MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

Université Akli Mohand Oulhadj de Bouira

Faculté des Sciences et Sciences Appliquées.

Département de Génie Electrique.

Mémoire de Master

Domaine : Sciences et Technologies

Filière : Télécommunication

Spécialité **:** Système des Télécommunications

THÈME :

Codage de la parole à bas débit

Réalisé par *:*

CHOULOU Imane ***OUCHENE Messaouda***

Soutenue publiquement le : 30 Octobre 2018

Devant le jury composé de :

Mr. REZKI Mohamed	M.C.B	Président	UAMOB
Mr. SAIDI Mohammed	M.A.A	Promoteur	UAMOB
Mr. DIB Riad	M.A.A	Examinateur	UAMOB
Mr. KASMI Réda	M.C.B	Examinateur	UAMOB

Année universitaire : 2017/2018

TABLE DES MATIERES

Page

REMERCIEMENTS

Nous commençons par remercier Allah tout puissant de nous avoir donné la volonté, l'amour du savoir et surtout le courage et la patience pour effectuer ce modeste travaille

Nous exprimons notre profonde gratitude à notre promoteur Mr Mohammed SAIDI d'avoir accepté de nous encadré pour notre projet de fin d'études, ainsi que pour son soutien, ses remarques pertinentes et son encouragement.

Nous remercions également les membres de jury pour l'intérêt qu'il ont bien voulu porter à notre travail en acceptant en acceptant de l'examiner et de le juger.

Enfin nos remerciements s'adressent plus particulièrement à nos familles, amis et toutes personnes qui ont participé de prés ou de loin à l'élaboration de ce travail.

DEDICACE

Je dédie ce modeste travail à :

Mes très chers parents qui m'ont aidé et soutenu tout au long de mes Études et mon succès :

A la source de mon bonheur, la flamme de mon cœur, celle qui s'est

Toujours sacrifié pour me voir réussir : Maman Saliha que j'adore

A l''homme de ma vie, mon exemple éternel, mon soutien moral et ma Source de joie : Mon père Rabah

Mes chères sœurs Selma et Hanane et mon frère Ramdhane pour leur amour et leur présence

dans ma vie

A mes profs surtout à ceux de l'université de brouira

et spécialement pour mon encadreur Mr Saidi

A mon binôme Lamis avec qui j'ai partagé tous les moments de stress de fatigue, mais aussi de fous rires

A toute ma famille et mes chers amis pour leur encouragement

Et pour tout le reste de ma famille

Imane

DEDICACES

A la lumière de mes jours, la source de mes efforts, a ma mère Ferroudja, qui a œuvré pour ma réussite, par son amour, son soutien, tous les sacrifices consentis et ses précieux conseils, pour toute son assistance et sa présence dans ma vie, reçois à travers ce travail aussi modeste soit-il, l'expression de mes sentiments et de mon éternelle gratitude.

A l'âme de mon père, A mes frères et sœurs

A mes neveux et nièces

Merci de m'avoir soutenu et témoigné votre affection durant tout ce temps. Je vous aime

A toute ma famille, mes amis, a chaque personne qui était la pour moi un jour

A tous mes enseignants qui mont marqués

Je vous dédier ce travail

Messaouda

Liste des figures

Figure ***page***

Liste des tableaux

Liste d'abréviations

A

AR : Auto-Regressif

APCM : Adaptive Pulse Code Modulation

ADPCM : Adaptive Differential Pulse Code Modulation

ADM : Adaptive Delta Modulation

APC : Adaptive Predictive Coding

ACELP: Algebraic Code Excited Linear Prediction

B

BSD : Bark Spectrum Distance

C

Codec : Coder / Decoder

CCITT : Comité Consultatif International de la Téléphonies et la Télégraphie

CELP : Code-Excited Linear Prediction

CDMA : Code Division Multiple Access

D

DAM : Diagnostic Acceptability Measure

dB : déciBel

DCT : Discrete Cosine Transform

DFT : Discrete Fourier Transform

DM : Delta Modulation

DSP : Digital Signal Processor

DRT : Diagnostic Rhyme Test

DPCM : Differential Pulse Code Modulation

F

FS-1016: US Federal Standard 1016 speech coder

F_e : Fréquence d'Echantillonnage

F_c : Fréquence de Coupure

I

ITU : International Telecommunications Union

ITU-T : ITU- Telecommunication standardization sector

IP : Internet protocol

G

GSM : Global System for Mobile communications

K

KHz : Kilos Hertz

Kbps : Kilos Bits Par Second

L

LPC : Linear Predictive Coding

LSF : Fréquences de Raies Spectrales

LSP: Line Spectral Pair

LTP: Long Term Prediction

LP: Linear Prediction

M

MBSD: Modified Bark Spectral Distortion

MOS: Mean Opinion Score

MIPS: Million Instructions per Second

P

PESQ: Perceptual Evaluation of Speech Quality

PSQM: Perceptual Speech Quality Measure

PCM: Pulse Code Modulation

PAPE : Phrases Arabes Phonétiquement Equilibrées

Q

VQ: Vector Quantization

S

SB-ADPCM: Sub Band **-** Adaptive Differential Pulse Code Modulation

SNR: Signal to Noise Ratio

SNRseg: Signal to Noise Ratio segmental

SMQ : Split Matrix Quantization

STU3 : Unité Téléphonique Sécurisée de troisième génération

T

TDMA : Time Division Multiple Access

TIMIT: Texas Instruments- Massachusetts Institute of Technology

U

UMTS : Universal Mobile Telecommunications System.

V

VoIP: Voice over IP

VoCoder : Voice Coder

W

WI: Waveform Interpolation

Introduction générale

Introduction générale

Le monde des télécommunications ne cesse de basculer vers le numérique. Les radio-mobiles, le multimédia et les communications globales font partie maintenant de notre langage quotidien. Les besoins de communication et de stockage s'accroissent rapidement. Cette croissance pousse les ingénieurs et les chercheurs à la mise en œuvre de techniques de compression de l'information de plus en plus performantes.

Dans les systèmes numériques modernes, le signal parole est représenté sous forme numérique (séquence d'éléments binaires, bits), il est nécessaire de représenter le signal par un nombre minimum de bits possible. Ainsi, pour le stockage de données, réduire le nombre de bits signifie l'économie de la mémoire. Pour les transmissions, réduire le débit binaire signifie l'économie de la bande passante. Il est donc nécessaire d'utiliser des algorithmes efficaces de compression de la voix. Le traitement qui permet d'effectuer une telle opération est les techniques de codage de la parole. Par exemple, dans les systèmes radio-mobiles la réduction du débit binaire de nos jours pour le codage de la parole a atteint de très bas débits.

Un système de codage de la parole comprend deux parties : le codeur et le décodeur (codec). Le codeur analyse le signal pour en extraire un nombre réduit de paramètres pertinents qui sont représentés par un nombre restreint de bits pour archivage (stockage) ou transmission. Le décodeur utilise ces paramètres pour reconstruire un signal de parole synthétique. L'objectif dans le codage de la parole est de représenter le signal vocal avec un nombre restreint de bits tout en gardant une qualité perceptuelle acceptable. Dans notre implémentation, nous nous sommes intéressés au codage de la parole à bas débit, qui est une technique efficace de compactage et produit la parole de qualité.

Actuellement, on dispose de plusieurs codeurs de parole efficaces ; parmi eux nous nous sommes intéressés particulièrement au célèbre codeur interpelé CELP (Code Exited Linear Prediction) comme outil d'application car ce dernier est largement utilisé dans le codage de la parole à bas débit. Il fut Introduit par B.S Atal et M.R. Schroeder qui a été le point de départ de nombreuses recherches dans le domaine de compression de la parole. Cette technique fait appelle à la quantification vectorielle et à la prédiction linéaire permet d'obtenir un signal de parole synthétique de bonne qualité pour un débit compris entre 4 et 8Kbps. Beaucoup de codeurs CELP ont été normalisées.

