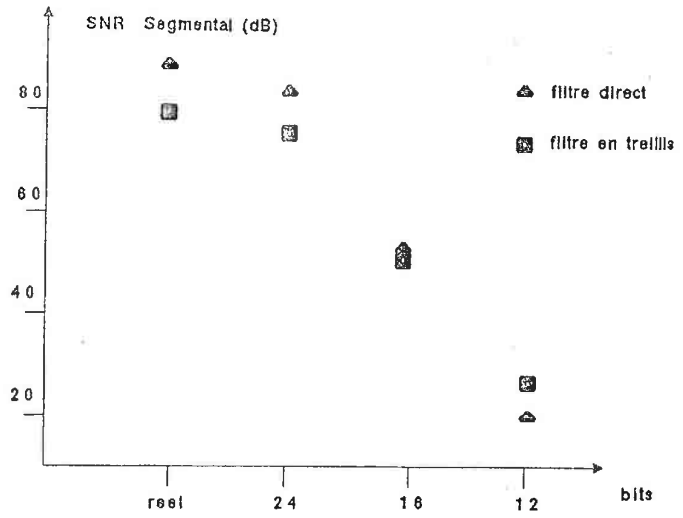


Figure 2.2: Evolution du rapport signal sur bruit segmental entre le signal original et le signal reconstruit par filtrage direct ou en treillis avec une réactualisation des filtres tous les 80, 128 et 160 échantillons.

II.4 DETERMINATION DU PREDICTEUR A COURT TERME:

Les redondances contenues dans le signal de parole peuvent être éliminées par des techniques de prédiction linéaire adaptative. Contrairement à l'analyse spectrale, elles offrent une représentation directe de l'onde de parole sous la forme d'un nombre limité de paramètres variables dans le temps. Ces paramètres se rapportent à la fonction de transfert du conduit vocal.

La modélisation par prédiction linéaire est très communément employée actuellement. Celle utilisée en traitement de la parole est une forme particulière de l'identification de processus qui ajuste un modèle linéaire autorégressif au signal. Elle est particulière du fait qu'aucune information sur l'entrée n'est disponible. On est réduit à admettre comme hypothèse que le signal d'entrée u_n est un bruit blanc à moyenne nulle.

La prédiction linéaire est également la généralisation du principe retenu dans les codeurs temporels tels que MIC différentiel ou MIC Prédictif Adaptatif que nous avons présentés dans le chapitre précédent. Un échantillon de signal s_n est prédit approximativement à partir de l'échantillon précédent s_{n-1}.

$$s_n = e_n + a.s_{n-1} \qquad (2.7a)$$

Dans le cas de la prédiction linéaire celle-ci porte sur la combinaison linéaire de plusieurs échantillons précédents, typiquement 8 à 16.

$$s_n = e_n + \sum_{i=1}^{MP} a_i.s_{n-i} \qquad (2.7b)$$

L'aspect, certainement le plus intéressant de la prédiction linéaire est qu'elle associe les avantages d'une analyse temporelle et ceux d'une analyse fréquentielle. L'erreur de prédiction, encore appelée résiduel, permet de retrouver toute information temporelle liée à des phénomènes transitoirs. Quant au prédicteur, il est représentatif de l'enveloppe spectrale moyenne à long terme du signal modélisé.

Les principaux auteurs qui ont contribué à l'application de la prédiction linéaire au signal de parole sont ATAL [1], ITAKURA [13] ainsi que MARKEL et GRAY [26]. Il existe une grande variété d'algorithmes que l'on peut classer en méthodes recursives en ordre d'une part et en temps d'autre part [12].

Comme nous l'avons déjà mentionné dans le paragraphe 3 du chapitre I, les méthodes récursives en temps réduisent le décalage temporel, mais leur modélisation est moins efficace. C'est pour cette raison que nous avons retenu des méthodes récursives en ordre, que nous allons décrire maintenant. L'intervalle sur lequel porte la modélisation varie entre 10 et 30 ms, compte-tenu de la faible stationnarité du signal de parole. Aussi les coefficients de prédiction sont renouvelés périodiquement au même rythme de 10 à 30 ms. Une revue complète sur la prédiction linéaire est proposée par MAKHOUL [19].

L'enveloppe spectrale à court terme d'un segment de signal stationnaire peut être approximée par la réponse impulsionnelle d'un filtre tout pôle. Les coefficients du filtre sont déterminés de façon à minimiser, au sens des moindres carrés, le signal résiduel r_n. Ce signal est obtenu en filtrant le signal de parole à travers le filtre A(z). Il est défini par:

$$r_n = s_n - \hat{s}_n = s_n - \sum_{i=1}^{MP} a_i.s_{n-i} \qquad (2.8)$$

L'erreur quadratique est donnée par la relation:

$$\epsilon = \sum_{n0}^{n1} r_n^2 = \sum_{n0}^{n1} (s_n - \hat{s}_n)^2 = \sum_{n0}^{n1} (s_n - \sum_{i=1}^{MP} a_i.s_{n-i})^2 \qquad (2.9)$$

La minimisation de ϵ est obtenue lorsque toutes les dérivées partielles par rapport aux coefficients a_i sont nulles:

$$\frac{\delta \epsilon}{\delta a_i} = 0 \qquad \text{pour } i = 1 \text{ à } P \qquad (2.10)$$

posons:

$$\phi_{i,j} = \sum_{n0}^{n1} s_{n-i}.s_{n-j} \qquad (2.11)$$

soit alors:

$$\sum_{i=1}^{P} a_i \phi_{i,j} = \phi_{0,j} \qquad \text{pour } j = 1 \text{ à } MP \qquad (2.12)$$

le système de P équations linéaires à MP inconnues.

L'erreur minimum ϵ_P à l'ordre MP est:

$$\epsilon^P = \sum_{n0}^{n1} s_n^2 - \sum_{i=1}^{MP} a_i.(\sum_{n0}^{n1} s_n.s_{n-i}) \qquad (2.13)$$

n_0 et n_1 sont les limites de sommation dont le choix conduit à des méthodes de covariance ou de corrélation. Les coefficients a_i ne sont pas exactement les mêmes dans les deux méthodes, comme nous allons le voir maintenant.

II.4.1 LA METHODE PAR COVARIANCE:

Cette méthode se base sur les hypothèses suivantes:

- le signal est défini par MP+N échantillons s_n où P est l'ordre du prédicteur et N la quantité d'échantillons prédits.

- un échantillon peut être prédit approximativement par ses P échantillons précédents. Ceci est valable exclusivement sur les N échantillons successifs.

- l'erreur quadratique totale (ou moyenne) entre le signal et le modèle est minimisée exclusivement sur les N échantillons.

On choisit n0 = 0 et n1 = N-1

Alors:

$$D_{i,j} = \sum_{n=0}^{N-1} s_{n-i}.s_{n-j} \qquad (2.14)$$

La minimisation est solution du système d'équations:

$$\sum_{i=1}^{MP} a_i D_{i,j} = D_{0,j} \qquad \text{pour } j = 1 \text{ à } P \qquad (2.15)$$

$$\epsilon^P = D_{0,0} - \sum_{i=1}^{MP} a_i.D_{0,i} \qquad (2.16)$$

Il s'écrit sous forme matricielle:

$$\begin{bmatrix} D_{1,1} & D_{1,2} & \cdots & D_{1,MP} \\ D_{2,1} & D_{2,2} & \cdots & D_{2,MP} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ D_{MP,1} & D_{MP,2} & \cdots & D_{MP,MP} \end{bmatrix} * \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_{MP} \end{bmatrix} = \begin{bmatrix} D_{1,0} \\ D_{2,0} \\ \cdot \\ \cdot \\ \cdot \\ D_{MP,0} \end{bmatrix} \qquad (2.17)$$

La résolution du système d'équation est obtenue par la méthode de CHOLESKY [4].

A l'intérieur d'un segment de signal, la méthode de covariance assume que le signal est non-stationnaire. Cette méthode ne garantit pas la stabilité du modèle. Toutefois celle-ci s'améliore lorsque la taille du segment analysé augmente.

Nous allons voir comment la méthode d'autocorrélation évite cette difficulté, tout en réduisant la quantité d'informations et le coût opératoire nécessaires à la résolution du système d'équation.

II.4.2 METHODE PAR AUTOCORRELATION:

Les hypothèses de cette méthode sont les suivantes:

- le signal est nul à l'extérieur du segment de signal considéré. Ceci est réalisé en multipliant le signal de parole par une fenêtre temporelle de largeur N, qui introduit une réduction de la résolution spectrale.

- tout échantillon de - infini à + infini peut être prédit approximativement à partir des P échantillons précédents.

- l'erreur quadratique totale entre le signal fenêtré et le modèle est minimisée de - infini à + infini.

Ces considérations donnent les équations normales d'autocorrélation.

On choisit n_0 = - infini et n_1 = + infini avec $s_n = 0$ si $0 < n < N-1$

Posons alors:

$$R_{|i-j|} = \sum_{-infini}^{+infini} s_{n-i}.s_{n-j} = \sum_{n=0}^{N-1-|i-j|} s_n.s_{n+|i-j|} \qquad (2.18)$$

d'où:

$$Ri = R-i \qquad (2.19)$$

On obtient alors le système d'équations:

$$\sum a_i R_{|i-j|} = R_j \qquad \text{pour } j = 1 \text{ à } MP \qquad (2.20)$$

$$\epsilon^P = R_0 - \sum_{i=1}^{MP} a_i.R_i \qquad (2.21)$$

Ce système s'écrit sous forme matricielle:

$$\begin{bmatrix} R_0 & R_1 & \cdots & R_{MP-1} \\ R_1 & R_0 & \cdots & R_{MP-2} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ R_{MP-1} & R_{MP-2} & \cdots & R_0 \end{bmatrix} * \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_{MP} \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \\ \cdot \\ \cdot \\ \cdot \\ R_{MP} \end{bmatrix} \qquad (2.22)$$

La matrice d'autocorrélation de dimension MPxMP est une matrice de TOEPLITZ particulière car elle est symétrique et tous les éléménts le long d'une diagonale sont identiques. Il suffit alors de MP+1 coefficients d'autocorrélation pour sa définition complète.

Compte-tenu des caractéristiques avantageuses de cette matrice, des méthodes rapides de résolution du système d'équations ont été trouvées. Parmi celles-ci, on peut noter une méthode itérative proposée par LEVINSON [17]. Elle utilise comme paramètres intermédiaires les coefficients de corrélation partielle K_i. Ces coefficients qui décrivent les filtres en treillis présentent également des propriétés remarquables. Aussi ces paramètres sont les plus aptes au codage et à la transmission.

Une autre méthode qui calcule directement les coefficients K_i a été proposée par LEROUX et GUEGUEN [16]. Elle n'utilise pas explicitement les paramètres ai, mais des variables intermédiaires qui sont en fait les coefficients d'intercorrélations entre l'entrée et la sortie des modèles d'ordre successif. Contrairement à la méthode décrite précédemment, celle-ci est très proche de la réalisation du modèle linéaire par une structure en treillis. On montre également que la dynamique de ces variables intermédiaires est bornée, ce qui permet une implantation aisée en virgule fixe.

CHOIX DE LA METHODE ET DES PARAMETRES:

La méthode de covariance ne permet pas d'obtenir de manière systématique un filtre stable. Ceci est un inconvénient majeur en codage. Notre choix s'est porté sur les méthodes par autocorrélation car elles garantissent la stabilité du filtre de synthèse. Notons toutefois que les résultats que donnent ces méthodes sont moins précis, compte-tenu du fenêtrage du signal original.

Des considérations liées aux propriétés des paramètres interviennent également. En effet, les coefficients K_i étant les plus aptes au codage la méthode de LEROUX et GUEGUEN [16] a été retenue pour estimer le prédicteur à court terme. Les filtres d'analyse et de synthèse ont une structure en treillis à deux multiplieurs.

Les valeurs standards des paramètres relevant de l'analyse LPC et du prédicteur à court terme sont précisées ci-dessous. Ces valeurs sont choisies en fonction des caractéristiques temporelles (signal de parole stationnaire sur 20ms) et spectrales (spectre entre 100 et 3400 Hz) du signal de parole:

N : fenêtre d'analyse LPC = 160
σ : facteur de préaccentuation = 0
MP: ordre du filtre de synthèse = 12

Le nombre total d'opérations par seconde, correspondant à l'analyse LPC est détaillé dans le tableau ci-dessous. Le facteur ϕ représente le nombre de fenêtres de modélisation LPC (de longueur N) par seconde.

Analyse LPC par Leroux/Gueguen	multi.,addit./s	div./s
autocorrélation du signal2)	(MP+1).N.ϕ	
transformation de Leroux/Gueguen	2.MP^2.ϕ	MI.ϕ
filtrage MA	2.MP.N.ϕ	
Complexité pour les valeurs standards	310400	600

Tableau 2.1: Complexité liée à l'analyse LPC

II.5 FONCTION DE PONDERATION:

En synthèse de la parole, certains défauts, comme le biais sur les coefficients des filtres ou encore une mauvaise définition de la source, nuisent à la qualité de la parole restituée. Le critère de minimisation des moindres carrés ne prend pas en compte les propriétés de la perception auditive. Il en résulte à l'écoute, un bruit de fond perceptible en dehors des plages formantiques.

L'intérêt de cette fonction de pondération est d'obtenir un masquage spectral de l'erreur. L'effet de masquage, mis en évidence par des expériences de psychoacoustique, permet de tolérer une erreur plus importante dans les régions spectrales à forte énergie. ATAL [2] réalise cette fonction de pondération sous la forme d'un filtre évolutif dérivé du filtre de synthèse. Il est caractérisé par la fonction de transfert suivante:

$$W(z) = \frac{A(z)}{A(z/\tau)} = \frac{1 + \sum_{j=1}^{MP} a_j . z^{-j}}{1 + \sum_{i=1}^{MP} a_i . z^{-i} . \tau^i} \qquad (2.23)$$

où τ représente le facteur perceptuel.

D'une manière très générale, le signal d'erreur e_n à minimiser est filtré à travers ce filtre perceptuel W(Z) de façon à produire un signal d'erreur perceptuel $\tilde{e}_n$ (fig. 2.3).

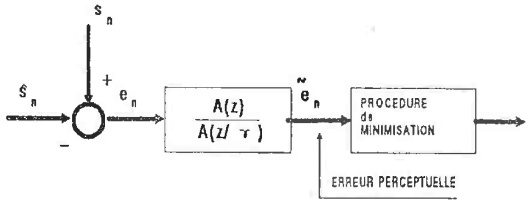


Figure 2.3: Forme générale d'un processus de modélisation avec fonction de pondération.

$$\tilde{E}(z) = E(z)\,W(z) = W(z)(S(z) - \hat{S}(z)) \qquad (2.24)$$

La valeur de τ qui est comprise entre 0 et 1 dépend de la désaccentuation que l'on souhaite apporter dans les régions formantiques. Etudions l'influence du facteur perceptuel sur le spectre de l'énergie du signal d'erreur pondéré.

Si $\tau = 1$: la pondération n'intervient pas car $W(z) = 1$.
Le spectre de l'énergie de l'erreur perpetuelle ϵ_{min} est plat, ce qui revient à blanchir le spectre de e_n (fig. 2.4a). Le bruit de codage est perçu dans les creux du spectre, c'est à dire entre les formants et dans les hautes fréquences.

Si $\tau = 0$: $W(z) = A(z)$, (fig. 2.4b).
Les spectres de l'énergie de l'erreur perceptuelle et de la réponse impulsionnelle du filtre $1/A(z)$ sont identiques à ϵ_{min} près.

Valeur retenue pour τ: Une bonne répartition de l'énergie de l'erreur perceptuelle entre les basses et les hautes fréquences est obtenue pour des valeurs de τ comprise entre 0.75 et 0.9, valeurs pour lesquelles le rapport signal sur bruit décroît légèrement en fonction de la fréquence (fig. 2.4c).

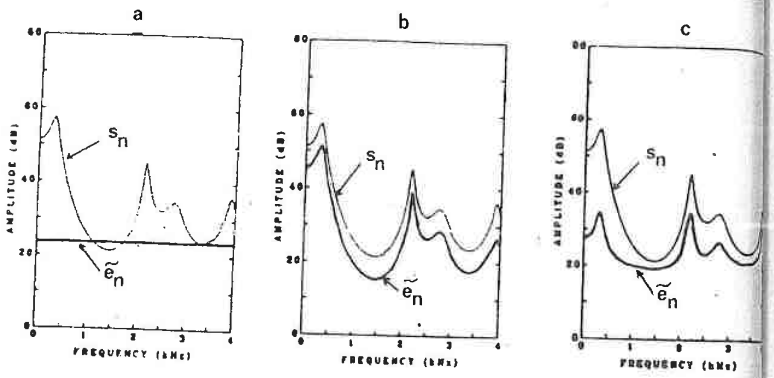


Figure 2.4: Spectre de l'énergie de l'erreur perceptuelle a)τ = 1, b)τ =0 et c)τ = 0,85

Soit h_n la réponse impulsionnelle du filtre de synthèse $1/A(z)$. Celle du filtre de pondération $F(z)$ s'écrit:

$$F(z) = \sum_{n=0}^{MA} h_n z^{-n} \tau^n = \sum_{n=0}^{MA} f_n z^{-n} \qquad (2.25)$$

Le filtre de pondération $F(z)$ présente une autre caractéristique particulière. En effet, τ étant compris entre 0 et 1, l'enveloppe de la réponse impulsionnelle de $F(z)$ décroît rapidement suivant la loi τ^n.

Donc f_n est proche de zéro pour $n \ll N$; N étant la taille de de la fenêtre de minimisation. La figure ci-dessous illustre l'influence du facteur perceptuel τ sur la durée de la réponse impulsionnelle de $F(z)$.

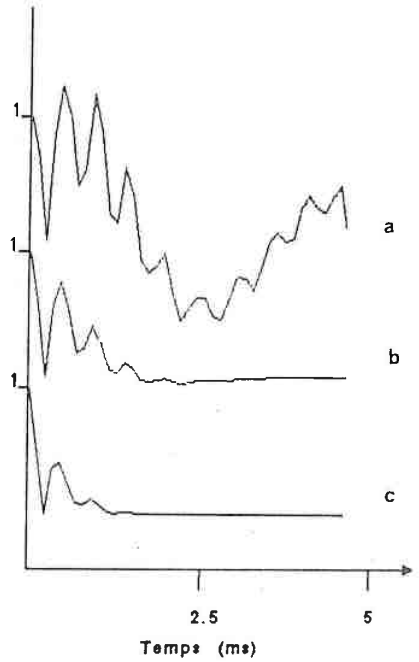


Figure 2.5: Influence du facteur perceptuel τ sur la durée de la réponse impulsionelle du filtre de pondération $F(z)$, a) τ = 1, b) τ = 0.9, τ =0.85

Soit MA la durée de f_n, on peut conclure que deux échantillons du signal d'erreur perceptuelle décalés de plus de MA échantillons ne sont plus corrélés. Cette propriété permet de réduire l'intervalle de sommation du produit de convolution.

II.6 DETERMINATION DU PREDICTEUR A LONG TERME:

Nous venons de voir dans le paragraphe II.4, que la prédiction à court terme, que nous avons retenue, est fondée sur la fonction d'autocorrélation. Il en est de même pour le prédicteur à long terme. L'autocorrélation est maximale en zéro et décroît ensuite. Or, dans le cas de sons voisés, la périodicité du signal d'excitation se traduit, au niveau de la fonction d'autocorrélation, par des pics décalés dont l'écart correspond à la période des impulsions. Ce décalage varie typiquement de 27 à 160 échantillons pour des "pitchs" respectifs de 300 à 50 Hz (échantillonnage à 8 kHz). On peut donc appliquer une prédiction à long terme qui s'écrit sous la forme générale:

$$r_n = e_n + \sum_{i=-L}^{L} b_i . r_{n-P+i} \qquad \text{soit en prenant la transformée en z:} \quad (2.26)$$

$$\frac{R(z)}{E(z)} = \frac{1}{1 - B(z)} = \frac{1}{1 - \sum_{i=-L}^{L} b_i . z^{-P+i}} \qquad (2.27)$$

où P représente le décalage du prédicteur à long terme qui est compris entre Pmin et Pmax. . b_i sont les coefficients de prédiction à long terme

Le nombre de coefficients du prédicteur à long terme vaut typiquement 1 ou 3.

Voyons maintenant les structures que peut prendre le prédicteur à long terme, ainsi que ses méthodes de calcul.

II.6.1 STRUCTURES DU PREDICTEUR A LONG TERME:

Le modèle de synthèse est schématisé sur la figure 2.6. Il est constitué des filtres modélisant respectivement la périodicité de la source et celle du conduit vocal. Il faut noter, que l'ordre dans lequel sont placés ces prédicteurs n'est pas quelconque, même si en théorie ils peuvent l'être. Ceci est lié à la faible stabilité du prédicteur à long terme. Lorsqu'il est à coefficient unique, la condition nécessaire et suffisante de stabilité est que le coefficient de prédiction soit strictement inférieur à 1 en module. Les pôles de ce filtre qui correspondent à la racine $p^{ième}$ de b sont très proches du cercle unité. Pour un filtre à 3 coefficients les conditions de stabilité sont plus compliquées. A ce propos, RAMACHANDRAM et KABAL [23] proposent des procédures de test de la stabilité ainsi que des procédures de stabilisation des filtres de prédiction à long terme. Celles-ci sont dérivées de la procédure générale de Shur-Cohn [25], dont l'application est trop complexe. Les calculs se faisant en précision finie ce filtre risque d'être instable même si b est strictement inférieur à 1 en module. Il est donc préférable de placer à la synthèse le prédicteur à long terme en amont du prédicteur à court terme car le signal d'exitation multi-impulsionnelle est peut énergétique. De plus le filtre 1/A(z) étant toujours stable et du type passe-bas, celui-ci absorbe tout risque d'instabilité.

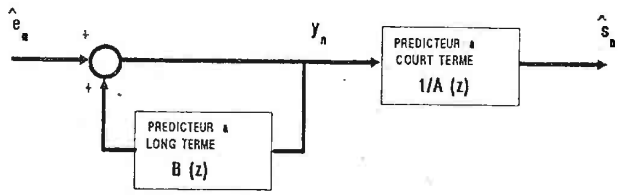


Figure 2.6: Structure du filtre de synthèse avec prédiction à long terme

A ce modèle de synthèse correspond le modèle symétrique d'analyse. Il est décrit dans la figure 2.7. La procédure de modélisation de l'excitation est placée en aval du filtre de déconvolution à long terme B(z) et porte sur l'erreur totale de prédiction e_n (à court terme et à long terme).

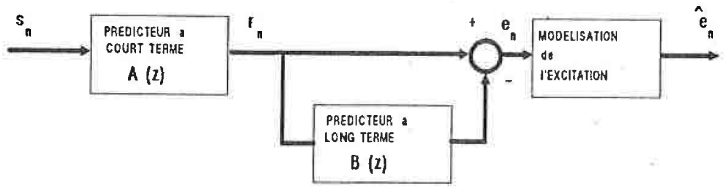


Figure 2.7: Structure de l'analyse avec prédicteur à long terme.

La structure du prédicteur à long terme peut être envisagée selon un autre schéma à l'analyse. En effet, au modèle de synthèse de la figure 2.6, il est possible d'associer le schéma d'analyse donné par la figure 2.8.

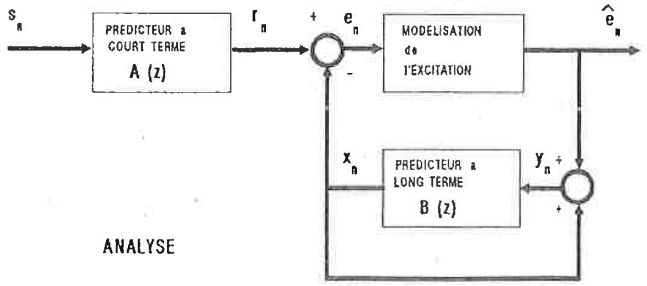


Figure 2.8: Structure de l'analyse avec prédicteur à long terme bouclé.

L'excitation n'est plus modélisée à la suite des deux déconvolutions, mais à l'intérieur de la boucle de prédiction à long terme. On passe ainsi d'un système de codage transversal (en anglais feedforward predictive coder) à un système de codage bouclé (en anglais feedback predictive coder). La formulation du prédicteur à long terme est:

$$r_n = e_n + \sum_{i=-L}^{L} b_i . y_{n-M+i} \qquad (2.28)$$

où

y_n est le signal d'excitation synthétique après modélisation.

Plusieurs auteurs comme LANCON [26] (dans le cas d'un codeur à excitation multi-impulsionnelle) mais également CHEN et GERSHO [6] (dans le cas d'un codeur à excitation par code) ont mis en application cette structure bouclée. Dans ce cas de figure, les traitements qui modélisent et encodent l'excitation jouent le rôle d'un quantificateur vectoriel adaptatif qui remplace le signal d'erreur par une séquence d'excitation particulière. Du même coup, les signaux d'excitation e_n sont identiques à l'analyse comme à la synthèse, garantissant ainsi la qualité la meilleure. Le décalage P du prédicteur à long terme peut correspondre à la période du fondamental ou à celle d'un de ses harmoniques.

Une telle approche impose quelques contraintes notamment sur la valeur minimale du décalage que peut prendre le prédicteur à long terme. Ce décalage doit être supérieur à la taille de la fenêtre de modélisation de l'excitation. Car la procédure de calcul du prédicteur à long terme et celle de l'excitation reposent sur des méthodes globales qui nécessitent la connaissance à priori des signaux sur tout l'intervalle de minimisation.

Voyons maintenant, comment calculer les paramètres du prédicteur à long terme.

II.6.2 DETERMINATION DU PREDICTEUR A LONG TERME:

De façon analogue à l'analyse par prédiction linéaire à court terme, les coefficients b_i sont déterminés par minimisation au sens des moindres carrés de l'erreur entre le signal original et son estimation à partir du signal décalé.

$$r_n = e_n + \sum_{i=-L}^{L} b_i . r_{n-M+i} \qquad (2.29)$$

METHODE SOUS-OPTIMALE:

Une solution sous-optimale est obtenue en minimisant l'erreur quadratique ϵ en deux étapes de façon à déterminer successivement le décalage P puis les coefficients de prédiction à long terme b_i.

$$\epsilon = \sum_{n\in\mathcal{P}'} e_n^2 = \sum_{n\in\mathcal{P}'} \left(r_n - \sum_{i=-L}^{L} b_i . r_{n-M+i} \right)^2 \qquad (2.30)$$

Minimiser l'erreur ϵ par rapport aux paramètres P et b_i du prédicteur à long terme revient à annuler la dérivée partielle de ϵ par rapport aux bi:

$$\frac{\delta \epsilon}{\delta b_i} = 0 \qquad \text{pour } i = -L \text{ à } L \qquad (2.31)$$

posons:

$$R'_i = \sum_{n\in\mathcal{P}'} r_n . r_{n-i} \qquad (2.32a)$$

et

$$R_i = \sum_{n\in\mathcal{P}'} r_{n-i} . r_{n-i} \qquad (2.32b)$$

pour un filtre à 3 coefficients, les b_i sont solutions du système:

$$\begin{vmatrix} R_P & R_{P+1} & R_{P+2} \\ R_{P+1} & R_P & R_{P+1} \\ R_{P+2} & R_{P+1} & R_P \end{vmatrix} * \begin{vmatrix} b_{-1} \\ b_0 \\ b_1 \end{vmatrix} = \begin{vmatrix} R'_{P-1} \\ R'_P \\ R'_{P+1} \end{vmatrix} \qquad (2.33)$$

pour un prédicteur à 1 coefficient, le calcul de b se réduit à:

$$b = \frac{R'_P}{R_P} \qquad (2.34)$$

La détermination de b_i est fonction du décalage P. Il faut donc préalablement calculer P. Dans le cas d'un coefficient unique, en substituant b par sa solution dans l'expression de ϵ, on obtient:

$$\epsilon(p) = \sum_{n\in\mathcal{P}'} r_n^2 - \frac{\left(\sum_{n\in\mathcal{P}'} r_n \, r_{n-p} \right)^2}{\sum_{n\in\mathcal{P}'} r_{n-p}^2} \qquad (2.35)$$

ϵ est minimum lorsque le second terme est maximum. La valeur optimale de P est définie par l'indice p qui maximise l'autocorrélation normalisée:

$$P = p \text{ pour } \theta_p \text{ maximum soit } \theta_p = \frac{\left(\sum_{n\in\mathcal{P}'} r_n \, r_{n-p} \right)^2}{\sum_{n\in\mathcal{P}'} r_{n-p}^2} \qquad (2.36)$$

p variant entre Pmin et Pmax.

La première étape consiste donc à déterminer P à l'aide de l'expression 2.36. Lorsque P est connu ,le ou les coefficients de prédiction b_i sont respectivement solutions des relations 2.33 ou 2.34.

METHODE OPTIMALE:

Dans le cas du prédicteur bouclé, cette méthode de calcul n'est pas optimale. DYMARSKI [20] a proposé une méthode de calcul des paramètres du prédicteur à long terme en minimisant le signal d'erreur pondérée entre le signal perceptuel original p_n et le signal perceptuel synthétique q_n.

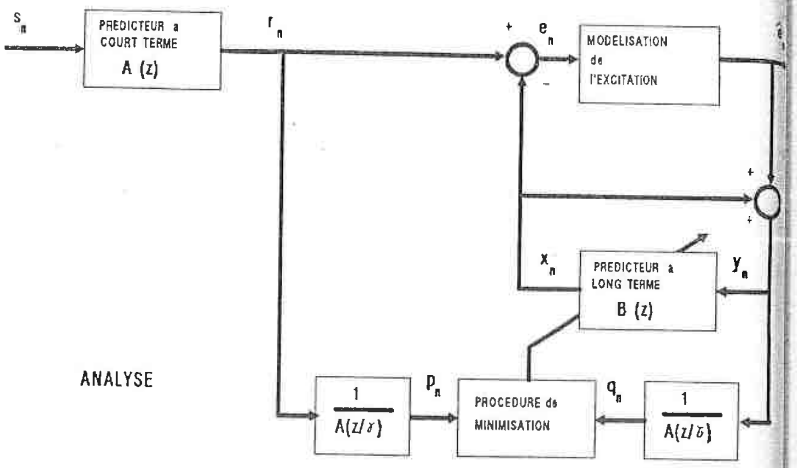


Figure 2.9: Calcul optimal des paramètres du prédicteur à long terme proposé par DYMARSKI.

Cette approche est optimale dans le sens où:

- l'introduction d'un filtre de pondération permet de bénéficier de l'effet de masquage au niveau du prédicteur à long terme.

- la prédiction à long terme est exacte car ses paramètres sont déterminés de manière à minimiser l'erreur pondérée entre le signal original et le signal synthétique.

Minimiser l'erreur perceptuelle $\tilde{e}_n$ entre le signal original et le signal synthétique revient à minimiser l'erreur de prédiction à long terme e_n pondérée par le filtre $1/A(z/\tau)$.

$$\tilde{E}(z) = (S(z) - \hat{S}(z))W(z) = P(z) - Q(z) = (R(z) - Y(z)) \frac{1}{A(z/\tau)} \qquad (2.37)$$

soit:

f_n : la réponse impulsionnelle du filtre $1/A(z/\tau)$
p_n : le signal perceptuel original défini par

$$p_n = \Sigma\, r_n . f_{n-i} \qquad (2.38)$$

q_n : le signal perceptuel synthétique défini par

$$q_n = \Sigma\, y_n\, f_{n-i} \qquad (2.39)$$

On minimise donc:

$$\epsilon = \sum_{n\in\mathcal{P}} (p_n - q_n)^2 = \sum_{n\in\mathcal{P}} \left(\sum_{j=n-p} (r_n - \sum_{i=-L}^{L} b_i . y_{p+i}) . f_{n-j} \right)^2 \qquad (2.40)$$

Comme pour la solution sous-optimale les paramètres du prédicteur à long terme sont déterminés en deux étapes en annulant la dérivée partielle de ϵ par rapport aux b_i:

$$\frac{\delta \epsilon}{\delta b_i} = 0 \quad \text{pour } i = -L \text{ à } L \qquad (2.40)$$

posons:

$$R'_i = \Sigma\, p_n . q_{n-i} \qquad (2.41)$$

et

$$R_i = \Sigma\, q_{n-i} . q_{n-i} \qquad (2.42)$$

Les valeurs optimales des coefficients b_i pour un décalage P sont solutions de:

$$\begin{vmatrix} R_P & R_{P+1} & R_{P+2} \\ R_{P+1} & R_P & R_{P+1} \\ R_{P+2} & R_{P+1} & R_P \end{vmatrix} * \begin{vmatrix} b_{-1} \\ b_0 \\ b_1 \end{vmatrix} = \begin{vmatrix} R'_{P-1} \\ R'_P \\ R'_{P+1} \end{vmatrix} \qquad (2.43)$$

dans le cas d'un prédicteur à 3 coefficients. Pour un prédicteur à 1 coefficient, le calcul de b se réduit à:

$$b = \frac{R'_P}{R_P} \qquad (2.44)$$

La valeur optimale du décalage est celle qui minimise 3.35 ce qui revient à maximiser:

$$P = p \text{ pour } \theta_p \text{ maximum soit } \theta_p = \frac{\left(\sum_{n \neq 0} p_n . q_{n-p} \right)^2}{\sum_{n \neq 0} q_{n-p}^2} \qquad (2.45)$$

pour p variant entre Pmin et Pmax.

La figure ci-dessous montre les performances du prédicteur à long terme qui sont obtenues en fonction de la fréquence à laquelle les paramètres de ce dernier sont actualisés. Notons également que le prédicteur à long terme optimal améliore de près de 3 dB le rapport signal sur bruit segmental.

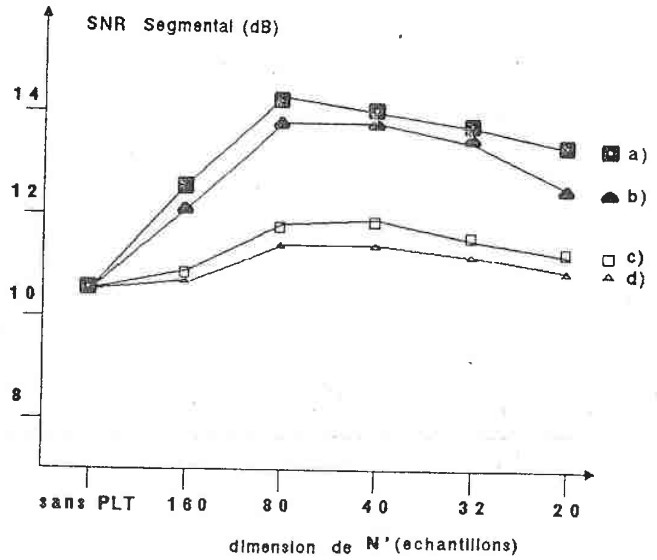


Figure 2.10: Evolution du rapport signal sur bruit segmental en fonction de la fréquence à laquelle les paramètres du prédicteur à long terme sont actualisés; a)voix masculine PLT bouclée optimale; b)voix féminine PLT bouclée optimale; c)voix masculine PLT non-bouclée; d)voix féminine PLT non-bouclée

L'évolution temporelle des différents signaux est proposée dans la figure 2.11. On peut constater que les filtrages à court terme et à long terme "blanchissent" le spectre des signaux résultants.

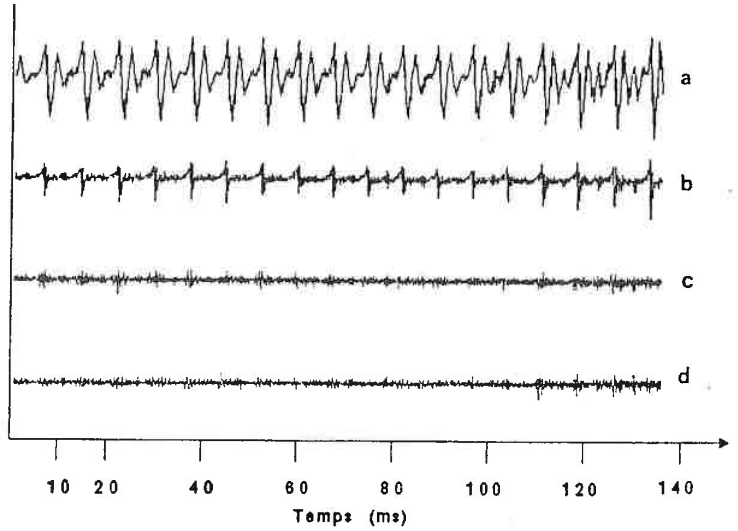


Figure 2.11: Visualisation des signaux: a) signal original; b) signal résiduel à court terme (MP=12); c) signal résiduel à long terme (PLT à 1 coefficient); d) signal résiduel à long terme (PLT à 3 coefficients).

Ce blanchiment s'accompagne d'une réduction de la distribution en amplitude des signaux (fig 2.12). Une prédiction à long terme à 3 coefficients n'apporte guère d'amélioration. De plus sa stabilité est plus complexe à contrôler. Nous avons donc choisi un prédicteur à long terme à coefficient unique.

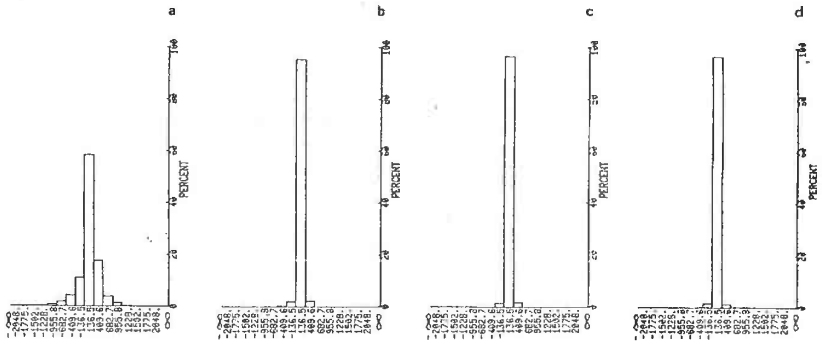


Figure 2.12: Visualisation de la distribution en amplitude des signaux: a) signal original; b) signal résiduel à court terme (MP=12); c) signal résiduel à long terme (PLT à 1 coefficient); d) signal résiduel à long terme (PLT à 3 coefficients).

La détermination des paramètres du prédicteur à long terme prend en compte le facteur perceptuel. Toutefois, la figure ci-dessous, montre que ce dernier n'améliore guère les performances du prédicteur à long terme.

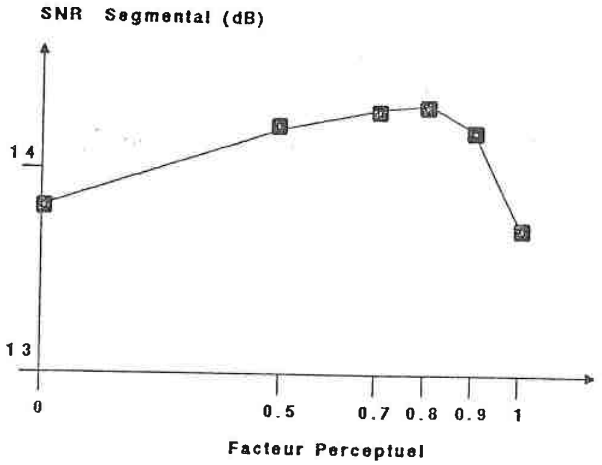


Figure 2.13: Influence du facteur perceptuel sur le prédicteur à long terme

Pour un facteur perceptuel nul, les paramètres du prédicteur à long terme sont directement déterminés à partir des signaux r_n et y_n , au lieu des signaux perceptuels p_n et q_n. Ainsi, le calcul des signaux p_n et q_n, qui représente un coût opératoire important, est éliminé. La figure 2.14 illustre la structure simplifié, mais optimale, que nous avons retenue.