Le jugement et l'évaluation de la qualité du signal de parole après un certain traitement (débruitage, codage, compression, transmission,...) ne peut se faire d'une manière satisfaisante

qu'à partir de tests objectifs et subjectifs. Dans ce cadre, l'Union International des Télécommunications (UIT) a développé des normes spécifiant les procédures expérimentales à suivre pour évaluer la qualité perceptuelle du signal vocal ; parmi eux, la mesure PESQ (Perceptual Evaluation of Speech Quality) qui permet l'évaluation de la distorsion due aux codecs vocaux et tenir compte des effets des canaux de transmission tels que les pertes de paquets et le bruit de fond. Cette mesure a été adoptée comme une recommandation P.862 de l'UIT.

Notre travail consiste à comparer entre les deux codeurs CELP à 8Kbps et 4.8 Kbps, nous avons donc organisé ce mémoire en quatre chapitres :

Le premier chapitre est une introduction aux techniques de codage de la parole : les codeurs de parole à bas débit, la modélisation paramétrique du signal de parole et les mesures d'évaluation des performances.

Le deuxième chapitre donne une description générale du codeur de la parole à excitation par code CELP.

Le troisième chapitre sera consacré à l'implantation d'un codeur de parole à prédiction linéaire excité par code (CELP) où nous exposerons le principe de base des deux standards dérivés de la technique CELP : le G.729 à 8 Kbps et le standard FS1016 à 4.8 Kbps.

Le quatrième chapitre contient une évaluation des deux codeurs ACELP et CELP est menée en utilisant l'évaluation perceptuelle de la qualité vocale (PESQ).

La conclusion générale et les perspectives possibles à notre travail sont présentées dans la conclusion générale.

Chapitre 1 :

Généralités sur le codage de la parole

1.1 Introduction

L'objectif du codage vocal est de représenter la parole sous forme numérique avec le moins de bits possible tout en conservant l'intelligibilité et la qualité requises pour une application particulière. Dans ce chapitre, nous décrivons en premier lieu les concepts fondamentaux du codage du signal de la parole suivis par les Critères de performances dans ce type de codage puis on introduit quelques codeurs de parole ; des notions de base y sont brièvement décrites afin de faciliter leur compréhension.

1.2 Le processus de codage de la parole

Afin de coder la parole, plusieurs étapes sont nécessaires. Le signal subit tout d'abord un filtrage anti-repliement, puis un échantillonnage suivi d'une quantification et enfin le codage. L'échantillonnage est le processus de représentation d'un signal continûment variable par une séquence de valeurs. La quantification consiste à représenter approximativement chaque échantillon dans un ensemble fini de valeurs. Enfin, le codage consiste à assigner un numéro réel à chaque valeur [1].

Avant l'échantillonnage, un filtre passe-bas de fréquence de coupure égale à la moitié de la fréquence d'échantillonnage est inséré pour éviter l'effet dénommé « repliement » ou « aliasing » postulé par le théorème de Nyquist-Shannon, ce filtre est appelé filtre« anti-repliement » ou « anti-aliasing » [1].

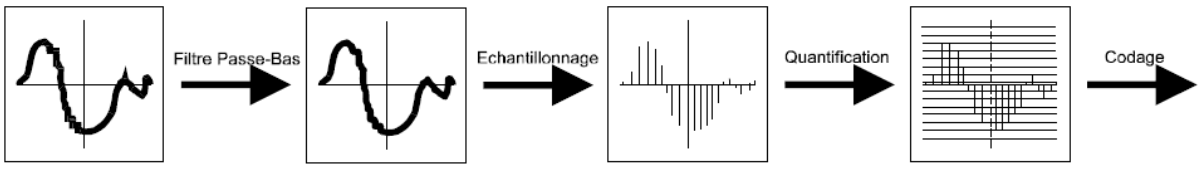


Figure 1.1 **:** Etapes suivies pendant le processus de codage de la parole [1].

1.2.1 L'échantillonnage

L'échantillonnage transforme le signal à temps continu x(t) en un signal à temps discret x(nTe) défini aux instants d'échantillonnage, multiples entiers de la période d'échantillonnage Te, celle-ci est elle-même l'inverse de la fréquence d'échantillonnage F_e. En ce qui concerne le signal vocal, le choix de F_e résulte d'un compromis. Son spectre peut s'étendre de 4 kHz jusque 12 kHz. Il faut donc en principe choisir une fréquence F_e égale à 24 kHz au moins pour satisfaire raisonnablement au théorème de Shannon.

Cependant, le coût d'un traitement numérique, filtrage, transmission, ou simplement enregistrement peut être réduit d'une façon notable si l'on accepte une limitation du spectre par un filtrage préalable. C'est le rôle du filtre de garde, dont la fréquence de coupure f_c est choisie en fonction de la fréquence d'échantillonnage retenue [2].

1.2.2 La Quantification

La quantification est une partie intégrante dans le codage. C'est l'opération de discrétisation d'une ou plusieurs variables, C'est aussi l'approximation de la valeur instantanée exact d'un signal par une valeur voisine tirée d'une association de N valeurs discrètes.

Si on désigne par x une variable aléatoire, un quantificateur est un appareil qui fait associer à l'entrée x comprise dans un intervalle, une sortie y comprise dans le même intervalle. Donc la quantification est l'opération de substitution des échantillons d'un signal analogique par des valeurs arrondies prises parmi un nombre fini de valeurs possibles [1].

1.3 Critères de performances dans le codage de la parole

Le codage de la parole consiste à réduire l'information contenue dans le signal parole tout en gardant une qualité satisfaisante du signal reconstitué. Afin de réaliser un codage performant, assurant un débit minimum avec une bonne qualité, il est nécessaire de prendre en considération certains critères de performances :

- **Débit binaire**

Lors du codage d'un signal de parole, le débit est défini comme le nombre de bits par unité de temps nécessaire pour coder la parole. Il est mesuré en bits par seconde (bps), ou généralement en kilobits par seconde. Il est important de faire la distinction entre kilobits par seconde (kbps) et kilo-octets par seconde [3].

- **Qualité du signal**

La qualité de la parole peut être déterminée par des tests d'écoute qui calculent l'opinion moyenne des auditeurs. La qualité de la parole peut être aussi déterminée par des mesures objectives comme la prédiction du gain, la distorsion spectrale logarithmique, etc. [4].

- **Complexité**

En réalité les algorithmes de codage sont exécutés sur des cartes DSP (Digital Signal Processor). Ces processeurs possèdent une mémoire de stockage et une vitesse (en MIPS Million Instructions per Second) limitées [4]. En général, plus le taux de compression est Elevé plus la complexité du codeur sera forte pour maintenir une certaine qualité du signal

transmis ; par conséquent, les algorithmes de codage de la parole ne doivent pas être complexes pour ne pas dépasser la capacité des cartes DSP modernes. D'autres mesures de complexité peuvent être signalées, telles que la taille physique du codeur ou du décodeur, son prix et sa consommation en puissance (en Watt ou en mW) qui constituent un important critère dans un système portable [5].

- **Retard de communication**

C'est la somme des délais algorithmique, délai de traitement, délai de transmission. Ce retard n'est pas tolérable surtout pour les applications en temps réel, telle que la téléphonie. Pour remédier au problème de retard il faut réduire la complexité des algorithmes, améliorer les protocoles de communications et augmenter les performances des processeurs [4].