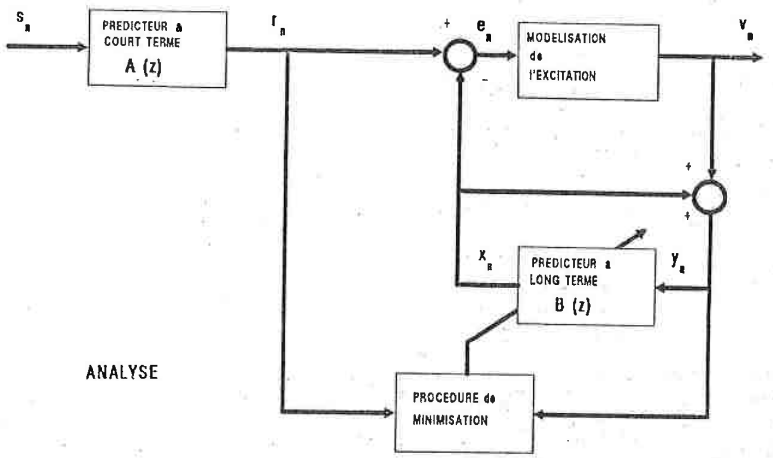


Figure 2.14: Calcul optimal simplifié des parmètres du prédicteur à long terme

Le nombre total d'opérations par seconde, correspondant à la modélisation du prédicteur à long terme est détaillé dans le tableau ci-dessous. Le facteur ø' représente le nombre de sous-fenêtres de modélisation du prédicteur à long terme (de longueur N ') par seconde.

A titre d'exemple, la puissante de traitement est donnée pour des valeurs standards des paramètres relevant du prédicteur à long terme:

N' : fenêtre de modélisation du prédicteur à long terme = 40 éch.
L : nombre de coefficients du prédicteur à long terme = 1
Pmin: décalage minimum du prédicteur à long terme = 40 éch.
Pmax: décalage maximum du prédicteur à long terme = 168 éch.
MP : ordre du filtre de synthèse = 12

forme optimale	multi.,addit./s	div./s
filtrage de r_n filtrage de y_n détermination du décalage et du facteur de gain prédiction à long terme	3.MP.N'.ø' 3.MP.(Pmax-Pmin+N').ø' 2.(Pmax-Pmin)(N'+1).ø' 2.N'.L.ø'	 (Pmax-Pmin).ø'
Complexité pour les valeurs standards	3612800	25600
forme optimale simplifiée	**multi.,addit./s**	**div./s**
détermination du décalage prédiction à long terme	2.(Pmax-Pmin)(N'+1).ø' 2.N'.L.ø'	(Pmax-Pmin).ø'
Complexité pour les valeurs standards	2115200	25600

Tableau 2.2: Complexité liée à la détermination optimale du prédicteur à long terme

II.7 DETERMINATION DE L'EXCITATION:

La modélisation de l'excitation par "analyse-synthèse", proposée par ATAL [1,3] consiste à trouver la séquence d'excitation du filtre de synthèse qui minimise au sens du critère retenu l'erreur entre le signal original et le signal synthétique. Dans cette description nous faisons l'hypothèse que le filtre de synthèse est constitué uniquement du prédicteur à court terme. La figure ci-dessous présente sous forme de bloc diagramme le processus à l'analyse.

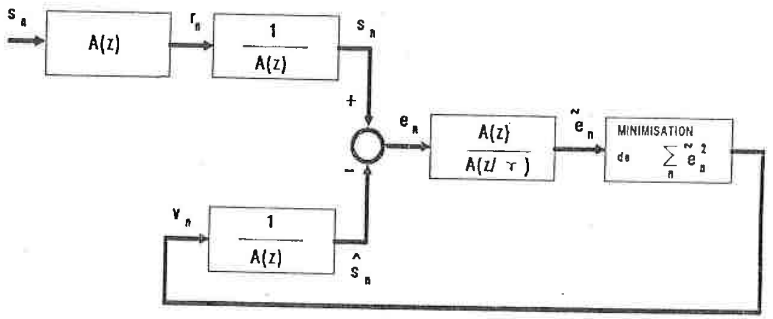


Figure 2.15: Bloc diagramme de l'analyse de la procédure de modélisation de l'excitation par analyse/synthèse.

Décrivons les différents blocs.

FILTRE DE SYNTHESE ET SOUSTRACTEUR:

La séquence d'excitation V(z) correspondant à un segment de signal est synthétisée à travers le filtre de synthèse 1/A(z). Ce signal est soustrait au signal original S(z) dans le soustracteur.

$$E(z) = S(z) - \hat{S}(z) = S(z) - V(z)\,\frac{1}{A(z)} \qquad (2.46)$$

PONDERATION DE L'ERREUR:

L'erreur entre le signal original et le signal synthétique est pondérée par le filtre perceptuel W(z) de manière à prendre en compte lors de la recherche de l'excitation l'effet de masquage lié à l'audition.

$$\tilde{E}(z) = E(z).W(z) = [\,S(z) - \hat{S}(z)\,]W(z) = [\,S(z) - \hat{S}(z)\,]\,\frac{A(z)}{A(z/\tau)} \qquad (2.47)$$

PROCEDURE DE MINIMISATION:

L'erreur perceptuelle, évaluée globalement sur un segment de signal est minimisée au sens des moindres carrés.

$$\epsilon = \sum_{n\neq\phi} \tilde{e}_n^2 = \sum_{n\neq\phi} [(s_n - \hat{s}_n) * w_n]^2 \qquad (2.48)$$

où w_n est la réponse impulsionnelle du filtre perceptuel.
* représente le produit de convolution

La forme particulière du filtre perceptuel permet de simplifier la relation ci-dessus. Il en découle 2 variantes.

soit R(z) le signal résiduel défini par:

$$R(z) = S(z)\, A(z) \qquad (2.49)$$

soit F(z) le filtre de pondération défini par:

$$F(z) = \frac{1}{A(z/\tau)} \qquad (2.50)$$

VARIANTE 1:

Cette variante fut proposée par LEFEVRE [22]. Elle isole le signal résiduel.

L'erreur perceptuelle s'écrit dans ce cas:

$$\tilde{E}(z) = F(z)\,[\,R(z) - V(z)\,] \qquad (2.51)$$

Il en découle le bloc diagramme simplifié suivant:

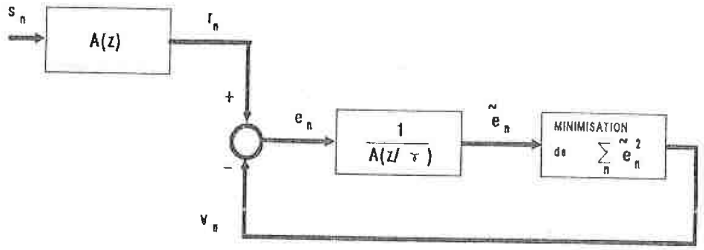


Figure 2.16: Bloc diagramme de la variante 1:

VARIANTE 2:

Cette variante qui isole le signal perceptuel P(z), fut proposée par BEROUTI[5].

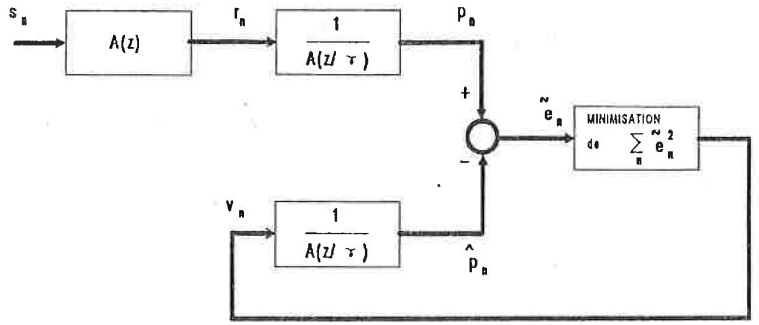


Figure 2.17: Bloc diagramme de la variante 2:

L'erreur perceptuelle s'écrit:

$$\tilde{E}(z) = F(z)\, R(z) - F(z)\, V(z) = P(z) - F(z)\, V(z) \qquad (2.52)$$

Ces deux variantes ont été proposées dans le cas particulier d'une excitation multi-impulsionnelle, mais elles peuvent être appliquées aux autres types d'excitations. Leurs performances sont très proches, mais pas identiques. Cette différence est due principalement aux effets de bords différents que procurent les deux formulations. Pour les codeurs qui utilisent le prédicteur à long terme bouclé, il est plus efficace de retenir la première variante car, comme nous le verrons, elle permet d'économiser des calculs tout en procurant également des performances un peu meilleures.

II.8 CONCLUSION:

Nous venons de décrire dans ce chapitre la structure d'une famille de codeurs hybrides performants qui ne diffèrent que par la nature du signal d'excitation. Ils mettent en oeuvre des processus de modélisation adaptative suivant la méthode de corrélation, qui est à base d'un critère des moindres carrés. Ces codeurs incorporent un filtre de pondération qui améliore leurs performances subjectives.

Compte-tenu des contraintes de stabilité et de robustesse aux traitements en précision finie, nous avons retenu la transformation de LEROUX/GUEGUEN pour déterminer les coefficients PARCOR du prédicteur à court terme. Le prédicteur à court terme a une structure en treillis.

La combinaison de la structure bouclée avec le calcul optimal des paramètres du prédicteur à long terme, proposée par DYMARSKY, procure une amélioration significative des performances subjectives et objectives du codeur. Ceci introduit quelques contraintes notamment sur le décalage minimum du prédicteur à long terme car les procédures de minimisation globales appliquées au calcul du prédicteur à long terme et au calcul de l'excitation sont interdépendantes. Néanmoins, nous retiendrons la solution optimale simplifiée, car elle offre une réduction significative de la complexité tout en limitant à environ 0.5 db la chute du rapport signal sur bruit segmental par rapport à la solution optimale

Dans le prochain chapitre, nous allons décrire un codeur particulier hybride dans lequel le signal d'excitation est approximé par des séquences d'impulsions.

Bibliographie:

Publications:

[1] "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave"
ATAL B.S., HANAUER S.L.
Journal of the Acoust. Soc. of Ame., 1971, pages 637-655

[2] "Predictive Coding of Speech and Subjective Error Criteria"
ATAL B.S., SCHROEDER M.R.
IEEE Trans. on ASSP, 1979, pages 247-254

[3] "A New Model of LPC Excitation for Producing Natural-Sounding Speech at Low Bit Rates"
ATAL B.S., REMDE J.R.
IEEE Proc. Int. Conf. on ASSP, 1982, pages 614-618

[4] "Numerical Solution of Linear Equations with Toeplitz and Vector Toeplitz Matrices"
BAREISS E.H.
Numer. Math., 1969, pages 404-424

[5] "Efficient Computation and Encoding of the Multipulse Excitation for LPC"
BEROUTI M., GARTEN H., KABAL P., MERMELSTEIN P
IEEE Proc. Int. Conf. on ASSP, 1984, pages 10.1.1-10.1.4

[6] "Vector Adaptive Predictive Coding of Speech at 9.6 kb/s"
CHEN J.H., GERSHO A.
IEEE Proc. Int. Conf. on ASSP, 1986, pages 1693-1696

[7] "Linear Prediction of Speech with a Least Absolute Error Criterion"
DENOEL E., SOLVAY J.P.
IEEE Trans. on ASSP, 1985, pages 1397-1403

[8] "Progres in Speech Recognition and Speech Synthesis"
FALLSIDE F.
Electronic Display '81 Conf. Proc., 1981, pages 47-62

[9] "Digital Lattice and Ladder Filter Synthesis"
GRAY A. H., MARKEL J.D.
IEEE Trans. on ASSP, 1980, pages 609-615

[10] "Time Dependent ARMA Modeling of Non Stationary signals"
GRENIER Y.
IEEE Trans. on ASSP, 1983, pages 899-911

[11] "Time-Frequency Analysis Using Time-Dependent ARMA Models"
GRENIER Y.
IEEE Proc. Int. Conf. on ASSP, 1984, 4185

[12] "Analyse de la Parole par les Méthodes de Modélisation Paramétrique"
GUEGUEN C.
Ann. Télécommun., 1985, pages 253-269

[13] "A Statistical Method for Estimation of Speech Spectral Density and Formant Frequencies"
ITAKURA F.,SAITO S.
Electron. Communication, Japan 1970, vol 53-A

[14] "Experimental Evaluation of Different Approches to the Multi-Pulse Coder"
KROON P., DEPRETERE E.F.,
IEEE Proc. Int. Conf. on ASSP, 1984, pages 10.4.1-10.4.4

[15] "Optimisation du Calcul des Coefficients de Corrélation Partielle"
LE ROUX J.
7 ème JEP, 1976

[16] "A Fixed point Computation of Partial Correlation Coefficients"
LE ROUX J., GUEGUEN C.
IEEE Trans. on ASSP, 1977, pages 257-259

[17] "The Wiener RMS (root mean square) Error Criterion in Filter Design and Prediction"
LEVINSON N.
Joun. of Math. and Phys., 1946, Vol 25.

[18] "Linear Prediction: A Tutorial Review"
MAKHOUL J.
IEEE Proc. Int. Conf. on ASSP, 1975, pages 561-580

[19] "Stable and Efficient Lattice Methods for Linear Prediction"
MAKHOUL J.
IEEE Trans. on ASSP, 1977, pages 423-428

[20] "Codeur Multi-Impulsionnel avec Prédiction Vectorielle à Long Terme: un Algorithme, une Procédure de Codage, l'Apport du Language ADA"
MOREAU N., DYMARSKI P., FRITSCH J.G.
Collo. GRETZI, 1987, pages 423-426

[21] "Sur la Norme L1 en Analyse Multi-Impulsionnelle"
OUAHABI A., HALILALI A.
Collo. GRETZI, 1987, pages 49-52

[22] "Codage de la Parole en Temps Réel à Débit Réduit Mettant en Oeuvre une Analyse Multi-Impulsionnelle"
LEFEVRE J.P., PASSIEN O.
13ème JEP, 1984

[23] "Stability and Performance Analysis of Pitch Filters in Speech Coders"
RAMACHANDRAN R.P., KABAL P.
IEEE Trans. on ASSP, 1987, pages 937-946

[24] "Application de la Prédiction Linéaire à l'Analyse du Signal de Parole"
SERIGNAT J.F.
Bulletin de l'Inst. de Phoné. de Grenoble, 1974, Vol. 3 pages 23-53

Thèse:

[25] "Etude, Simulation et Mise en Oeuvre sur Microprocesseur de Codeurs Prédictifs Multi-Impulsionnels"
LANCON E.
Université de Nice, 1986

Ouvrages:

[27] "Linear Prediction of Speech"
MARKEL J., GRAY A.
Springer Verlag New-York, 1976

[28] "Digital Signal Processing"
OPPENHEIM A.V., SCHAFER R.W.
Prentice Hall, Englewood,1975

CHAPITRE III

MODELISATION MULTI-IMPULSIONNELLE DE L'EXCITATION

III.1 INTRODUCTION:

Dans le cas d'une excitation multi-impulsionnelle, le signal résiduel est approximé de manière scalaire et globale par quelques impulsions bien placées dont les positions et les amplitudes sont à transmettre. En pratique, on constate qu'une impulsion par milliseconde en moyenne est suffisante pour restituer une qualité presque téléphonique.

Nous allons voir dans un premier temps, les procédures de calcul de l'excitation multi-impulsionnelle [2] dans le cas d'un prédicteur à court terme uniquement. Nous verrons ensuite les améliorations, mais également les contraintes, qu'apporte le prédicteur à long terme bouclé optimal.

III.2 PRINCIPE DE LA MODELISATION MULTI-IMPULSIONNELLE:

On cherche à représenter un segment du signal d'excitation v_n par K impulsions:

$$v_n = \sum_{k=1}^{K} A^k \, \delta_{n,M_k} \qquad \text{avec } \delta_{n,M_k} = 1 \text{ si } n = M_k, \; = 0 \text{ si } n = M_k \tag{3.1}$$

où M_k et A^k représentent respectivement la position et l'amplitude de la $k^{ème}$ impulsion.

Minimiser sans connaître à priori la position des impulsions est d'une grande complexité, compte-tenu de la non-linéarité des équations à résoudre.

La solution proposée par la plus part des auteurs passe par une minimisation partielle [1],[3],[15]. Elle consiste à déterminer de manière itérative la position puis l'amplitude de chacune des impulsions. A chaque itération, la contribution apportée par l'impulsion calculée est retirée au signal d'erreur. Cette approche est très avantageuse car elle permet d'une part de linéariser la solution, d'autre part à chaque itération elle limite à deux le nombre d'inconnues. Bien que sous-optimale, cette méthode procure des résultats très satisfaisants.

Nous avons vu dans le chapitre précédent (fig. 2.15 et fig. 2.16), que le schéma fonctionnel de la procédure d'analyse par synthèse pouvait se mettre sous deux formes selon que le soustracteur est placé avant ou après le filtre de pondération F(z). Les critères quadratiques minimisés s'écrivent pour ces deux variantes:

VARIANTE 1: modélisation à partir du signal résiduel r_n [14]:

$$\epsilon = \sum_{n \in \mathcal{B}} \left[\left(r_n - \sum_{k=1}^{K} A^k \, \delta_{n,M_k} \right) * f_n \right]^2 \tag{3.2}$$

VARIANTE 2: modélisation à partir du signal perceptuel p_n [3]:

$$\epsilon = \sum_{n \in \mathcal{B}} \left[p_n - \left(\sum_{k=1}^{K} A^k \, \delta_{n,M_k} \right) * f_n \right]^2 \tag{3.3}$$

f_n représente la réponse impulsionnelle du filtre de pondération $F(z) = 1/A(z/\tau)$, le symbole * représente ici le produit de convolution.

Ces équations mettent en évidence la non-linéarité de l'estimation simultanée des positions M_k et des amplitudes A^k.

Comme pour l'estimation des paramètres du filtre de synthèse, nous retenons la méthode d'autocorrélation, qui étend les limites de l'intervalle de sommation de l'erreur quadratique de **-infini à +infini**. En effet, BEROUTI [3] montre que pour du signal de parole, la méthode d'autocorrélation, qui simplifie la recherche des impulsions, donne des performances subjectives identiques à celles de la méthode par covariance.

III.3 DETERMINATION UNE A UNE DES IMPULSIONS:

La détermination de la position M_k et de l'amplitude A^k de l'impulsion d'ordre k suppose que les k-1 impulsions déterminées précédemment sont optimales.

La méthode consiste donc à déterminer la position M_k de l'impulsion d'ordre k en minimisant l'erreur quadratique ϵ^k par rapport à son amplitude A^k. r_n^k est le signal résiduel auquel la contribution des k-1 impulsions placées précédemment a été soustraite. De même p_n^k est le signal perceptuel auquel la contribution des K-1 impulsions placées précédemment a été soustraite.

VARIANTE 1: modélisation à partir du signal résiduel r_n:

$$\epsilon^k = \sum_{n=0}^{N+MA-1} \left[\sum_{j=n-MA}^{n} r_n^k . f_{n-j} \right]^2 = \sum_{n=0}^{N+MA-1} \left[\sum_{j=n-MA}^{n} (r_n^{k-1} - A^k \delta_{n,M_k}) . f_{n-j} \right]^2 \qquad (3.4)$$

où MA est la durée de la réponse impulsionnelle du filtre de pondération F(z) et r_n^k le signal d'erreur résiduel lorsque k impulsions ont été placées. r_n^0 représente le signal résiduel initial issu du filtre blanchissant A(z).

VARIANTE 2: modélisation à partir du signal perceptuel p_n:

$$\epsilon^k = \sum_{n=0}^{N+MA-1} \left[p_n^k \right]^2 = \sum_{n=0}^{N+MA-1} \left[p_n^{k-1} - A^k . f_{n-M_k} \right]^2 \qquad (3.5)$$

où MA est la durée de la réponse impulsionnelle du filtre de pondération F(z) et p_n^k le signal d'erreur perceptuel lorsque k impulsions ont été placées. p_n^0 est le signal perceptuel initial, précédent le placement des impulsions.

III.3.1 CALCUL DE L'AMPLITUDE:

L'expression de A^k est déterminée en annulant la dérivée partielle de ϵ^k par rapport à A^k:

$$\frac{\delta \epsilon^k}{\delta A^k} = 0 \qquad (3.6)$$

on trouve:

pour la VARIANTE 1:

$$A^k = \frac{\sum_{j \neq \phi} r_j^{k-1} \sum_{n=m}^{m+MA} f_{n-m} . f_{n-j}}{\sum_{n=m}^{m+MA} f_{n-m}^2} = \frac{\sum_{j \neq \phi} r_j^{k-1} . \phi_{m,j}}{\phi_{m,m}} \qquad (3.7a)$$

Selon la méthode d'autocorrélation [3], $\phi_{m,m}$ est une constante indépendante de m. Dans ce cas la relation ci-dessous se simplifie et s'écrit:

$$A^k = \sum_{j=m-MA}^{m+MA} r_j^{k-1} \frac{C_{|m-j|}}{C_0} = \sum_{j=m-MA}^{m+MA} r_j^{k-1} . C'_{|m-j|} = t_m^k \qquad (3.7b)$$

t_m^k peut encore s'écrire:

$$t_m^k = r_m^{k-1} + \sum_{j=1}^{MA} (r_{m-j}^{k-1} + r_{m+j}^{k-1}) C'_j \qquad (3.8)$$

où t_m^k est l'intercorrélation entre le signal résiduel r_m^1 et l'autocorrélation normalisée C'_i de la réponse impulsionnelle du filtre de pondération F(z).

pour la VARIANTE 2:

$$A^k = \frac{\sum_{n=m}^{m+MA} f_{n-m}\, p_n^{k-1}}{\sum_{n=m}^{m+MA} f_{n-m}} = \frac{\sum_{n=m}^{m+MA} f_{n-m}\, p_n^{k-1}}{C_0} = \sum_{n=m}^{m+MA} f'_{n-m}\, p_n^{k-1} = \alpha_m^k \qquad (3.9)$$

où α_m^k est l'intercorrélation entre le signal perceptuel p_n^{k-1} et la réponse impulsionnelle normalisée du filtre de pondération f'_n.

III.3.2 RECHERCHE DE LA POSITION:

En substituant l'expression de l'amplitude optimale A^k, quelque soit la variante, dans les relations de l'erreur quadratique on obtient:

$$\epsilon^k = \sum_{n=0}^{N+MA-1} [\, r_n^{k-1}]^2 - \frac{t_M^{k-1}}{k-1}^2 \qquad (3.10)$$

ou

$$\epsilon^k = \sum_{n \neq \emptyset} [\, p_n^{k-1}]^2 - \frac{C_o\, \alpha_M^{k-1}}{k-1}^2 \qquad (3.11)$$

Les produits t_m^2 et $C_o\alpha_m^2$ sont toujours positifs. La meilleure position M_k pour l'impulsion d'ordre k est donnée par la valeur de m qui minimise ϵ^k par conséquent qui maximise les fonctions de localisation t_m ou α_m.

VARIANTE 1:

$$M_k = m \text{ pour } |t_m^k| \text{ maximum} \qquad (3.12)$$
$$A^k = t_m^k$$

VARIANTE 2:

$$M_k = m \text{ pour } |\alpha_m^k| \text{ maximum} \qquad (3.13)$$
$$A^k = \alpha_m^k$$

III.3.3: REACTUALISATION DU SIGNAL:

Lorsque la position et l'amplitude de l'impulsion sont connues, la contribution de celle-ci doit être soustraite, selon la variante, au signal résiduel r_n ou au signal perceptuel p_n:

pour la VARIANTE 1: actualisation du résiduel

$$r_n^k = r_n^{k-1} - A^k \qquad \text{pour } n = M_k \qquad (3.14)$$

pour la VARIANTE 2: actualisation du perceptuel

$$p_n^k = p_n^{k-1} - A^k . f_{n-M_k} \qquad \text{pour } n = M_k \text{ à } M_k + MA \qquad (3.15)$$

La figure 3.1a visualise le signal résiduel initial r_n^o et l'autocorrélation normalisée C'$_k$ de la réponse impulsionnelle du filtre de pondération F(z). A droite, apparaît le signal résiduel r_n^k (pour k=1 à 4) actualisé après le placement successif des impulsions. Rappelons que l'actualisation du signal résiduel consiste à soustraire, à chaque itération, l'impulsion à ce signal. La figure 3.1.b représente, la même procédure itérative dans le cas de la variante 2 qui modélise l'excitation à partir du signal perceptuel initial p_n^o et la réponse impulsionnelle normalisée f'$_n$ du filtre de pondération F(z).

L'actualisation du signal perceptuel consiste à soustraire, à ce signal, la contribution de l'impulsion convoluée par le filtre de pondération, qui n'est autre que la réponse impulsionnelle du filtre F(z) multipliée par l'amplitude de l'impulsion.

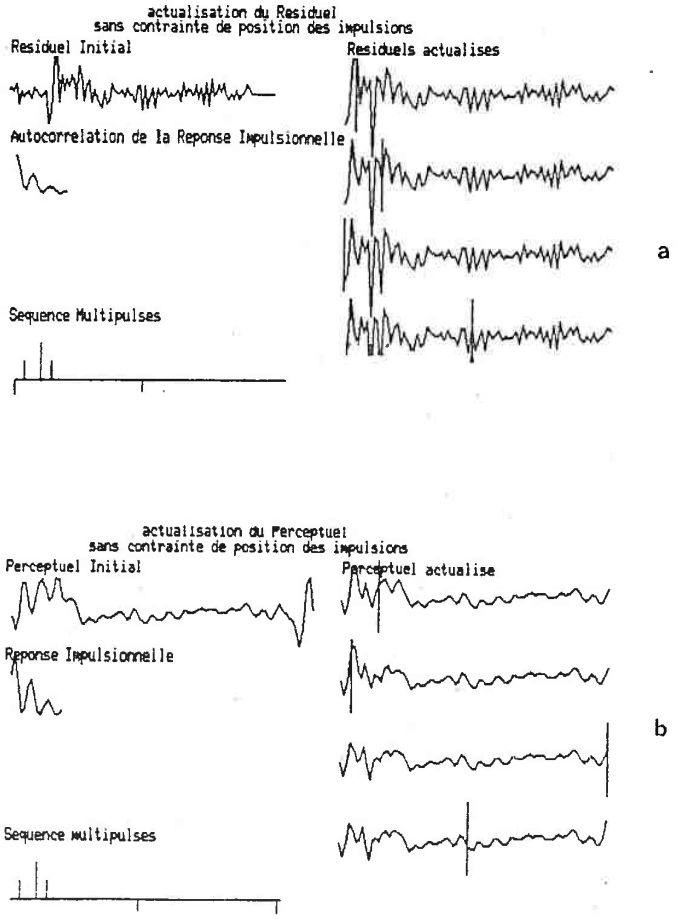


Figure 3.1: Visualisation de la détermination itérative des impulsions; a)avec réactualisation du résiduel r_n, b)avec réactualisation du perceptuel p_n

Comme on peut le constater, les séquences d'excitation que donnent les deux variantes ne sont pas tout à fait identiques car les effets de bord qu'engendrent les deux formulations sont légèrement différentes. De manière globale, on constate que la variante 1 donne systématiquement un rapport signal sur bruit segmental meilleur de 0.5 dB. Ceci s'explique par le fait que

dans la variante 2, le signal perceptuel est obtenu par filtrage, tandis que son actualisation est réalisée, en lui soustrayant la réponse impulsionnelle tronquée sur MA échantillons. Pour faire disparaître cette disymétrie, le signal perceptuel doit être construit à partir de la réponse impulsionnelle du filtre de pondération, à savoir:

$$p_n = \sum_{j=0}^{MA-1} r_{n-j}.f_j \qquad (3.16)$$

Dans la variante 1, ce phénomène n'existe pas, car l'excitation multi-impulsionnelle est déterminée directement à partir du résiduel, qui rappelons le est homogène au signal d'excitation.

III.3.4) REACTUALISATION DE LA FONCTION DE LOCALISATION:

Des variantes au niveau de la réactualisation proposent de retirer la contribution de l'impulsion directement aux fonctions de localisation, qui dans ce cas doivent être mémorisées sous forme de vecteurs.

La contribution de l'impulsion est retirée à t_n dans la VARIANTE 1:

$$t_n^k = t_n^{k-1} - A^k.C'_{|n-M_k|} \qquad \text{pour } n = M_k - MA \text{ à } M_k + MA \qquad (3.17)$$

La contribution de l'impulsion est retirée à α_n dans la VARIANTE 2:

$$\alpha_n^k = \alpha_n^{k-1} - A^k.C_{|n-M_k|} \qquad \text{pour } n = M_k - MA \text{ à } M_k + MA \qquad (3.16)$$

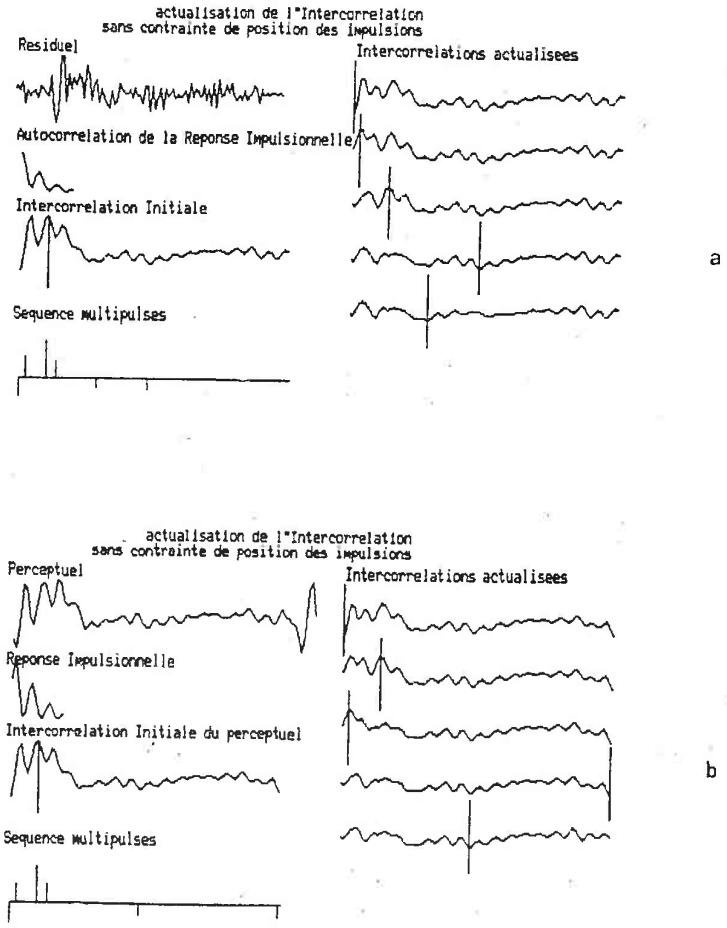


Figure 3.2: Visualisation de la détermination itérative des impulsions avec réactualisation directe des fonctions de localisation; a) t_n ; b) α_n

Les remarques faites précédement sont également valables pour les formulations où les fonctions de localisation sont directement actualisées. L'intérêt de ces expressions est de diminuer de manière significative le nombre de multiplications et d'additions (Tab 3.1). Ceci est souhaitable pour une implantation sur des microprocesseurs de traitement du signal. En revanche dans le cas d'une réalisation VLSI, la réactualisation des signaux, dont la complexité est directement proportionnelle au nombre d'impulsions, est préférable car elle privilégie la régularité du traitement.

III.3.5 OPTIMISATION GLOBALE:

La succession des minimisations partielles ne donnent pas la solution optimale globale. Il est également possible de réestimer à chaque itération l'ensemble des amplitudes des k impulsions, dont les positions sont supposées optimales en résolvant le système matriciel à l'ordre K .

$$\begin{vmatrix} C'_{|m_1-m_1|} & C'_{|m_1-m_2|} & \cdots & C'_{|m_1-m_K|} \\ C'_{|m_2-m_1|} & C'_{|m_2-m_2|} & \cdots & C'_{|m_2-m_K|} \\ \cdot & \cdot & \cdots & \cdot \\ C'_{|m_K-m_1|} & C'_{|m_K-m_2|} & \cdots & C'_{|m_K-m_K|} \end{vmatrix} \times \begin{vmatrix} A^1 \\ A^2 \\ \cdot \\ A^K \end{vmatrix} = \begin{vmatrix} t_{m_1} \\ t_{m_2} \\ \cdot \\ t_{m_K} \end{vmatrix} \qquad (3.19)$$

Cette approche donne de meilleurs résultats, mais le coût en opérations supplémentaires est élevé. Un bon compromis semble-t-il consiste à ne résoudre qu'une seule fois le système d'équation lorsque les positions de toutes les impulsions sont connues. Les performances des trois solutions sont données dans la figure ci-dessous.

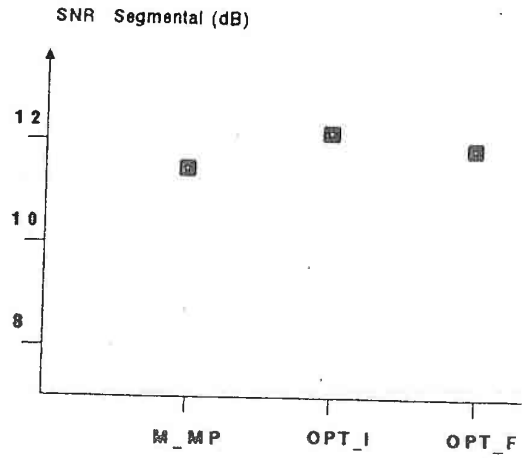


Figure 3.3: Evolution du rapport signal à bruit segmental; MMP modélisation par fenêtre, OPTI avec optimisation partielle pour toute nouvelle impulsion, OPTF avec une optimisation globale lorsque les positions des K impulsions sont connues.

On peut également citer de nombreuses variantes supplémentaires, qui moyennant certaines contraintes [4], permettent de limiter le débit important lié aux positions des impulsions . C'est le cas du codeur que propose KOYAMA [11]. Il réalise une quantification de la fonction de localisation, déportant ainsi la procédure de recherche des impulsions dans la synthèse. Toutefois, les performances de cette approche sont insuffisantes. A la synthèse, la fonction de localisation reconstruite présente des discontinuités qui sont dues à la juxtaposition des vecteurs quantifiés. Ceci perturbe la convergence

du calcul des impulsions.

SINGHAL [24] propose une optimisation des paramètres du filtre de synthèse, après que l'excitation multi-impulsionnelle a été déterminée. Celle-ci permet de réduire le biais lié à l'estimation des paramètres du filtre pour les sons voisés. Cette optimisation, n'apporte toutefois pas d'amélioration significative au niveau de la distance cepstrale entre le signal original et le signal synthétique.

D'autres approches cherchent à simplifier la recherche des positions des impulsions. C'est le cas de la solution proposée par CAELEN [6], qui place les impulsions aux positions où le résiduel présente les amplitudes les plus fortes. Mais une telle solution présente des performances limitées, car elle ne prend pas en considération l'intercorrélation qui existe entre deux impulsions proches.

La variante la plus intéressante est certainement celle proposée par KROON [13],[27] qui porte le nom de "Regular Pulse Coding". Elle réduit le nombre de positions à transmettre par fenêtre à une seule, en plaçant toutes les impulsions à intervalle régulier. Cette approche a été retenue comme norme pour le radio téléphone numérique à 16 kbits/s.

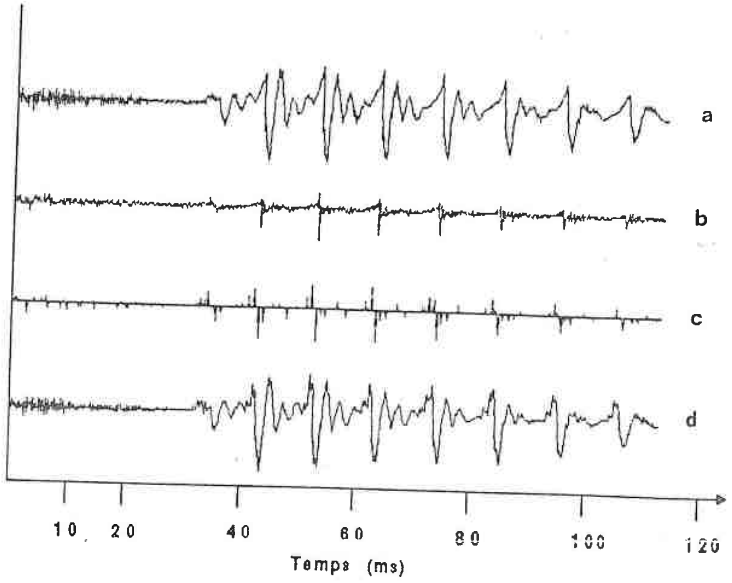


Figure 3.4: Visualisation des différents signaux de la variante 1, a) signal original, b) signal résiduel, c) signal d'excitation multi-impulsionnelle, e) signal synthétique

III.4 EFFETS DE BORD:

La modélisation de l'excitation est réalisée sur des fenêtres contiguës. Toutefois, le calcul des fonctions de localisation α_n et t_n, ainsi que les opérations de réactualisation (du signal ou de la fonction de localisation) nécessitent un débordement sur les fenêtres adjacentes.

Dans le cas de la variante 1, le calcul des MA premiers et MA derniers termes de t_n utilise respectivement les MA derniers échantillons de résiduel r_n de la fenêtre précédente et les MA premiers échantillons de la fenêtre suivante. Pour la variante 2, il n'existe qu'un débordement en aval, mais celui-ci est le plus pénalisant car il introduit un décalage supplémentaire d'une fenêtre d'analyse. Ce décalage peut être évité en prolongeant la fenêtre courante par les échantillons de la réponse libre du filtre. Dans ce cas, la mémoire du filtre doit être préalablement mémorisée de façon à conserver la continuité des signaux. Cependant cette opération pénalise fortement la régularité du traitement.

Les simulations mettent en évidence (fig. 3.5), un autre effet de bord, qui se traduit par de fortes distorsions du signal aux fontières des fenêtres. Il est caractéristique de l'actualisation systématique, mais brutal, des coefficients des filtres, car d'une fenêtre à l'autre le gain du filtre peut varier dans des proportions importantes. Ce phénomène se traduit au niveau du signal d'excitation multi-impulsionnelle par une concentration exagérée d'impulsions à la frontière des fenêtres.

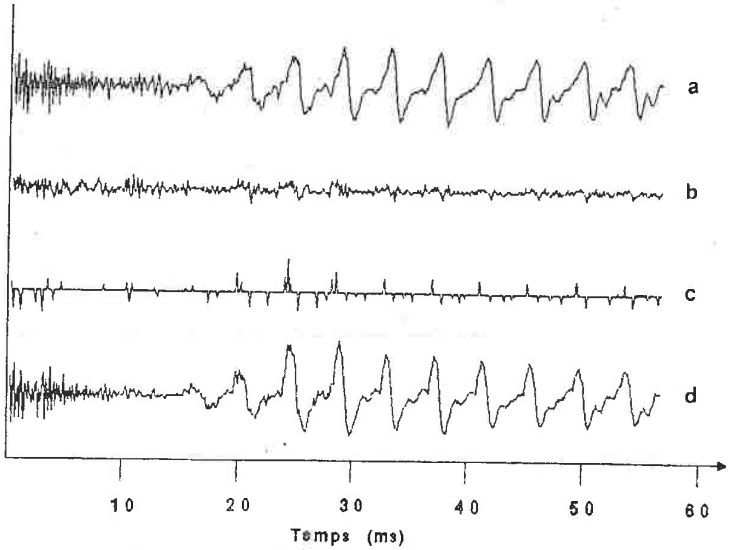


Figure 3.5: Visualisation des différents signaux perturbés localement par les effets de bord, a)signal original, b)signal d'excitation multi-impulsionnelle, c)signal synthétique

III.5 MODELISATION MULTI-IMPULSIONNELLE AVEC PREDICTION A LONG TERME:

Le système de codage à excitation multi-impulsionnelle présenté dans le paragraphe précédent permet de restituer un signal synthétique de bonne qualité. On constate toutefois que, pour une qualité comparable, la synthèse d'une voix féminine ou d'enfant requiert plus d'impulsions, donc un débit plus élevé, qu'une voix masculine. Ceci est lié à la période du fondamental.

De plus, dans un même segment de signal, on constate que l'amplitude des impulsions synchrones avec le fondamental est plus importante que celle des autres impulsions. En effet on observe des rapports d'amplitudes qui fluctuent entre 8 et 32, voire plus. Ceci est pénalisant au niveau de l'encodage comme nous le verrons dans le chapitre V.

L'introduction d'un prédicteur à long terme optimal, atténue les points critiques décrits précédemment. Il permet de conserver le même nombre d'impulsions par unité de temps quelque soit le locuteur. D'autre part le rapport en amplitude des impulsions ne dépasse guère 8, ce qui favorise un encodage efficace.

Le principe du prédicteur à long terme bouclé a été décrit en détail dans le paragraphe 2.6. La procédure de calcul de l'excitation multi-impulsionnelle requiert comme entrée le signal d'erreur de prédiction à long terme. La figure ci-dessous donne la configuration du système à l'analyse où le bloc de minimisation de u_n réalise le calcul de l'excitation multi-impulsionnelle.

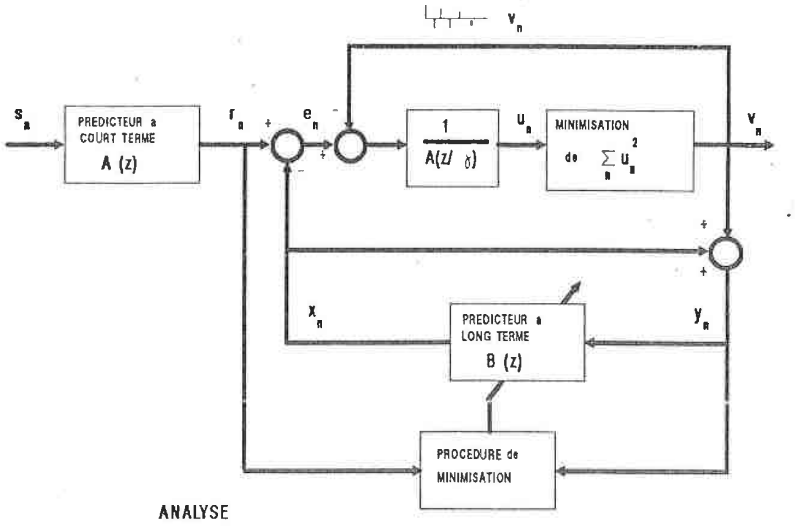


Figure 3.6: Bloc diagramme à l'analyse du codeur à excitation multi-impulsionnelle avec prédiction à long terme optimale

La détermination de l'excitation multi-impulsionnelle s'applique à des sous-fenêtres ø du signal d'erreur e_n, dont la taille est dictée par la durée de l'intervalle durant lequel les paramètres du prédicteur à long terme sont figés. Pour chaque sous-fenêtre à analyser, on dispose d'une part du signal résiduel r_n de la sous-fenêtre courante et précédente, d'autre part du signal résiduel reconstruit y_n des sous-fenêtres traitées précédemment.

Pour un prédicteur à long terme bouclé, les procédures de calcul des paramètres du prédicteur et de l'excitation multi-impulsionnelle, sont imbriquées. En effet, pour que l'on puisse réaliser la modélisation de l'excitation il est nécessaire de disposer du signal e_n qui est le résultat du filtrage du signal résiduel r_n par le prédicteur à long terme:

$$e_n = r_n - x_n \qquad (3.20)$$

où

$$x_n = \sum_{i=-L}^{i=-1} b_i . y_{n-p+i} \quad \text{est la partie prédictible du signal } e_n \qquad (3.21)$$

D'autre part pour déterminer les paramètres optimaux du prédicteur à long terme il est nécessaire de disposer du signal résiduel synthétique y_n qui est reconstruit à partir de l'excitation multi-impulsionnelle $\tilde{v}_n$.

$$y_n = \tilde{v}_n + x_n \qquad (3.22)$$

On commence par estimer le décalage P du prédicteur à long terme en respectant la contrainte que P doit être supérieur à Pmin. Pmin est défini par rapport à la dimension de la sous-fenêtre de minimisation, dont la dimension est N'. Si la prédiction à long terme et la modélisation de l'excitation sont réalisées successivement dans la même sous-fenêtre, Pmin= N'. En revanche si ces deux modélisations sont réalisées en parallèle, mais avec une sous-fenêtre de décalage, Pmin=2.N'. La valeur optimale du décalage P est celle qui maximise 3.23.

$$\theta_p = \frac{\left[\sum_{n=0}^{N'-1} r_n . y_{n-p} \right]^2}{\sum_{n=0}^{N'-1} y_{n-p}^2} \qquad (3.23)$$

$$P = p \quad \text{pour } \theta_p \text{ maximum} \quad \text{avec} \quad Pmin < p < Pmax \qquad (3.24)$$

La valeur de Pmax est choisie de façon à couvrir les sons voisés de locuteurs masculins, dont la période de l'excitation peut descendre jusqu'à 60 Hz, soient environ 128 échantillons (pour un échantillonnage à 8 kHz). Pmax vaut donc Pmin + 128.

Ensuite le ou les coefficients de prédiction sont solutions de:

$$\begin{vmatrix} R_p & R_{p+1} & R_{p+2} \\ R_{p+1} & R_p & R_{p+1} \\ R_{p+2} & R_{p+1} & R_p \end{vmatrix} * \begin{vmatrix} b_{-1} \\ b_0 \\ b_1 \end{vmatrix} = \begin{vmatrix} R'_{p-1} \\ R'_p \\ R'_{p+1} \end{vmatrix} \qquad (3.25)$$

dans le cas d'un prédicteur à 3 coefficients. Pour un prédicteur à coefficient unique, le calcul de b se réduit à:

$$b = \frac{R'_p}{R_p} \qquad (3.26)$$

avec :

$$R'_i = \sum_{n=0}^{N'-1} r_n . y_{n-i} \qquad (3.27)$$

et

$$R_i = \sum_{n=0}^{N'-1} y_{n-i} . y_{n-i} \qquad (3.28)$$

Lorsque le prédicteur à long terme est défini, le signal d'erreur de prédiction à long terme e_n est obtenu en soustrayant au signal résiduel r_n sa partie prédictible (rel. 3.20).

De façon à simplifier la détermination de l'excitation, le prédicteur à long terme peut être exclu de la procédure de modélisation multi-impulsionnelle. Ceci est sans effet sur la qualité de la parole, sauf pour les sons dont la période du voisement est inférieure à l'intervalle de minimisation. C'est donc au signal e_n qu'est appliqué un des algorithmes de détermination de l'excitation multi-impulsionnelle. A priori, les deux variantes décrites précédemment peuvent être utilisées.

VARIANTE 1: modélisation à partir du signal résiduel e_n:

$$\epsilon = \sum_{n \in \phi} \left[(e_n - \sum_{k=1}^{K} A^k \, \delta_{n,M_k}) * f_n \right]^2 \qquad (3.29)$$

VARIANTE 2: modélisation à partir du signal perceptuel u_n du signal résiduel e_n:

le signal perceptuel u_n est défini par:

$$u_n = \sum_{i=n-MA}^{n} e_n \, f_{n-i} \qquad (3.30)$$

l'erreur à minimiser s'ecrit:

$$\epsilon = \sum_{n \in P} [\, u_n - (\sum_{k=1}^{K} A^k \, \delta n, M_k) * f_n \,]^2 \qquad (3.31)$$

En pratique, la variante 1 est la plus adaptée au prédicteur à long terme bouclé. De plus, comme nous avons pu le signaler précédemment elle donne des performances légèrement meilleures que la variante 2. Ainsi, les impulsions sont directement déterminées à partir du signal d'erreur e_n.

Soit C'_j l'autocorrélation normalisée de la réponse impulsionnelle du filtre de pondération F(z):

$$C'_j = \frac{\sum_{n=0}^{N'-1} f_n . f_{n-j}}{\sum_{n=0}^{N'-1} f_n . f_n} \qquad \text{pour } j = 0 \text{ à } MA \qquad (3.32)$$

les impulsions sont calculées par les relations ci-dessous:

$$t_m^k = A^k = e_m^{k-1} + \sum_{j=1}^{MA} (e_{m-j}^{k-1} + e_{m+j}^{k-1}) \, C'_j \qquad (3.33)$$

$$M_k = m \text{ pour } |t_m^k| \text{ maximum avec } 0 \le m \le N' \qquad (3.34)$$

$$A^k = t_{M_k}^k \qquad (3.35)$$

$$e_n^k = e_n^{k-1} - A^k \qquad \text{pour } n = M_k \qquad (3.36)$$

Lorsque l'excitation multi-impulsionnelle v_n est calculée, le signal résiduel synthétique y_n est obtenu par la relation:

$$y_n = v_n + x_n = v_n + \sum_{i=-L}^{L} b_i \, y_{n-P+i} \qquad (3.37)$$

La figure 3.7 visualisent l'évolution temporelle des signaux significatifs dans le cas de la variante 1 avec prédiction à long terme bouclée optimale simplifiée.

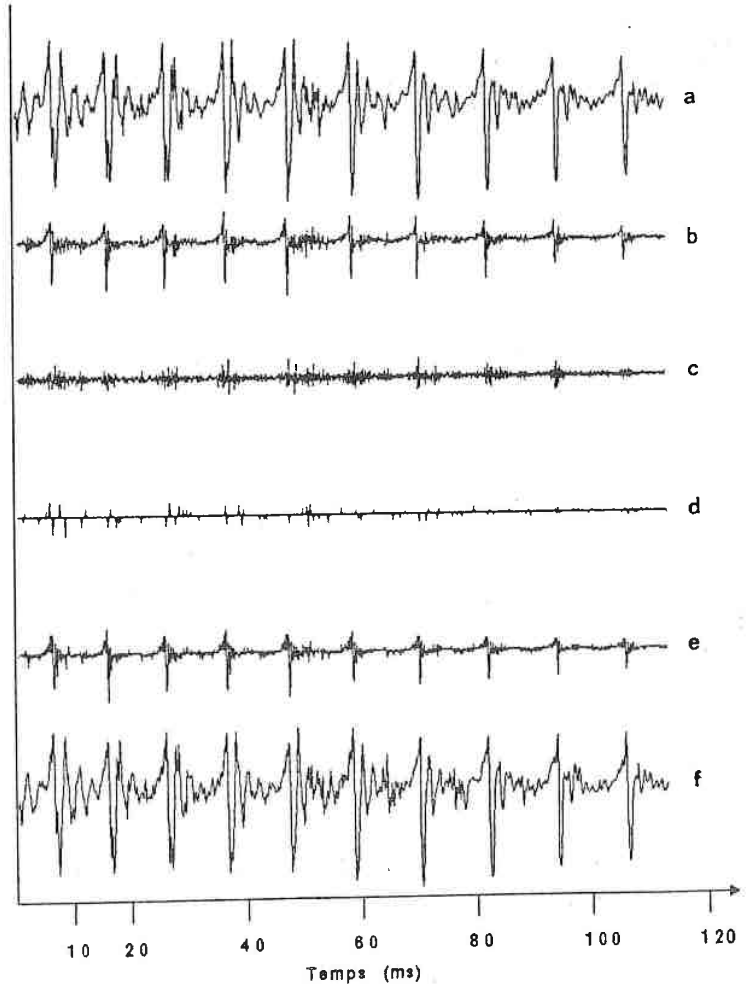


Figure 3.7: Visualisation des signaux significatifs pour une modélisation avec prédiction à long terme bouclée optimale simplifiée et le calcul de l'excitation multi-impulsionnelle suivant la variante 1; a) signal original; b) signal résiduel; c) signal résiduel à long terme; d) excitation multi-impulsionnelle; e) signal résiduel reconsruit; f) signal synthétique

III.6 PARAMETRES STANDARDS ET COMPLEXITE:

Les performances d'un codeur multi-impulsionnel, comme on peut le voir ci-dessous, sont fortement liées au nombre d'impulsions qui constituent le signal d'excitation.

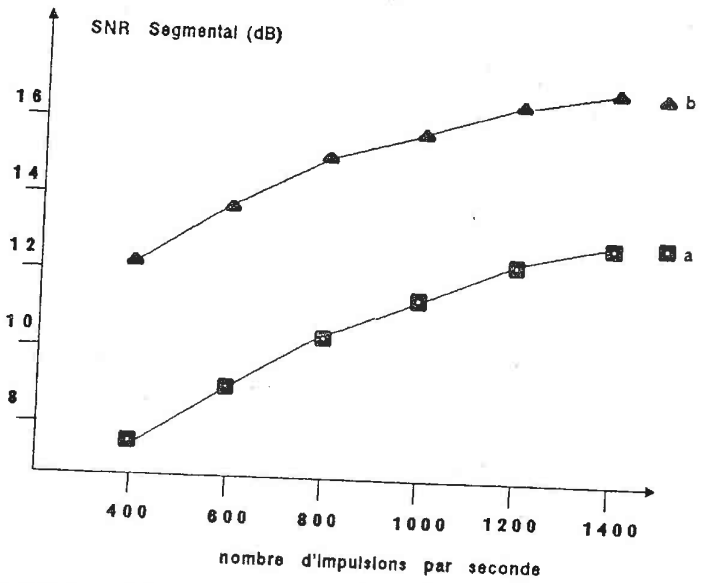


Figure 3.8: Evolution du rapport signal sur bruit segmental en fonction du nombre d'impulsions par unité de temps; a) sans prédiction à long terme; b) avec prédiction à long terme optimale.

La complexité et le débit dépendent directement de la taille de la fenêtre d'analyse, une réduction de celle-ci, sans modifier la taille de la fenêtre d'analyse LPC, est souhaitable. La dégradation que cela engendre n'est pas significative. La modélisation de l'excitation par sous-fenêtre répond de manière satisfaisante à cette exigence.

Pour des valeurs de paramètres standards:

N : fenêtre d'analyse LPC = 160
N': sous-fenêtre de prédiction à long terme et d'analyse multi-impulsionnelle
MP: l'ordre des filtres en treillis = 40
MA: durée de la rép. imp. de F(z) ou de son autocorrélation = 12
K : le nombre d'impulsions par sous-fenêtre = 25
= 5

La complexité des différentes variantes est résumée dans le tableau ci-dessous. Le facteur ϕ représente le nombre de fenêtres de modélisation LPC (de longueur N) par seconde. Quant au facteur ϕ', il représente le nombre de sous-fenêtres de modélisation de l'excitation (de longueur N').

variante 1, actualisation du signal	multi.,addit./s	div./s
réponse impulsionnelle de F(z)	3.MP.MA.ϕ	
autocorrélation de F(z)	2.MA.ϕ	MA.ϕ
critère + amplitudes	K.2.MA.N'.ϕ'	
actualisation du signal	K.ϕ'	
Complexité pour les valeurs standards	2077250	1250
variante 1, actualisation de la fonction de localisation	**multi.,addit./s**	**div./s**
réponse impulsionnelle de F(z)	3.MP.MA.ϕ	
autocorrélation de F(z)	2.MA.ϕ	MA.ϕ
critère + amplitudes	2.MA.N'.ϕ'	
actualisation du signal	K.2.MA.ϕ'	
Complexité pour les valeurs standards	526250	1250
variante 2, actualisation du signal	**multi.,addit./s**	**div./s**
calcul du signal perceptuel	3.MP.N'.ϕ	
réponse impulsionnelle de F(z)	3.MP.MA.ϕ	
normalisation de la rép. imp. de F(z)	MA.ϕ	MA.ϕ
critère + amplitudes	K.MA.N'.ϕ'	
actualisation du signal	K.MA.ϕ'	
Complexité pour les valeurs standards	1359250	1250
variante 2, actualisation de la fonction de localisation	**multi.,addit./s**	**div./s**
calcul du signal perceptuel	3.MP.N'.ϕ	
réponse impulsionnelle de F(z)	3.MP.MA.ϕ	
normalisation de la rép. imp. de F(z)	MA.ϕ	MA.ϕ
critère + amplitudes	MA.N'.ϕ'	
actualisation du signal	K.2.MA.ϕ'	
Complexité pour les valeurs standards	584250	1250
résolution globale finale	**multi.,addit./s**	**div./s**
inversion matricielle	2.K^2.ϕ'	
Complexité pour les valeurs standards	10000	1250

Tableau 3.1: Complexité des différentes procédures de détermination de l'excitation multi-impulsionnelle

III.7: CONCLUSION:

Quelque soit la variante, l'actualisation de la fonction de localisation réduit dans un rapport 3 à 4 la puissance de traitement nécessaire à la modélisation multi-impulsionnelle. Plus le nombre d'impulsions est élévé plus ce rapport est grand. Une gestion efficace de la mémoire permet de stocker la fonction de localisation sans augmenter la taille de la mémoire. En effet, une zone mémoire de dimension N'+MA échantillons est suffisante pour mémoriser alternativement le signal résiduel, puis la fonction de localisation.

L'étude comparative des codeurs à excitation multi-impulsionnelle, avec et sans prédiction à long terme, met en évidence que l'introduction du prédicteur à long terme bouclé optimal simplifié permet de gagner environ 2 dB en rapport signal sur bruit segmental, et ceci pour un débit approximativement identique.

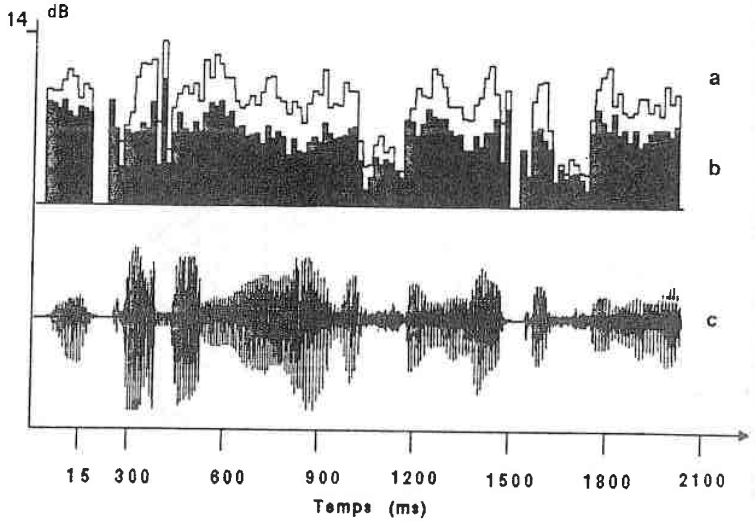


Figure 3.9: Evolution temporelle du rapport signal sur bruit segmental, a) avec prédicteur à long terme, b) sans prédicteur à long terme optimal, c) signal synthétique

Dans le cas d'une modélisation sans prédiction à long terme, le signal d'excitation du filtre de synthèse est constitué de quelques 1200 impulsions par secondes. L'introduction du prédicteur à long terme, enrichi en échantillons non nuls ce signal. Ceci permet de réduire très sensiblement la raucité du signal synthétique, qui par ailleurs est déja atténuée par la pondération perceptuelle de l'erreur. De plus il améliore la restitution des basses fréquences.

L'augmentation du débit qu'entraine la transmission des paramètres du prédicteur à long terme, peut être compensée par une légère diminution du nombre d'impulsions. A raison de 1000 impulsions par seconde, ce codeur à excitation multi-impulsionnelle procure un rapport signal sur bruit segmental moyen de 14 dB. Comme nous le verrons dans le chapitre V, consacré au codage des paramètres, l'introduction du prédicteur à long terme permet également de réduire la précision avec laquelle les amplitudes sont quantifiées.

Dans le chapitre V, les résultats concernant le codage des paramètres d'un codeur MPLPC mettent en évidence que la transmission de l'excitation représente plus des deux tiers du débit global. Les procédures de modélisation que nous allons décrire dans le prochain chapitre ont justement pour objectif de réduire le débit associé à l'excitation.

Bibliographie:

Publications:

[1] "Multi-Pulse Excited Speech Coder Based on Maximum Crosscorrelation Search Algoritm"
ARASEKI T., OZAWA K., ONO S., OCHIAI K.
IEEE Proc. Int. Conf. on ASSP, 1983, pages 23.3.1-23.3.5

[2] "A New Model of LPC Excitation for Producing Natural-Sounding Speech at Low Bit Rates"
ATAL B.S., REMDE J.R
IEEE Proc. Int. Conf. on ASSP, 1982, pages 614-617

[3] "Efficient Computation and Encoding of the Multipulse Excitation for LPC"
BEROUTI M., GARTEN H., KABAL P., MERMELSTEIN P
IEEE Proc. Int. Conf. on ASSP, 1984, pages 10.1.1-10.1.4

[4] "MPLPC Algorithms for Integration on Silicon"
BUSSOD D., FRITSCH J.G.
Rapport ESPRIT projet 64, 1986

[5] "Specification of the MPLPC Algotithm"
BUSSOD D.,FRITSCH J.G.
Rapport ESPRIT projet 64, 1987

[6] "Synthèse de la Parole par Codage Multi-Impulsionnel"
CAELEN J., SONG J.M.
Coll. GRETZI, 1985, pages 881-886

[7] "A New Approach to Multipulse LPC Coder Design"
DEMBO A., MALAH D.
IEEE Proc. Int. Conf. on ASSP, 1985, pages 949-952

[8] "Modifying the Multipulse Excitation in LPC Speech Coding"
GARCIA-GOMEZ L., FIGUEIRAS-VIDAL A.R., SANTOS-SUAREZ J.M., VERGARA-DOMINGUEZ L., CASAR-CORREDERA J.R.
Signal Processing II, 1983, pages 375-377

[9] "Perceptual Wheightings and Optimal Pulse Positioning in Multipulse LPC Speech Coding"
HAWKINS H.A., WILKES D.M., CLEMENTS M.A., HAYES M.H.
IEEE Proc. Int. Conf. on ASSP, 1985, pages 477-480

[10] "Efficient Algorithm for Multi-Pulse LPC Analysis of Speech"
JAIN V.K., HANGARTNER R.
IEEE Proc. Int. Conf. on ASSP, 1984, pages 1.4.1-1.4.4

[11] "Fully Vector-Quantized Multipulse LPC at 4800 bps"
KOYAMA H., GERSHO A.
IEEE Proc. Int. Conf. on ASSP, 1986, pages 445-448

[12] "Experimental Evaluation of Different Approaches to the Multi-Pulse Coder"
KROON P., DEPRETTERE E.F.
IEEE Proc. Int. Conf. on ASSP, 1984, pages 10.4.1-10.4.4

[13] "Regular-Pulse Excitation A Novel Aproach to Effective and Efficient Multipulse Coding of Speech"
KROON P., DEPRETTERE E.F., SLUYTER R.J.
IEEE Trans. on ASSP, 1986, pages 1054-1063

[14] "Efficient Algorithms for Obtaining Multipulse Excitation for LPC Coders"
LEFEVRE J.P., PASSIEN O.
IEEE Proc. Int. Conf. on ASSP, 1985, pages 957-960

[15] "Codage de la Parole en Temps Réel à Débit Réduit Mettant en Oeuvre une Analyse Multi-Impulsionnelle"
LEFEVRE J.P., PASSIEN O.
13ème JEP, 1984

[16] "A Lattice-Ladder Structure for Multipulse Linear Predictive Coding of Speech"
MANOLAKIS D., CARAYANNIS G.
IEEE Trans. on ASSP, 1987, pages 228-231

[17] "Codeur Multi-Impulsionnel avec Prédiction Vectorielle à Long Terme: un Algorithme, une Procédure de Codage, l'Apport du Language ADA"
MOREAU N., DYMARSKI P., FRITSCH J.G.
Coll. GRETSI, 1987, pages 423-426

[18] "Codage de la Parole a Moyen Débit: Etude Bibliographique des Méthodes de Déconvolution"
MOREAU N.
Rapport Interne, TELIC-ENST

[19] "Codage Multi-Impulsionnel pour la Restitution de la Parole par Modèles Evolutifs"
OMNES-CHEVALIER M.C., GRENIER Y., CHOLLET G.
Coll. GREZTI, 1985, pages 887-892

[20] "Low Bit Rate Speech Enhancement Using a New Method of Multiple Impulse Excitation"
PARKER A., ALEXENDER S.T., TRUSSEL H.J.
IEEE Proc. Int. Conf. on ASSP, 1984, pages 1.5.1-1.5.4

[21] "Amélioration des Performances du Codage à Excitation Multi-Impulsionnelle"
PASSIEN O., LEFEVRE J.P.
COll. GRETZI, 1985, pages 875-880

[22] "All-Pole Speech Modeling with a Maximaly Pulse-Like Residual"
RICHARD C.R., CLEMENTS M.A.
IEEE Proc. Int. Conf. on ASSP, 1985, pages 481-484

[23] "A Non-Iterative Algorithm for Obtaining Multi-Pulse Excitation for Linear-Predictive Speech Coders"
SENENSIEB G.A., MILBOURN A.J., LLOYD A.H., WARRINGTON I.M.
IEEE Proc. Int. Conf. on ASSP, 1984, pages 10.2.1-10.2.4

[24] "Optiminzing LPC Filter Parameters for Multi-Pulse Excitation"
SINGHAL S., ATAL B.S.
IEEE Proc. Int. Conf. on ASSP, 1983, pages 781-784

[25] "Improving Performance of Multi-Pulse LPC Coders at Low Bit Rates"
SINGHAL S., ATAL B.S.
IEEE Proc. Int. Conf. on ASSP, 1984, pages 1.3.1-1.3.4

[26] "Pole-Zero Multipulse Speech Representation Using Harmonic Modeling in the Frequency Domain"
TRANCOSO I.M., ALMEIDA L.B., TRIBOLET J.M.
IEEE Proc. Int. Conf. on ASSP, 1985, pages 260-263

Thèses:

[27] "Time-Domain Coding of (Near) Toll Quality Speech at Rates Below 16 kb/s"
KROON P.
1985, Delft University

[28] "Etude, Simulation et Mise en Oeuvre sur Microprocesseur de Codeurs Prédictifs Multi-Impulsionnels
LANCON E.
1985, Université de Nice

[27] "Analyse et Restitution du Signal de Parole par Modèles Evolutifs"
OMNES-CHEVALIER M.C.
1986, ENST Paris

Ouvrages:

[28] "Computer Speech Processing"
Edited by FALLSIDE F., WOODS W.A.

CHAPITRE IV

MODELISATION VECTORIELLE

DE L'EXCITATION

CHAPITRE IV

IV.1 INTRODUCTION:

Nous venons de voir dans le chapitre précédent, que la modélisation multi-impulsionnelle opère une réduction d'information sur le signal d'excitation en forçant à zéro environ 7 échantillons sur 8. Malgré cela, le débit associé à l'excitation multi-impulsionnelle représente une part importante dans le débit global. Aussi, réduire le débit global passe par un encodage plus sévère du signal d'excitation. L'étude de ceci a fait et fait encore l'objet de nombreux travaux de recherche dans le domaine de la quantification vectorielle [2,3,4,8,9].

Dans ce chapitre, nous allons décrire différentes procédures de modélisation de l'excitation, plus ou moins complexes, dont la modélisation et le codage du signal d'excitation sont basés sur la procédure d'analyse synthèse, dans sa forme vectorielle.

Nous commencerons par le codeur à excitation par code, introduit par ATAL [3,10], dont la procédure de modélisation de l'excitation peut être présentée comme la formulation stochastique de la procédure d'analyse synthèse.

Nous décrirons ensuite un codeur à excitation stochastique optimale, dont la formulation de la procédure de modélisation de l'excitation est également déduite de la méthode d'analyse synthèse. Cette formulation est la plus générale dans le sens où la procédure minimise l'erreur par rapport aux 3 paramètres qui décrivent l'excitation, à savoir:

- l'index de la séquence d'excitation
- la phase de la séquence d'excitation
- le facteur de gain à appliquer à la séquence d'excitation

Enfin, nous proposons un codeur à excitation multi-impulsionnelle vectorielle, dont le signal d'excitation est modélisé par une procédure qui combine l'analyse multi-impulsionnelle et l'analyse stochastique.

IV.2 MODELISATION DE L'EXCITATION PAR CODE AVEC PREDICTEUR A LONG TERME BOUCLE

Le codeur à excitation par code consiste à appliquer une quantification vectorielle sphérique [1] au signal d'excitation. Mais sa particularité réside avant tout dans le choix de la métrique. En effet celle-ci prend en compte l'effet de masquage qu'introduit le filtre perceptuel.

Le principe du codeur à excitation par code s'explique aisément à partir de la synthèse. Le signal synthétique est reconstruit à partir de séquences d'excitations filtrées successivement à travers les prédicteurs à long terme et à court terme. Chaque séquence d'excitation est caractérisée par son index i dans le dictionnaire et par le facteur de gain A^i qu'il faut lui appliquer. Notons également que les séquences d'excitation v^i sont normalisées, c'est à dire que:

$$\sum_n v_n^{i\,2} = 1 \qquad (4.1)$$

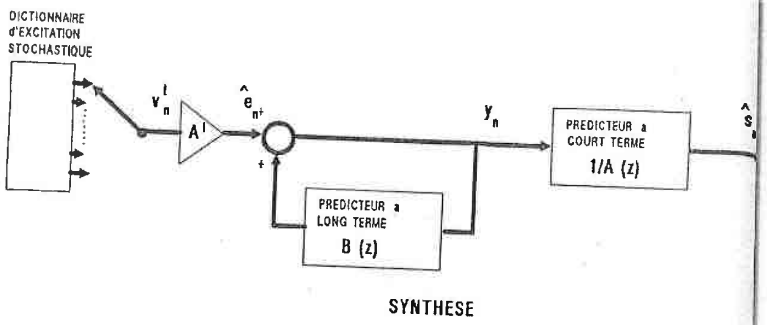


Figure 4.1: Partie synthèse du codeur à excitation par code

A l'analyse, la procédure de modélisation proposée par ATAL [3] met en oeuvre un prédicteur à long terme transversal. Elle approxime le signal d'erreur e_n, qui est segmenté en blocs de 5 ms (40 échantillons), par des séquences d'excitations v_n^i de même longueur. Ces séquences sont choisies dans le dictionnaire de façon à minimiser au sens des moindres carrés le signal d'erreur perceptuel u_n représentatif de la différence entre le signal original et le signal synthétique reconstruit à partir de celles-ci.

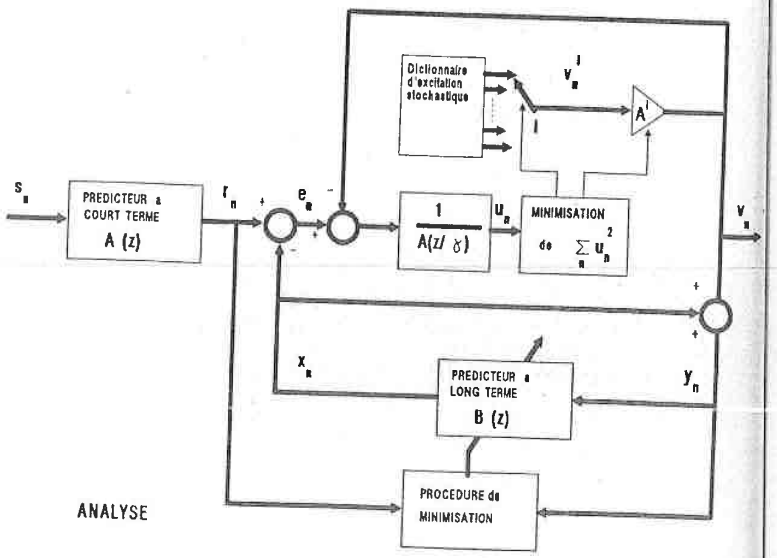


Figure 4.2: Partie analyse du codeur à excitation par code

Dans la suite de ce paragraphe, nous proposons de décrire un codeur à excitation par code qui intègre d'une part un prédicteur à long terme bouclé optimal simplifiée d'autre part un dictionnaire déterminé de manière statistique. Les avantages du prédicteur à long terme bouclé ont été décrits en détail dans le chapitre II. L'utilisation d'un dictionnaire statistique, procure à performance égale, une réduction sensible de la taille du dictionnaire, offrant du même coup une réduction de la complexité.

IV.2.1 PROCEDURE DE MODELISATION DE L'EXCITATION STOCHASTIQUE:

La séquence d'excitation optimale est caractérisée par l'index i de l'innovation v^i et du facteur de gain A^i associé. Pour la séquence d'excitation i, la détermination de ses paramètres revient à minimiser au sens des moindres carrés l'erreur perceptuelle entre le signal original s_n et le signal synthétique $\hat{s}_n$:

$$\epsilon = \sum_{n \in \mathcal{P}} [(s_n - \hat{s}_n) * w_n]^2 \qquad (4.2)$$

où * représente le produit de convolution.

Comme pour le codeur à excitation multi-impulsionnelle, le prédicteur à long terme bouclé est exclu de la procédure de modélisation, de façon à simplifier celle-ci.

Il en découle les deux variantes que nous avons présentées dans le paragraphe 5 du chapitre III. L'erreur quadratique ϵ dépend de la séquence d'excitation i. Modéliser l'excitation revient donc à déterminer l'index i de la séquence d'excitation ainsi que le facteur de gain A^i associé qui minimise l'erreur quadratique ϵ. Sachant que le dictionnaire d'excitation compte D séquences, les deux variantes s'écrivent:

VARIANTE 1:

$$\epsilon^i = \sum_{n \in \mathcal{P}} [(e_n - A^i . v_n^i) * f_n]^2 \quad \text{pour } i = 1 \text{ à } D \qquad (4.3)$$

VARIANTE 2:

$$\epsilon^i = \sum_{n \in \mathcal{P}} [u_n - A^i . v_n^i * f_n]^2 \quad \text{pour } i = 1 \text{ à } D \qquad (4.4)$$

u_n est le signal perceptuel du signal original s_n auquel a été enlevée la contribution du prédicteur à long terme. u_n correspond également au résiduel à long terme e_n reconstruit à travers le filtre de pondération F(z).

$$u_n = e_n * f_n = \sum_{i=n-MA}^{n} e_n . f_{n-i} \qquad (4.5)$$

La plupart des auteurs [10,11] retiennent la variante 2, car elle se prête bien au prédicteur à long terme transversal. Par contre, notre approche retient la variante 1 car, comme pour le codeur à excitation multi-impulsionnelle, celle-ci est plus adaptée au prédicteur à long terme bouclé.

L'erreur quadratique ϵ^i à minimiser peut encore s'écrire sous forme matricielle, à savoir:

$$\epsilon^i = [\ (\ {}^T e - A^i.{}^T v^i)\ F\ ]^2 \qquad (4.6)$$

où ^{T} signifie le transposé.

avec ${}^T e = e_0, e_1, \ldots, e_{N'-1}$ le segment de résiduel à modéliser

${}^T v^i = v^i_0, v^i_1, \ldots, v^i_{N'-1}$ une séquence d'excitation d'index i

$$F = \begin{vmatrix} f_0 & f_1 & f_2 & . & . & f_{N'-2} & f_{N'-1} \\ 0 & f_0 & f_1 & . & . & f_{N'-3} & f_{N'-2} \\ 0 & 0 & f_0 & . & . & f_{N'-4} & f_{N'-3} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & . & . & . & . & f_0 & f_1 \\ 0 & . & . & . & . & 0 & f_0 \end{vmatrix} \qquad (4.7)$$

La séquence d'excitation optimale est celle qui minimise ϵ^i:

$$\frac{\delta\ \epsilon^i}{\delta\ A^i} = 0 \Rightarrow - {}^T e.F.{}^T F.v^i + A^i.{}^T v^i.F.{}^T F.v^i = 0 \qquad (4.8)$$

Soit C la matrice d'autocorrélation de la réponse impulsionnelle définie:

$$C = \begin{vmatrix} c_0 & c_1 & c_2 & . & . & . & c_{N-1} \\ c_1 & c_0 & c_1 & . & . & . & c_{N-2} \\ c_2 & c_1 & c_0 & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ c_{N-2} & c_{N-3} & . & . & . & c_1 \\ c_{N-1} & c_{N-2} & . & . & . & c_0 \end{vmatrix} \qquad (4.9)$$

où

$$c_i = \Sigma\ f_n.f_{n-i}$$

Cette matrice, qui est carrée et symétrique, est de la dimension du segment de signal à modéliser. S'agissant de la matrice d'autocorrélation de la réponse impulsionnelle du filtre de pondération, les termes de corrélation c_i pour i supérieur à MA-1 (la durée de la réponse impulsionnelle du filtre de pondération) sont nuls. Ainsi le triangle supérieur droit et le triangle inférieur gauche de la matrice sont nuls.

La prise en compte des effets de bord, dans cette formulation, à des incidences sur la taille de la matrice d'autocorrélation. En effet, le traitement exacte nécessite la prise en compte des échantillons adjacents en amont et en aval au segment modélisé. Par conséquent, la dimension de la matrice passe de N'xN' à N'+MA x N'+MA. Cependant, une approximation raisonnable consiste à négliger ces effets de bord. Dans ce cas la dimension de la matrice d'autocorrélation se limite à N'xN'.

La valeur optimale du facteur de gain de l'excitation d'ordre i est donnée par:

$$A^i = \frac{{}^T e.C.v^i}{{}^T v^i.C.v^i} \qquad (4.10)$$

La matrice C étant symétrique ϵ^i s'écrit:

$$\epsilon^i = {}^T e.C.e - \frac{[\ {}^T e.C.v^i\ ]^2}{{}^T v^i.C.v^i} \qquad (4.11)$$

Minimiser ϵ^i revient donc à maximiser Γ^i par rapport à v^i:

$$\Gamma^i = \frac{[\ {}^T e.C.v^i\ ]^2}{{}^T v^i.C.v^i} \qquad (4.12)$$

Etudions l'influence du facteur perceptuel sur Γ^i.

Si le facteur $\tau = 0$, la matrice C se réduit à une matrice diagonale unitaire. Γ^i s'écrit alors:

$$\Gamma^i = \frac{[\ {}^T e.v^i\ ]^2}{{}^T v^i.v^i} \qquad (4.13)$$

La modélisation de l'excitation par code est alors équivalente à une quantification vectorielle directe du signal résiduel à long terme e_n.

Si le facteur $\tau = 1$, la matrice C correspond à la matrice d'autocorrélation du filtre de synthèse. La modélisation de l'excitation par code est équivalente à une quantification vectorielle non-pondérée du signal original s_n.

Pour les valeurs standards des paramètres, à savoir:

- N': dimension des blocs d'excitation = 40 échantillons
- D : dimension du dictionnaire = 1024 séquences d'excitation

la modélisation de l'excitation à partir des relations 4.10 et 4.12 ci-dessus procure une complexité de calcul colossale de 320 millions x,+/s. Réduire cette complexité passe par la diagonalisation de la matrice d'autocorrélation C. Pour cela une transformation orthogonale est appliquée à la matrice F. Cette procédure rapide est décrite dans le prochain paragraphe.

IV.2.2 MODELISATION A L'AIDE D'UNE PROCEDURE RAPIDE:

La complexité de cette procédure provient du produit $C.v^i$ au numérateur et au dénominateur. TRANCOSO [11], qui a mis en oeuvre la variante 2, propose de substituer la matrice F par une matrice diagonale D, sachant que toute matrice peut être décomposée en deux matrices orthogonales M, L et une diagonale. Cette substitution peut être également appliquée à la variante 1.