- **Sensibilité aux erreurs de canal**

Ce paramètre mesure la robustesse du codeur de la parole par rapport aux erreurs de canal, les erreurs qui sont souvent provoquées par la présence du bruit dans le canal, de la perte de paquets de signal et de l'interférence inter-symboles [4].

- **Largeur de bande du codeur**

Pour mieux exploiter la bande passante, il faut bien choisir la largeur de bande du codeur, en effet, on utilise des codeurs à bande étroite dans les applications où la haute qualité n'est pas exigée, pourvu que le signal soit intelligible [1].

1.4 Classification des codeurs de parole

De nombreux travaux relatifs au codage tendent à maximiser le compromis entre l'efficacité, le cout et la qualité des systèmes de communication en fonction des débits disponibles. Traditionnellement les codeurs de parole sont divisés en trois grandes classes, les codeurs en formes d'ondes, les codeurs paramétriques ou vocodeurs et les codeurs hybrides.
Les codeurs en formes d'ondes opèrent à des débits hauts, néanmoins, ils fournissent une très bonne qualité de parole. Les codeurs paramétriques opèrent à de très bas débits mais produisent des signaux de qualité synthétique. Enfin, les codeurs hybrides combinent les deux techniques de codage, le codage en formes d'ondes et le codage paramétrique et fournissent une bonne qualité de parole pour des débits moyens [6].

1.4.1 Les codeurs en formes d'ondes

Ce type de codeurs essaye de reproduire la forme d'onde du signal d'entrée à coder. Ils sont conçus pour être indépendants du signal, ainsi, ils peuvent être employés pour coder une large variété de signaux. Généralement, ils sont de faible complexité, ils fournissent des signaux de parole de bonne qualité à des débits au-dessus de 16 kbps. Le codage en formes d'ondes peut être effectué aussi bien dans le domaine temporel que dans le domaine fréquentiel [6].

1.4.1.1 Codeurs dans le domaine temporel

Ces codeurs réalisent le processus de codage sur des échantillons temporels du signal. Les méthodes de codage les plus connues dans le domaine temporel sont : le codage PCM (Pulse Code Modulation), le codage APCM (Adaptive Pulse Code Modulation), le codage DPCM (Differential Pulse Code Modulation), le codage ADPCM (Adaptive Differential Pulse Code Modulation), le codage DM (Delta Modulation), le codage ADM (Adaptive Delta Modulation) et le codage APC (Adaptive Predictive Coding) [6].

- **Les codeurs PCM**

C'est un processus de quantification échantillon par échantillon. N'importe quelle quantification scalaire peut être utilisée avec ce schéma, mais la forme de quantification la plus utilisée est la quantification logarithmique, dans laquelle on prend en considération la nature du signal parole qui possède une densité de probabilité proche d'une gaussienne, autrement dit, les faibles amplitudes du signal parole sont plus fréquentes par rapport aux grandes amplitudes (Figure1.1).

Deux variantes (incluses dans la norme CCITT G.711) se sont répandues dans la téléphonie : La norme américaine (μ-Law), utilisée aux États-Unis et au Japon et la norme européenne (A-Law) utilisée dans le reste du monde et dans les communications internationales. Elles sont toutes deux des variations d'une correspondance exponentielle [7].

- **Les codeurs DPCM et ADPCM**

La technique PCM ne fait aucune supposition sur la nature des formes d'ondes à coder, par conséquent, elle fonctionne bien pour des signaux différents de ceux de la parole. Cependant, en codant la parole, il existe une très forte corrélation entre les échantillons successifs obtenus. Cette corrélation peut être exploitée pour réduire le débit binaire. Une méthode simple de le faire est de transmettre uniquement la différence entre deux échantillons. Le signal différence possèdera alors une gamme dynamique plus réduite que le signal original, et peut être alors quantifié moyennant un nombre de niveaux de reconstitution plus réduit. Dans la méthode citée plus haut, l'échantillon précédent est utilisé pour prédire la valeur de

l'échantillon présent. La prédiction sera améliorée si un bloc plus large de la parole est utilisé pour la prédiction. Cette technique est connue sous le nom de DPCM. Dans la figure 1.2le codeur sur la gauche, et le décodeur sur la droite. Le quantificateur inverse, convertit les codes transmis en la valeur û(n) [8].

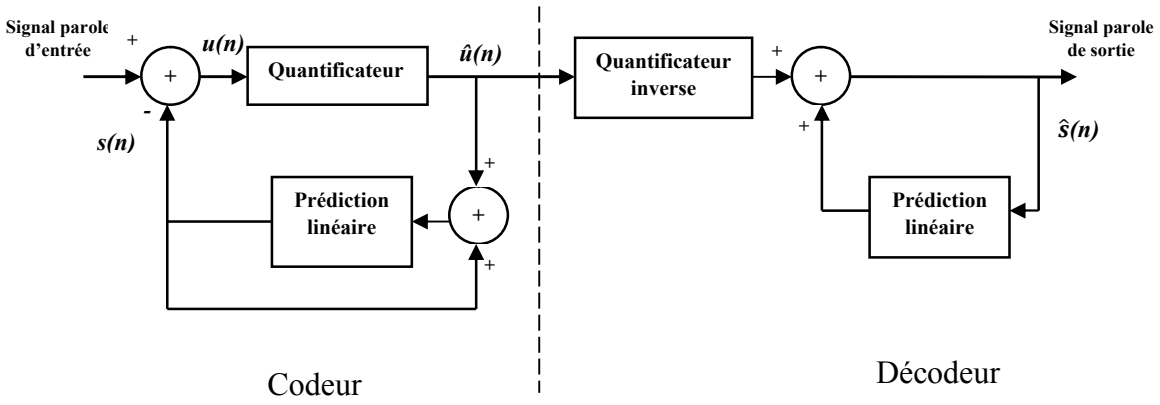


Figure 1.2 : Schéma de principe du codeur DPCM [8].

Une version améliorée de la DPCM est la DPCM Adaptative (ADPCM) dans laquelle le prédicteur et le quantificateur sont adaptés aux caractéristiques locales du signal d'entrée. Il existe un bon nombre de recommandations de l'ITU basées sur les algorithmes ADPCM pour la bande étroite (fréquence d'échantillonnage de 8 kHz) et le codage audio comme le G.726 opérant à 40, 32, 24 et 16 kbps. La complexité de l'ADPCM est légèrement faible [8].

1.4.1.2 Codeurs dans le domaine fréquentiel

Les codeurs de formes d'ondes dans le domaine fréquentiel divisent le signal en un nombre de composantes fréquentielles et code chacune d'elles séparément. Le nombre de bits utilisé pour coder chaque composante fréquentielle peut varier de manière dynamique. Les codeurs dans le domaine fréquentiel sont divisés en deux groupes : Les codeurs en sous bande (sub band coders) et les codeurs par transformée (transform coders) [6].

- **Les codeurs en sous bande**

Les codeurs en sous bande emploient des filtres passe bande pour diviser le signal en un nombre de signaux passe bande (sub band signals) qui sont codés séparément. Au niveau du récepteur, les signaux en sous bande sont décodés et additionnés pour reconstruire le signal de sortie. L'avantage principal du codage en sous bande est que la quantification du bruit produit dans une bande est confiné uniquement dans cette bande. L'organisme ITU a standardisé en codage sous bande le codeur audio G.722, SB-ADPCM (Sub Band- Adaptive Differential Pulse Code Modulation), qui code les signaux audio à large bande de 7 kHz échantillonnés à 16 kHz, pour une transmission à 48, 56 ou 64 Kbps [9].

- **Les codeurs par transformée**

Cette technique transforme par bloc, un segment du signal d'entrée dans le domaine fréquentiel ou un domaine similaire. Le codage adaptatif est réalisé en attribuant plus de bits aux coefficients de transformation les plus importants. Au niveau du récepteur, le décodeur fait la transformation inverse pour obtenir le signal reconstruit. Plusieurs transformées comme la DFT (Discrete Fourier Transform) ou DCT (Discrete Cosine Transform) peuvent être utilisées [6].

1.4.2 Les codeurs paramétriques ou vocodeurs

Les performances des codeurs paramétriques, connus aussi sous le nom de Vocodeurs, sont fortement dépendantes de la précision des modèles de production de la parole. Ces codeurs sont conçus spécifiquement pour des applications à bas débit et sont principalement destinés à maintenir une qualité satisfaisante de la parole. Les vocodeurs les plus efficaces sont basés sur la prédiction linéaire LP (Linear Prediction). Une qualité des communications peut être obtenue à des débits inférieurs à 2.4 kbps avec les vocodeurs LP [6].

1.4.3 Les Codeurs hybrides

Les codeurs hybrides sont conçus pour fournir une qualité aussi bonne à des débits relativement faibles ou moyens, ce sont donc, des codeurs intermédiaires entre les codeurs en formes d'ondes et les vocodeurs. Cependant, ces codeurs ont tendance à nécessiter un nombre d'opérations plus élevé. Virtuellement, tous les codeurs hybrides reposent sur l'analyse LP pour l'obtention des paramètres du modèle de synthèse. Les techniques de formes d'ondes utilisées pour coder le signal d'excitation et les modèles de production du pitch peuvent être incorporées pour améliorer les performances. A partir des années 80, l'intérêt pour les codeurs CELP (Code-Excited Linear Prediction) ne cesse d'augmenter [10].

Dans les codeurs CELP, l'analyse LP est utilisée pour obtenir le signal d'excitation. La modélisation du pitch est utilisée pour coder efficacement le signal d'excitation. Le standard G.729 de l'ITU est un codeur CELP qui produit une qualité téléphonique (toll quality) de la parole à 8 kbps [10].

Les codeurs de formes d'ondes par interpolation WI (Waveform interpolation) modélisent le signal résiduel par des formes d'ondes caractéristiques qui peuvent être interpolées aussi bien dans le domaine temporel que fréquentiel pour la reconstitution du signal. Pour des débits inférieurs à 4 kbps, les codeurs WI donnent de meilleures performances, comparés à d'autres codeurs opérant à des débits similaires. Cependant, les codeurs WI sont actuellement alourdis par leur complexité élevée et par leur retard (typiquement 40 ms) [4].

1.5 Modèle prédictif de production de la parole

La parole peut être considérée comme étant un signal pseudo-stationnaire, c.-à-d. stationnaire sur de courtes durées allant en général de 5 jusqu' à 30 ms. Sur cette période, il est possible de caractériser le spectre du signal par deux attributs :

- L'enveloppe spectrale.
- La structure fine du spectre.

Le codage linéaire de prédiction se fonde sur un modèle fortement simplifié pour la production de la parole [11].

Un signal voisé peut être modélisé par le passage d'un train d'impulsion $u(n)$ à travers un filtre numérique récursif de type tous pôles. Cette modélisation reste valable dans le cas des sons non voisé, à condition que $u(n)$ soit un bruit blanc. Le modèle final est illustré à la fig1.3. Il est souvent appelé modèle auto-régressif (AR), parce qu'il correspond dans le domaine temporel à une régression linéaire de la forme [11] :

$$\hat{s}(n) = G.u(n) + \sum_{i=1}^{p} -a_i \hat{s}(n-i) \qquad (1.1)$$

Où $u(n)$ est le signal d'excitation, ce qui exprime que chaque échantillon est obtenu en ajoutant un terme d'excitation à une prédiction obtenue par combinaison linéaire de *P* échantillons précédents. Les coefficients du filtre sont appelés coefficients de prédiction et le modèle AR est souvent appelé modèle de prédiction linéaire [12].

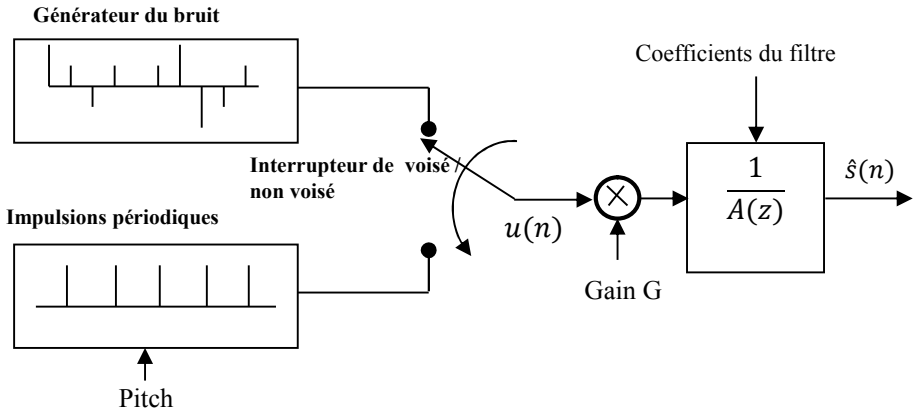


Figure 1.3 : Modèle simplifié de la production de la parole [12].

Ce modèle comprend :

- Un générateur périodique d'impulsions ;
- Un générateur de bruit blanc ;
- Un interrupteur servant à choisir les sons voisés ou non voisés ;
- Un gain G proportionnel à la valeur efficace du signal ;
- Un filtre tous pôles $H(z) = 1/A(z)$.

Par analogie entre le modèle physique et le modèle mathématique, on peut donner les relations d'équivalence suivantes [11] :

Conduit vocal	⟺	Filtre LP $1/A(z)$
Flux d'air	⟺	Signal d'excitation u(n) ou signal résiduel
Vibration des cordes vocales	⟺	Son voisé
Période de vibration des cordes vocales	⟺	Période du pitch
Volume d'air	⟺	Gain

1.6 Mesures d'évaluation des performances

Après avoir effectué la conception d'un algorithme de codage, il est nécessaire de mettre ce dernier sous tests afin de l'évaluer et de voir s'il répond aux normes et aux critères de codage.

Les algorithmes de codage de la parole sont évalués selon plusieurs critères dont les plus importants sont : la qualité du signal, le débit binaire, la complexité de l'algorithme et le retard de communications, nous consacrons cette partie au premier critère qui est la qualité du signal. Dans les communications numériques, la qualité du signal parole est évaluée selon quatre catégories [4] :

- ***Qualité diffusion ou broa*dcast**

Qui se réfère aux larges bandes (typique 50-7000 Hz et20-20000 Hz pour disques compacts) c'est la plus haute qualité qu'on peut atteindre, elle nécessite des débits au moins de 32 à 64 kbps.

- **Qualité réseau ou toll**

C'est la qualité qui permet d'entendre la parole sur un réseau téléphonique (pour une bande de 200-3200 Hz avec un rapport signal sur bruit de 30 dB et une distorsion moins de 2 à 3 %).

- **Qualité de communications**

Elle implique une certaine dégradation de la qualité de la parole, néanmoins, elle présente une qualité naturelle et hautement intelligible. Cette qualité peut être atteinte à des débits supérieurs à 4 kbps.

- **Qualité synthétique**

La parole synthétique est intelligible, néanmoins, elle n'est pas naturelle et perd la reconnaissance de locuteur.

Le but actuel dans le codage de la parole est d'atteindre la qualité toll pour des débits de 4 kbps. Actuellement, les codeurs opérant en dessous de 4 kbps de débit, fournissent une qualité synthétique. La mesure de qualité est une tâche importante mais très difficile. Il y a deux manières pour mesurer la qualité de la parole, on distingue la mesure subjective et la mesure objective.