Soit C la matrice d'autocorrélation définie par:

$$C = M\,D\,{}^{T}L \qquad (4.14)$$

où M et L sont des matrices de transformation et D la matrice diagonale de C [13].

En substituant la matrice C par sa forme décomposée dans les relations 4.10 et 4.11, il résulte que:

$$A^i = \frac{{}^{T}e.M.D.{}^{T}L.v^i}{{}^{T}v^i.M.D.{}^{T}L.v^i} \qquad (4.15)$$

ainsi que:

$$\epsilon^i = {}^{T}e.M.D.{}^{T}L.e - \frac{[\,{}^{T}e.M.D.{}^{T}L.v^i\,]^2}{{}^{T}v^i.M.D.{}^{T}L.v^i} \qquad (4.16)$$

Le vecteur v^i et la matrice L sont indépendants du temps. Aussi le produit ${}^{T}v^i.L$ peut être réaliser une fois pour toute. En substituant:

$$x = {}^{T}M.e \quad \text{et} \quad g^i = {}^{T}L.v^i \qquad (4.17)$$

les relations 4.20 et 4.21 s'écrivent:

$$A^i = \frac{{}^{T}x.D.g^i}{{}^{T}g^i.D.g^i} \qquad (4.18)$$

et:

$$\epsilon^i = {}^{T}x.D\,x - \frac{[\,{}^{T}x.D.g^i\,]^2}{{}^{T}g^i.D.g^i} \qquad (4.19)$$

Ainsi, le dictionnaire d'excitation ne contient pas les séquences d'excitation v^i mais g^i qui sont le résultat du produit de v^i par L. Grâce à cette transformation la complexité de la modélisation de l'excitation, pour des paramètres de valeur standard, est ramenée à environ 24 millons x,+/s. De plus elle utilise directement le signal résiduel e_n. Cette procédure simplifiée illustrée par la figure ci-dessous.

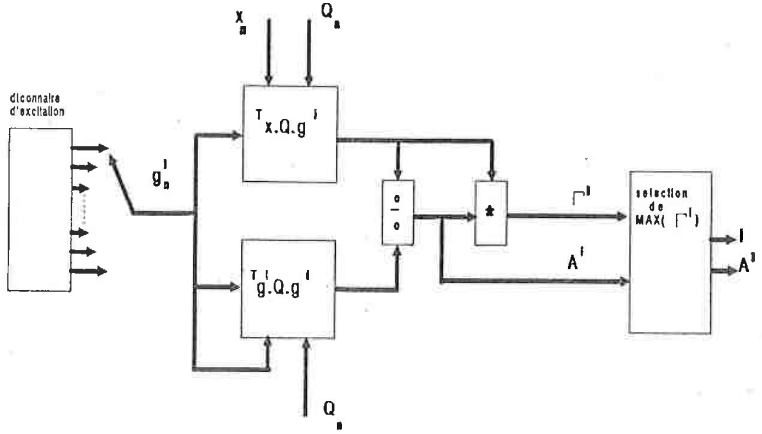


Figure 4.3: Schéma bloc du codeur à excitation par code simplifié

IV.2.3 EXTRACTION STATISTIQUE DU DICTIONNAIRE D'EXCITATION:

Le nombre de vecteurs standard que contient le dictionnaire d'excitation est de 1024 vecteurs de dimension 40. Sachant qu'un échantillon est représenté sur 10 bits, l'effort de mémorisation engendré uniquement par ce dictionnaire est de 400 kbits. De façon à éviter la mémorisation du dictionnaire, plusieurs auteurs proposent, l'utilisation d'un dictionnaire, dont les séquences d'excitations sont calculées à partir de fonctions aléatoires. Néanmoins cette solution a l'inconvient de procurer un coût supplémentaire en opérations qui s'élève à plusieurs dizaines de millions de multiplications et additions par seconde. En d'autres termes, la génération des vecteurs d'excitation, représente une complexité approximativement égale à celle de la procédure de modélisation rapide.

Nous proposons d'utiliser un dictionnaire d'excitation statistique. Il a l'avantage d'être de taille plus restreinte, dans la mesure où son extraction est faite sur une base de données de signal de parole dont les caractéristiques spectrales sont identiques à celles du signal à coder. L'inconvénient de cette approche est que les performances du dictionnaire se

dégradent dès que les caractéristiques du signal à coder et celles du signal qui a servi à l'extraction du dictionnaire sont différentes.

L'extraction statistique du dictionnaire des séquences d'excitation fait appel à un algorithme à seuil, que nous avons présenté dans le paragraphe 3 du chapitre I. Notre choix c'est porté sur la méthode à seuil, compte-tenu du compromis intéressant en terme de complexité et de convergence, qu'elle présente. Pour une telle méthode, les éléments déterminants sont, la mesure de distance ou métrique retenue, et la valeur du seuil de classification.

CHOIX DE LA MESURE DE DISTANCE:

La mesure utilisée est celle définie par le relation 4.12. La classification se faisant à partir d'un seuil fixe, il est nécessaire de normaliser cette métrique, qui est homogène à une intercorrélation, de telle sorte qu'elle varie entre 0 et 1. Les séquences d'excitation v^i du dictionnaire étant normalisées, il reste à normaliser uniquement le signal résiduel à long terme e_n. La relation définissant la métrique s'écrit alors:

$$\Gamma'^{\,i} = \frac{[\ {}^{T}e.C.v^i]}{{}^{T}v^i.C.v^i} \times \frac{{}^{T}v^i.C.v^i}{{}^{T}e.C.v^i} \qquad (4.20)$$

$\Gamma'^{\,i} = 1$, lorsque le segment de signal résiduel e_n et la séquence d'excitation $v_n{}^i$ sont colinéaires. En revanche, $\Gamma'^{\,i} = 0$, lorsque e_n et $v_n{}^i$ sont orthogonaux. On peut donc prévoire, que plus la valeur du seuil se rapproche de 1, plus le nombre de classes ou de vecteurs du dictionnaire sera élévé.

CHOIX DU DICTIONNAIRE:

La démarche retenue pour déterminer le dictionnaire, consiste à faire varier le seuil par petits pas (de manière croissante). A chaque expérimentation, des mesures de rapport signal sur bruit segmental, ainsi que des tests d'écoute permettent de mesurer les performances du dictionnaire extrait.

La figure ci-dessous illustre la compression qu'offre la procédure de classification en fonction de la valeur du seuil. Ainsi passer d'un seuil de 0.2 à 0.3 multiplie approximativement par un facteur 3 la taille du dictionnaire. On peut noter également, que la taille du dictionnaire, pour un seuil donné, dépend également du facteur perceptuel τ. Diminuer τ augmente de manière significative la taille du dictionnaire. Ceci s'explique du fait que plus le facteur perceptuel est faible, plus le signal perceptuel se rapproche du signal résiduel, de sorte que l'information de phase ou de périodicité qu'il contient est prépondérante.

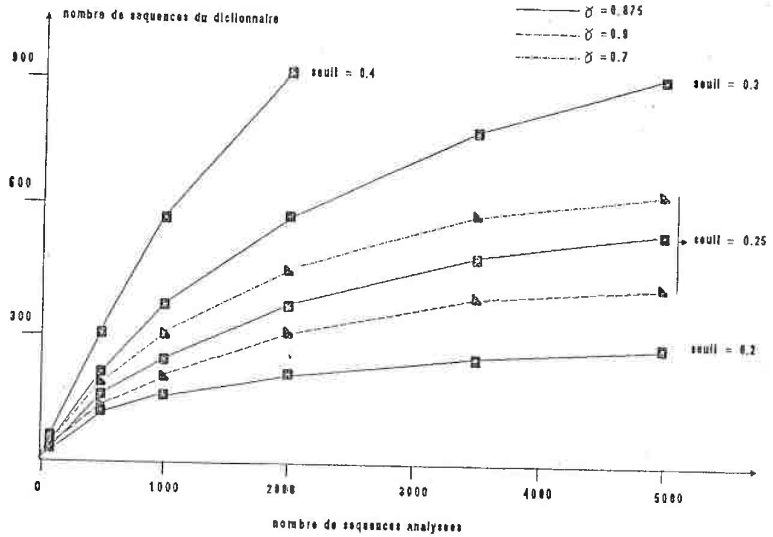


Figure 4.4: Evolution de la taille du dictionnaire en fonction du nombre de trames analysées avec pour paramètres, le seuil de classification et le facteur perceptuel

La figure ci-dessous, met en évidence que les performances, en terme de rapport signal sur bruit segmental du codeur à excitation par code sont intimement liées à la dimension D du dictionnaire d'excitation. Pour un seuil de classification de 0.25, la dimension du dictionnaire peut être limitée à 512 vecteurs. Le rapport signal sur bruit segmental que procure le codeur à excitation par code avec un tel dictionnaire est de 12.5 dB. Des évaluations objectives montrent que la qualité que procure le codeur à excitation multi-impulsionnelle, qui est de 14 dB, est supérieure de 1.5 dB. Mais ces 1.5 dB d'écart ne sont pas très significatifs. En effet pour le codeur à excitation par code l'excitation est implicitement quantifiée. Aussi, pour mettre en parallèle les performances auditives des deux codeurs, il est nécessaire de prendre comme référence le rapport signal sur bruit segmental que procure le codeur à excitation multi-impulsionnelle lorsque le signal d'excitation est quantifié. Il est 13.5 dB. La différence qui reste, soit 1 dB est liée à la minimisation globale du facteur de gain de l'excitation par code, car ce facteur est ajusté globalement pour chaque segment de 40 échantillons.

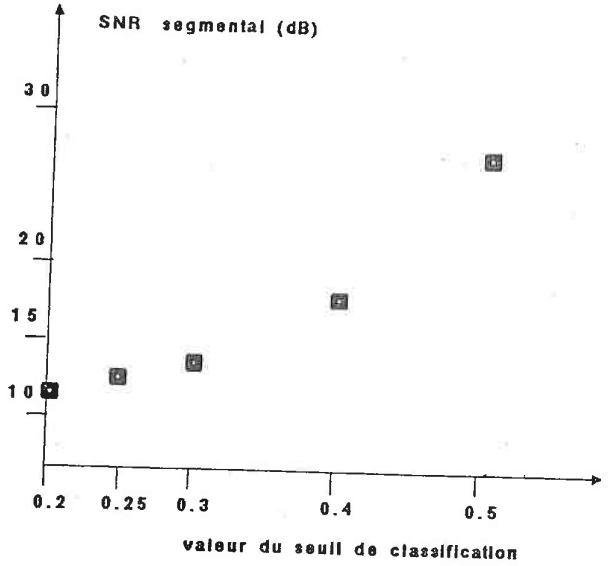


Figure 4.5: Evolution du rapport signal sur bruit segmental en fonction du seuil de classification

La figure 4.6 visualise pour un segment de signal voisé, l'évolution de la distance Γ en fonction de l'index i du dictionnaire sur un segment de signal voisé. La procédure de classification retient la séquence d'excitation qui procure l'intercorrélation maximum. On constate que pour ce segment d'excitation, il n'y a que 3 ou 4 séquences qui emergent réellement. Ceci est signe que le dictionnaire offre une bonne couverture de l'excitation.

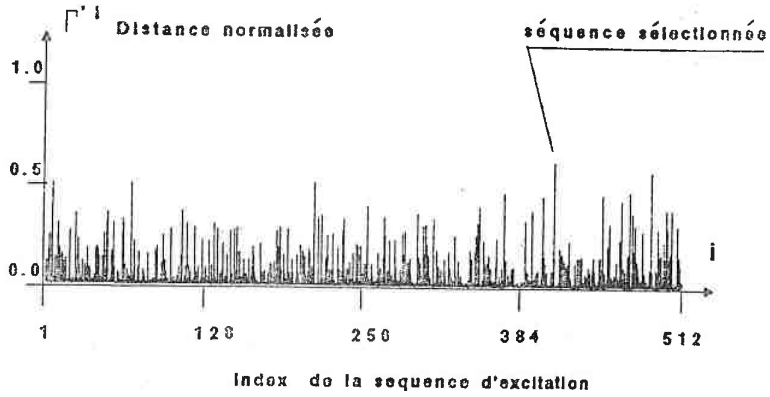


Figure 4.6: Evolution de la distance de classification Γ^i en fonction de l'index du dictionnaire de 512 séquences excitation

De ces expérimentations, nous pouvons conclure qu'un dictionnaire de quelques 512 vecteurs de dimension 40 procure une qualité comparable à une excitation multi-impulsionnelle de 1000 impulsions par secondes, soient respectivement des débits de 2.6 kbits/s (1.8 kbits/s pour l'index i, 0.8 kbits/s pour le facteur de gain A^i) et de 8 kbits/s pour l'excitation.

La figure 4.7 visualise les différents signaux significatifs du codeur à excitation par code. Elle fait apparaître que le signal d'excitation est visuellement différent du signal résiduel à long terme, en particulier pour les instants compris entre 70 et 120 ms. Ceci est lié au facteur perceptuel qui a pour rôle de pondérer la répartition spectrale de l'erreur en pénalisant les régions spectrales à forte energie, à savoir les fréquences correspondant au fondamental. Néanmoins, le prédicteur à long terme bouclé, permet de reconstruire convenablement le signal résiduel y_n de façon à restituer un signal synthétique proche du signal original.

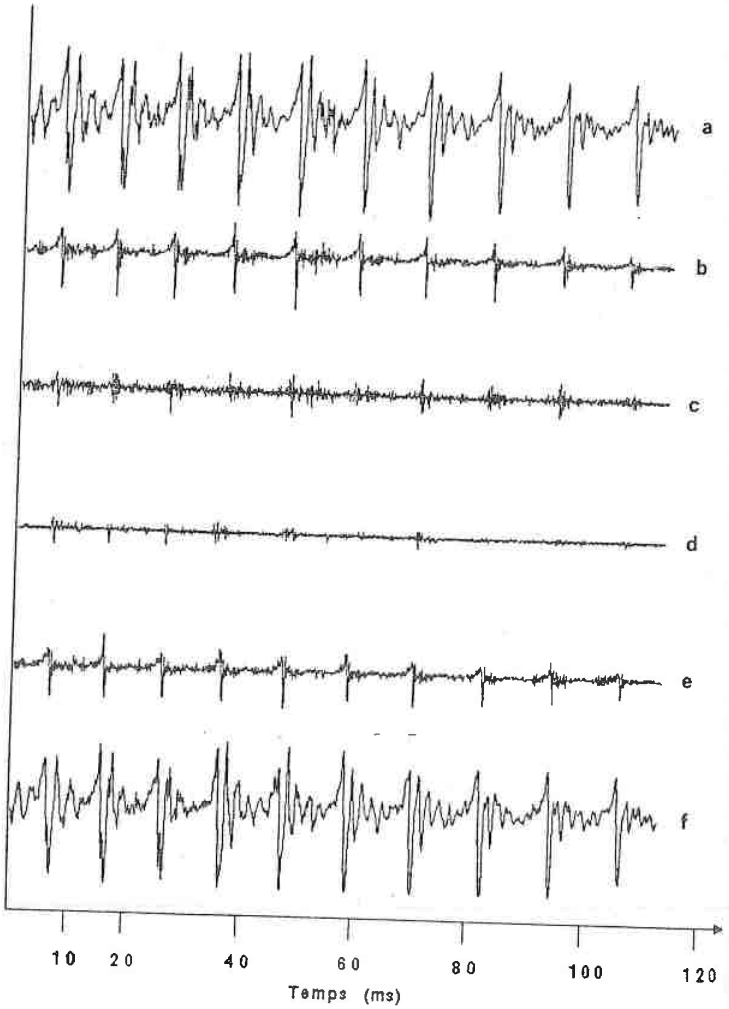


Figure 4.7: Evolution temporelle des signaux caractéristiques, a) signal original, b) signal résiduel à court terme, c) signal résiduel à long terme, d) signal d'excitation par code, e) signal résiduel à court terme reconstruit, f) signal synthétique

IV.2.4 LIMITATION DE LA MODELISATION PAR CODE DE L'EXCITATION:

Comme on peut s'y attendre ce codeur restitue difficilement les segments de signal, qui sont de nature impulsionnelle, comme les occlusives, compte-tenu du critère quadratique et de la minisation globale qui caractérisent la méthode d'analyse de l'excitation (fig 4.8). Ce codeur permet toutefois de restituer de la parole de qualité sub téléphonique tout en limitant le débit

de l'excitation à 0.325 bits par échantillon d'excitation. Rappelons que pour le codeur à excitation multi-impulsionnelle ce dernier se situe plutôt à 1 bit par échantillon.

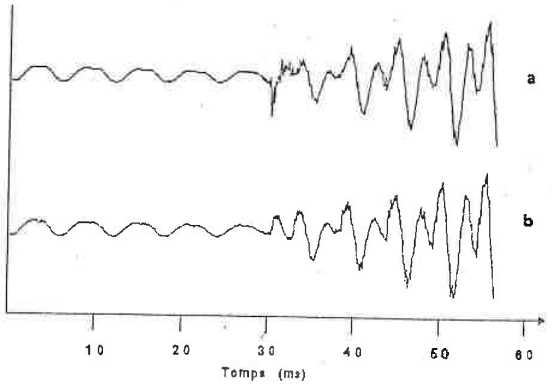


Figure 4.8: Visualisation a) du signal original, b) du signal synthétique

Le nombre d'opérations nécessaires par seconde, est détaillé dans le tableau ci-dessous. Le facteur ϕ représente le nombre de fenêtres de modélisation LPC (de longueur **N**) par seconde. Quant au facteur ϕ', il représente le nombre de fenêtres de modélisation de l'excitation (de longueur N').

Les valeurs des paramètres que cette étude a permis de dégager sont précisées ci-dessous:

- N : dimension des segments d'analyse LPC = 80 échan. (10 ms)
- MP: ordre du prédicteur à court terme = 16
- MA: durée de l'autocorrélation de la réponse impulsionnelle du filtre F(z) = 20 éch. (1.25 ms)
- N': dimension des segments de la modélisation du prédicteur à long terme et de la modélisation de l'excitation = 40 éch. (5 ms)
- D :nombre de vecteurs du dictionnaire d'excition = 512

Forme directe	multi.,addit./s	div./s
réponse impulsionnelle de F(z)	$3.MP.MA.\phi$	
autocorrélation de F(z)	$MA^2.\phi$	$MA.\phi$
critère + facteur de gain	$D.(N'^2+ 2.N'+ 1).\phi'$	$D.\phi'$
complexité pour les valeurs standards	$172\ 10^6$	104400
forme rapide	**multi.,addit./s**	**div./s**
réponse impulsionnelle de F(z)	$3.MP.MA.\phi$	
autocorrélation de F(z)	$MA^2.\phi$	$MA.\phi$
diagonalisation de la matrice	$2N'^2+N'^3.\phi$	
critère + facteur de gain	$N'.(3.D + 1).\phi'$	$D.\phi'$
complexité pour les valeurs standards	$19\ 10^6$	104400

Tableau 4.1: Complexité liée à la modélisation de l'excitation par code

L'introduction de la diagonalisation permet de réduire d'un facteur N'/3 environ la complexité à l'analyse. En revanche, à la synthèse il est nécessaire de déduire, à partir des séquences g^i du dictionnaire, la séquence d'excitation v^i. La séquence v^i est donnée par la relation:

$$v^i = M.g^i \qquad (4.21)$$

Ce traitement représente à la synthèse un coût en opérations supplémentaires de $N'^2.\phi'$ multiplications et additions par seconde, soit 320000 x,+/s. Le tableau ci-dessus montre que pour la procédure rapide, la complexité est approximativement proportionnelle au nombre de vecteurs du dictionnaire D ainsi qu'à la dimension des vecteurs N'.

Il reste que, la complexité de traitement ainsi que l'effort de mémorisation qu'engendre ce codeur, même dans la formulation rapide avec diagonalisation, sont à la limite de la technologie actuelle. En effet la nouvelle génération de microprocesseurs de traitement de signal réalisent jusqu'à 10 millions de multiplications et additions par seconde. En réalité, comme les procédures de traitement de signal ne consistent pas uniquement à réaliser des produits scalaires, cette puissance de traitement chute dans un rapport 1.5 à 2. C'est le cas en particulier pour les traitements aussi complexes et diversifiés que ceux qu'intègrent les procédures de codage de la parole.

Compte-tenu des limitations que nous venons de présenter, la réduction de la complexité du codeur à excitation par code passe par la réduction de la taille du dictionnaire. Une première approche qui répond à cette contrainte, consiste à réduire la dimension de N' . Ceci est possible par un filtrage suivi d'un sous-échantillonnage du signal résiduel à modéliser. Une autre possibilité consiste à réduire la dimension de N' aux échantillons les plus significatifs dont la répartition des positions sur le segment de signal n'est pas régulière. On obtient ainsi, une réduction de la complexité et de la taille du dictionnaire d'excitation, d'un facteur au moins égal au rapport de réduction de N'. Une deuxième approche consiste à structurer le dictionnaire d'excitation en plusieurs sous-dictionnaires [4,8]. C'est à l'aide d'un critère simple que le sous-dictionnaire est sélectionné puis parcouru partiellement afin de trouver la séquence d'excitation.

Dans la suite de ce chapitre, nous présentons deux procédures de modélisation mixte de l'excitation, qui sont dérivées de ces deux approches. Elles offrent d'une part une réduction significative de la taille du dictionnaire (donc de la complexité), d'autre part elles permettent de couvrir de manière continue des débits compris entre 6 et 12 kbits/s.

IV.3 MODELISATION MIXTE DE L'EXCITATION:

Nous venons de décrire jusqu'à présent, le codeur à excitation multi-impulsionnelle et le codeur à excitation par code. Il apparaît dans le tableau ci-dessous, que ces deux codeurs sont complémentaires.

	modélisation multi-impulsionnelle de l'excitation	**modélisation stochastique de l'excitation**
points forts	faible complexité de l'ordre de 0.6 10^6 x,+/s	débit lié à l'excitation de l'ordre de 0.325 bit/échan. soit environ 2.6 kbits/s
points faibles	débit minimum pour l'excitation 1 bit/échan. soit 8 kbits/s	complexité superieure à 12 millions de x,+/s

Tableau 4.2: Points forts et points faibles des codeurs à excitation multi-impusionnelle et à excitation par code

Ce tableau peut se résumer de la manière suivante: le codeur à excitation multi-impulsionnelle ne permet pas de restituer de la parole de qualité sub-téléphonique pour des débits inférieurs à 8 kbits/s, mais il a l'avantage d'être d'une complexité qui n'excède pas 1 millions de x,+/s. En revanche, le codeur à excitation par code permet de réduire ce débit à moins de 3 kbits/s, mais au prix de quelques 19 10^6x,+/s.

C'est pourquoi, nous avons envisagé deux codeurs, à savoir:

- le codeur à excitation optimale par code
- le codeur à excitation multi-impulsionnelle vectorielle

qui permettent de couvrir de manière continue les débits intermédiaires en combinant les modes d'excitation multi-impulsionnelle et par code.

IV.3.1 MODELISATION OPTIMALE PAR CODE DE L'EXCITATION:

Dans un codeur à excitation par code tel que celui décrit dans le paragraphe précédent, l'information de phase de l'excitation est contenue de manière implicite dans le dictionnaire d'excitation. On peut observer qu'un nombre important de séquences d'excitation v^i sont peu différentes à un décalage près. Ceci est illustré, pour deux séquences d'excitation, par la figure 4.9 . De ce fait, en faisant abstraction de l'information de phase les vecteurs v^i peuvent être rangés dans un nombre très limité de classes.

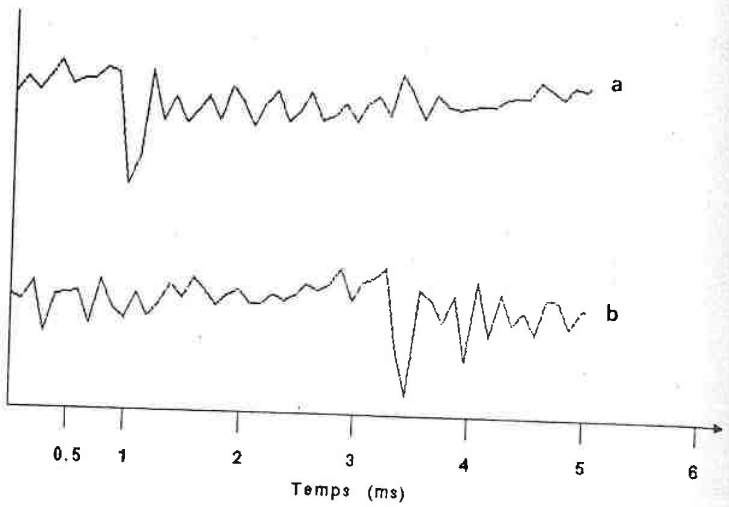


Figure 4.9: visualisation de deux séquences d'excitation semblables à un décalage près

Le calcul et la transmission explicite de cette information de phase, comme pour le codeur à excitation multi-impulsionnelle, permet de limiter chaque classe à une seule séquence d'excitation. Par ce biais, l'extraction statistique du dictionnaire d'excitation permet de limiter à moins d'une dizaines le nombre de séquences d'excitation.

IV.3.1.1 PROCEDURE DE MODELISATION OPTIMALE PAR CODE:

La procédure de modélisation optimale par code, que nous proposons, met en oeuvre la formulation la plus générale de la procédure d'analyse par synthèse, d'où son nom de procédure optimale par code. Par conséquent, la recherche de l'excitation optimale revient à minimiser l'erreur quadratique par rapport aux 3 paramètres qui sont:

- I le numéro de la séquence d'excitation
- A le facteur de gain à appliquer à la séquence d'excitation
- M la position par rapport à l'origine de la séquence d'excitation

Cette procédure a été étudiée dans un premier temps sans prédiction à long terme. Ce n'est que dans un deuxième temps que le prédicteur à long terme bouclé a été introduit dans le codeur à excitation optimale par code.

Le signal résiduel est approximé par plusieurs séquences d'excitation qui peuvent se superposées partiellement. Dans le cas d'une modélisation sans prédiction à long terme, l'erreur quadratique à minimiser s'écrit pour la variante 1:

$$\varepsilon = \sum_{n=0} \left[s_n - \left(\sum_{k=1}^{K} A^k . v^k_{n-M_k} \right) * f_n \right]^2 \qquad (4.27)$$

Comme pour le codeur à excitation multi-impulsionnelle les séquences d'excitation sont déterminées de manière itérative. La formulation itérative s'écrit:

$$\varepsilon^k = \varepsilon^{k-1} - \left[\left(A^k . v^k_{M_k} \right) * f_n \right]^2 \qquad \text{pour } i = 1 \text{ à } D \qquad (4.28)$$

A l'itération k, la séquence d'excitation optimale est celle qui minimise ε^k par rapport à A, M et I. On obtient alors sous forme matricielle:

$$A^k = \frac{{}^t r^{k-1} . C . v^i_m}{{}^t v^i_m . C . v^i_m} \qquad (4.29)$$

La matrice C étant symétrique, l'erreur ε^k à l'itération k s'écrit:

$$\varepsilon^k = {}^t r^{k-1} . C . r^{k-1} - \frac{\left[{}^t r^{k-1} . C . v^i_m \right]^2}{{}^t v^i_m . C . v^i_m} \qquad (4.30)$$

Minimiser ε^k à l'itération k revient à maximiser $\Gamma^k_{i,m}$ qui s'écrit:

$$\Gamma^k_{i,m} = \frac{\left[{}^t r^{k-1} . C . v^i_m \right]^2}{{}^t v^i_m . C . v^i_m} \qquad (4.31)$$

$I^k = i$ pour

$|\Gamma^k_{i,m}|$ maximum avec $1 \leq i \leq D$ et $0 \leq m \leq N$ (4.32)

$M_k = m$ pour

$$A^k = \frac{{}^{t}r^{k-1}.C.v^{k}_{m_k}}{{}^{t}v^{k}_{m_k}.C.v^{k}_{m_k}} \qquad (4.33)$$

$\Gamma^k_{i,m}$ est en fait une matrice de dimension NxD où N et D représentent respectivement la dimension du segment de signal à modéliser et la dimension du dictionnaire. A l'itération k, la détermination de l'excitation revient à rechercher le maximum dans cette matrice $\Gamma^k_{i,m}$ dont les indices ligne et colonne caractérisent respectivement la position M_k et l'index I_k de l'excitation, qui minimise l'erreur quadratique ϵ^k.

Lorsque les 3 paramètres A^k, M_k, I_k de la séquence d'excitation correspondant à l'itération k sont connus, la contribution de celle-ci doit être soustraite au signal résiduel, à savoir:

$$r_n^k = r_n^k - v^{i}_{n}{}^{k} \qquad \text{pour } n = M_k \text{ à } M_k + MA \qquad (4.34)$$

La figure ci-dessous illustre sous forme de bloc diagramme, le principe de la procédure de modélisation par code optimal.

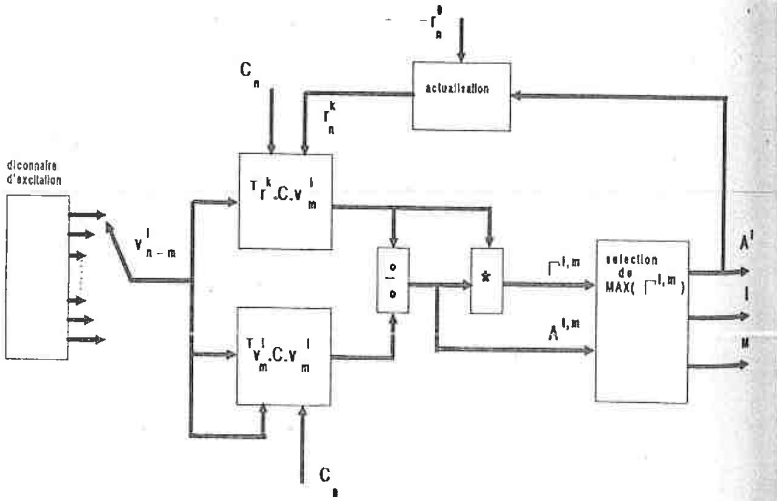


Figure 4.10: principe de la détermination de l'excitation optimale par code à l'itération k

La durée des séquences d'excitation a été choisie de façon à ce qu'elle puisse contenir la partie centrale du résiduel correspondant à l'impulsion glottique.

Le coût opératoire qu'engendre cette procédure de modélisation est détaillé dans le tableau ci-dessous. Les facteur ϕ et ϕ' représentent respectivement le nombre de fenêtres et sous-fenêtres de modélisation par seconde.

Les valeurs standards des paramètres que cette étude a permis de dégager sont précisées ci-dessous:

- N : dimension du segment d'analyse LPC = 160 éch.
- N': dimension du segment de modélisation de l'excitation = 80 éch.
- MP: ordre du filtre de synthèse en treillis = 12
- K : nombre de séquences d'excitation par segment N' = 10
- MA: durée des séquences d'excitations = 25

opérations	multi.,addit./s	div./s
réponse impulsionnelle de F(z)	3.MP.MA.ϕ	
autocorrélation de F(z)	MA^2.ϕ	MA.ϕ
critère + facteur de gain	2.K.D.MA.N'.ϕ'	K.D.ϕ'
actualisation	K.MA.ϕ'	

Tableau 4.3: Complexité correspondant à la modélisation par code optimal

On peut noter que la complexité de cette procédure de modélisation de l'excitation est approximativement proportionnelle au produit K.D.

IV.3.1.2 COMPARAISON DE LA MODELISATION PAR CODE OPTIMAL AVEC LES MODELISATIONS MULTI-IMPULSIONNELLE ET PAR CODE:

La modélisation multi-impulsionnelle et la modélisation par code sont des formes restrictives de la procédure décrite précédemment. En effet, celles-ci réduisent à deux le nombre de paramètres à déterminer en figeant de manière implicite le troisième.

Dans le cas du codeur à excitation multi-impulsionnelle, le paramètre i correspondant à l'index est éliminé, du fait que le dictionnaire d'excitation se limite à une seule séquence qui est l'impulsion de dirac. A l'itération k la matrice se réduit à un vecteur Γ^k_m de la dimension du segment de signal modélisé N.

Pour le codeur à excitation par code, une seule séquence d'excitation est déterminée par segment de signal à modéliser. D'autre part, la position relative M_k est éliminée, du fait que les séquences d'excitation succesives sont calées systématiquement en début de la fenêtre à modéliser. La matrice se réduit à un vecteur Γ_i dont la dimension correspond au nombre de séquences d'excitation du dictionnaire D.

IV.3.1.3 EXTRACTION DU DICTIONNAIRE D'EXCITATION:

GERSHO [7] a montré que le signal de parole se classe en 3 catégories selon son degré de corrélation. On trouve ainsi:

- les signaux fortement corrélés proches d'un signal sinusoïdal.
- les signaux moyennement corrélés
- les signaux décorrélés

Des exemples typiques de signal de parole correspondant à ces trois classes sont illustrés par la figure ci-dessous.

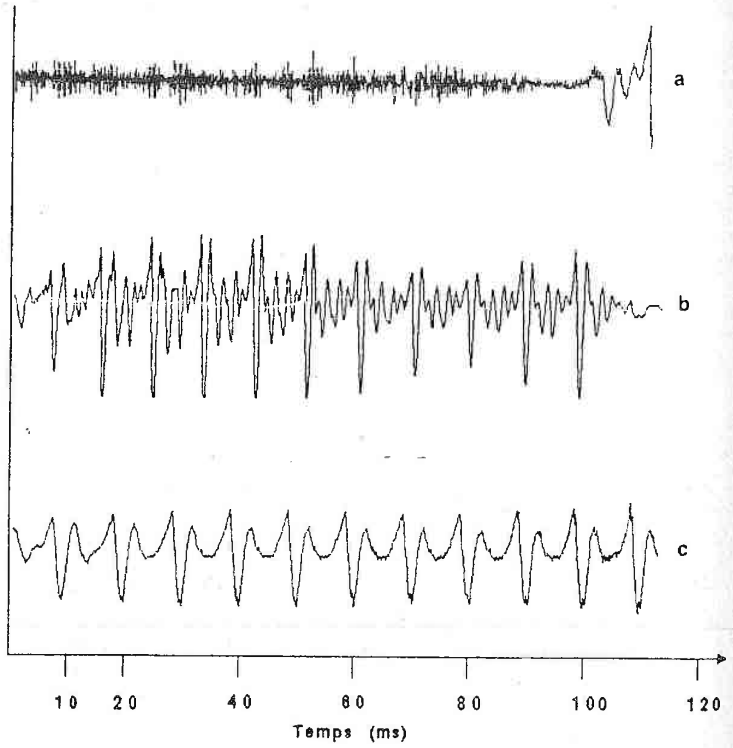


Figure 4.11: visualisation du signal de parole a)signal décorrélé, b)signal moyennement corrélé, c)signal fortement corrélé

Dans la mesure où le signal de parole peut être classé en ces trois catégories, il est peut être possible d'appliquer une telle classification au signal résiduel. Afin de le vérifier, nous avons appliquer au signal résiduel un algorithme de classification à seuil, qui doit nous permettre de constituer un dictionnaire de taille très limité.

La métrique, qui est appliquée dans l'algorithme de classification, correspond à l'intercorrélation $\Gamma_{i,m}$ normalisée, à savoir:

$$\Gamma'_{i,m} = \frac{\left[{}^{\tau_r.C.v_m^i} \right]^2}{{}^{\tau_{v_m^i}.C.v_m^i}} \times \frac{{}^{\tau_{v_m^i}.C.v_m^i}}{{}^{\tau_r.C.v_m^i}} \qquad (4.35)$$

L'extraction du dictionnaire d'excitation est une étape importante car elle conditionne pour une grande part les performances du codeur, en terme de qualité, de compexité et de débit.

Plusieurs dictionnaires de taille différente ont été extraits. Toutefois, de façon à faciliter la convergence de l'algorithme de classification, un premier dictionnaire est constitué manuellement avec une séquence d'excitation représentative de chacune des classes.

C'est ensuite, pour chacun des dictionnaires, que des mesures objectives de rapport signal sur bruit segmental ont permis de définir le nombre de séquences d'excitation par unité de temps de façon à produire un signal synthétique de même qualité sub-téléphonique, soit un rapport signal sur bruit de 14 dB environ. En observant le contenu des différents dictionnaires, on constate que les séquences d'excitation supplémentaires, qui se sont ajoutées aux 3 de bases, ressemblent beaucoup à des impulsions de dirac superposées à du bruit.

La tableau ci-dessous illustre l'évolution du nombre de séquences d'excitation par unité de temps ainsi que la complexité et le débit (se rapportant à l'excitation) que procure la modélisation par code optimal en fonction de la taille du dictionnaire. Le débit par séquence d'excitation est défini par la relation suivante:

$$\text{débit} = x + \log_2 D \qquad (4.36)$$

Le coefficent x représente le nombre de bits nécessaire au codage de la position et du facteur d'amplitude de la séquence d'excitation. Ce facteur qui fluctue entre 7 et 10 peut être, à première approximation, considéré comme une constante, égale à 9, indépendante du nombre de séquences d'excitations.

taille du dictionnaire	nombre de séquences par seconde	complexité en x,+/s	débit bits/s
3	1000	$12.2\ 10^6$	$11\ 10^3$
8	700	$22.6\ 10^6$	$8.4\ 10^3$
16	600	$38.6\ 10^6$	$7.8\ 10^3$
32	500	$64.2\ 10^6$	$7\ 10^3$

Tableau 4.4: Le nombre de séquences d'excitation par unité de temps, la complexité et le débit en fonction de la taille du dictionnaire, à rapport signal sur bruit constant de 14 dB environ

On peut conclure que l'augmentation de la taille du dictionnaire permet effectivement de réduire le nombre de séquences d'excitation par unité de temps. En effet pour un dictionnaire de 32 vecteurs, le nombre de séquences d'excitation par unité de temps peut être réduit dans un rapport 2 par rapport au nombre de séquences pour un dictionnaire de 3 vecteurs. Mais la complexité qu'engendre la solution avec un dictionnaire de 32 vecteurs n'est pas réaliste. Notons également qu'il existe un nombre minimum de séquences d'excitation, qui est de l'ordre de 500 par seconde, au dessous du quel le signal d'excitation n'est pas assez entretenu en échantillons non nuls pour fournir une qualité sub-téléphonique. Néanmoins en augmentant la dimension des séquences d'excitation à 30 échantillons par exemple, il est possible de corriger partiellement ce défaut, mais au détriment de la complexité.

Notre choix s'est donc porté sur le dictionnaire à 3 vecteurs auquel à été ajoutée une quatrième séquence d'excitation qui est l'impulsion de dirac car elle permet de compenser des erreurs impulsives très ponctuelles. Cette solution semble la plus intéressante en terme de complexité et de dimension du dictionnaire, même si le débit qu'elle procure n'est pas très bas. La figure 4.12 visualise ces quatres séquences d'excitation que nous avons retenues.

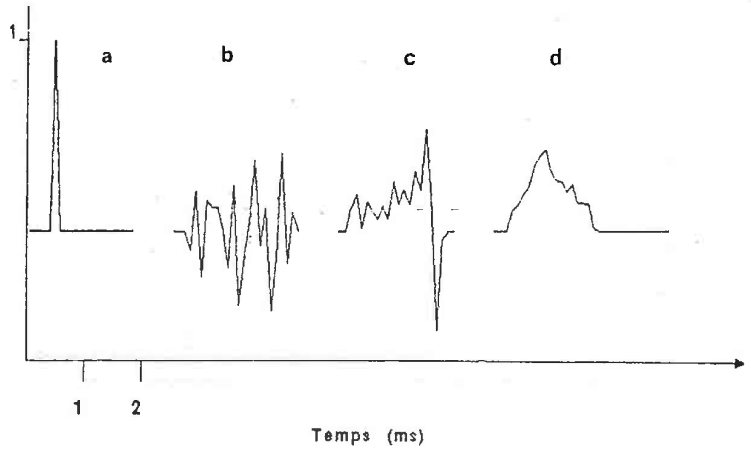


Figure 4.12: visualisation des séquences d'excitation a)dirac, b)excitation décorrélée; c)excitation faiblement corrélée, d)excitation fortement corrélée

Le signal d'excitation issu de cette modélisation sans prédiction à long terme peut être comparé au signal d'excitation multi-impulsionnel qui procure une qualité comparable.

La figure ci-dessous visualise un segment de signal d'excitation optimal par code, ainsi que le segment de signal d'excitation multi-impulsionnel équivalent. On peut constater que là où l'excitation multi-impulsionnelle nécessite 3 ou 4 impulsions pour restituer l'impulsion glotale, l'excitation optimale par code en nécessite 1 ou 2 séquences.

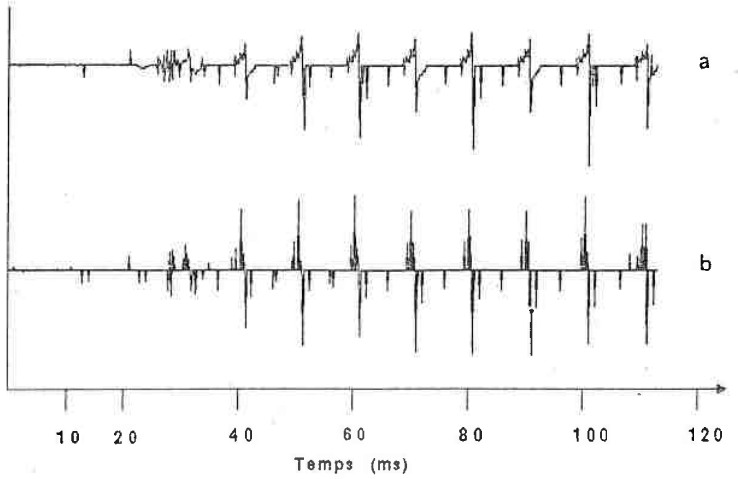


Figure 4.13: Visualisation a) du signal d'excitation optimale, b) du signal d'excitation multi-impulsionnelle équivalent

La figure ci-dessous visualise, pour des sons voisés, l'évolution de l'erreur quadratique normalisée au cours des itérations successives, dans le cas d'une modélisation multi-impulsionnelle et d'une modélisation par code optimal.

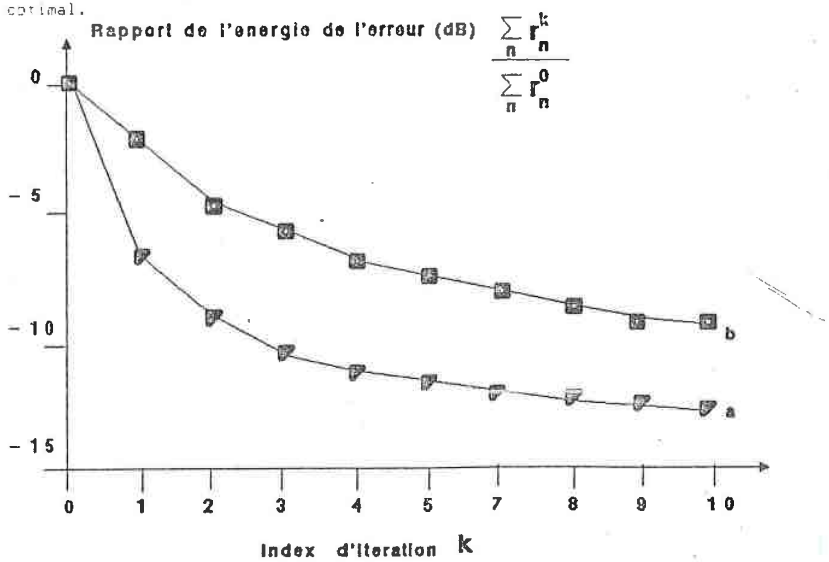


Figure 4.14: Evolution moyenne de l'erreur quadratique normalisée sur des segments de 10 ms pour des sons voisés, a)modélisation optimale par code, b)modélisation multi-impulsionnelle

Il apparaît clairement que la modélisation optimale par code est plus efficace que la modélisation multi-impulsionnelle surtout pour les 3 à 4 premières itérations.

A titre de comparaison la courbe ci-dessous visualise les performances (pour le dictionnaire à 4 séquences d'excitation) en fonction du nombre de séquences d'excitation par unité de temps, ceci pour les codeurs à excitation multi-impulsionnelle et à excitation optimale par code. On peut constater qu'à qualité égale, il suffit de 700 excitations par seconde, là ou le codeur à excitation multi-impulsionnelle en nécessite 1000. Néanmoins la transmission du numéro de la séquence d'excitation anéantit partiellement cette réduction de débit.

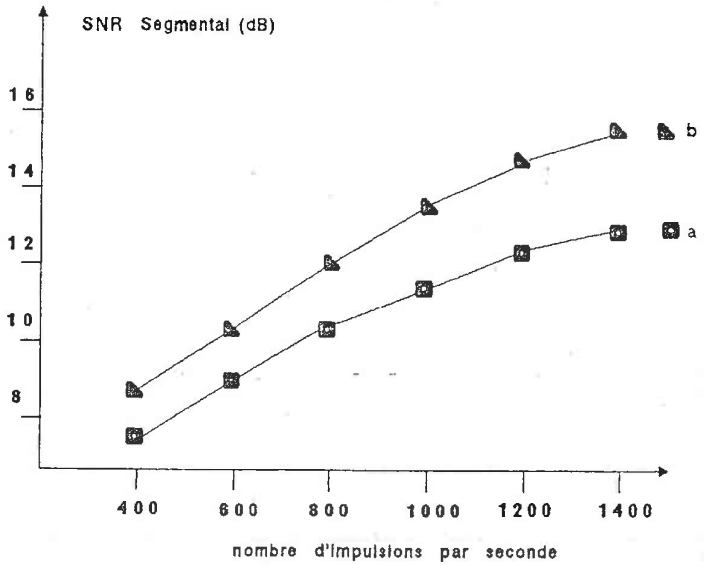


Figure 4.15: Evolution du rapport signal sur bruit segmental en fonction du nombre de séquences d'excitation par unité de temps avec le dictionnaire d'excitation à 4 séquences d'excitation.

IV.3.1.4 MODELISATION OPTIMALE PAR CODE AVEC PREDICTEUR A LONG TERME BOUCLE:

Comme pour la modélisation multi-impulsionnelle, l'introduction du prédicteur à long terme, augmente l'efficacité de la modélisation de l'excitation, notamment au niveau de l'entretien du signal d'excitation.
Dans ce cas, la procédure de modélisation par code optimal est appliquée au signal résiduel à long terme e_n.

Une démarche similaire à celle présentée dans le paragraphe précédent a permis d'extraire un dictionnaire d'excitation. La taille du dictionnaire a également été limitée à 4 séquences d'excitation, dont une est une impulsion de dirac.

L'introduction du prédicteur à long terme permet de réduire à 700, le nombre de séquenses d'excitation par seconde, ce qui porte d'une part la complexité à 11.4 10^6 multiplications et additions par seconde, d'autre part le débit à 7.7 kbits/s.

En observant de près l'occurrence des séquences d'excitation au cours des itérations successives (fig 4.17), on peut noter que l'impulsion de dirac apparaît d'autant plus fréquemment que l'ordre d'itération est élevé car le signal d'erreur se rapproche d'un bruit uniforme dans lequel il ne reste que quelques pics que les autres séquences d'excitation plus spécifiques n'ont pas pu modéliser. Ce constat nous permet de simplifier la procédure de modélisation, en limitant, pour les itérations d'ordre supérieur à 3, le dictionnaire d'excitation à une seule séquence d'excitation qui est l'impulsion de dirac. Ainsi le parcour global de tout le dictionnaire d'excitation peut être limité aux 3 premières itérations. Il en résulte d'une part une réduction significative de la complexité qui passe de 11.4 10^6 à 6.6 10^6 multiplications et additions par seconde. D'autre part le débit associé à l'excitation passe également de 7.7 à 6.9 kbits/s, car pour les séquences d'excitation correspondant aux itérations supérieures à 3, il est inutile de transmettre l'index.

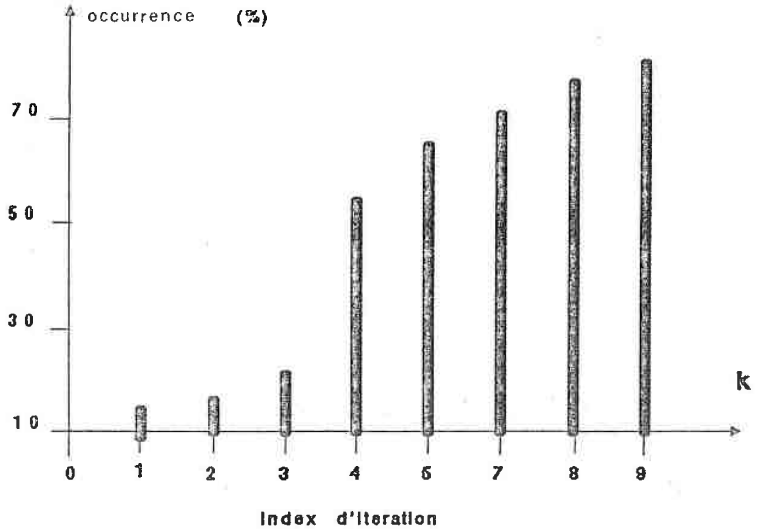


Figure 4.16: Occurrence de la séquence d'excitation correspondant à l'impulsion de dirac en fonction de l'index d'itération

Finalement, la procédure de modélisation par code optimal apparaît comme étant d'une complexité très conséquente, compte-tenu du débit qu'elle permet d'atteindre. En effet au 6.9 kbits/s correspondant à l'excitation, il faut ajouter quelques 2 kbits/s qui sont pris par les paramtres du filtre de synthèse et du prédicteur à long terme. Ceci porte donc le débit global à environ 9 kbits/s. Cette procédure a toutefois l'avantage de réduire l'effort de mémorisation du dictionnaire d'excitation à 4x25 mots de 10 bits, soit environ 1 kbits.

La figure 4.17 visualise les signaux caractéristiques du codeur à excitation optimale par code, à raison de 700 excitations par seconde.

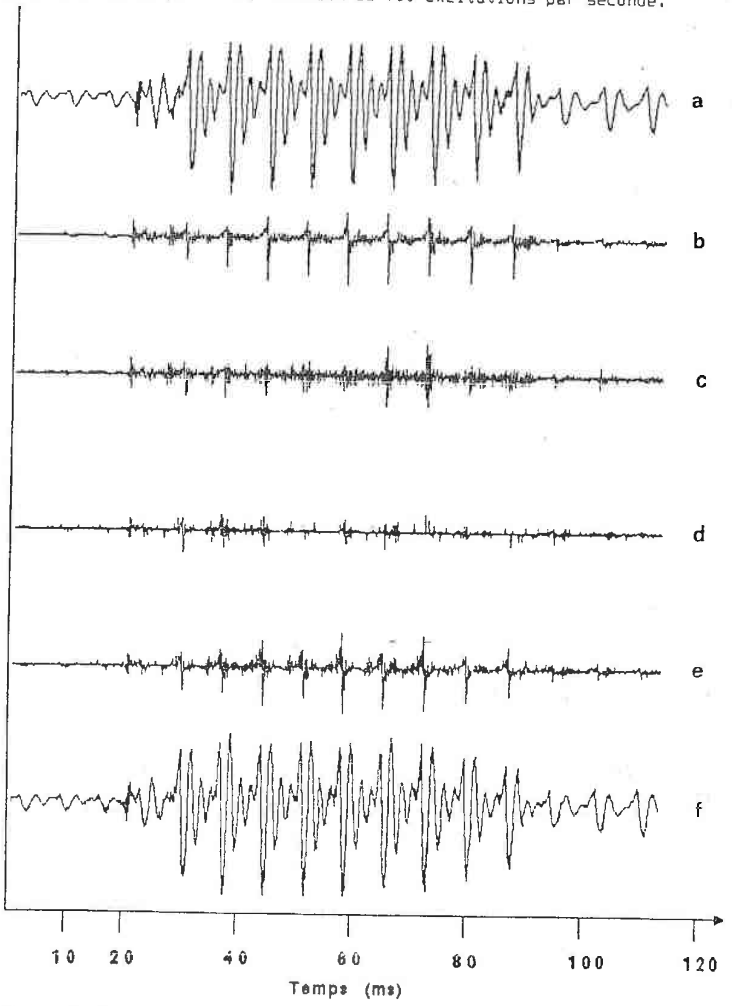


Figure 4.17: Evolution temporelle, a) signal original, b) signal résiduel, c) signal résiduel à long terme, d) signal d'excitation optimal par code, e) signal résiduel à court terme reconstruit, f) signal synthétique.

La procédure de modélisation que nous nous proposons de décrire dans le paragraphe suivant, offre un rapport complexité débit nettement plus intéressant.

IV.3.2 MODELISATION MULTI-IMPULSIONNELLE VECTORIELLE DE L'EXCITATION:

Le modélisation à excitation multi-impulsionnelle décrite dans le chapitre III réalise une modélisation et une quantification scalaire du signal résiduel par des séquences d'impulsions dont les amplitudes et positions sont encodées séparément. Le débit associé à ce signal d'excitation multi-impulsionnelle représente environ les deux tiers du débit global. L'application d'une quantification vectorielle aux amplitudes des impulsions permet de réduire à moins d'un bit le débit par amplitude d'impulsion.

Dans ce cas la dimension des vecteurs est conditionnée par le nombre d'impulsions par séquence d'excitation. On passe ainsi de vecteurs de dimension 40 pour la modélisation par code à des vecteurs de dimension 5 à 8 pour le codeur VMPLP.

L'application à posteriori, de la quantification vectorielle aux amplitudes des impulsions, n'est pas une solution satisfaisante car elle ne permet pas de prendre en compte l'effet de masquage que procure le filtre perceptuel.

La combinaison des procédures de modélisation multi-impulsionnelle et de modélisation par code permettent d'envisager une procédure multi-impulsionnelle vectorielle qui a l'avantage de prendre en compte l'effet de masquage, lors de la quantification vectorielle des amplitudes des impulsions. La figure 4.18 représente sous-forme de bloc diagramme le principe du codeur à excitation multi-impulsionnelle vectorielle, qui intègre également un prédicteur à long terme bouclé.

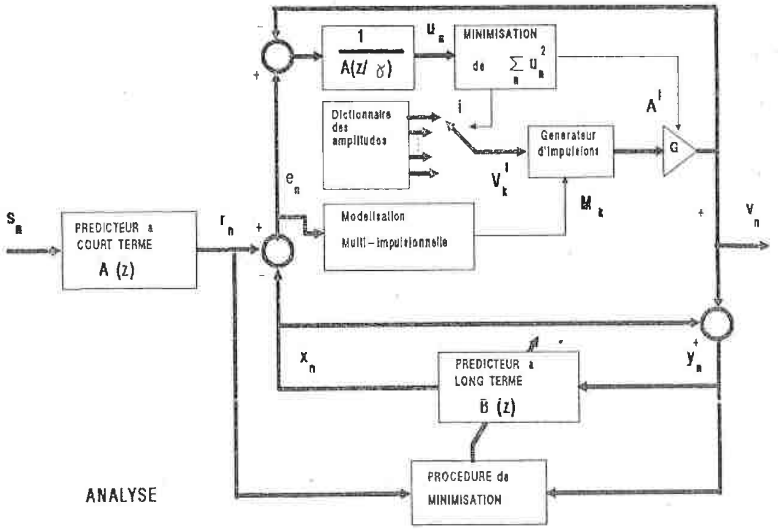


Figure 4.18: Bloc diagramme du codeur à excitation multi-impulsionnelle vectorielle

On peut observer que cette modélisation est équivalente à la modélisation par code lorque le nombre d'impulsions placées dans un segment de signal est égal au nombre d'échantillons du segment.

IV.3.2.1 PROCEDURE DE MODELISATION MULTI-IMPULSIONNELLE VECTORIELLE:

Cette procédure de modélisation se décompose de la manière suivante:

1: détermination itérative de la position M_k des impulsions comme cela a été décrit dans le chapitre III. La combinaison du dictionnaire des amplitudes avec la position des impulsions permet de constituer les vecteurs d'excitations u^i:

$$u^i = \sum_{k=1}^{K} v^{i,k}.\delta_{M_k} \qquad (4.37)$$

qui s'écrit sous-forme vectorielle:

$${}^T u^i = [\, u^{i,1}_{m_1}, 0, u^{i,2}_{m_2}, 0, .. 0, u^{i,3}_{m_3}, 0, 0, u^{i,K}_{m_K} \,]$$

2: détermination des amplitudes des impulsions à partir de la procédure de modélisation par code décrite dans le paragraphe 2. Compte-tenu de la forme particulière des séquences d'excitations, la procédure de modélisation par code se simplifie. On obtient respectivement:

$$\Gamma^i = \frac{[\, {}^T e.\Omega.u^i \,]^2}{{}^T u^i.Q.u^i} \qquad (4.38)$$

et

$$A^i = \frac{{}^T e.\Omega.u^i}{{}^T u^i.Q.u^i} \qquad (4.39)$$

avec Ω la matrice d'autocorrélation de dimension NxK, définie par:

$$\Omega = \begin{bmatrix} c_{|m_1|} & c_{|m_2|} & \cdot & \cdot & \cdot & c_{|m_3|} & c_{|K|} \\ c_{|m_1-1|} & c_{|m_2-1|} & \cdot & \cdot & \cdot & c_{|m_3-1|} & c_{|K-1|} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ c_{|m_1-N+2|} & c_{|m_2-N+2|} & \cdot & \cdot & \cdot & c_{|m_3-N+2|} & c_{|K-N+2|} \\ c_{|m_1-N+1|} & c_{|m_2-N+1|} & \cdot & \cdot & \cdot & c_{|m_3-N+1|} & c_{|K-N+1|} \end{bmatrix} \qquad (4.40)$$

Q la matrice d'autocorrélation symétrique de dimension KxK, définie par:

$$Q = \begin{bmatrix} c_{|m_1-m_1|} & c_{|m_2-m_1|} & \cdot & \cdot & c_{|m_3-m_1|} & c_{|K-m_1|} \\ c_{|m_1-m_2|} & c_{|m_2-m_2|} & \cdot & \cdot & c_{|m_3-m_2|} & c_{|K-m_2|} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ c_{|m_1-K+1|} & c_{|m_2-K+1|} & \cdot & \cdot & c_{|m_3-K+1|} & c_{|K-K+1|} \\ c_{|m_1-K|} & c_{|m_2-K|} & \cdot & \cdot & c_{|m_3-K|} & c_{|K-K|} \end{bmatrix} \qquad (4.41)$$

On passe ainsi d'un système de dimension N' à un système de dimension K. Le rapport de réduction de la complexité est alors au moins égal à N'/K. En réalité ce rapport est encore plus important car la réduction de la dimension des vecteurs s'accompagne d'une réduction du nombre de séquences du dictionnaire. Il en resulte une réduction de la complexité qui ne dépasse pas 5 10^6 *,+/s. L'introduction de la diagonalisation de la matrice d'autocorrélation Q apporte, comme pour le codeur à excitation par code, une réduction supplémentaire du coût opératoire. Ceci est détaillé dans le tableau ci-dessous où le facteur ϕ représente le nombre de fenêtres de modélisation LPC (de longueur N) par seconde. Quant au facteur ϕ', il représente le nombre de fenêtres de modélisation de l'excitation (de longueur N').

Le nombre d'impulsions que nous avons retenu est déduit des simulations proposées dans le paragraphe 7 du chapitre III. Le codeur à excitation multi-impulsionnelle décrit dans ce chapitre place 1000 impulsions/s. La quantification vectorielle introduit une distorsion supérieure à la quantification scalaire. De façon à obtenir des performances très proches de celles du codeur à excitation multi-impulsionnelle, il est nécessaire de compenser cette distorsion liée à la quantification verctorielle par une augmentation du nombre d'impulsions. Aussi, nous retenons 1200 impulsions/s pour le codeur à excitation multi-impulsionnelle vectorielle.

Rappelons que le prédicteur à long terme bouclé impose que le signal v_n qui actualise le signal y_n doit être le signal d'excitation quantifié, à savoir $\hat{v}_n$. De façon à augmenter l'efficacité de la quantification vectorielle en terme de réduction de débit, nous optons pour une actualisation des paramètres du prédicteur à long terme toute les 10 ms. Ainsi la dimension des vecteurs d'amplitude est de 12.

forme directe	multi.,addit./s	div./s
réponse impulsionnelle de F(z)	3.MP.MA.ϕ	
autocorrélation de F(z)	$MA^2.\phi$	MA.ϕ
modélisation multi-impulsonnelle	2.MA(N' + K).ϕ'	
critère + facteur de gain	D.(K^2 + 2.K + 1).ϕ'	D.ϕ'
forme simplifiée	**multi.,addit./s**	**div./s**
réponse impulsionnelle de F(z)	3.MP.MA.ϕ	
autocorrélation de F(z)	$MA^2.\phi$	MA.ϕ
modélisation multi-impulsonnelle	2.MA(N' + K).ϕ'	
diagonalisation de la matrice	$2.K^2+K^3.\phi$	
critère + facteur de gain	K.(3.D + 1).ϕ'	D.ϕ'

Tableau 4.5: Complexité liée à la détermination de l'excitation par la procédure de modélisation multi-impulsionnelle vectorielle

IV.3.2.2 EXTRACTION STATISTIQUE DU DICTIONNAIRE D'EXCITATION:

L'extraction du dictionnaire des amplitudes des séquences d'excitation fait appel, comme pour le codeur à excitation par code, à des algorithmes de classification.

CHOIX DE LA MESURE DE DISTANCE:

La mesure de distance est définie par la relation 4.34 normalisée:

$$\Gamma'^{i} = \frac{[\,{}^{T}e.O.u^{i}\,]^{e}}{{}^{T}u^{i}.K.u^{i}} \times \frac{{}^{T}u^{i}.K.u^{i}}{{}^{T}e.O.u^{i}} \qquad (4.42)$$

CHOIX DU SEUIL:

Les figures ci-dessous visualisent, l'évolution de la taille du dictionnaire et du rapport signal sur bruit segmental en fonction du seuil de classification.

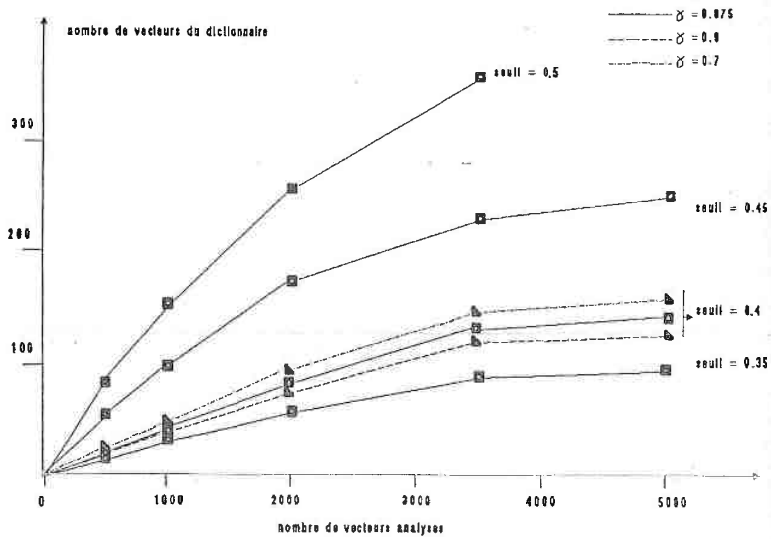


Figure 4.19: Evolution de la taille du dictionnaire en fonction du nombre de trames analysées avec pour paramètre, le seuil de classification et le facteur perceptuel.

La figure ci-dessous visualise l'évolution du rapport signal sur bruit segmental en fonction du seuil de classification. Le rapport signal sur bruit segmental tend asymptotiquement vers la valeur qui correspond à une

modélisation multi-impulsionnelle sans quantification de l'amplitude des impulsions.

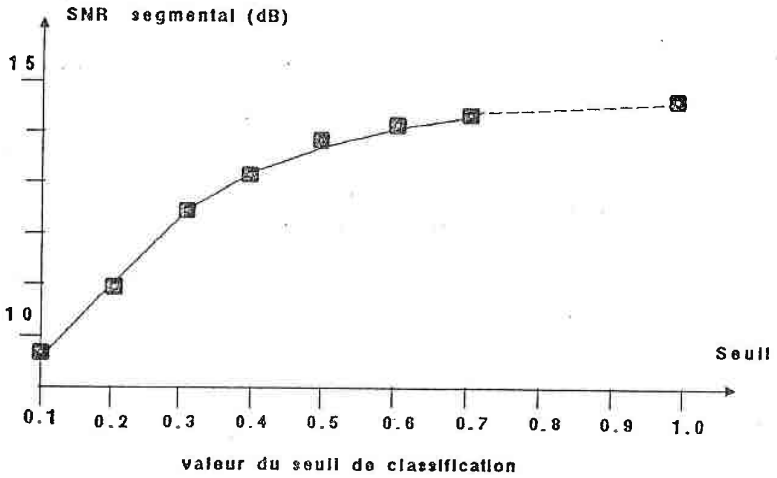


Figure 4.20: Evolution du rapport signal à bruit segmental en fonction du seuil de classification

Compte-tenu des résultats présentés par la figure ci-dessus, un dictionnaire de 256 vecteurs de dimension 12 (seuil de classification 0.4), permet de restituer un signal de parole de qualité sub-téléphonique, dont le rapport signal sur bruit est de 13.5 dB environ.

Les valeurs des paramètres que cette étude a permis de dégager sont précisés ci-dessous:

- N: dimension des segments d'analyse LPC = 160 éch.(20 ms)
- MP: ordre du filtre de synthèse en treillis = 12
- MA: durée de l'autocorrélation de filtre F(z) = 20 éch. (1.25 ms)
- N':dimension des segments de la modélisation du prédicteur à long terme et de la modélisation de l'excitation = 80 éch. (10 ms)
- D :nombre de vecteurs du dictionnaire d'excitaion = 256

forme directe	multi.,addit./s	div./s
complexité pour les valeurs standards	4.75 10^6	26600
forme simplifiée	multi.,addit./s	div./s
complexité pour les valeurs standards	1.53 10^6	26600

Tableau 4.6: Complexité de la procédure de modélisation multi-impulsionnelle vectorielle

Le débit nécessaire au codage de l'excitation se limite à 5.6 kbits/s, dont 4.4 kbits/s sont pris par l'encodage des positions des impulsions. A cela, il faut ajouter le débit correspondant aux paramètres du filtre de synthèse et du prédicteur à long terme, qui représente un débit supplémentaire de 2 kbits/s. Le débit global qu'offre ce codeur est donc inférieur à 8 kbits/s. La complexité, de la procédure de modélisation de l'excitation est inférieure à 5 10^6 multiplications et additions par seconde. Une réduction supplémentaire de cette complexité peut être obtenue en optant, comme pour le modélisation stochastique, pour la solution avec décomposition de la matrice d'autocorrélation. Ainsi la complexité chute à 1.5 10^6 *,+/s. Notons toutefois, que l'effort de mémorisation qu'engendre le dictionnaire des amplitudes, n'est pas négligeable, car il représente quelques 256x12 mots de 10 bits, soit 30 kbits.

La figure ci-dessous illustre, l'évolution temporelle des différents signaux significatifs du codeur à excitation multi-impulsionnelle vectorielle avec prédiction à long terme bouclée.

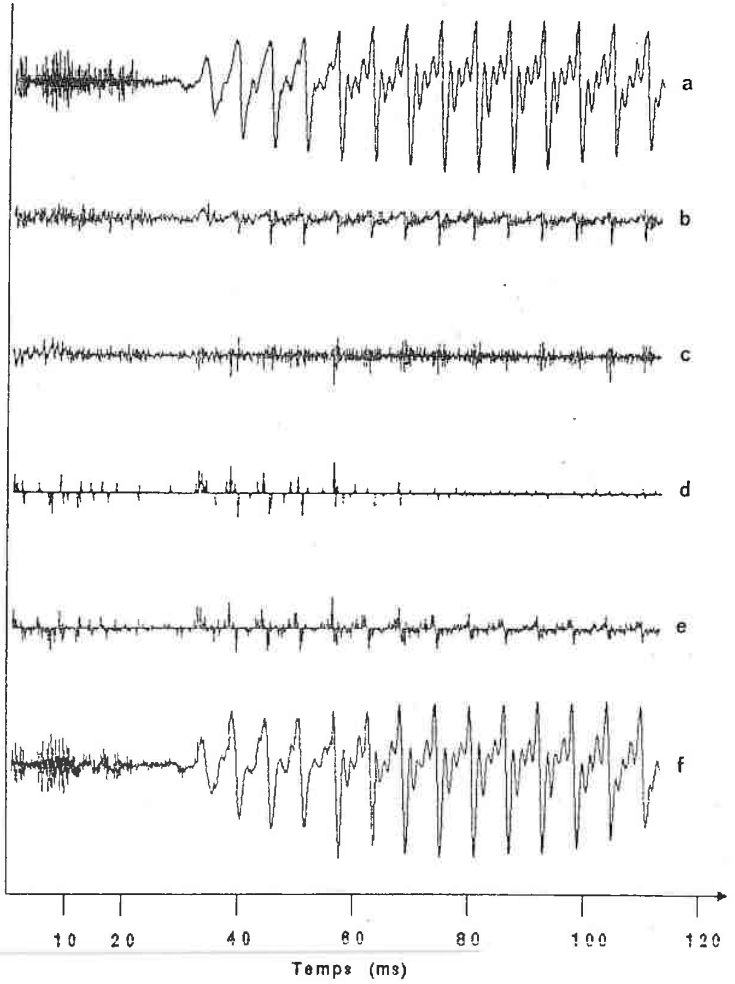


Figure 4.21: Evolution des signaux a) signal original, b) signal résiduel à court terme, c) signal résiduel à long terme, d) signal d'excitation multi-impulsionnelle vectorielle avec positionnement uniforme, e) signal résiduel à court terme reconstruit, f) signal synthétique.

Les performances de la procédure de modélisation multi-impulsionnelle vectorielle sont intéressantes tant en qualité du signal restitué qu'en complexité et débit. Il reste que le débit associé aux positions des

impulsions reste inchangé par rapport à la procédure de modélisation multi-impulsionnelle. L'utilisation d'une technique similaire au "regular pulse coding" proposé par KROON [12], permet de réduire le débit correspondant aux positions des impulsions. La position de l'impulsion de la première itération définit, à un décalage près, le positionnement relatif de toute la séquence multi-impulsionnelle, dont toutes les impulsions secondaires sont réparties de manière régulière autour de l'impulsion principale. Cependant, comme le positionnement des impulsions secondaires est sous-optimal, il est nécessaire, de façon à restituer un signal de parole de qualité sub-téléphonique (rapport signal sur bruit de 13 à 14 dB), d'augmenter dans un rapport 2 le nombre d'impulsions à placer.

IV.4 CONCLUSION:

Nous venons de décrire dans ce chapitre plusieurs procédures de modélisation du signal d'excitation, qui découlent de la procédure d'analyse synthèse, mais dans sa forme vectorielle. Toutes ces procédures ont l'avantage de s'intégrer dans un même schéma de codeur. Il comporte une analyse LPC et une prédiction à long terme bouclée qui favorise la compensation partielle de l'erreur de quantification qu'introduisent les différentes procédures de modélisation de l'excitation.

Le codeur à excitation par code met en oeuvre une technique de codage pleine d'avenir. Grâce à la décomposition matricielle et de l'utilisation d'un dictionnaire statistique, la complexité peut être réduite à quelques 19 10^6 multiplications et additions par seconde. Il reste que la mémorisation du dictionnaire d'excitation est très conséquente car elle atteind 200 kbits.

Outre l'étude de la procédure de modélisation stochastique, nous avons introduit deux autres techniques de modélisation vectorielle de l'excitation. Celles-ci sont le résultat de la combinaison de la modélisation multi-impulsionnelle et de la modélisation stochastique.

La procédure de modélisation optimale par code procure des performances comparables à un codeur à excitation multi-impulsionnelle tout en réduisant dans un rapport 10/7 le nombre d'excitation. Toutefois , à qualité sub-téléphonique le rapport complexité débit n'est pas très favorable, car la transmission de l'index des séquences d'excitation annule partiellement, le débit économisé par la réduction du nombre de séquences d'excitation. Notons néanmoins, que cette procédure de modélisation a l'avantage de limiter à moins de 1 kbits la taille du dictionnaire d'excitation. Il reste que son utilisation est très limitée dans le domaine du codage de la parole.

Le codeur à excitation multi-impulsionnelle vectorielle, permet grâce à la quantification vectorielle pondérée des amplitudes des impulsions de limiter le débit à moins de 2 bits/amplitude. En revanche cette procédure ne permet pas de réduire le débit lié aux positions des impulsions. Toutefois, l'introduction du principe de "regular pulse coding" proposé par KROON [12] permet de réduire également le débit correspondant aux positions des impulsions.

Finalement, pour l'ensemble des procédures de modélisation de l'excitation que nous avons décrites, le sous-échantillonnage du signal résiduel est une solution qui offre une réduction de la complexité. A première approximation, celle-ci est dans un rapport égal au rapport de sous-échantillonnage. Cette diminution de la complexité s'accompagne également, mais dans une moindre mesure, d'une réduction du débit.

Bibliographie

Publications:

[1] "Baseband Speech Coding at 2400 bps using "Spherical Vector Quantization"
ADOUL J.P., LAMBLIN C., LEGUYADER A.
IEEE Proc. Int. Conf. on ASSP, 1984,

[2] "La Quantification Vectorielle des Signaux: Approche Algébrique"
ADOUL J.P.
Ann. Télécomm., 1986, pages 158-177

[3] "High-Quality Speech at Low Bit Rates Multi-Pulse and Stochastically Excited Linear Predictive Coders"
ATAL B.S.,
IEEE Proc. Int. Conf. on ASSP, 1986, pages 1681-1684

[4] "Multiple Stage Vector Quantization for Speech Coding"
BIING-HWANG JUANG, GRAY A.H.
IEEE Proc. Int. Conf. on ASSP, 1982, pages 597-600

[5] "Gain-Adaptative Quantization for Medium-Rate Speech Coding"
CHEN J.H., GERSHO A.
IEEE Proc. Int. Conf. on ASSP, 1985, pages 1456-1460

[6] "Vector Adaptative Predictive Coding of Speech at 9.6 kb/s"
CHEN J.H., GERSHO A.
IEEE Proc. Int. Conf. on ASSP, 1986, page 1693-1696

[7] "Vector Quantization: A Pattern-Matching Technique for Speech Coding"
GERSHO A., CUPERMAN V.
IEEE Trans. on ASSP, 1983, pages 15-21

[8] "Full Search and tree Searched Vector quantization of Speech Waveforms"
GRAY R.M., ABUT H.
IEEE Proc. Int. Conf. on ASSP, 1982, pages 593-596

[9] "Speech Coding Using Efficient Block Codes"
SCHRODER M.R, ATAL B.S.
IEEE Proc. Int. Conf. on ASSP, 1982, pages 1668-1671

[10] "Stochastic Coding of Speech Signals at Very Low Bit Rates: The Importance of Speech Perception"
SCHROEDER M.R., ATAL B.S.
Speech Commu., 1985, pages 155-162

[11] "Efficient Procedures for Finding the Optimum Innovation in Stochastic Coders"
TRANCOSO I.M., ATAL B.S.
IEEE Proc. Int. Conf. on ASSP, 1986, pages 2375-2378

Thèses:

[12] "Time-Domain Coding of (Near) Toll Quality Speech at Rates Below 16 kb/s"
KROON P.
1985, Delft University (Hollande)

Ouvrages:

[13] "Computer Science and Applied Mathematics"
STEWART G.W.
Academic Press, New-York, 1973

CHAPITRE V

CODAGE ET QUANTIFICATION DES PARAMETRES DES CODEURS

V.1 INTRODUCTION:

Les chapitres précédents ont été consacrés à la description de procédures de modélisation qui extraient les paramètres aptes à la transmission ou au stockage. Ces paramètres doivent être représentés par un nombre fini de symboles de façon à respecter le débit envisagé. Dans ce chapitre nous allons décrire les techniques d'encodage que nous avons mises en oeuvre dans les différents codeurs. A ce titre, on peut observer que les procédures de codage relèvent plus de l'analyse statistique des données que du traitement du signal. L'opération d'encodage doit prendre en compte la spécificité des paramètres. Ainsi, les techniques d'encodage peuvent être classées comme suit:

- **encodage par quantification;** qui discrétise les valeurs des paramètres. A chaque niveau de quantification correspond un mot binaire apte au stockage ou à la transmission. L'encodage par quantification peut être fait de manière scalaire ou vectorielle, selon que les paramètres sont traités individuellement ou en séquence. Cet encodage introduit une erreur de quantification, qui se traduit par une distorsion entre la valeur des paramètres, avant et après codage-décodage. Les paramètres tels que les coefficients du filtre de synthèse et les amplitudes des séquences d'excitation multi-impulsionnelle, par exemple, sont adaptés à un tel encodage.

- **encodage sans quantification;** qui réalise un encodage, ou plutôt un décodage, sans erreur. Un tel encodage est nécessaire pour les paramètres temporels représentatifs d'un décalage ou d'une position. C'est le cas des positions des impulsions, ou du décalage du prédicteur à long terme. Si ces paramètres sont encodés par quantification, il est indispensable de réinjecter la distorsion dans la procédure de modélisation. Mais il y a un risque de non convergence des algorithmes. Comme on peut s'y attendre un tel encodage ne permet pas d'atteindre des débits aussi bas que l'encodage par quantification.

Le tableau ci-dessous fournit les caractéristiques des techniques de codage, qui sont applicables aux différents paramètres.

paramètre	encodage
coefficients du filtre de synthèse	par quantification
coefficients du prédicteur à long terme	par quantification
amplitude des impulsions	par quantification
position des impulsions	sans erreur
facteur de gain des séquences d'excitation stochastique	par quantification
décalage du prédicteur à long terme	sans erreur

Tableau 5.1: Technique de codage pour les paramètres des différents codeurs

Certains codeurs procurent de manière inhérente un nombre limité de symboles. C'est le cas de la quantification vectorielle.

Les résultats du codage sont intimement liés aux caractéristiques temporelles et spectrales du signal de parole qu'on souhaite encoder. Aussi, il est important de définir préalablement les caractéristiques du signal de parole qui est choisi pour optimiser et évaluer l'encodage des paramètres, à savoir:

- la base de données de signal de parole est constituée de phrases phonétiquement équilibrées qui sont données en annexe D. Ces phrases sont prononcées par des locuteurs masculins et féminins. Elle se décompose en deux parties. La première partie qui représente environ 5 minutes de parole sert à la constitution des tables ou dictionnaires de quantification. La deuxième partie, sert aux évaluations objectives et subjectives.

- les conditions d'enregistrement sont pour un certain nombre de phrases du type "haute qualité", tandis que pour d'autres du type "téléphonique"

- le signal de parole est échantillonné à 8 kHz puis filtré dans la bande téléphonique (100 - 3400 Hz).

- par simulation une dégradation correspondant à une conversion loi linéaire -> loi A -> loi linéaire est appliquée au signal de parole.

- le niveau moyen du signal de parole est fixé à -15 dB par rapport à la saturation d'un signal sinusoïdal en sortie du convertisseur analogique-numérique "cofidec".