1.6.1 Mesures objectives

Le système auditif humain est l'évaluateur le plus adéquat de la qualité et des performances d'un codeur de la parole. Il permet de préciser l'intelligibilité et la sonorité naturelle des sons. Bien que les tests d'écoute subjectifs donnent une bonne évaluation des codeurs de la parole, ils exigent beaucoup de temps et sont inconsistants. Les mesures objectives peuvent donner une évaluation immédiate et efficace de la qualité d'un algorithme de codage.

Les mesures objectives de distorsion peuvent être calculées aussi bien dans le domaine temporel (calcul du rapport signal sur bruit) que fréquentiel (mesure de distorsions) [11].

1.6.1.1 Mesures Objectives dans le Domaine Temporel

Les mesures objectives les plus importantes dans le domaine temporel sont les suivantes :

- **Le rapport signal sur bruit SNR (Signal to Noise Ratio)**

C'est la mesure objective de la qualité la plus commune pour l'évaluation des performances des algorithmes de compression. Le SNR est défini comme un rapport de l'énergie moyenne du signal parole sur l'énergie moyenne du signal d'erreur, le SNR est généralement exprimé en décibel dB et défini par :

$$\text{SNR} = 10log_{10}\left(\frac{Energie\ moyenne\ du\ signal\ parole}{Energie\ moyenne\ du\ signal\ d'erreur}\right) dB = 10log_{10}\frac{\sum_{n=-\infty}^{\infty} s^2[n]}{\sum_{-\infty}^{\infty}(s[n]-\hat{s}[n])^2}\, dB \quad (1.2)$$

Où ŝ[n] est la version codée du signal parole original s[n]. La mesure SNR n'est par une estimation exacte de la qualité, en effet, le SNR ne donne qu'une seule évaluation pendant

toute la durée du signal, on traite le signal parole en tant qu'un seul vecteur, alors qu'en réalité, l'auditeur effectue plusieurs comparaisons pour un signal parole donné. C'est pourquoi on préfère utiliser le SNR segmental [4].

- **Le SNR segmental (SNRseg)**

Le SNR (Signal to Noise Ratio) segmental est la mesure de qualité objective la plus utilisée dans le domaine temporel. Il définit la moyenne des SNRseg issus de plusieurs segments de courte durée (15 à 20 ms) :

$$SNR_{seg} = \frac{1}{M}\sum_{m=0}^{M-1} 10log_{10}\left[\frac{\sum_{i=mN}^{mN+N-1} s^2(i)}{\sum_{i=mN}^{mN+N-1}(s(i)-\hat{s}(i))^2}\right] dB \quad (1.3)$$

Où $s(i)$,$\hat{s}(i)$, N et M sont respectivement le signal de référence, le signal débruité, la longueur d'un segment et le nombre total de segments.

Le SNRseg est meilleur que le SNR. Cependant, si le signal de parole contient des segments de silence, ce qui est très probable, le s(i) sera nul et n'importe quelle quantité de bruit entrainera un SNR en dB négatif pour ce segment. Ce problème peut être résolu partiellement en choisissant un seuil d'énergie au délai duquel le SNR segmental sera calculé [13].

1.6.1.2 Mesure objective dans le domaine fréquentiel

La différence entre l'enveloppe spectre du signal parole original et celle du signal codé, qui peut être traduite par une différence entre les fréquences des formants ou entre leur largeur, conduit à des sons phonétiquement différents. C'est pourquoi on fait recours à la distorsion spectrale [4]**.** Une brève description des différentes mesures de distorsion dans le domaine fréquentiel est présentée dans ce qui suit :

- **Distorsion d'Itakura-Saito**

La mesure d'Itakura-Saito repose sur l'analyse LPC. Son expression fait intervenir le modèle tout pôle du signal de référence s et celui du signal testé y. Soient $p(w)$,$\hat{p}(w)$ les densités spectrales de puissance du modèle AR du signal de référence et du signal de test [13].

La distance d'Itakura-Saito est donnée par :

$$d_{IS}(p(w),\hat{p}(w)) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{p(w)}{\hat{p}(w)} - log\frac{p(w)}{\hat{p}(w)} - 1\right] dw \quad (1.4)$$

1.6.1.3 Mesures objectives perceptuelles

Les mesures objectives de distorsion sont souvent sensibles aux variations du gain et du délai, en plus elles ne prennent pas en considération les propriétés perceptuelles de l'oreille. D'autre part, les mesures subjectives sont lentes et coûteuses [11]. Parmi les mesures perceptuelles les plus utilises nous pouvons citer : la distance Bark Spectrum Distance (BSD), la mesure BSD modifiée (MBSD), la mesure PSQM (Perceptual Speech Quality Measure) et la mesure PESQ (Perceptual Evaluation of Speech Quality). Pour notre cas, nous allons utiliser le PESQ comme un logiciel d'évaluation de la qualité synthétique des codeurs [14].

- **BSD et MBSD**

La mesure BSD (Bark Spectral Distortion) est parmi les premiers critères à avoir incorporé des notions en relation avec notre système d'audition dans l'évaluation de la qualité de la parole. Le BSD a pour objectif de mesurer la distorsion entre le signal de référence et celui codé, dans le domaine de Bark. La sensation de force sonore connue sous le nom de sonie est mise en jeu pour calculer cette distorsion. En effet, la distorsion totale est la moyenne de la distance euclidienne entre la sonie du signal de référence et celle du signal débruité.

Le MBSD (Modified Bark Spectral Distortion) introduit le seuil de masquage du bruit pour calculer la distorsion dans le BSD ; l'idée est de ne tenir compte que de la distorsion audible. Effectivement, tout ce qui est au-dessous du seuil de masquage du bruit est imperceptible à l'oreille humaine. Par conséquent, la distorsion totale est la moyenne de la différence entre les sonies du signal de référence et du signal débruité pondérée par un paramètre s'annulant lorsque la distorsion est inaudible [13].

1.6.2 Mesures Subjectives

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

Il existe trois types de mesures subjectives [11] de la qualité généralement utilisées :

- Le test DRT (Diagnostic Rhyme Test) : Il s'agit d'une mesure d'intelligibilité dont la tâche est de reconnaître un ou deux mots possibles parmi un ensemble de paires de rimes, par exemple (meat – heat).
- Le test DAM (Diagnostic Acceptability Measure) : Elle sert pour l'évaluation des systèmes de communications, elle est basée sur l'acceptabilité de la parole par des auditeurs normatifs qualifiés

- Le test MOS (Mean Opinion Score) : Les auditeurs évaluent l'échantillon de parole sous test dans l'une des cinq catégories de qualité, montrées dans le tableau1. Chaque catégorie se voit attribuer une valeur numérique. La valeur MOS résultante est la valeur moyenne de tous les écouteurs pour chacun des discours testés. Il existe divers aspects de la dégradation trouvée dans la parole testée, par exemple la limitation de la bande passante, le bruit additif, l'écho, la distorsion non linéaire, etc.

La valeur du MOS est calculée par : $\text{MOS} = \frac{\sum_i N_i i}{N}$ i=1,…,5 (1.5)

N : nombre d'auditeurs ayant participé au test.

Ni : nombre d'auditeurs qui ont choisi la catégorie i.

Tableau 1.1 : Description du test MOS [4].

Valeur MOS	**Qualité de la parole**	**Niveau de distorsion**
5	excellent	Imperceptible
4	Bon	Juste perceptible mais pas gênant
3	Assez bon	Perceptible légèrement gênant
2	Médiocre	Gênant mais pas désagréable
1	Mauvais	Très gênant et désagréable

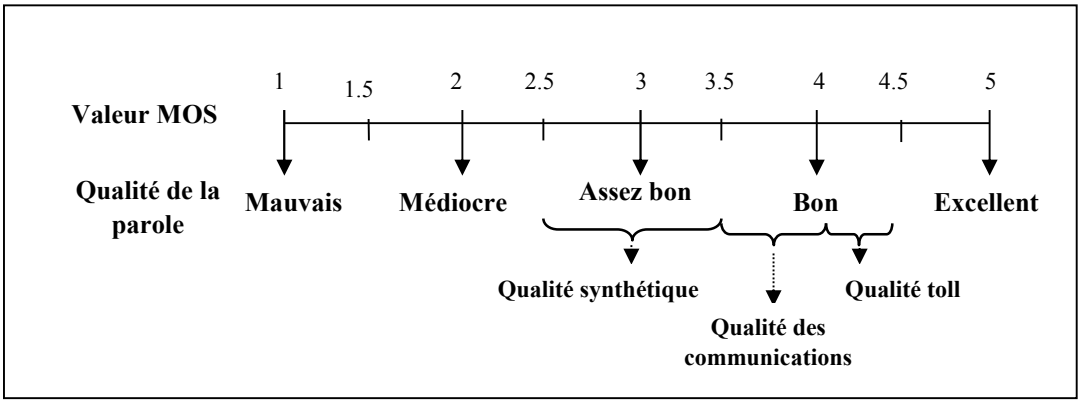


Figure 1.4 : Relations entre les valeurs MOS et la qualité de la parole [4].

1.7 Conclusion :

Le codage de la parole a connu un formidable essor lié au besoin de préserver la bande passante des réseaux de transmission et de diffusion et du besoin de limiter l'espace de stockage occupé par le signal de parole. Dans ce premier chapitre, nous avons présenté les différents types de codeurs de parole, le codage basé sur la prédiction linéaire et en dernier point on a cité les mesures d'évaluation des performances. Dans le chapitre qui suit nous allons se focaliser sur le Codeur de parole à excitation par code CELP.

Chapitre 2 :

Etude descriptive du codeur CELP

2.1 Introduction :

Les codeurs de type CELP dominent le codage de la parole à bas débit ; ils offrent de bonnes performances à un débit aussi faible jusqu'à 4.8 Kbps et sont classés parmi les meilleurs codeurs de parole à bas débit. Ils sont basés sur le principe d'analyse par synthèse, qui fait appel à la prédiction linéaire et à la quantification vectorielle [15, 16].

Aujourd'hui, la majorité des systèmes de codage bas débit utilisent ce type de codage comme en témoignent les nombreuses normes qui l'utilisent en téléphonie et dans la transmission faible et moyen débit de la parole : norme GSM (Global system for Mobile communications) demi-débit, norme UMTS (Universal Mobile Telecommunications System)… [17].

2.2 Principe d'un codeur CELP :

En 1985, Atal et Schroeder définissent le codeur à prédiction linéaire excité par code CELP (Code Excited Linear Prediction), qui détermine une forme d'onde optimale du signal résiduel de prédiction en utilisant la technique d'analyse par synthèse. Le codage CELP réalise une modélisation paramétrique du signal de parole sous la forme d'un signal d'excitation, généralement issu d'un dictionnaire (codebook) de formes d'ondes prédéterminées, passant au travers d'un ou plusieurs filtres. La technique CELP est parmi une des idées les plus influentes dans le codage de la parole et ses principes constituent la base de beaucoup de codeurs normalisés [18]. Globalement les codeurs de type CELP modélisent le système de production de la parole en trois étages :

- Un étage d'excitation ;
- Un étage modélisant l'effet des cordes vocales ;
- Un étage modélisant la fonction de transfert du conduit vocal.

CELP est une méthode efficace d'analyse par synthèse en boucle fermée pour les systèmes de codage de la parole à bande étroite et moyenne. Dans les codeurs CELP, la parole est segmentée en trames (généralement 10-30 ms de long) pour rester dans l'hypothèse de stationnarité du signal de parole [19] et pour chaque trame, un ensemble optimal de paramètres de prédiction linéaire et de pitch est déterminé et quantifié. La fenêtre d'analyse est divisée en plusieurs sous-fenêtres et l'excitation est calculée pour chaque sous-fenêtre d'analyse par quantification vectorielle. Le signal d'excitation est modélisé par une combinaison linéaire de vecteurs, extraits des dictionnaires adaptatif et stochastique de taille bien définie [20].

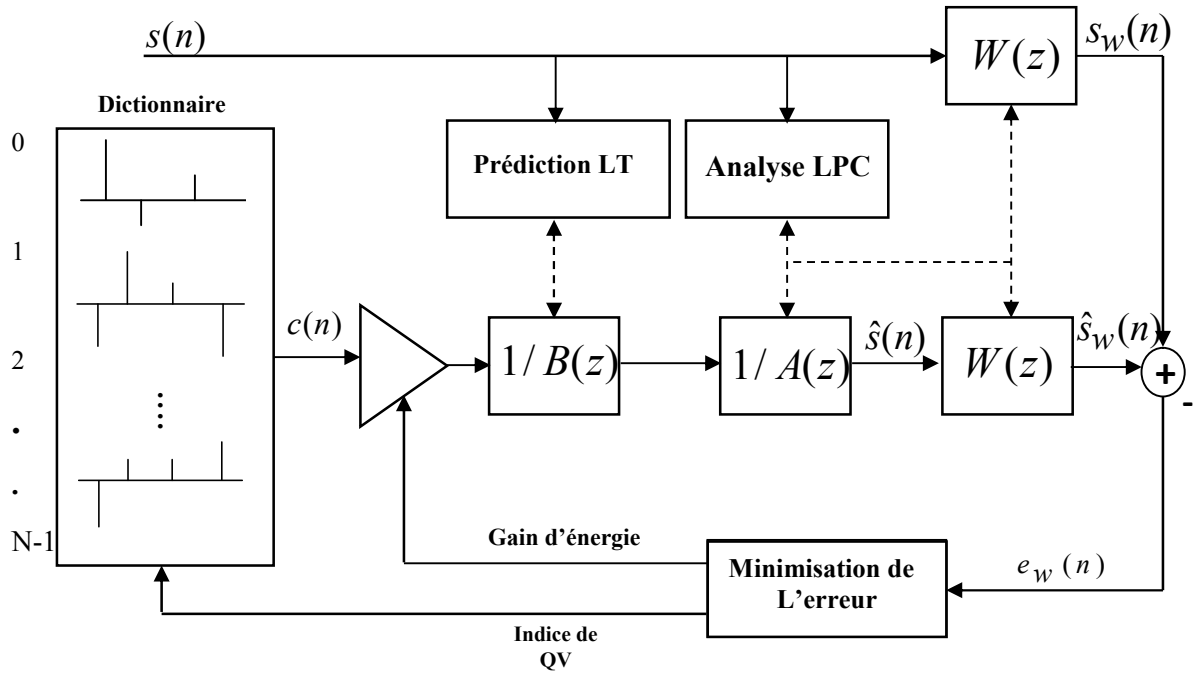


Figure 2.1 : Schéma de principe du codeur CELP [21].

Le filtre de synthèse est représenté par la fonction de transfert$1/A(z)$. La compression est réalisée parce que la transmission des échantillons du signal de parole s(n) est substituée par la transmission du signal d'erreur (qui est représenté sur un nombre réduit de bits) et des coefficients du filtre de synthèse [22].