Avant d'aborder la description des techniques de codage, nous rappelons les propriétés que doivent posséder les paramètres à quantifier, ainsi que l'incidence des modules de codage sur les performances globales des codeurs.

V.1.1 PROPRIETES DES PARAMETRES A ENCODER PAR QUANTIFICATION:

Les paramètres à quantifier doivent posséder des propriétés de stabilité et d'agencement dans un ordre naturel.

La première propriété, qui s'applique plus particulièrement aux paramètres des filtres, implique que la quantification ne déplace pas, dans le plan des z, les pôles du filtre en dehors du cercle unité.

La deuxième propriété suppose qu'une séquence de paramètres est ordonnée de manière inhérente. En effet prenons les coefficients PARCORS $K_1, K_2 \ldots$, l'échange de K_1 avec K_2 modifie la réponse du filtre. Si un tel ordre existe, une étude statistique spécifique à chaque paramètre peut être utilisée pour développer des techniques de quantification plus efficaces.

Il reste que les techniques de codage doivent prendre en compte la spécificité de chaque paramètre, en terme de dynamique, de sensibilité et de distribution statistique. Ainsi, les procédures de quantification des coefficients des filtres sont différentes de celles appliquées au signal d'excitation. Nous les examinerons donc séparément.

V.1.1.1 QUANTIFICATION SCALAIRE [11]:

Quantifier de manière scalaire des paramètres, consiste à encoder ces derniers de manière isolée, en partant de l'hypothèse qu'il n'existe aucune corrélation entre les paramètres d'une même trame ou de deux trames successives. On peut noter également que la quantification scalaire ne permet pas d'encoder ces paramètres à moins d'un bit par paramètre.

V.1.1.2 QUANTIFICATION VECTORIELLE [4,5,6,9,15]:

A priori, la quantification vectorielle peut s'appliquer à tout jeu de paramètres, quelque soit sa nature spectrale ou temporelle. Les performances ainsi que la complexité de cette technique résident dans le choix de la métrique ainsi que dans la taille et la nature du dictionnaire. Rappelons qu'un quantificateur vectoriel est défini d'une part par la dimension **M** des vecteurs à quantifier, d'autre part par le nombre de vecteurs **D** qui composent le dictionnaire. Si S est la fréquence à laquelle les vecteurs sont quantifiés, le débit par unité de temps s'écrit:

$$Déb = \log_2(D).S \text{ bits/s} \qquad (5.1)$$

et le débit par coefficient est défini par:

$$R = \frac{1}{M} \log_2(D) \qquad (5.2)$$

Cette relation met en évidence que le codage vectoriel permet d'encoder un coefficient avec une fraction de bit.

Cependant, l'utilisation de la quantification vectorielle est pénalisée par sa complexité et son coût en stockage. Le nombre d'opérations, qui dépend évidemment de la métrique, est de l'ordre de $M.D.S^a$ opérations par seconde pour la norme L2, d'autre part l'effort de mémorisation qu'engendre le stockage du dictionnaire s'élève à **M.D** mots.

On observe donc que la taille du dictionnaire, ou plus précisément le nombre de vecteurs qui le composent **D** , est déterminant dans la complexité du traitement, dans l'effort de mémorisation et dans le débit.

V.1.2 INCIDENCE DU CODAGE PAR QUANTIFIATION SUR LE CODEUR:

L'introduction des modules de quantification dans la chaîne d'analyse, n'est pas sans incidence sur les performances globales du codeur. Pour les codeurs de qualité sub-téléphonique, le placement de ces modules nécessite un soin tout particulier de façon à ne pas anéantir les performances des procédures de modélisation et d'extraction des paramètres. En effet, selon l'endroit où les modules de quantification sont placés, la dégradation inhérente au codage par quantification peut être partiellement compensée.

Une règle simple, qui permet de minimiser cette dégradation, consiste à placer, dans la mesure du possible, les modules de quantification de telle manière que les modules de traitement situés en aval utilisent les paramètres identiques à ceux de la synthèse, à savoir les valeurs quantifiées des paramètres. Ceci est illustré par la figure 5.1. Cette règle introduit

toutefois quelques contraintes notemment temporelles, qui ont pour répercussion, selon le cas, d'augmenter le délai de restitution du codeur, ou d'accelérer les traitements situés en aval.

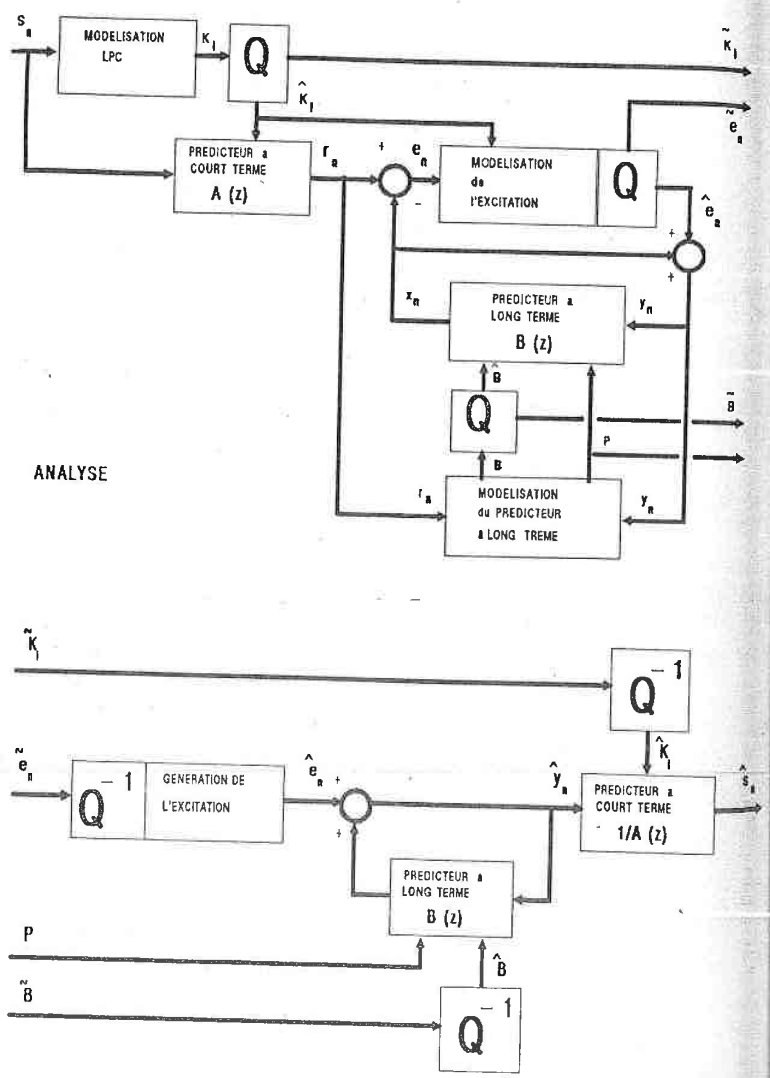


Figure 5.1: Placement des modules a) de codage dans le codeur, b) de décodage dans le décodeur

V.2 CODAGE DES PARAMETRES DU PREDICTEUR A COURT TERME:

Le filtre de synthèse H(z) peut être décrit par les paramètres suivants:

- les coefficients d'autocorrélation R_i
- les coefficients de prédiction linéaire a_i
- les coefficients cepstraux C_i
- les pôles P_i
- les coefficients de corrélation partielle (PARCOR) K_i

Viswanathan et Makhoul [19] ont montré que les coefficients de corrélation partielle K_i encore appelés PARCOR sont les paramètres les plus aptes à la transmission ou au stockage, car il garantissent la stabilité du filtre et leur arrangement est implicite.

Le prédicteur à court terme, reproduit l'enveloppe spectrale du signal de parole. La quantification de ses coefficients ne doit pas altérer de manière significative cette enveloppe spectrale.

V.2.1 QUANTIFICATION SCALAIRE:

La quantification modifie la valeur des coefficients Parcors. De façon à mesurer la distorsion qu'introduit la quantification, il est nécessaire de déterminer la sensibilité spectrale du filtre de synthèse aux faibles variations des K_i. Celle-ci est définie par:

$$\frac{\delta S}{\delta K_i} = \lim_{\delta K_i \to 0} \left| \frac{\delta S}{\delta K_i} \right| \qquad (5.3)$$

où δS est la variation du spectre du filtre due à la variation δK_i du coefficient K_i.

Quelque soit l'approche, expérimentale [19] ou analytique [12], il apparaît que cette sensibilité spectrale est élévée pour les valeurs de K_i proches de +1 et -1. On peut donc conclure que la sensibilité des K_i est non-uniforme. Aussi, avant quantification, est-il préférable de transformer les K_i par une fonction qui rend uniforme la sensibilité spectrale. Les fonctions de transformation sont:

$$G_i = \mathrm{Log} \left| \frac{1 + K_i}{1 - K_i} \right| \quad \text{proposée par Viwanathan et Makhoul,} \qquad (5.4)$$

et

$$\theta_i = \mathrm{Arcsin}\ (K_i) \ \text{proposée par Gray et Markel.} \qquad (5.5)$$

Les G_i sont encore appelés les Rapport d'Aire Logarithmique (en anglais LAR Logarithmic Area Ratio).

De façon à simplifier ces fonctions de transformation, UN [18] propose d'approximer celles-ci par quelques segments de droite. Cependant pour les valeurs de K_i proches des bornes +1 et -1, cette technique n'est pas assez précise.

De manière intuitive, on peut penser que, compte-tenu de la distribution en amplitude des coefficients transformés, il est préférable de réduire le pas de quantification pour les amplitudes présentant une forte occurrence. La fonction représentative de cette distribution non-uniforme est l'histogramme à partir du quel il est possible de déduire la fonction de quantification à erreur minimale, spécifique à chaque coefficient, proposée par MARKEL et GRAY [12]. Les histogrammes ainsi que les fonctions de quantification à erreur minimale des 12 premiers K_i et G_i sont présentées en annexe A. On peut observer que ce sont principalement les 3 premiers coefficients qui présentent des distributions très spécifiques.

L'étude statistique des K_i et des G_i a permis de définir l'intervalle de variation de chaque coefficient. Ces résultats ont été obtenus selon la méthode d'autocorrélation de Le Roux-Gueguen [10] avec des fenêtres de 20 ms et sans préaccentuation. Ils sont présentés dans le tableau ci-dessous pour une probabilité de 99 %. Il apparaît que la largeur de l'intervalle de variation des coefficients K_i, donc des G_i, diminue fortement pour les coefficients d'index supérieur à 3.

indice	coefficients K_i		coefficients G_i	
	min	max	min	max
1	-0.9918	0.8281	-2.3854	1.0267
2	-0.6885	0.9639	-0.7340	1.7356
3	-0.8293	0.7072	-1.0301	0.7657
4	-0.5821	0.8233	-0.5782	1.0136
5	-0.6973	0.5488	-0.7487	0.5356
6	-0.4342	0.7937	-0.4039	0.9393
7	-0.6069	0.4559	-0.6115	0.4275
8	-0.5172	0.6271	-0.4973	0.6398
9	-0.6525	0.3698	-0.6772	0.3372
10	-0.3398	0.4736	-0.3074	0.4934
11	-0.4512	0.3125	-0.4223	0.2808
12	-0.2915	0.4389	-0.2607	0.4090

Tableau 5.2: Valeurs minimales et maximales qui délimitent les intervalles de variation des coefficients K_i et G_i pour une probabilité de 99 %.

En résumé, des deux fonctions de transformations f_1 et f_2, qui compensent respectivement la sensibilité non-uniforme et la distribution non-uniforme, il est possible de déduire une fonction de transformation globale f_3, à savoir:

$$f_3(K_i) = D_i = f_2[\, f_1(K_i)\,] \qquad (5.6)$$

Une quantification linéaire avec arrondi peut ensuite être appliquée à D_i en subdivisant l'intervalle de variation en 2^{NB+1} sous-intervalles de longueur égale où NB représente le nombre de bits alloués au codage d'un coefficient. Les valeurs quantifiées des K_i sont ensuite obtenues en appliquant la transformation inverse f_3^{-1} aux valeurs quantifiées de D_i.

La procédure de quantification à erreur minimale réalise la répartition des bits de façon à minimiser la distorsion. La sensibilité de tous les coefficients étant supposée identique, il est nécessaire d'allouer plus de bits aux coefficients qui présentent les intervalles de variation les plus grands.

Les performances des tables de codage extraites pour différents débits, sont évaluées à l'aide d'un critère objectif qui est la distance ceptrale. La figure ci-dessous visualise l'évolution de la distance cepstrale en fonction du nombre de bits alloués à l'ensemble de la trame de 12 coefficients K_i.

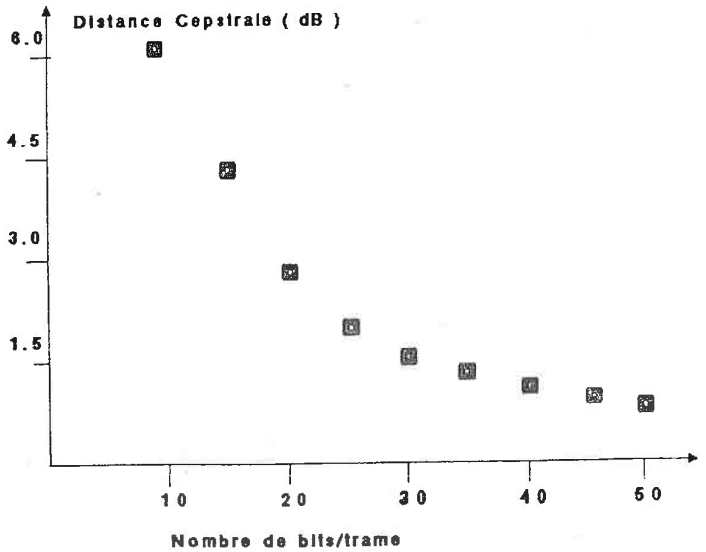


Figure 5.2: Evolution de la distance cepstrale en fonction nombre de bits alloués aux K_i (filtre d'ordre 12)

La distance cepstrale doit être inférieure à 1.5 dB, limite au delà de laquelle, la quantification n'introduit pas uniquement une atténuation des formants, mais s'accompagne également d'un déplacement des formants. En allouant 40 bits à la trame de 12 K_i la distorsion moyenne ne dépasse pas 1.2

dB. La figure 5.3 illustre pour les K_i la distribution des pas de quantification distribués de manière non-uniforme entre -1 et +1.

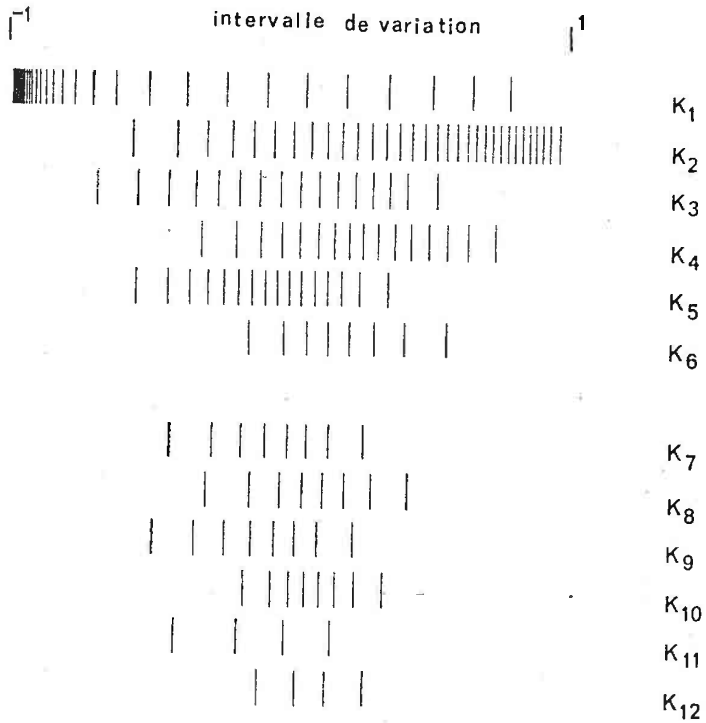


Figure 5.3: Distribution non-uniforme des valeurs dans la table de codage

V.2.1.1 QUANTIFICATION TABULEE NON-UNIFORME:

La quantification par table consiste à approximer la valeur des paramètres à quantifier par la valeur la plus proche contenue dans la table de codage. Cette table contient les valeurs quantifiées dont la distribution non-uniforme sur l'intervalle de variation est déduite de la fonction de transformation globale f_g. La valeur la plus proche et son index représentent respectivement la valeur quantifiée et la valeur codée du paramètre. On peut par analogie considérer que cette table correspond au dictionnaire d'un codeur vectoriel, dans le cas particulier où les vecteurs sont de dimension unitaire.

Le principe des procédure de codage et de décodage par table est illustré par la figure ci-dessous.

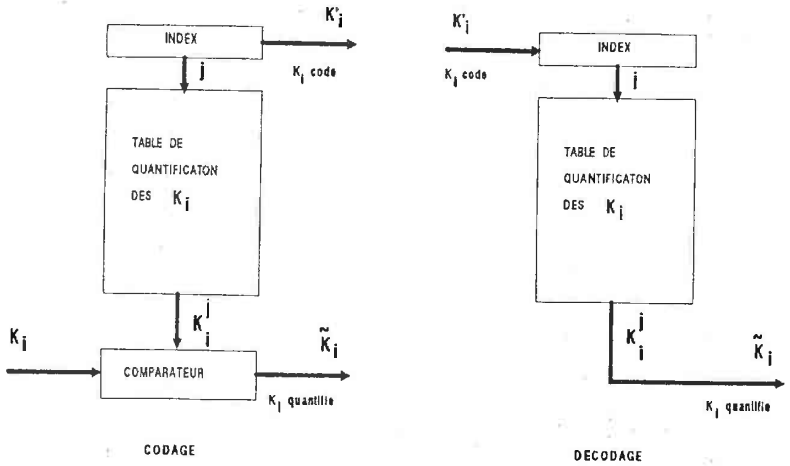


Figure 5.4: Principe du codage et décodage scalaire par table

V.2.1.2 QUANTIFICATION CALCULEE NON-UNIFORME:

La technique de quantification scalaire par table est très certainement la plus utilisée. Elle a toutefois l'inconvénient d'utiliser une table de codage dont la taille atteind 140 mots de 12 bits, soit 1.6 kbits environ.

La technique de quantification scalaire par calcul, a l'avantage de réduire très fortement la taille de cette table. Elle consiste à approximer la fonction de transformation f_g ainsi que son inverse f_g^{-1} par une fonction polynomiale du type:

$$D_i = f_g^{-1}(K_i) = a_0 + a_1.K_i + a_2.K_i^2 \ldots + a_K.K_i^J = \sum_{j=0}^{J} a_j.K_i^j \qquad (5.7)$$

Les coefficients a_j du polynome sont déterminés de façon à minimiser l'erreur au sens des moindres carrés. Ces fonctions de transformation globales spécifiques à chaque K_i sont présentées en annexe C.

On constate, que les fonctions de transformations globales correspondant aux K_i d'index supérieur à 3 sont fort semblables et peuvent être approximées par une unique fonction de transformation. Ainsi le nombre de fonction de transformation est limité à 4, ce qui représente quelques 2x20 coefficients sur 12 bits à mémoriser.

La quantification uniforme peut ensuite être appliquée aux coefficients D_i. Nous avons retenu une quantification avec arrondi car elle offre un pas de quantification supplémentaire. La figure ci-dessous illustre le principe de la quantification non-uniforme par calcul.

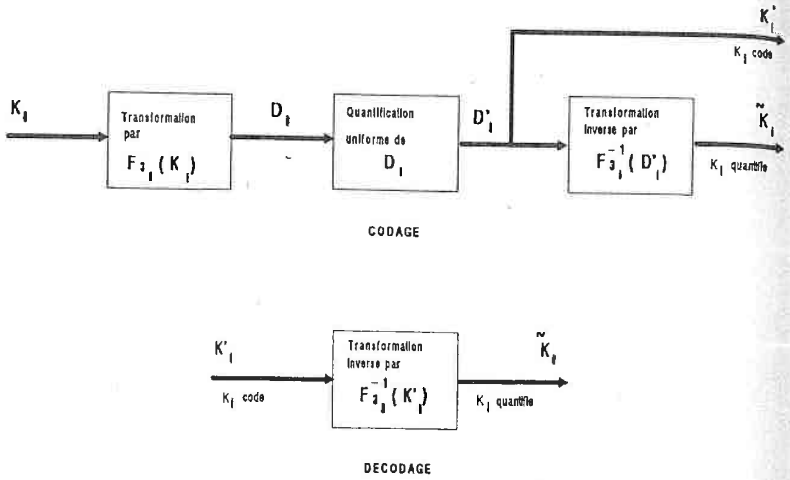


Figure 5.5: Principe des modules d'encodage et de décodage par calcul

V.2.2 QUANTIFICATION VECTORIELLE:

La quantification vectorielle, qui est, rappelons le, la généralisation de la quantification scalaire à un espace de dimension supérieur à 1, permet de remédier à la limitation du codage scalaire, à savoir un débit par coefficient supérieur ou égale 1 bit.

Le vecteur de M coefficients K_i, représentatif du filtre de synthèse sur un intervalle de temps donné, est considéré comme une entité. Lors du codage, chaque vecteur est approximé, au sens d'une certaine métrique (norme L2), par un représentant du dictionnaire. Un mot binaire, représentatif de l'index, permet d'identifier à la synthèse le vecteur du dictionnaire choisi.

L'utilisation du codage vectoriel, suppose qu'un dictionnaire à été préalablement constitué. Pour cela nous retenons également une procédure de classification à seuil. Nous avons choisi, comme métrique, la distance euclidienne que nous appliquons aux coefficients G_i. Cette métrique s'écrit:

$$d(\hat{G},\tilde{G}^j) = \left[\sum_{i=1}^{MP} (G_i - \tilde{G}_i^j)^2 \right]^{1/2} \quad \text{pour } j = 1 \,..\, D \qquad (5.8)$$

Le meilleur candidat du dictionnaire est celui qui fournit la distance d la plus faible. Cette métrique peut encore s'écrire:

$$d(\hat{G},\tilde{G}^j) = \left[\sum_{i=1}^{MP} (\hat{G}_i^2 + (\tilde{G}_i^j)^2 - 2\hat{G}_i.\tilde{G}_i^j) \right]^{1/2} \quad \text{pour } j = 1 \,..\, D \qquad (5.9)$$

Les deux premiers termes sont toujours positifs, aussi minimiser d revient à choisir le vecteur du dictionnaire qui maximise le produit scalaire suivant:

$$d'(\hat{G},\tilde{G}^j) = \sum_{i=1}^{MP} \hat{G}_i.\tilde{G}_i^j \quad \text{pour } j = 1 \,..\, D \qquad (5.10)$$

Cette solution a l'avantage de réduire la complexité liée à la quantification vectorielle, comme cela est décrit dans le tableau ci-dessous. Le facteur ϕ représente le nombre de fenêtres de modélisation LPC (de longueur N) par seconde.

quantification vectorielle	multi.,addit./s	bits/s
distance euclidienne	2.MP.D.ϕ	Log_2D
distance euclidienne simplifiée	MP.D.ϕ	"

Tableau 5.3: Complexité de la procédure de quantification vectorielle des coefficients K_i

La figure 5.6a représente sous forme de projection dans le plan l'évolution de G_1 en fonction de G_2 de quelques 5000 vecteurs d'appentissage. La figure 5.6b présente de manière similaire, la projection de G_1 en fonction de G_2 du dictionnaire constitué de 1024 vecteurs. On constate que le dictionnaire offre une couverture satisfaisante.

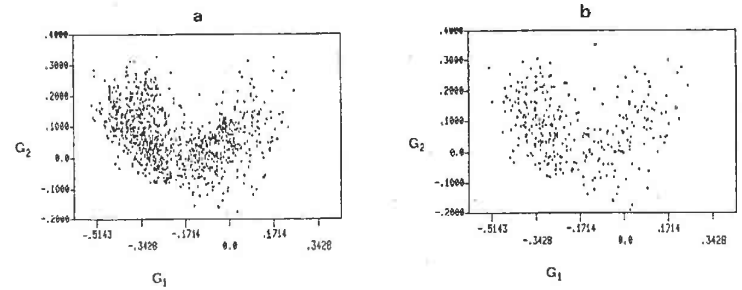


Figure 5.6: Projection dans le plan de G_1 en fonction de G_2, a) de 5000 vecteurs d'appentissage, b) des 1024 vecteurs du dictionnaire

Les résultats expérimentaux démontrent l'efficacité du codage vectoriel, qui permet de limiter à 1024 vecteurs la taille du dictionnaire pour un filtre d'ordre 12. Ceci est illustré par la figure 5.7, où plusieurs dictionnaires offrant des débits de 5 à 11 bits/trame ont été évalués.

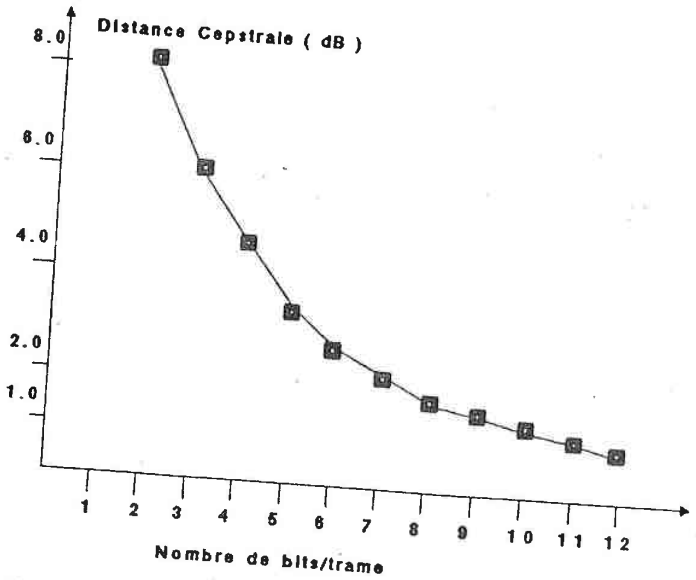


Figure 5.7: Distance cepstrale moyenne pour un codage vectoriel d'un filtre d'ordre 12

Le dictionnaire de 1024 vecteurs offre une distance cepstrale moyenne de 1.3 dB, qui est du même ordre que celle que procure la quantification scalaire avec 40 bits par trame.

V.2.3 QUANTIFICATION MIXTE:

La quantification scalaire ne permet pas de réduire le débit à moins de 35 à 40 bits/trame pour un filtre d'ordre 12. Par contre, le codage vectoriel permet de diviser par un facteur 4 ce débit, mais au prix d'un effort de mémorisation et d'un coût en calcul important.

Comme K_1 et K_2 représentent 90% de l'inertie (fig 5.8), nous proposons de les quantifier de manière scalaire. Les K_i restant sont encodés de manière vectorielle. Ainsi, on réduit de deux la dimension des vecteurs, en éliminant les axes suivant lesquels l'inertie est la plus forte.

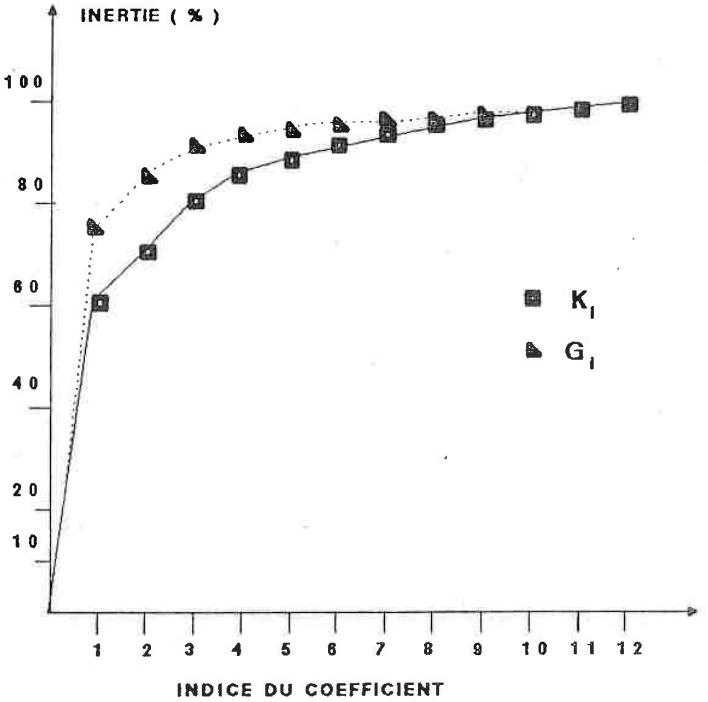


Figure 5.8: Inertie des K_i et G_i pour les index 1 à 12

Il en résulte une diminution du nombre de vecteurs du dictionnaire. Ceci est dû à la réduction à 10 de la dimension des vecteurs, mais également à l'élimination des axes qui procurent l'inertie la plus forte. La figure ci-dessous illustre l'évolution du dictionnaire en fonction de la dimension des vecteurs.

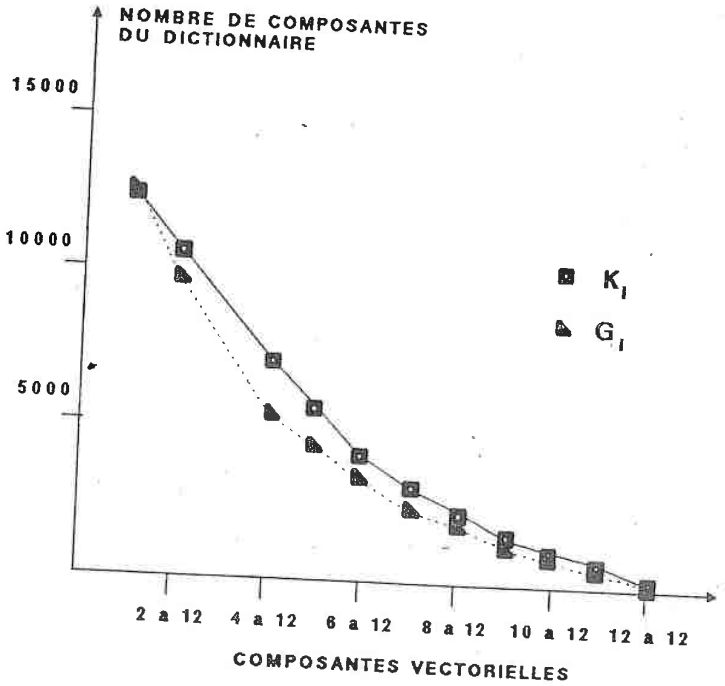


Figure 5.9: Evolution de la taille du dictionnaire en fonction de la dimension des vecteurs a)K_i, b)G_i

Une distance cepstrale du même ordre que celle d'une quantification scalaire sur 40 bits est obtenue pour un dictionnaire de 512 vecteurs.

V.2.4 RESULTATS COMPARATIFS ET COMPLEXITE:

Nous venons de décrire 3 procédures de quantifications des paramètres K_i du filtre de synthèse. Leur performances en terme de réduction du débit, mais également en terme de complexité, sont éminemment variables.

Les performances des quantificateurs scalaires, vectoriels et mixtes sont comparées à débit égal et distance ceptrale égale (fig 5.10). Il apparaît que la quantification vectorielle permet de réduire le débit à 10 bits par vecteur, là où la quantification scalaire en nécessite 40. La quantification mixte semble être un bon compromis, car pour un débit de 18 bits par vecteur,

la dimension du dictionnaire à mémoriser est réduite de moitié. Il en résulte également une réduction de la complexité du traitement d'un facteur 2.

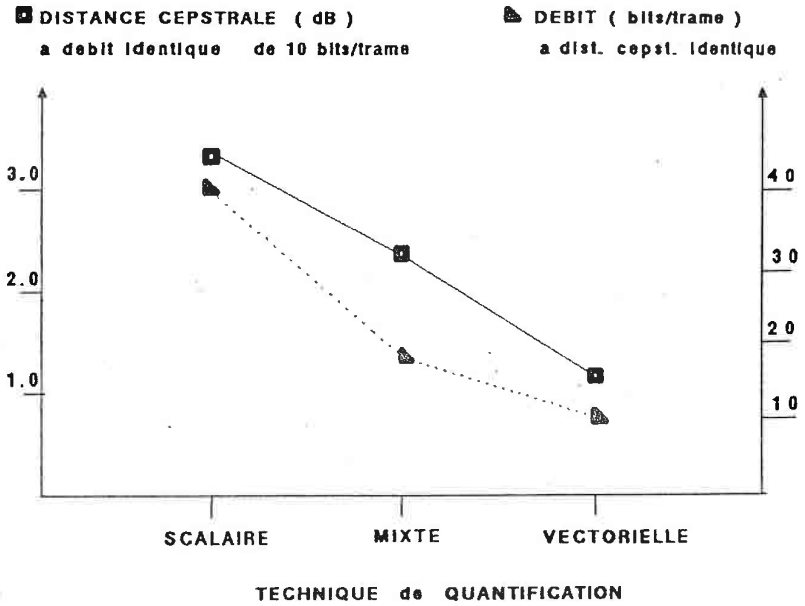


Figure 5.10: Comparaison entre les quantificateurs scalaires, mixtes et vectoriels pour un filtre d'ordre 12.

Nous avons mentionné dans les chapitres précédents que l'ordre du filtre de synthèse pour un codeur à excitation par code doit être plus élévé que pour un codeur à excitation multi-impulsionnelle, car l'efficacité de la procédure de modélisation par code dépend beaucoup de la blancheur du signal résiduel.

V.2.4.1 PREDICTEUR A COURT TERME DU CODEUR A EXCITATION MULTI-IMPULSIONNELLE:

L'ordre retenu pour le filtre de synthèse est de 12. Ce choix est le résultat de tests d'écoute et d'évaluations objectives de la distance cepstrale. Dans ce cas, l'application des 3 procédures de quantification décrites précédemment permet d'envisager les débits suivants:

quantification	débit bits/trame	répartition du débit sur k_1 à k_{12}	complexité x,+/trame	mémorisation ROM kbits
scalaire	40 bits	5,5,4,4,3,3,3,3,3,3,2	140	tab ->1.37 cal ->0.4
mixte	18 bits	5,4 pour K_1 et K_2 9 pour K_3 à K_{12}	5168	50.5
vectoriel	10 bits	10 pour K_1 à K_{16}	12288	120

Tableau 5.4: Débit et complexité du filtre de synthèse d'ordre 12 suivant les techniques de quantification retenues pour le codeur à excitation multi-impulsionnelle.

V.2.4.2 FILTRE DE SYNTHESE DU CODEUR A EXCITATION PAR CODE:

Se basant sur des simulations et tests d'écoute, nous avons retenu un filtre d'ordre 16. De façon à conserver un certain équilibre entre le débit nécessaire au filtre de synthèse et à l'excitation, le codage vectoriel des paramètres du filtre s'impose. Néanmoins le tableau ci-dessous fournit, à titre indicatif, les débits atteints par les autres techniques de codage.

quantification	débit bits/trame	répartition du débit sur k_1 à k_{12}	complexité x,+/trame	mémorisation ROM kbits
scalaire	50 bits	5,5,4,4,3,3,3,3,3,3,2,3,2,3,2,2,2	176	tab ->2.06 cal ->0.4
mixte	18 bits	5,4 pour K_1 et K_2 9 pour K_3 à K_{16}	7232	70.6
vectoriel	11 bits	11 pour K_1 à K_{16}	32768	320

TABLEAU 5.5: Débit du filtre de synthèse d'ordre 16.

V.3 CODAGE DES PARAMETRES DU PREDICTEUR A LONG TERME:

Les paramètres inhérents au prédicteur à long terme sont:

- le coefficient de prédiction à long terme B, qui peut être encodé par quantification.

- le décalage P, qui nécessite un encodage sans erreur.

V.3.1) ENCODAGE DU COEFFICIENT DE PREDICTION:

Comme le prédicteur à long terme est à coefficient unique, une quantification scalaire de B s'impose. La stabilité de ce prédicteur est garantie si B est strictement inférieur à 1 en module. La quantification de B met en oeuvre une procédure de quantification à erreur minimale, qui prend en compte la distribution en amplitude du coefficient B. La figure ci-dessous visualise sa distribution.

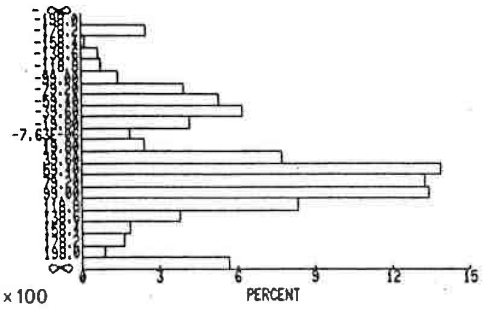


Figure 5.11: Histogramme du coefficient B du prédicteur à long terme

Les évaluations objectives (fig 5.12) montrent une rupture très nette du rapport signal à bruit segmental lorque le coefficient B est encodé sur moins que 4 bits. Aussi, le nombre minimum de bits à accorder à B est de 4.

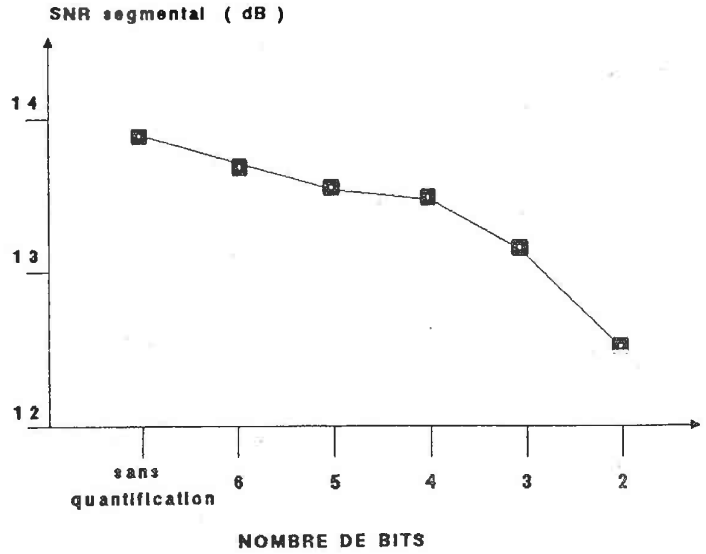


Figure 5.12: Evolution du rapport signal sur bruit segmental

V.3.2 CODAGE DU DECALAGE:

Ce paramètre nécessite un encodage sans erreur. En effet, un décodage non exacte de P, peut provoquer le décalage voire l'anéantissement de l'impulsion glottique, qui excite le filtre de synthèse. P évolue entre les bornes définies par Pmin et Pmax. La valeur à transmettre est P-Pmin car Pmin est une constante qui est égale à la dimension de la fenêtre de modélisation. L'histogramme ci-dessous illustre la distribution de P, dans le cas d'une modélisation multi-impulsionnelle.

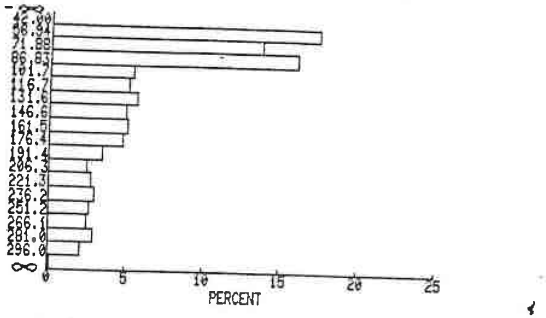


Figure 5.13: Histogramme de P pour Pmin=40 et Pmax=296

La valeur de Pmax n'est pas sans influence sur les performances de la prédiction à long terme (fig 5.14). Les résultats des évaluations objectives montrent que Pmin-Pmax peut être limité à 128.