Tout d'abord, la corrélation à court terme sur une trame du signal est réduite par prédiction linéaire, c'est-à-dire qu'un échantillon de la trame est estimé par combinaison linéaire des échantillons précédents, ceci pour un nombre fini d'échantillons. Le filtre de synthèse modélise le conduit vocal et il est formé par l'ensemble des coefficients de prédiction. Un résidu de prédiction est obtenu par différence entre le signal d'entrée et son estimée par prédiction linéaire. Ce résidu est quantifié par une combinaison linéaire de deux mots de code, provenant de deux dictionnaires. Le résidu quantifié joue le rôle d'excitation du filtre de synthèse [19].

Le premier dictionnaire, dit adaptatif, modélise la corrélation à long terme présente dans le résidu, résultant de la vibration des cordes vocales. Ce dictionnaire contient un ensemble d'excitations quantifiées des dernières trames codées. Les mots de code sont indicés par une valeur appelée pitch, qui caractérise la périodicité du signal à la trame courante. Une fois le mot de code optimal trouvé, son gain associé est également calculé. Le deuxième dictionnaire, qualifié de fixe, contient un ensemble de séquences prédéfinies et code l'information non prédictible, appelé innovation. Le codeur détermine le mot de code optimal ainsi que son gain

associé. Dans les deux cas, le mot de code ainsi que son gain sont obtenus en minimisant l'erreur quadratique moyenne entre le signal original et le signal reconstruit. Cette méthode est appelée analyse par synthèse [19].

L'excitation consiste en la somme des deux mots de code, pondérés par leur gain quantifié respectif. Le dictionnaire adaptatif est mis à jour en concaténant cette excitation aux excitations des trames précédentes. Les propriétés de masquages du système auditif peuvent être prises en compte en pondérant l'erreur par une fonction dépendant des coefficients de prédiction à court terme [19].

L'idée est d'utiliser [17] :

- un filtre prédictif qui décorrèle (à court terme) les échantillons et fournit une erreur de prédiction (ou résiduel).

- une quantification vectorielle (de type forme-gain) pour coder ce résiduel.

La notoriété de la méthode est due à la qualite de la voix obtenue pour des debits allant de 4,8 kbps à 16 kbps. Un des inconvenients majeur de la méthode est le délai qu'elle implique (entre 50 ms et 100 ms) en raison des calculs. De tels délais peuvent engendrer et propager un echo [23].

2.3 Masquage spectral

Introduction d'un facteur perceptuel

Les fonctions de coût quadratiques se prêtent bien aux calculs : elles possèdent la bonne propriété de fournir un système linéaire lorsque l'on dérive ce critère par rapport aux paramètres inconnus. Par contre, ce critère n'est pas forcément bien adapté à notre système auditif. Une correction perceptive est très largement utilisée pour pallier cet inconvénient. On rajoute une fonction de pondération, sous la forme d'un filtre de fonction de transfert $\mathrm{W(z)} = A(z)/A(z/\gamma)$, avant le critère de minimisation comme l'indique la figure 2.2 [24].

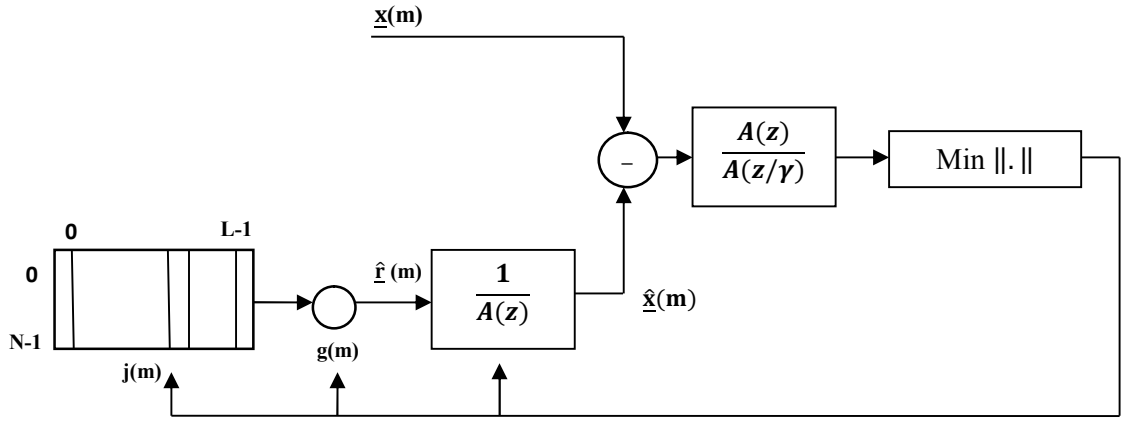


Figure 2.2 : Introduction d'une fonction de pondération.

Maintenant le bruit dû à la quantification est moins perceptible lorsque le signal a beaucoup d'énergie. On dit que le signal masque le bruit. Il n'est pas possible de jouer sur la puissance totale du bruit de quantification. Par contre il est possible de modifier la forme spectrale du bruit. On cherche donc une fonction de pondération qui attribue moins d'importance aux zones fréquentielles énergétiques c'est-à-dire aux zones formantiques [24]. On montre que la fonction de transfert $W(z) = A(z)/A(z/\gamma)$ avec $0 < \gamma < 1$ joue ce rôle. En effet, si on note

$$A(z) = 1 + a_1 z^{-1} + \cdots + a_p z^{-p} = \prod_{i=1}^{P}(1 - p_i z^{-1}) \qquad (2.1)$$

Où p_i spécifie la $i^{ème}$ racine du polynôme A(z), on remarque que

$$A\left(\frac{z}{\gamma}\right) = 1 + a_1 \gamma z^{-1} + \cdots + a_P \gamma^P z^{-P} = \prod_{i=1}^{P}(1 - \gamma p_i z^{-1}) \qquad (2.2)$$

Le module de la réponse en fréquence du filtre $1/A(z/\gamma)$ présente des pics moins accentués que celui du filtre 1/A(z) puisque les pôles du filtre $1/A(z/\gamma)$ sont ramenés vers le centre du cercle unité par rapport à ceux du filtre 1/A(z). Le module de la réponse en fréquence du filtre $W(z) = A(z)/A(z/\gamma)$ a donc la forme souhaitée comme le montre la figure 2.3, dans cet exemple $\gamma = 0.8$.

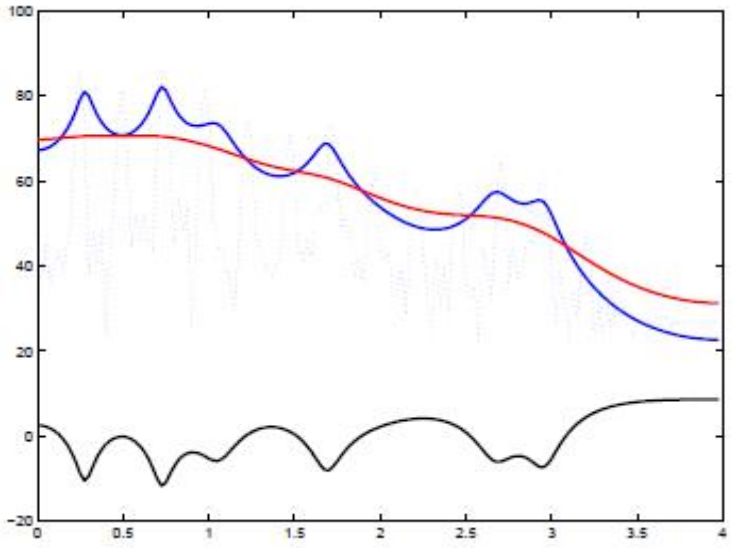


Figure 2.3 : Réponses en fréquences des filtres 1/A(z), $1/A(z/\gamma)$ et $A(z)/A(z/\gamma)$ pour un son voisé dont le spectre est visualisé en pointillés [24].