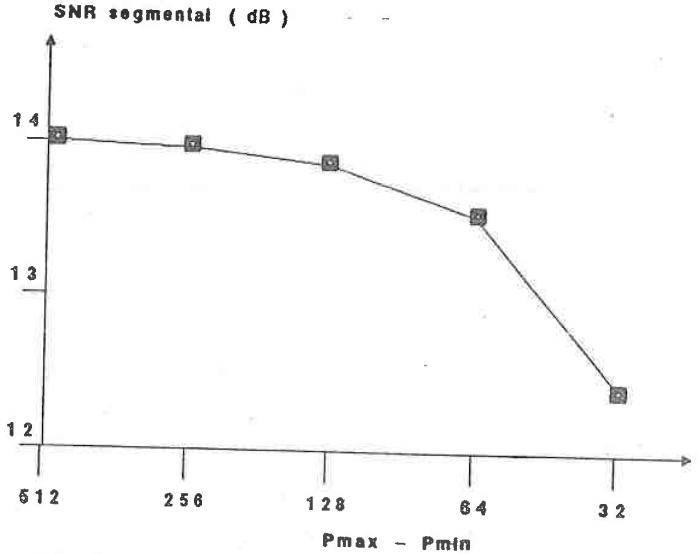


Figure 5.14: Evolution du rapport signal sur bruit en fonction de Pmax

Réaliser un simple encodage binaire de P-Pmin est pénalisant d'un point de vue du débit. Aussi, nous avons opté pour l'introduction de contraintes lors du calcul de P, de telle sorte que le coefficient P de la sous-fenêtre courante est localisé dans un intervalle de ± 32 par rapport au P de la sous-fenêtre précédente. Il en résulte que le décalage P peut être encodé de manière exacte sur 6 bits.

V.3.3 RESULTATS:

Nous venons de décrire les procédures de codage des paramètres du prédicteur à long terme. On peut constater que c'est le paramètre P qui nécessite le débit le plus élevé. En effet l'encodage de ce dernier doit être sans erreur. L'introduction de contraintes permet toutefois de limiter son débit à 6 bits.

Le débit qu'il est nécessaire d'allouer aux paramètres du prédicteur à long terme est résumé dans le tableau ci-dessous.

paramètre codé	débit bits/trame
le coefficients de prédiction linéaire B	4 bits
le décalage P	6 bits
le prédicteur à long terme globalement	10 bits

Tableau 5.6: Débit correspondant aux paramètres du prédicteur à long terme

V.4 CODAGE DE L'EXCITATION:

Les excitations élémentaires qui sont mises en oeuvre dans les différentes procédures de modélisation que nous avons décrites dans les chapitres III et IV sont:

- l'excitation multi-impulsionnelle
- l'excitation par code

V.4.1 CODAGE DE L'EXCITATION MULTI-IMPULSIONNELLE:

Encoder le signal d'excitation multi-impulsionnelle, échantillon par échantillon, n'est pas efficace compte-tenu du nombre considérable d'échantillons nuls. Aussi, il est préférable de considérer uniquement les impulsions en terme de position et amplitude. Si les amplitudes peuvent être encodées par quantification, il n'en va pas de même pour les positions, qui nécessitent un encodage sans erreur. En effet l'approximation de la position introduit dans le signal synthétique des déphasages qui provoquent une diminition sensible du rapport signal sur bruit segmental.

V.4.1.1 CODAGE DES AMPLITUDES:

L'étude statistique des amplitudes des impulsions met en évidence l'apport du prédicteur à long terme qui réduit la dynamique des impulsions. Ceci favorise leur encodage.

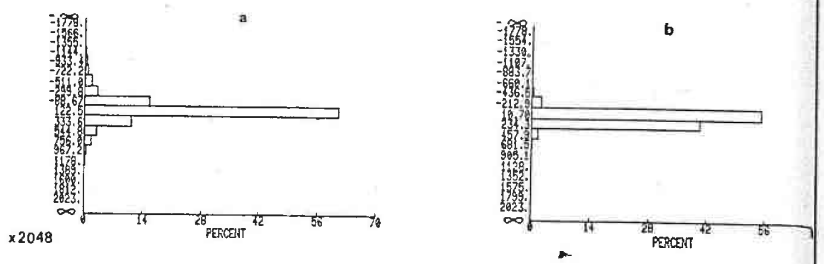


Figure 5.15: Histogramme des amplitudes des impulsions a) sans prédiction à long terme, b) avec prédiction à long terme

Compte-tenu de la dispersion en amplitude des impulsions, nous avons développé une solution tabulée non-uniforme à écart minimum. Cette procédure réalise une pseudo-normalisation en sélectionnant une table de quantification parmi I en fonction d'un critère simple, défini par:

$$Crit = MAX \mid A_i \mid \quad \text{pour } i = 1 \text{ à } K \qquad (5.11)$$

K étant le nombre d'impulsions par sous-fenêtre

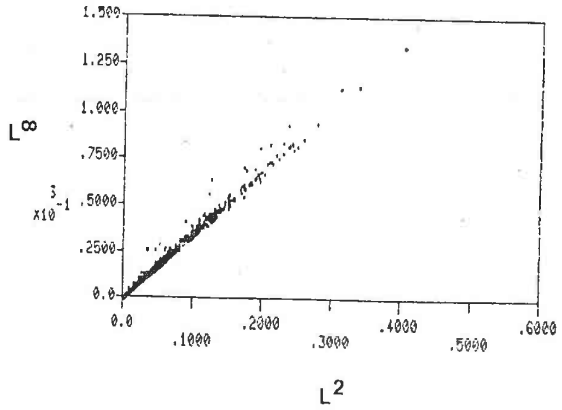


Figure 5.16: Projection dans le plan de la norme L2 en fonction de la norme L infini

Comme l'illustre la figure 5.16, le comportement de ce critère est très proche de la norme L2 . Ceci s'explique par le fait que la contribution des impulsions secondaires dans la norme L2 est négligeable, car celles-ci sont souvent atténuées dans un rapport supérieur à 4 par rapport à l'impulsion principale.

C'est à partir de la distribution statistique du critère L infini , que les I plages de quantifications sont déduites. A chacune de ces plages est associée une table de quantification. Chaque table est spécifique aux séquences multi-impulsionnelles qui entrent dans la plage correspondante.

Le principe de la procédure de quantification est décrit dans la figure ci-dessous. Après l'évaluation du critère, la table de codage séléctionnée encode toutes les amplitudes de la séquence. Les paramètres encodés que fournit cette procédure sont d'une part l'index correspondant à la table sélectionnée, d'autre part les valeurs codées des amplitudes.

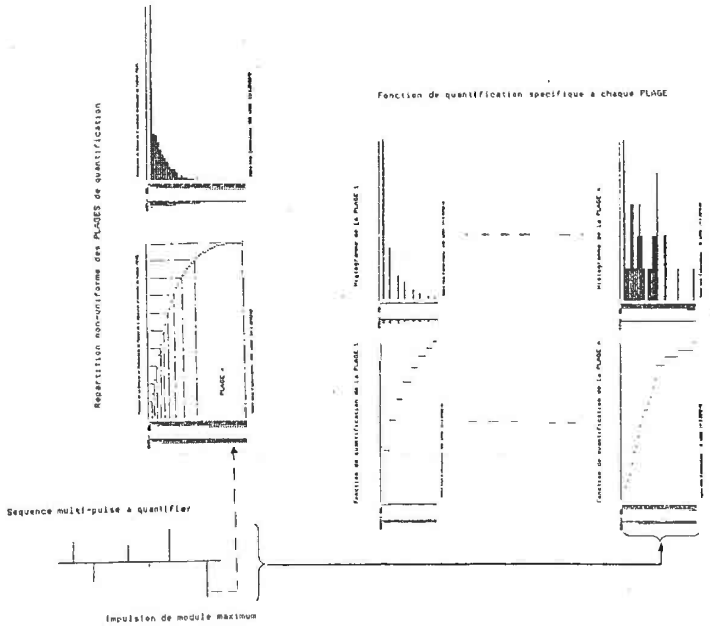


Figure 5.17: Quantification non-uniforme à erreur minimale des amplitudes des impulsions

Les évaluations objectives, en terme de rapport signal sur bruit segmental, mettent en évidence que la subdivision en 8 plages (soit 8 tables de codage) est satisfaisante. En revanche, l'encodage scalaire des amplitudes ne permet pas de réduire à moins de 3 bits le débit par amplitude (fig 5.18).

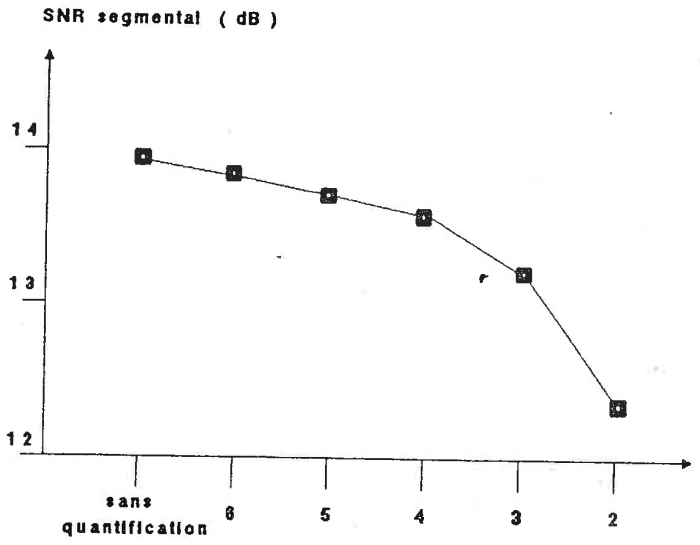


Figure 5.18: Evolution du rapport signal sur bruit segmental en fonction du nombre de bits alloués aux amplitudes des impulsions

V.4.1.2 CODAGE DES POSITIONS:

La distribution statistique des positions des impulsions est quasi-uniforme, sauf aux limites de l'intervalle de modélisation. Ceci est lié aux effets de bord, pour cela il faut revenir à la relation 3.8 du chapitre III. Comme le signal résiduel en aval de l'intervalle courant est nul, l'estimation des impulsions dans cette région est faussée. C'est pourquoi, il apparaît plus d'impulsions en début d'intervalle pour compenser effet.

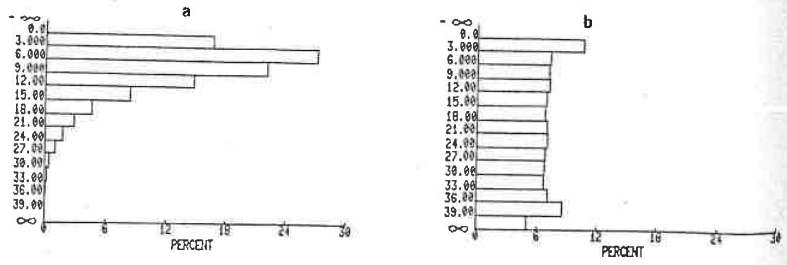


Figure 5.19: Histogramme a) des écarts entre deux impulsions consécutives, b) des positions des impulsions, sur un intervalle de modélisation de longueur 40 échantillons

L'encodage brutal des positions par rapport à la longueur de l'intervalle de minimisation procure un débit exorbitant de $\log_2 N'$ bits par position. Ceci se traduit pour une fenêtre de longueur 40 à 6 bits par impulsion. A raison de 5 impulsions par fenêtre on obtient un débit de 6 kbits/s uniquement pour les positions.

Une solution plus économique est obtenue par un encodage différentiel des positions des impulsions. En théorie, cette solution ne permet pas de réduire le débit, car le nombre maximum de bits nécessaires est de $\log_2 N\text{-}MI$ bits (MI est le nombre d'impulsions dans l'intervalle N) par impulsion. En pratique, une réduction à 5 bits par position peut être obtenue. Car comme le montre la figure ci-dessous, l'écart entre deux impulsions consécutives reste inférieur à 31 avec une probabilité supérieure à 99%. Néanmoins, ce débit qui est de 5 kbits/s est encore trop élevé.

CODAGE ENUMERATIF DES POSITIONS DES IMPULSIONS:

Schalwijk [16] et Cover [3] ont proposé une procédure de codage, qui peut être appliquée aux positions des impulsions si celles-ci sont ordonnées dans le temps. Cet algorithme est obtenu en traversant, au fur et mesure qu'on parcourt l'intervalle, un arbre binaire dans lequel un 1 indique la présence d'une impulsion. Chaque fois qu'un 1 est rencontré dans l'arbre binaire un compteur est incrémenté de:

$$C_m^n = \frac{n!}{(k)!(n-k)!} \qquad (5.12)$$

où n représente le nombre d'échantillons restant jusqu'à la fin de l'intervalle

et k le nombre de positions plus une qu'il reste encore à encoder.

La valeur finale du compteur contient le code unique Ω qui représente la valeur encodée des positions des impulsions.

$$\Omega = \sum_{k=1}^{K} C_{K+1-k}^{N-M_k} \qquad (5.13)$$

La procédure d'encodage s'écrit:

```
Ω = 0
m = K
pour k = 1 à K
      n = N - Mk
      Ω = Ω + C(n,m)
      m = m - 1
fin
```

A l'opposé de l'encodage, la procédure de décodage teste si pour toutes les positions possibles l'incrément C est inférieur au code Ω. Si ceci est vérifié, la position courante correspond à une position d'impulsion et C est

soustrait de Π. Cette opération est répétée pour toutes les positions.

La procédure de décodage s'écrit:

```
k = K
pour n = N à 1
    si  C(n,k) < Π  alors
        Π = Π - C(n,k)
        M_k = N - n
        k = k - 1
    fin
fin
```

La valeur maximum que prend Π est:

$$C_K^N - 1$$

En plaçant 5 impulsions par intervalle de dimension 40, la valeur maximum du code est de 658008. Il peut être représenté sur 20 bits. Ainsi le débit correspondant à la position d'une impulsion est de 4 bits (fig 5.20)

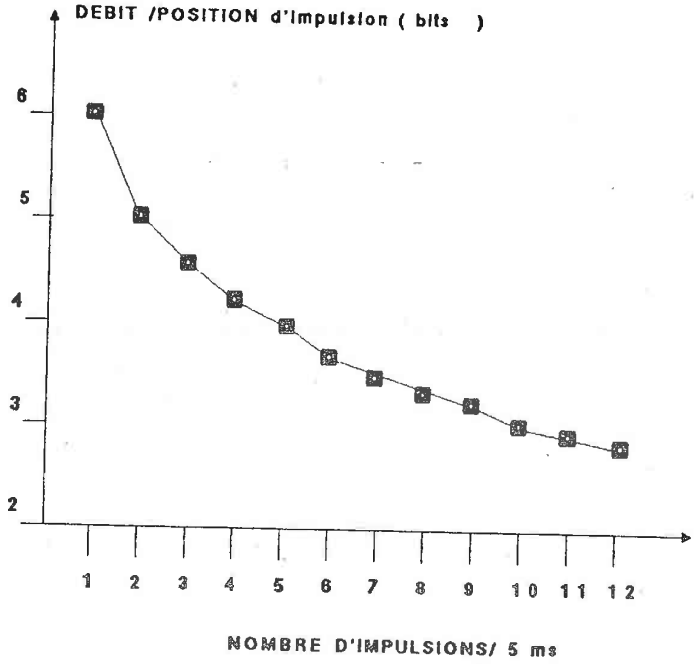


Figure 5.20: Evolution du débit des positions en fonction du nombre d'impulsions pour un codage énumératif sur un intervalle de dimension 40

De façon à bénéficier au maximum de l'efficacité de ce codage, nous avons opté pour une solution qui consiste à positionner alternativement, toutes les impulsions d'une même séquence d'excitation aux positions paires puis impaires. Ainsi la dimension de l'intervalle est divisé par un facteur 2. En reprenant l'exemple précédent, la valeur maximum du code est alors de 15504, soit un débit par position de 2.8 bits. Cette solution apporte toutefois une légère dégradation, que des tests d'écoute ont révélé comme non significative.

V.4.1.3 RESULTATS:

Nous venons de décrire les procédures de codage des amplitudes et positions des impulsions.

L'encodage des amplitudes des impulsions est réalisé par une procédure de quantification scalaire non-uniforme à erreur minimale, qui intègre une pseudo-normalisation. Cette procédure permet de limiter à 3 bits le débit par amplitude. Il faut ajouter à ce débit les 3 bits liés à l'index de la table de quantification qui est identique pour toutes les impulsions d'une même séquence d'excitation multi-impulsionnelle.

L'encodage des positions des impulsions, met en oeuvre une procédure de codage énumératif, qui permet, pour les paramètres standards (1000 impulsions/s et dimension de l'intervalle = 40), de réduire à 4 bits le débit par position d'impulsion.

Le débit qu'il est nécessaire d'allouer à l'excitation multi-impulsionnelle, à raison de 1000 impulsions par seconde, est résumé dans le tableau ci-dessous.

paramètre codé	débit par impulsion
amplitude de l'impulsion	3 bits
index de la table de quantification	3/5 bits
position de l'impulsion	4 bits
une impulsion globalement	7 + 3/5 bits

Tableau 5.7: Débit correspondant à l'excitation multi-impulsionnelle, à raison de 1000 impulsions par seconde

V.4.2 CODAGE DE L'EXCITATION PAR CODE:

Le signal d'excitation par code est caractérisé par l'index du dictionnaire et le facteur de gain qu'il faut appliquer à la séquence d'excitation lors de la synthèse. L'index représente une information encodée. Il ne reste donc que le facteur de gain, auquel nous avons appliqué un codage différentiel, compte-tenu de l'évolution lente de ce paramètre par rapport à la période de calcul de l'excitation. Une quantification scalaire non-uniforme à erreur minimale est ensuite appliquée à l'erreur de prédiction.

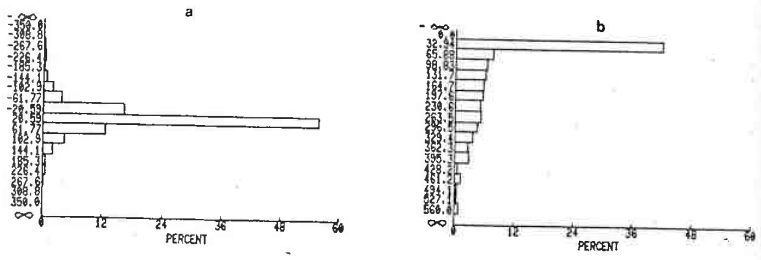


Figure 5.21: Histogrammes a)du facteur de gain, b) de la différence entre deux facteurs de gain consécutifs.

Les évaluations objectives ainsi que les tests d'écoute, montrent que 4 bits sont suffisants pour encoder, sans dégradation significative, le facteur de gain.

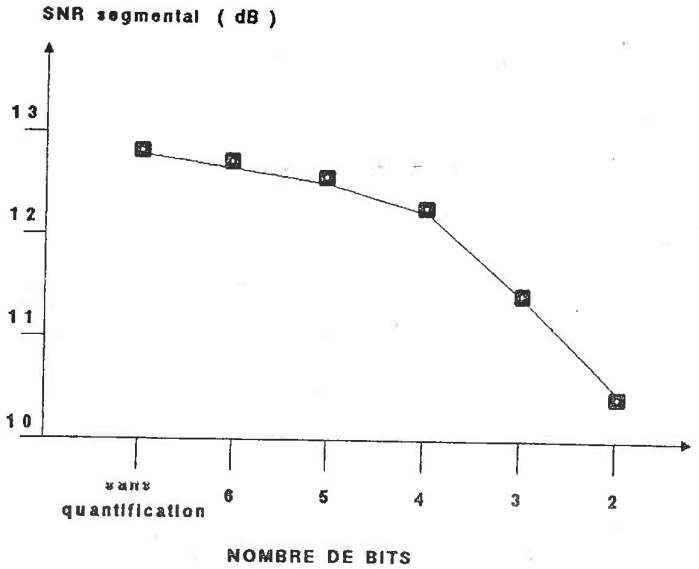


Figure 5.22: Evolution du rapport signal sur bruit segmental en fonction du nombre de bits alloués au facteur de gain des séquences d'excitation stochastique.

V.5 CONCLUSION:

Nous venons de détailler dans ce chapitre des procédures de codage adaptées aux différents paramètres que fournissent les codeurs que nous avons décrits dans les chapitres III et IV. Parmi ces procédures de codage très diverses, on peut citer les techniques de codage scalaire ou vectorielle qui s'opposent aux procédures de codage sans erreur. Ces dernières sont d'ailleurs indispensables à l'encodage des paramètres représentatifs du temps telles que les positions des impulsions et le décalage du prédicteur à long terme. Il existe sans doute encore d'autres procédures de codage, notament les procédures de codage semi-vectorielles.

Le principal enseignement, que nous pouvons tirer des résultats que nous avons obtenus dans ce chapitre, est que les performances de ces procédures de codage conditionnent largement la qualité du signal synthétique restitué. Les effets de ces procédures de codage, sur l'évolution temporelle et spectrale du signal de parole, ont été examinés.

Comme on peut s'y attendre, le taux de compression qu'offrent les procédures de codage sans erreur, est moins élévé que celui des procédures mettant en oeuvre une quantification. C'est le cas pour l'excitation multi-impulsionnelle, où l'encodage des positions requiert un débit plus élévé (4 bits par impulsion), que les amplitudes (3 bits par impulsion).

Il faut également aborder les aspects concernant la complexité et l'effort de mémorisation qu'engendrent ces procédures de codage. Là encore il n'y a pas de surprise, plus les procédures de codage sont efficaces en terme de débit, plus leur complexité est élévée. Pour les procédures de quantification vectorielle, cette complexité élévée s'accompagne d'un effort de mémorisation important, qui est engendré par le stockage du dictionnaire. L'utilisation de dictionnaires stucturés [6,14] permet de réduire cette complexité, au détriment de la qualité, car la quantification dans ce cas est sous-optimale.

Bibliographie:

Publications:

[1] "La quantification Vectorielle des Signaux: Approche Algébrique"
ADOUL J.-P.
Ann. Télecommun., 1986, pages 158-177

[2] "Predictive Coding of Speech at Low Bit Rates"
ATAL B.S.
IEEE Trans. on Comm., 1982, pages 600-614

[3] "Enumerative Source Coding"
COVER T.M.
IEEE Trans. IT, 1973,pages 73-77

[4] "On the Structure of Vector Quantizers"
GERSHO A.
IEEE Trans. on IT, 1982, pages 157-166

[5] "Vector Quantization: A Pattern-Matching Technique for Speech Coding"
GERSHO A., CUPERMAN V.
IEEE Trans. on ASSP, 1983, pages 15-21

[6] "Full Search and Tree Searched Vector Quantization of Speech Waveforms"
GRAY R.M., ABUT H.
IEEE Proc. Int. Conf. on ASSP, 1982, pages 593-596

[7] "Quantization and Bit Allocation in Speech Processing"
GRAY A.H., MARKEL J.D.
IEEE Trans. on ASSP, 1976, pages 459-473

[8] "Comparison of Optimal Quantizations of Speech Reflection Coefficients"
GRAY A.H., MARKEL J.D.,GRAY R.M.
IEEE Trans. on ASSP, 1977, pages 9-23

[9] "Distortion Performance of Vector Quantization for LPC Voice Coding"
JUANG B.-H., WONG D.Y., GRAY A.H.
EEE Trans. on ASSP, 1982, pages 294-304

[10] "A Fixed Point Computation of Partial Correlation Coefficients"
LE ROUX J., GUEGUEN C.
IEEE Trans. on ASSP, 1977, pages 257-259

[11] "Least Squares Quantization in PCM"
Lloyd S.P.
IEEE Trans. on IT, 1957, pages 129-137

[12] "Implementation and Comparison of Two Transformed Reflection Coefficinet Scalar Quantization Methods"
MARKEL J.D., GRAY A.H.
IEEE Trans. on ASSP, 1980, pages 575-583

[13] "Codeur Multi-impulsionnel avec Prediction Vectorielle à Long Terme, un Algorithme, une Procédure de Codage, l'Apport du Language ADA"
MOREAU N., DYMARSKI P., FRITSCH J.-G.
COllo. GRETSI, 1987, pages 423-426

[14] "Les Techniques de Numérisation de la Parole à Bas Débit Applicables aux Liaisons Navales"
POTAGE J., ROCHETTE D., MATHEVON G.
Revue Technique Thomson-CSF, vol. 18, 1986, pages 171-205

[15] "Product Code Vector Quantizers for Waveform and Voice Coding"
SABIN J.M., GRAY R.M.
IEEE Trans. on ASSP, 1984, pages 474-488

[16] "An Algorithm for Source Coding"
SCHALWIJK J.P.M
IEEE Trans. on IT, 1972, pages 395-399

[17] "Bit Allocation and Encoding for Vector Sources"
SEGALL A.
IEEE Trans. on IT, 1976, pages 162-169

[18] "Piecewise Linear Quantization of LPC Reflection Coefficients"
UN C.K., YANG S.C.
IEEE Proc. Int. Conf. on ASSP, 1986, pages 417-420

[19] "Quantization Properties of Transmission Parameters in Linear Prediction Systems"
VISWANNATAN R., MAKHOUL J.
IEEE Trans. on ASSP, 1975, pages 309-321

[20] "An 800 bits/s Vector Quantization LPC Vocoder"
WONG D.Y., JUANG B.-H., GRAY A.H.
IEEE Trans. on ASSP, 1982, pages 770-780

CHAPITRE VI

4 CODEURS DE QUALITE SUB-TELEPHONIQUE

CHAPITRE VI

VI.1 INTRODUCTION:

Les caractéristiques déterminantes dans le choix d'un codeur sont la qualité, le débit, la complexité et l'effort de mémorisation. Dans cet effort de mémorisation, on peut d'ailleurs distinguer, d'une part la mémorisation RAM liée au stockage des signaux et des paramètres évoluant au cours du temps, d'autre part la mémorisation ROM qui résulte du stockage des tables ou dictionnaires de codage.

Nous avons décrit dans les chapitres précédents des procédures de modélisation et de codage efficaces. Aussi nous nous proposons de réaliser quatre codeurs qui pour un débit compris entre 6.1 et 11.8 kbits/s procurent une qualité sub-téléphonique approximativement identique. Notons également, que les 4 codeurs possèdent le même structure de base (fig 6.1 et fig 6.2), à savoir: un prédicteur à court terme, un prédicteur à long terme bouclé dans lequel s'intègrent les différentes procédures de modélisation de l'exitation.

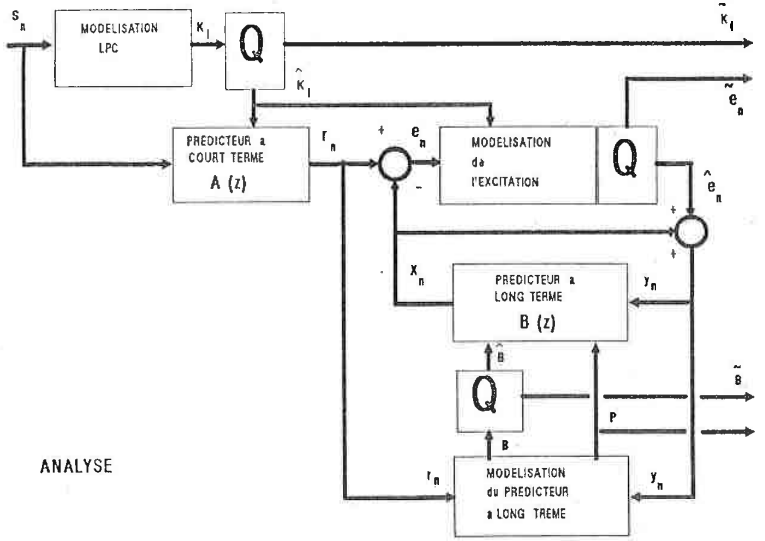


Figure 6.1: Structure de base des 4 codeurs proposés à l'analyse

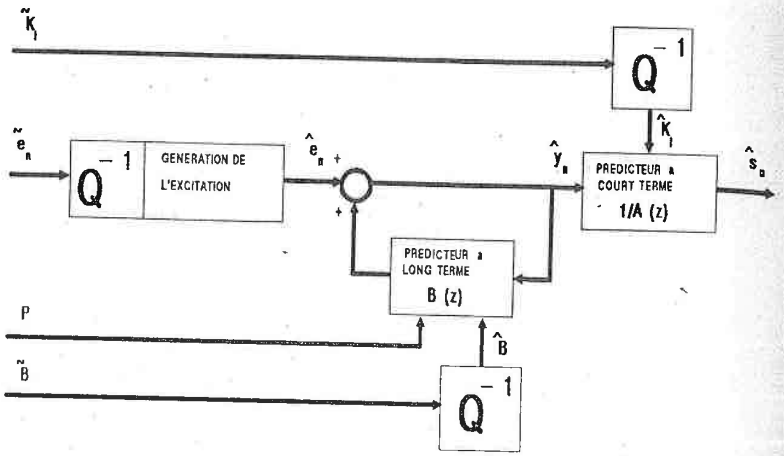


Figure 6.2: Structure de base des 4 codeurs proposés à la synthèse

Chacun des codeurs met en oeuvre une des techniques de modélisation de l'excitation que nous avons décrite dans les chapitres 3 et 4. Notons également, que les procédures de quantification et de codage ont été choisies en rapport avec notamment la complexité des procédures de modélisation de l'excitation. Ainsi, le codeur à excitation multi-impulsionnelle met en oeuvre uniquement des techniques de quantification scalaire, de façon à procurer une complexité relativement réduite. A l'opposé, le codeur à excitation par code réalise une quantification vectorielle des paramètres du filtre de synthèse. Les caractéristiques des quatres codeurs sont évaluées à qualité égale en terme de débit, de complexité et d'effort de mémorisation.

Un jeu de paramètre unique pour l'ensemble des quatres codeurs ne peut être envisagé, compte-tenu de leur particularité. Aussi, un travail de synthèse important a été nécessaire, pour définir les valeurs des paramètres qui permettent d'atteindre avec une dégradation, la plus faible possible, une qualité objective de 12.5 dB ±1dB en rapport signal sur bruit segmental.

VI.2 CODEUR A EXCITATION MULTI-IMPULSIONELLE A 11.8 KBITS/S:

Le codeur à excitation multi-impulsionnelle que nous proposons procure le débit le plus élevé, car il met en oeuvre uniquement des techniques de quantification scalaire. Il en résulte une complexité relativement limitée, qui ne dépasse guère les 2.1 10^6 de multiplications et additions par seconde. Les valeurs des paramètres utilisés sont résumées dans le tableau ci-dessous.

Prédicteur à court terme:

algorithme	Le Roux-Gueguen
actualisation	20 ms
ordre du filtre	12
méthode de codage	quantification scalaire calculée à erreur minimale
débit par trame	40 bits/trame

Prédicteur à long terme:

méthode	optimale simplifiée
actualisation	5 ms
nombre de coefficients	1
décalage maximum Pmax-Pmin	128
méthode de codage des coefficients	quantification scalaire calculée à erreur minimale
débit par trame des coefficients	4 bits/trame
méthode de codage du décalage	codage binaire sans erreur
débit par trame du décalage	7 bits/trame

Excitation:

méthode	multi-impulsionnelle variante 1 avec actualisation de la fonction de localisation
intervalle de minimisation	5 ms
durée de la réponse impulsionnelle	2.5 ms
nombre d'impulsions	5 impulsions/5ms
méthode de codage des amplitudes	quantification scalaire à erreur minimale avec pseudo-normalisation
débit par amplitude d'impulsion	3 + 3/5 bits/impulsion
méthode de codage des positions	codage énumératif
débit par position d'impulsion	4 bits/impulsion

Tableau 6.1: Paramètres et techniques de modélisation et de quantification retenus pour le codeur à excitation multi-impulsionnelle

La répartition des 11.8 kbits donnée dans le tableau ci-dessus n'est pas unique. En effet, si pour le prédicteur à court terme, les choses sont claires, il n'en va pas de même pour le prédicteur à long terme et pour l'excitation multi-impulsionnelle qui sont intimement liés. Une actualisation moins fréquente des paramètres du prédicteur à long terme (toutes les 10 ms uniquement) réduit dans un rapport 2 le débit nécessaire a ce dernier. Les 1.1 kbits/s disponibles, soient 11 bits/10ms, peuvent être affectés à l'excitation multi-impulsionnelle. Ceci permet, pour chaque fenêtre, de transmettre deux impulsions supplémentaires. Néanmoins, ceci n'améliore pas de manière significative les performances.

Le tableau ci-dessous illustre les caractéristiques du codeur à excitation multi-impulsionnelle complet (analyse et synthèse) pour un débit de 11.8 kbits.

Prédicteur à court terme:

débit ..... 2000 bits/s
complexité à l'analyse ..... 315000 x,+/s
complexité à la synthèse ..... 192000 x,+/s
effort de mémorisation RAM ..... 210x13 bits
effort de mémorisation ROM ..... 24X12 bits

Prédicteur à long terme:

débit du prédiceur à long terme ..... 2200 bits/s
complexité à l'analyse..... 1042000 x,+/s
complexité à la synthèse ..... 8000 x,+/s
effort de mémorisation RAM ..... 376x13 bits
effort de mémorisation ROM ..... 8x12 bits

Excitation:

débit global pour l'excitation ..... 7600 bits/s
complexité à l'analyse ..... 745000 x,+/s
complexité à la synthèse ..... 1000 x,+/s
effort de mémorisation RAM ..... 80x13 bits
effort de mémorisation ROM ..... 70x12 bits

CARACTERISTIQUES GLOBALES:

débit ..... 11800 bits/s
complexité à l'analyse ..... 2100000 x,+/s
complexité à la synthèse ..... 200000 x,+/s
effort de mémorisation RAM .... 666x13 bits
effort de mémorisation ROM .... 102x12 bits

Tableau 6.2: Caractéristiques du codeur à excitation multi-impulsionnelle

Cette configuration du codeur à excitation multi-impulsionnelle, permet de restituer un signal de parole dont le rapport signal sur bruit segmental moyen est de 13.2 dB. Cette qualité a servi de référence lors de la réalisation des trois autres codeurs.

VI.3 CODEUR A EXCITATION OPTIMAL PAR CODE A 8.9 KBITS/S:

Les valeurs des paramètres retenues sont le résultat d'évaluations objectives. Les valeurs présentées dans le tableau ci-dessous permettent de restituer un signal de parole dont le rapport signal sur bruit segmental moyen est de 12.7 dB.

Prédicteur à court terme:

méthode ..... Le Roux-Gueguen
actualisation ..... 20 ms
ordre du filtre 12
méthode de codage ..... quantification mixte
taille du dictionnaire ..... 512 vecteurs de dimension 10
débit par trame ..... 18 bits/trame

Prédicteur à long terme:

méthode ..... optimale simplifiée
actualisation ..... 10 ms
nombre de coefficients 1
décalage maximum Pmax-Pmin 128
méthode de codage des coefficients ..... quantification scalaire calculée à erreur minimale
débit par trame des coefficients ..... 4 bits/trame
méthode de codage du décalage ..... codage binaire sans erreur
débit par trame du décalage ..... 7 bits/trame

Excitation:

méthode ..... par code optimal simplifiée
intervalle de minimisation..... 10 ms
durée de la rép. imp. de F(z) ..... 2.5 ms
nombre d'impulsions ..... 700 séquences/s
taille du dictionnaire ..... 4 séquences de 3.125 ms
débit correspondant à l'index ..... 2 bits/séquence
méthode de codage du facteur de gain ... quantification scalaire à erreur minimale avec pseudo-normalisation
débit pour le facteur de gain ..... 3 + 4/7 bits/séquence
méthode de codage des positions ..... codage énumératif
débit pour les positions ..... 32 bits/(7 séquences)

Tableau 6.3: Paramètres et techniques de codage pour le codeur à excitation par code optimal

Le tableau ci-dessous illustre les caractéristiques du codeur à excitation par code optimal complet (analyse et synthèse) pour un débit de 8.9 kbits.

Prédicteur à court terme:

débit ..... 900 bits/s
complexité à l'analyse ..... 568000 x,+/s
complexité à la synthèse ..... 192000 x,+/s
effort de mémorisation RAM ..... 280x13 bits
effort de mémorisation ROM ..... 5130X12 bits

Prédicteur à long terme:

débit du prédiceur à long terme ..... 1100 bits/s
complexité à l'analyse ..... 1029000 x,+/s
complexité à la synthèse ..... 8000 x,+/s
effort de mémorisation RAM ..... 376x13 bits
effort de mémorisation ROM ..... 8x12 bits

Excitation:

débit global pour l'excitation ..... 6.900 bits/s
complexité à l'analyse ..... 6600000 x,+/s
complexité à la synthèse ..... 17500 x,+/s
effort de mémorisation RAM ..... 120x13 bits
effort de mémorisation ROM ..... 80x12 bits

CARACTERISTIQUES GLOBALES:

débit ..... 8900 bits/s
complexité à l'analyse ..... 8200000 *,+/s
complexité à la synthèse ..... 217500 x,+/s
effort de mémorisation RAM .... 400x13 bits
effort de mémorisation ROM .... 5210x12 bits

Tableau 6.4: Caractéristiques du codeur à excitation par code optimal

Le constat est que ce codeur est d'une complexité excessive par rapport au débit qu'il offre.

VI.4 CODEUR A EXCITATION MULTI-IMPULSIONELLE VECTORIELLE 7.6 KBITS/S:

L'efficacité, en terme de débit, de ce codeur est d'autant plus élevée que la dimension des vecteurs d'amplitudes est grande. Néanmoins, comme celà a été décrit dans le chapitre II, une contrainte importante est que le prédicteur à long terme bouclé impose que la procédure de quantification de l'excitation soit placée à l'intérieur de la boucle de prédiction. La solution, qui répond à ces deux exigences, consiste à réactualiser toutes 10 ms le prédicteur à long terme. La dimension des vecteurs d'amplitudes est de 12, sachant que 1200 impulsions par seconde constituent le signal d'excitation.

Cette configuration, dont les valeurs de paramètres sont résumées dans le tableau ci-dessous, restitue un signal de parole de qualité sub-téléphonique. Le rapport signal sur bruit segmental moyen est de 13 dB.

Prédicteur à court terme:

méthode ..... Le Roux-Gueguen
actualisation ..... 20 ms
ordre du filtre 12
méthode de codage ..... quantification mixte
taille du dictionnaire ..... 512 vecteurs de dimension 10
débit par trame ..... 18 bits/trame

Prédicteur à long terme:

méthode ..... optimale simplifiée
actualisation ..... 10 ms
nombre de coefficients 1
décalage maximum Pmax-Pmin 128
méthode de codage des coefficients ..... quantification scalaire calculée à erreur minimale
débit par trame des coefficients ..... 4 bits/trame
méthode de codage du décalage ..... codage binaire sans erreur
débit par trame du décalage ..... 7 bits/trame

Excitation:

méthode ..... multi-impulsionnelle vectorielle simplifiée variante 1 avec actulisation de l'intercorrélation
intervalle de minimisation..... 10 ms
durée de la réponse impulsionnelle ..... 2.5 ms
nombre d'impulsions ..... 1200 impulsions/s
taille du dictionnaire des amplitudes... 256 vecteurs de dimension 12
débit correspondant à l'index ..... 8 bits/séquence
méthode de codage du facteur de gain ... quantification scalaire à erreur minimale
débit pour le facteur de gain ..... 4 bits/séquence d'impulsions
méthode de codage des positions ..... codage énumératif
débit pour les positions des impulsions 44 bits pour 12 impulsions

Tableau 6.5: Paramètres et techniques de modélisation et de quantification retenus pour le codeur à excitation multi-impulsionnelle vectorielle

Le tableau ci-dessous illustre les caractéristiques du codeur à excitation multi-impulsionnelle vectorielle complet (analyse et synthèse) pour un débit de 7.6 kbits.

Prédicteur à court terme:

débit du filtre de synthèse	900 bits/s
complexité	568000 x,+/s
complexité à la synthèse	192000 x,+/s
effort de mémorisation RAM	280x13 bits
effort de mémorisation ROM	5130X10 bits

Prédicteur à long terme:

débit du prédiceur à long terme	1100 bits/s
complexité à l'analyse	1029000 x,+/s
complexité à la synthèse	8000 x,+/s
effort de mémorisation RAM	416x13 bits
effort de mémorisation ROM	16x10 bits

Excitation:

débit global pour l'excitation	5600 bits/s
complexité à l'analyse	1530000 x,+/s
complexité à la synthèse	1200 x,+/s
effort de mémorisation RAM	280x13 bits
effort de mémorisation ROM	3072x10 bits

CARACTERISTIQUES GLOBALES:

débit	**7600 bits/s**
complexité à l'analyse	**3328000 x,+/s**
complexité à al synthèse	**201200 x,+/s**
effort de mémorisation RAM	**956x13 bits**
effort de mémorisation ROM	**8192x10 bits**

Tableau 6.6: Caractéristiques du codeur à excitation multi-impulsionnelle vectorielle

On constate également que la complexité est répartie de manière relativement uniforme sur les trois modélisations. Comme pour le codeur à excitation multi-impulsionnelle, une réduction supplémentaire du débit peut être obtenue en plaçant toutes les impulsions d'une même sous-fenêtre aux positions paires ou impaires. Ainsi la dimension de l'intervalle de modélisation est divisée par deux. Le débit correspondant aux positions se limite alors à 33 bits pour 12 impulsions au lieu de 44. Dans ce cas le débit global de 6.5 kbits/s est obtenu. La dégradation qu'introduit cette simplification, n'est pas significative car elle n'exède 0.7 dB. On peut noter également que la complexité liée à la modélisation multi-impulsionnelle s'en trouve réduite de quelques 160 10^3 multiplications et additions par seconde.

VI.5 CODEUR A EXCITATION STOCHASTIQUE A 6.1 KBITS/S:

Le codeur à excitation par code est d'autant plus efficace que le signal résiduel, auquel est appliqué la modélisation stochastique, est proche d'un bruit blanc. Ceci nous conduit d'une part à augmenter l'ordre du prédicteur à court terme, d'autre part à actualiser ce dernier plus fréquemment. En revanche, comme nous le mentionnions dans le chapitre II, l'augmentation du nombre de coefficients du prédicteur à long terme n'apporte pas d'amélioration significative. Ce prédicteur reste donc à coefficient unique. Le rapport signal sur bruit segmental moyen du signal de parole que permet de restituer ce codeur est de 12.1 dB.

Les valeurs des paramètres utilisées sont résumées dans le tableau ci-dessous.

Prédicteur à court terme:

méthode	Le Roux-Gueguen
actualisation	10 ms
ordre du filtre	16
méthode de codage	quantification vectorielle
débit par trame	11 bits/trame

Prédicteur à long terme:

méthode	optimale simplifiée
actualisation	5 ms
nombre de coefficients	1
décalage maximum Pmax-Pmin	128
méthode de codage des coefficients	quantification scalaire à erreur minimale
débit par trame des coefficients	4 bits/trame
méthode de codage du décalage	codage binaire sans erreur
débit par trame du décalage	7 bits/trame

Excitation:

méthode	modelisation par code simplifié variante 1
intervalle de minimisation	5 ms
taille du dictionnaire d'excitation	512 séquences de 5 ms
débit correspondant à l'index	9 bits/séquence
méthode de codage du facteur de gain	quantification scalaire à erreur minimale avec pseudo-normalisation
débit par facteur de gain	5 bits/séquence

Tableau 6.7: Paramètres du codeur à excitation par code

Le tableau ci-dessous illustre les caractéristiques du codeur à excitation multi-impulsionnel complet (analyse et synthèse) pour un débit de 12 kbits.

Prédicteur à court terme:

débit du filtre de synthèse ..... 1100 bits/s
complexité à l'analyse ..... 3564000 x,+/s
complexité à la synthèse ..... 256000 x,+/s
effort de mémorisation RAM ..... 240x13 bits
effort de mémorisation ROM ..... 32768X10 bits

Prédicteur à long terme:

débit du prédiceur à long terme ..... 2200 bits/s
complexité à l'analyse ..... 1042000 x,+/s
complexité à la synthèse ..... 8000 x,+/s
effort de mémorisation RAM ..... 396x13 bits
effort de mémorisation ROM ..... 18x10 bits

Excitation:

débit global pour l'excitation ..... 2800 bits/s
complexité à l'analyse ..... 19000000 x,+/s
complexité à la synthèse ..... 328000 x,+/s
effort de mémorisation RAM ..... 100x13 bits
effort de mémorisation ROM ..... 14000x10 bits

CARACTERISTIQUES GLOBALES:

débit ..... 6100 bits/s
complexité à l'analyse ..... 23600000 x,+/s
complexité à la synthèse ..... 592000 x,+/s
effort de mémorisation RAM .... 736x13 bits
effort de mémorisation ROM .... 46786x10 bits

Tableau 6.8: Caractéristiques du codeur à excitation par code

Ce codeur dont le débit est fort attrayant est néanmoins très difficile à mettre en oeuvre compte-tenu de la complexité et de l'effort de mémorisation ROM qu'il engendre. L'introduction d'un sous-échantillonnage du signal résiduel, auquel est appliquée la modélisation par code, permet de réduire la taille du dictionnaire d'excitation ainsi que la complexité de traitement (se rapportant à la modélisation de l'excitation) dans un rapport approximativement égal au rapport de sous-échantillonnage.

VI.6 CONCLUSION:

Nous avons proposé dans ce chapitre quatre codeurs qui pour une complexité variant entre 2.1 10^6 et 23.6 10^6 x,+/s, permettent de restituer de la parole de qualité sub-téléphonique, pour des débits compris respectivement entre 11.8 et 6.1 kbits/s.

La figure ci-dessous visualise, pour les quatres codeurs, les dégradations cumulées liées à la quantification des paramètres.

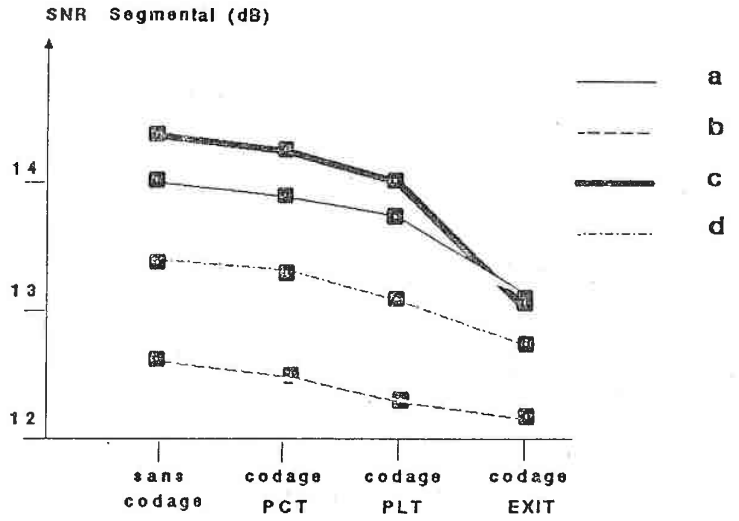


Figure 6.3: Evolution du rapport signal sur bruit segmental, au fur et à mesure de l'encodage des différents paramètres des codeurs a)codeur à excitation multi-impulsionnelle, b)codeur à excitation optimale par code, c)codeur à excitation multi-impulsionnelle vectorielle, d)codeur à excitation par code

La figure ci-dessous illustre les performances de ces codeurs en terme de rapport signal sur bruit segmental, de débit, de complexité, d'effort de mémorisation RAM et effort de mémorisation ROM.

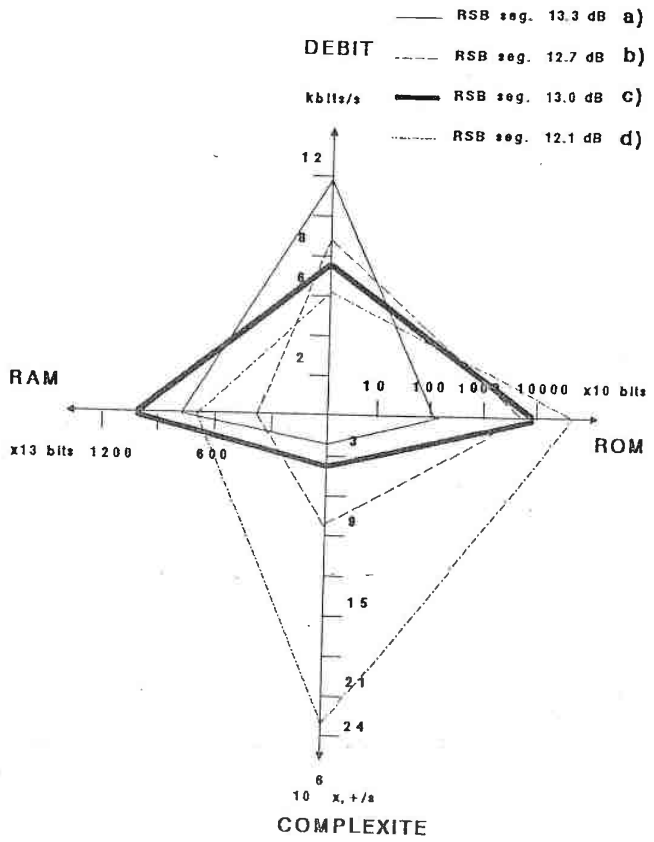


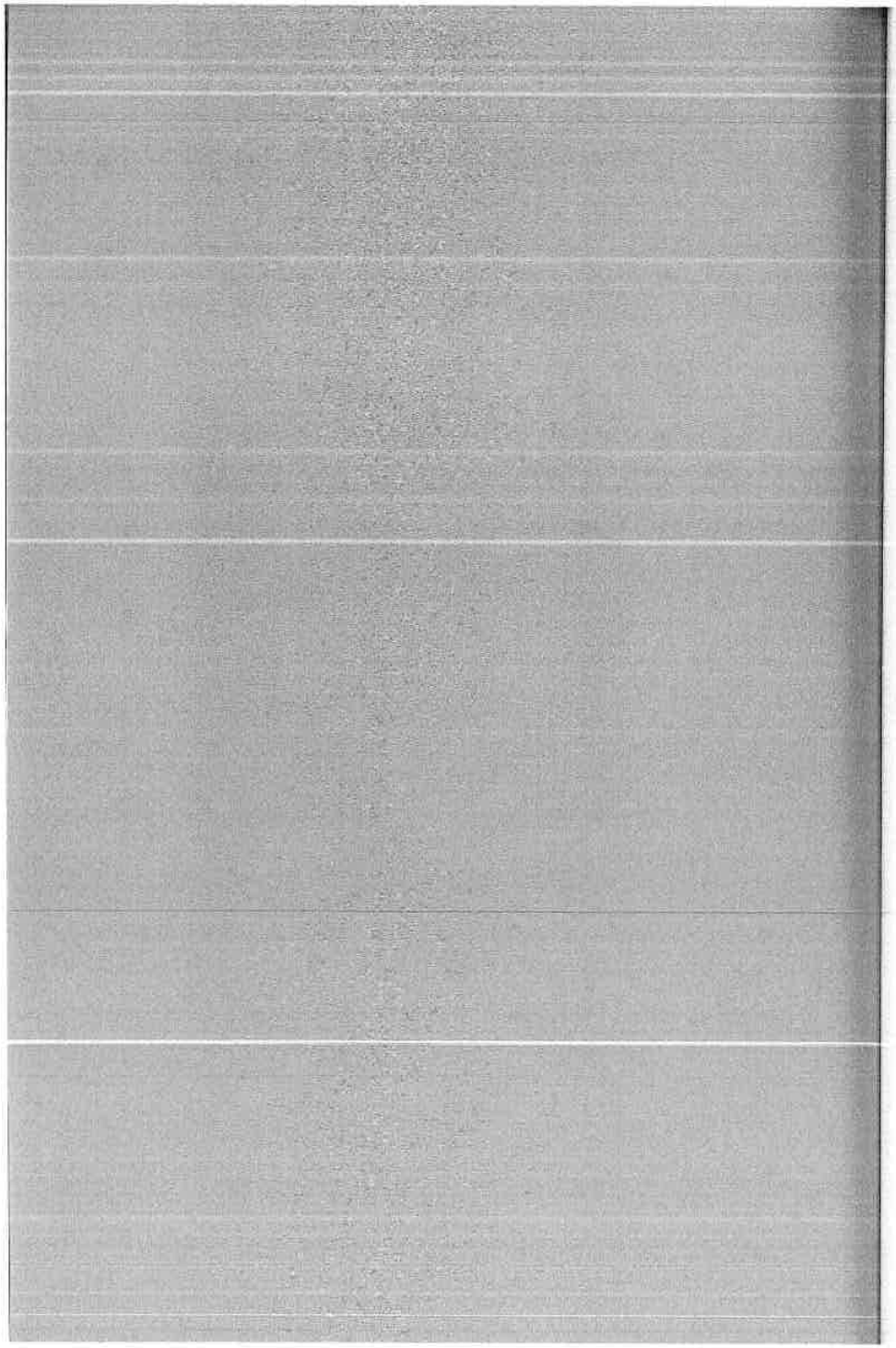
Figure 6.4: Performances des codeurs proposés: a) codeur à excitation multi-impulsionnelle; b) codeur à excitation par code optimal; c) codeur à excitation multi-impulsionnelle vectorielle; d) codeur à excitation par code

Finalement, le codeur à excitation multi-impulsionnelle vectorielle apparaît comme étant le résultat d'un bon compromis par rapport aux autres codeurs, sur le plan de la complexité et du débit principalement. En effet pour une complexité de 35% supérieure à celle du codeur à excitation multi-impulsionnelle, ce codeur permet d'atteindre un débit nettement inférieur à 8 kbits/s. Compte-tenu des capacités des mémoires disponibles actuellement, le stockage des dictionnaires ne représente pas un obstacle à la réalisation d'un

tel codeur. De plus ce codeur est d'une grande souplesse. Il peut évoluer de manière quasi continue vers le codeur à exitation multi-impulsionnelle, ou vers le codeur à excitation par code, dont tous les échantillons des séquences d'excitation sont non nuls.

CHAPITRE VII

CONCLUSION GENERALE



VII: CONCLUSION GENERALE:

Cette thèse a été consacrée à l'étude et la simulation d'une famille de 4 codeurs hybrides, offrant des débits compris entre 6 et 12 kbits/s, pour des applications sub-téléphoniques. Tous les 4 codeurs présentent la même structure de base. Ils combinent une méthode APC (Adaptive Predictive Coding), qui estime l'enveloppe spectrale du signal sous forme de filtre AR, avec des méthodes de déconvolution itérées (scalaire ou vectorielle) qui déterminent l'excitation du filtre de synthèse. La modélisation de l'excitation inclut un prédicteur à long terme bouclé dont la détermination des paramètres [MOREAU, DYMARSKY, FRITSCH] est le résultat de la minimisation globale de l'erreur quadratique perceptuelle entre le signal original et le signal synthétique. Il en est de même pour la séquence d'excitation optimale qui est déterminée de façon à minimiser l'erreur quadratique perceptuelle entre le signal original et le signal synthétique. La pondération perceptuelle a pour rôle de distribuer de manière non-uniforme le spectre du signal d'erreur de façon à mettre à profit l'effet de masquage auditif.

Le signal d'excitation, peut être le résultat d'une approche scalaire ou vectorielle. L'approche scalaire modélise le signal d'excitation par des séquences d'impulsions de dirac. Pour l'approche vectorielle, la recherche de l'excitation, consiste à selectionner parmi un nombre fini de séquences celle qui procure l'erreur la plus faible, au sens du critère perceptuel.

La méthode scalaire, la plus connue est celle proposée par [ATAL 1982], qui porte le nom de modélisation à excitation multi-impulsionnelle. Celle-ci a été étudiée en détail. Un effort important a été porté sur l'étude, la simulation et l'évaluation de procédures qui offrent une réduction significative de la complexité. On peut citer notamment celle qui permet de réduire d'un rapport deux l'intervalle de minimisation en plaçant les impulsions uniquement aux positions paires ou impaires, ou encore celle qui introduit des contraintes de positionnement des impulsions. L'influence des effets secondaires liés à l'actualisation des paramètres du filtre de synthèse sur ces procédures, a également été évaluée. Des procédures de codage scalaire efficaces ont été développées pour encoder notamment l'excitation. Ainsi, ce codeur restitue une parole de qualité sub-téléphonique (13.2 dB de rapport signal sur bruit segmental), pour un débit de 11.8 kbits/s et une complexité globale de l'ordre de 2 10^6 multiplications et additions par seconde. La qualité que procure ce codeur nous a servi de référence lors de la réalisation des autres codeurs.

Un codeur, issu des méthodes vectorielles, est le codeur à excitation par code qui fut proposé récemment par [ATAL 1986]. Notre approche a été de développer une procédure de modélisation rapide déduite de la variante 1, qui met à profit les propriétés de décomposition matricielle [TRANCOSO 1986]. De plus, comme la complexité de cette méthode vectorielle est directement proportionnelle à la dimension du dictionnaire d'excitation, notre approche a également été de constituer ce dernier à l'aide d'un algorithme de classification. Ceci a permis de réduire la complexité à quelques 12 10^6 multiplications et additions par seconde. D'autre part la taille du dictionaire ne compte que 512 vecteurs de dimension 40. Le débit correspondant à l'excitation est de 2.8 kbits/s. L'application d'un codage vectorielle aux coefficients du filtre de synthèse porte, d'une part le débit global à 6.1 kbits/s, d'autre part la complexité à 17 10^6 multiplications et additions par seconde. La parole restituée est de qualité sub-téléphonique, dont le rapport signal sur bruit segmental est de 12.1 dB.

Deux nouvelles procédures de modélisation vectorielle, appellées modélisation optimale par code et modélisation multi-impulsionnelle vectorielle, ont également été développées. Ces procédures combinent les principes des procédures de modélisation multi-impulsionnelle et par code.

Par un traitement et un encodage explicite de l'information de phase contenue dans le signal d'excitation, la procédure de modélisation optimalepar code a permis de réduire à moins d'une dizaine le nombre de séquences d'excitation du dictionnaire. Cette procédure permet, à qualité comparable, de réduire d'un facteur 1.5 le nombre de séquences d'excitation par rapport à une excitation multi-impulsionnelle. Néanmoins, la transmission de l'information correspondant à l'index de l'excitation anéantit partiellement cette réduction d'information. Cette procéduce s'avère relativement complexe et n'apporte pas de réduction significative du débit lié à l'excitation. Notons toutefois que le comportement de ce codeur est plus intéressant pour les très faibles débits. L'utilisation du quantificateur mixte pour encoder les coefficients du filtre de synthèse permet de limiter le débit du codeur à moins de 9 kbits/s. La complexité globale qu'il engendre est de 8.4 10^6 multiplications et additions par seconde.

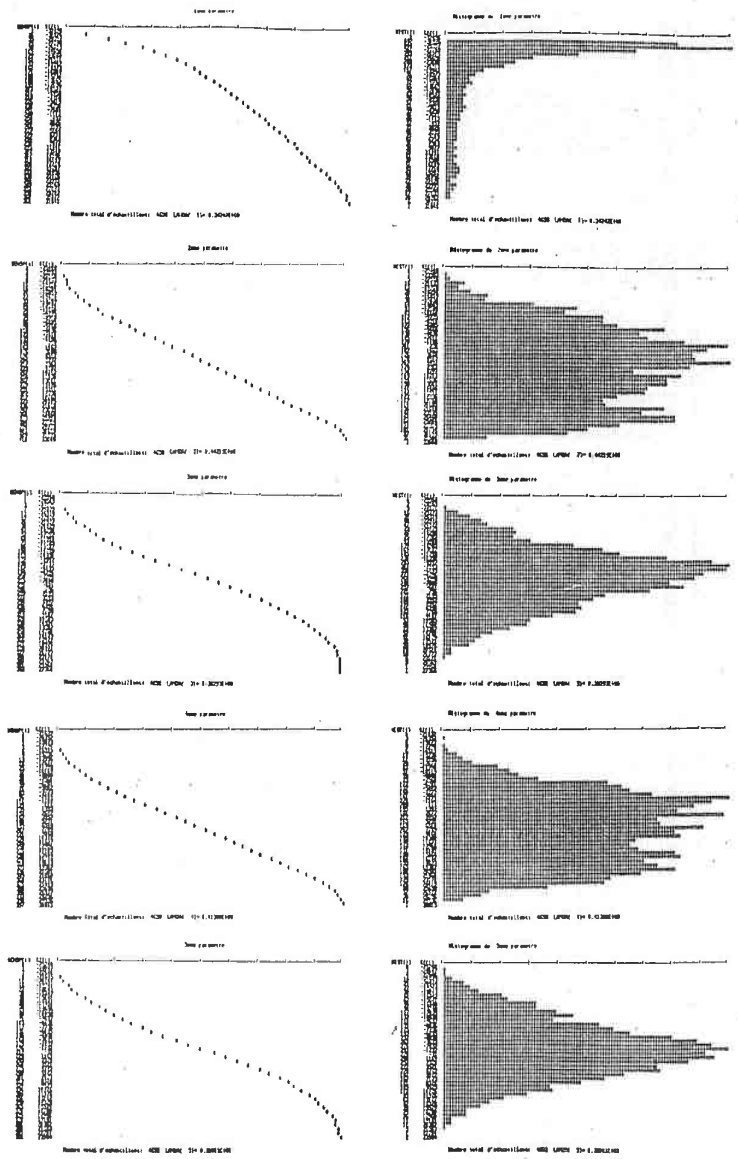
Le codeur à excitation multi-impulsionnelle vectorielle, met en oeuvre une quantification vectorielle pondérée des amplitudes des impulsions. Il procure des performances intéressantes tant sur le plan de la complexité que du débit. En effet pour un signal de parole synthétique de qualité sub-téléphonique, la complexité et le débit atteignent respectivement 3.3 10^6 multiplications et additions par seconde et 7.6 kbits par seconde. De plus, il a l'avantage d'être très souple à l'utilisation, ce qui lui permet de couvrir de manière continue les débits compris entre 7 kbits/s et 10 kbits/s. Il peut ainsi répondre à un grand nombre d'applications telles que les messageries vocales, ou les répondeurs enregistreurs solides.

Finalement, les résultats que nous avons obtenus en simulation, montrent que cette famille de codeurs hybrides permet de restituer de la parole de qualité sub-téléphonique pour des débits compris entre 6 et 12 kbits.

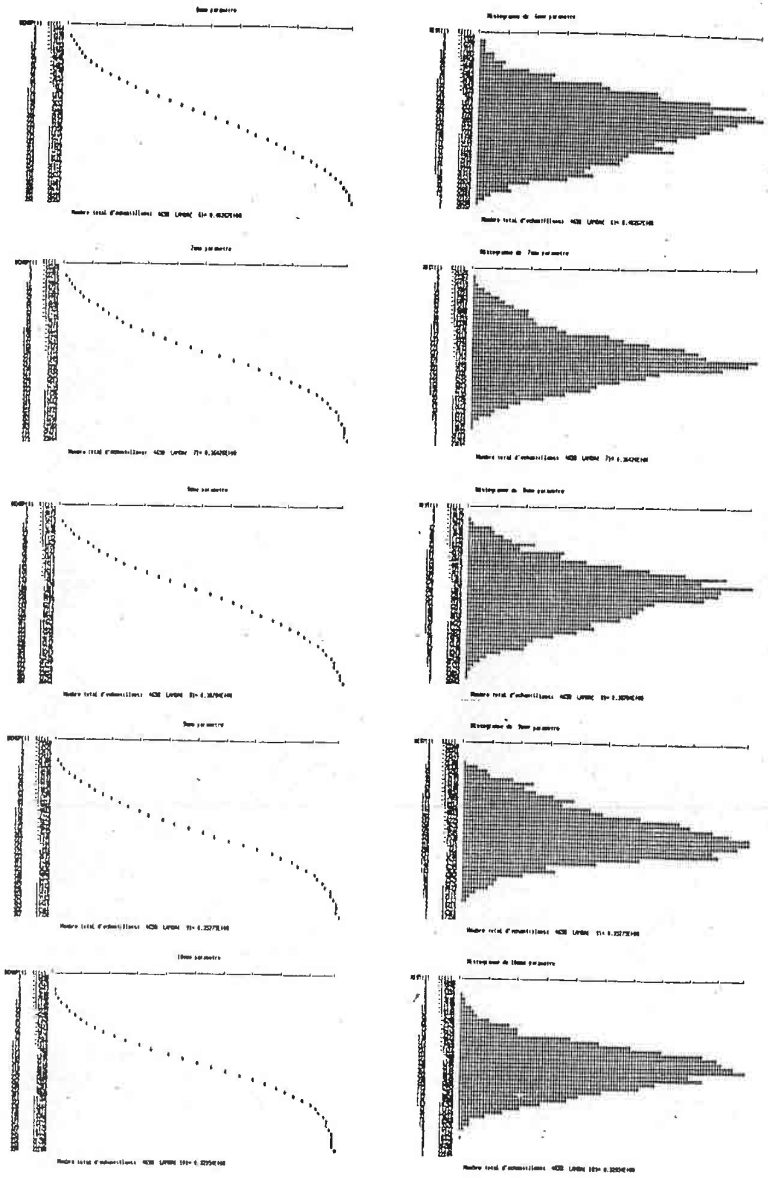
Un certains nombre d'améliorations peuvent être apportées à ces codeurs notamment sur le plan de la réduction de la complexité et de l'effort de mémorisation. L'application d'un sous-échantillonnage au signal d'excitation à modéliser permet de réduire la complexité des algorithmes d'un facteur approximativement égal au rapport de sous-échantillonnage. Un autre point intéressant à étudier consiste à utiliser, pour le codeur à excitation multi-impulsionnelle vectorielle et le codeur à excitation par code, des quantificateurs vectoriels algébriques, qui font usage des propriétés des réseaux réguliers, pour constituer les dictionnaires.

ANNEXE

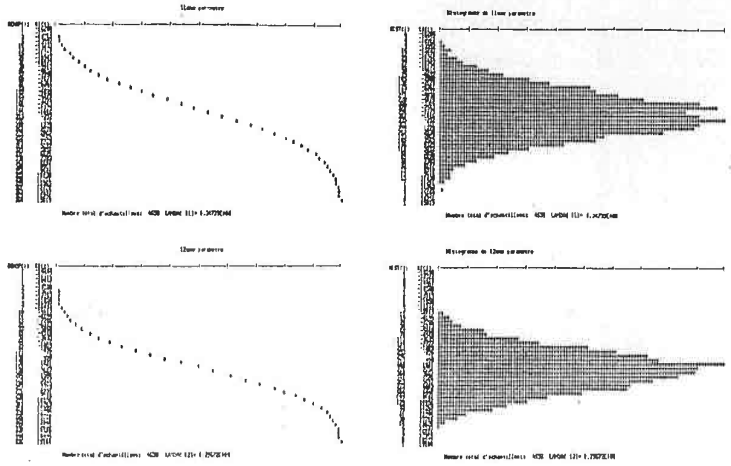
ANNEXE A: Histogrammes et fonction de quantification à erreur minimale des coefficients K_i ($1 \leq i \leq 5$)



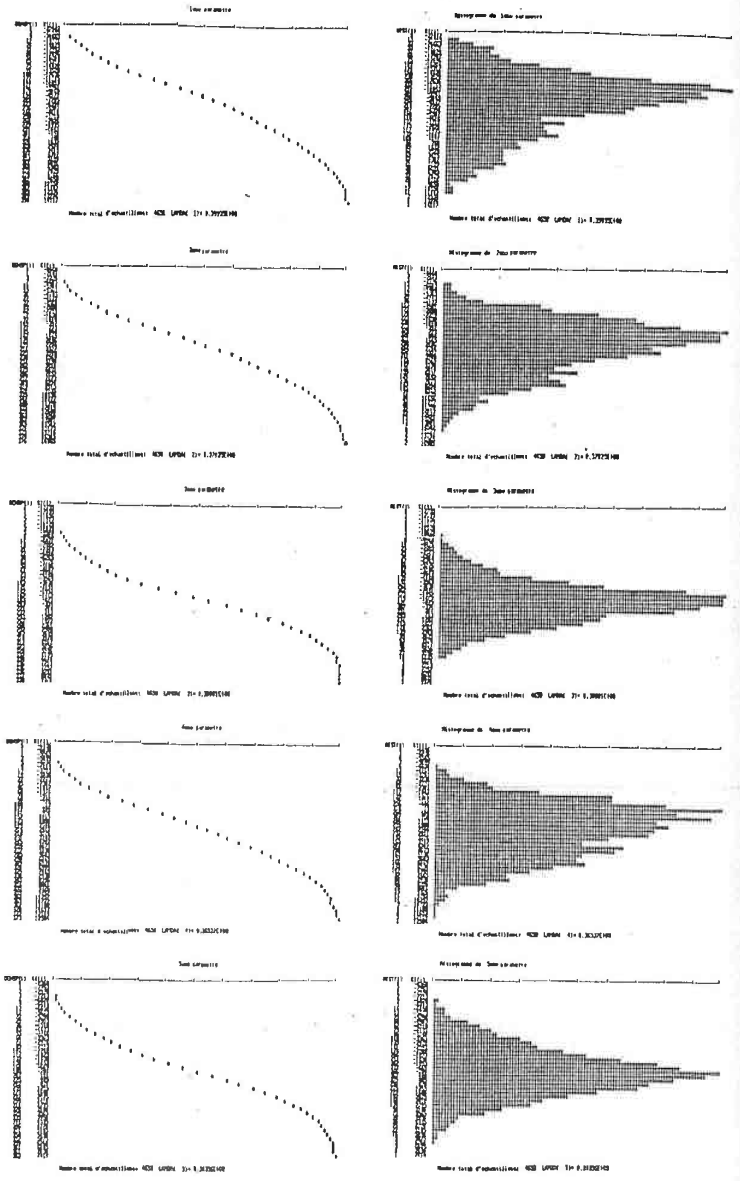
ANNEXE A: Histogrammes et fonction de quantification à erreur minimale des coefficients K_i ($6 \leq i \leq 10$)



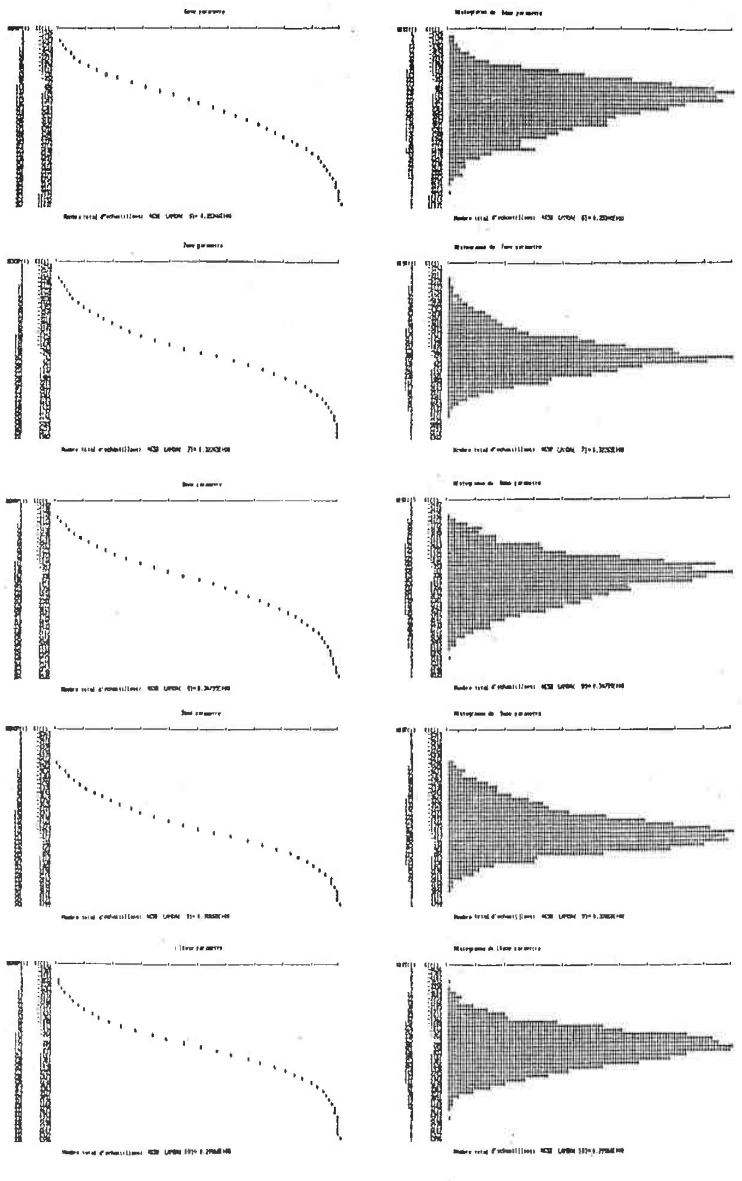
ANNEXE A: Histogrammes et fonction de quantification à erreur minimale des coefficients K_i ($10 \leq i \leq 12$)



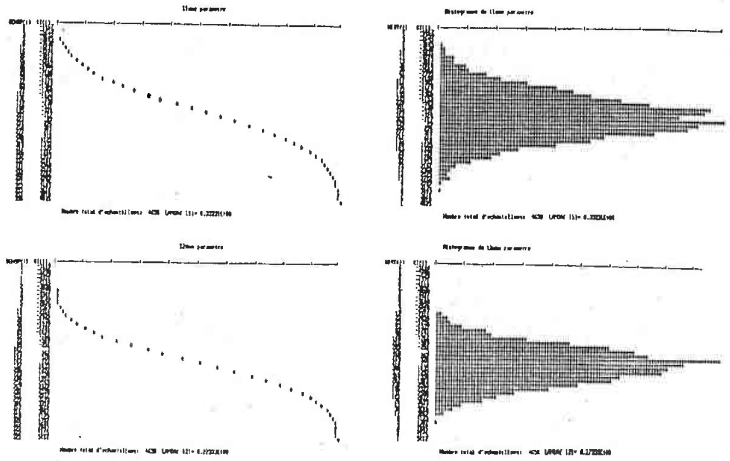
ANNEXE B: Histogrammes et fonction de quantification à erreur minimale des coefficients G_i $(1 \leq i \leq 5)$



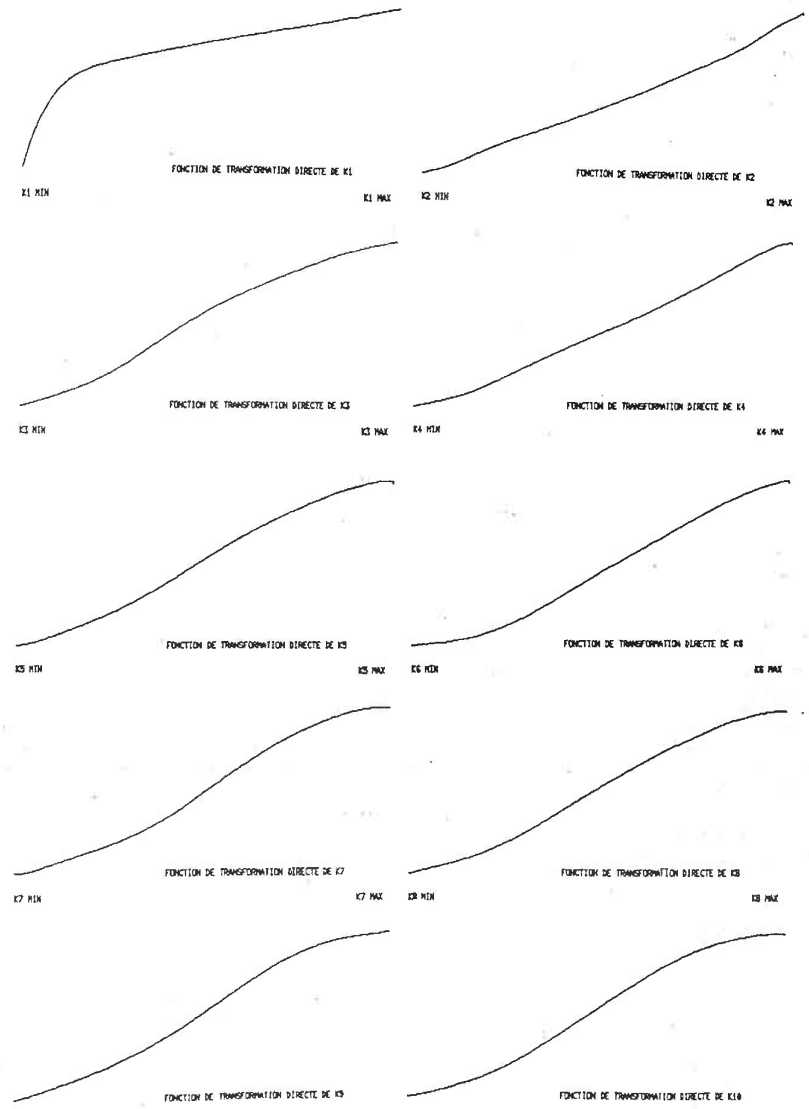
ANNEXE B: Histogrammes et fonction de quantification à erreur minimale des coefficients G_i $(6 \leq i \leq 10)$



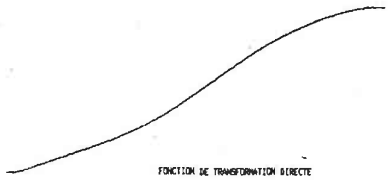
ANNEXE B: Histogrammes et fonction de quantification à erreur minimale des coefficients G_i $(10 \leq i \leq 12)$



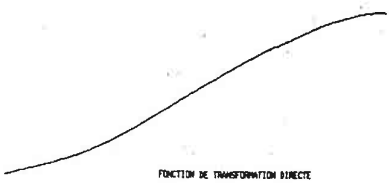
ANNEXEX C: Fonctions de transformation globale des K_i $(1 \leq i \leq 10)$ approximées par des polynomes de degré inférieur ou égal à cinq



ANNEXE C: Fonctions de transformation globale des K_i ($11 \leq i \leq 12$) approximées par des polynomes de degré inférieur ou égal à cinq



FONCTION DE TRANSFORMATION DIRECTE



FONCTION DE TRANSFORMATION DIRECTE

Résumé:

On désigne sous le terme de codage (abrégé de "codage réducteur de débit binaire") l'opération par laquelle l'information provenant d'une source de parole numérisée est transformée en une autre de moindre débit, telle que par une opération inverse, appelée décodage, la parole d'origine plus ou moins approximée peut être retrouvée.

Nous décrivons dans ce mémoire 4 codeurs hybrides temporels de qualité sub-téléphonique. Le schéma de base de ces codeurs est constitué d'une part d'un Prédiction Adaptative, d'autre part une modélisation par Analyse-Synthèse.

Le prédicteur adaptatif modélise, sous forme de filtre linéaire, l'enveloppe spectrale à court terme du signal. Le signal d'excitation correspondant à ce filtre est obtenu par la modélisation par d'analyse-synthèse qui inclue une fonction de pondération perceptuelle. 4 procédures de modélisation de l'excitation sont proposées.

Les 4 codeurs résultant offrent à qualité égale des débits compris entre 6 et 12 kbits/s.

Mots-Clés:

Codage de la parole à bas débit

Qualité sub-téléphonique

Filtre perceptuel

Codeur à excitation multi-impulsionnelle

Codeur à excitation par code

Quantification scalaire

Quantification vectorielle