Le diagramme donnant le principe de la modélisation devient celui de la Figure 2.4.

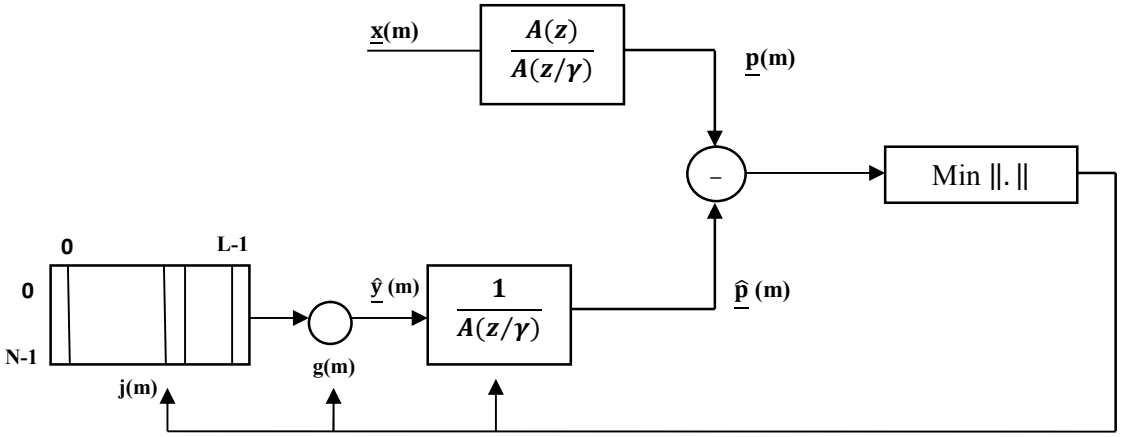


Figure 2.4 : Modélisation du signal perceptuel [24].

Ce diagramme met en évidence le fait que l'on cherche à modéliser le signal perceptuel $\underline{p}$ par $\underline{\hat{p}}$. Par la suite le filtre perceptuel est le filtre caractérisé par la fonction de transfert $1/A(z/\gamma)$. Le choix de la valeur numérique du facteur perceptuel permet de moduler la fonction de pondération à sa convenance. Pour γ = 1, tout se passe comme si on n'utilisait pas de fonction de pondération ; on effectue une modélisation du signal original. L'erreur de reconstruction $\underline{x} - \underline{\hat{x}}$ sera approximativement blanche. Pour γ = 0, on réalise une modélisation du signal résiduel. L'erreur de reconstruction aura la forme spectrale du signal original. On choisit généralement γ voisin de 0.8.

2.4 Formalisation du problème

Dans un codeur CELP à bas débit, le problème essentiel est de réaliser une bonne modélisation du signal d'excitation du filtre de synthèse. Pour ce faire, on cherche les vecteurs *cj*(1),..., *cj*(*M*) dans un dictionnaire d'excitation et les gains respectifs *gj*(1) ,..., *gj*(*M*) de façon à ce que le vecteur d'excitation $\hat{r} = [\hat{r}(0), \dots, \hat{r}(N'-1)]$ filtré par le filtre perceptuel $1/A(z/\gamma)$ donne le vecteur perceptuel modélisé $\hat{p}$ le plus ressemblant possible au vecteur perceptuel *p*. Pour cela, nous devons minimiser l'erreur quadratique entre $\hat{p}$ et *p*. Le critère devient alors la minimisation de l'énergie de l'erreur perceptuelle sur l'ensemble de la sous fenêtre d'analyse [16].

2.4.1 Expression du critère

La forme du modèle de l'entrée étant fixée, il faut maintenant, déterminer l'expression de $\|p - \hat{p}\|^2$, où le vecteur perceptuel *p* est un vecteur connu et le vecteur $\hat{p}$ dépend nonseulement des paramètres inconnus *j*(1),...,*j*(*M*) et *g j*(1) ,..., *g j*(*M*) mais aussi de l'excitation $\hat{r}$correspondante à la fenêtre précédente [16] . Le signal perceptuel modélisé a pour expression [25] :

$$\hat{p}(n) = \sum_{i=0}^{\infty} h(i)\,\hat{r}\,(n-i), \qquad n = 0, \dots, N'-1 \tag{2.3}$$

Où *h*(*i*) est la réponse impulsionnelle du filtre $1/A(z)/\gamma$(considéré causal : *h*(*i*) = 0 pour *i* < 0).

L'expression précédente se décompose en deux termes :

$$\hat{p}(n) = \sum_{i=0}^{n} h(i)\,\hat{r}\,(n-i) + \sum_{i=n+1}^{\infty} h(i)\,\hat{r}\,(n-i), \qquad n = 0, \dots, N'-1 \tag{2.4}$$

Le premier terme est a priori inconnu mais le deuxième terme est connu puisqu'il fait intervenir $\hat{r}(n)$ pour *n* <0 ; c'est-à-dire, l'excitation du filtre de synthèse déterminée dans les sous-fenêtres d'analyse précédentes. Notons que, *n* = 0 caractérise le premier échantillon de la sous-fenêtre d'analyse [16]. Le signal perceptuel modélisé (synthétique) sur l'intervalle [0… *N*′−1] s'écrit :

$$\hat{p}(n) = \sum_{k=1}^{M} \mathrm{g}_{j(k)} \sum_{i=0}^{n} h(i) \cdot c_{j(k)}(n-i) + \sum_{i=n+1}^{\infty} h(i) \cdot \hat{r}(n-i) \tag{2.5}$$

Adoptons une notation vectorielle, où l'on appelle :

$\hat{p}_0 = [\hat{p}_0(0) \dots \hat{p}_0(N'-1)]^T$, avec : $\hat{p}_0(n) = \sum_{i=n+1}^{\infty} h(i)\hat{r}(n-i)$ (2.6)

La contribution de l'excitation provenant des sous-fenêtres précédentes dans la sous-fenêtre courante ; et :

$$f_j = \left[f_j(0) \dots f_j(N'-1)\right]^T, \quad \text{avec :} \quad f_j(n) = \sum_{i=0}^{n} h(i)c_j(n-i) \tag{2.7}$$

Le résultat du filtrage du vecteur c_j par le filtre perceptuel partant de conditions initiales nulles.

L'ensemble des vecteurs fj (j = 0,..., L−1) constituera le *dictionnaire filtré*. Ainsi, sous forme vectorielle, l'expression (2.5) s'écrit :

$$\hat{p} = \hat{p}_0 + \sum_{k=1}^{M} \mathrm{g}_{j(k)} f_{j(k)} \tag{2.8}$$

Où $f\,j(k)$ est le vecteur issu du dictionnaire filtré à l'indice $j(k)$. Finalement le critère d'erreur perceptuel qui doit être minimisé, s'écrit de la façon suivante :

$$E = \|p - \hat{p}\|^2 = \left\| p - \hat{p}_0 - \sum_{k=1}^{M} \mathrm{g}_{j(k)} f_{j(k)} \right\|^2 \tag{2.9}$$

2.4.2 Minimisation du critère

Le problème de la détermination de l'excitation dans un codeur CELP se résume ainsi : Connaissant p, $\hat{p}_0$ et le dictionnaire filtré, trouver les indices $j(1)$,..., $j(M)$ et les gains $g\,j(1)$,.., $g\,j(M)$ de façon à minimiser le critère (2.9). Pour simplifier les notations, on appelle p le vecteur perceptuel auquel on a enlevé la contribution des sous-fenêtres précédentes, Ainsi la relation (2.9) devient [16] :

$$E = \left\| p - \sum_{k=1}^{M} \mathrm{g}_{j(k)} f_{j(k)} \right\|^2 \tag{2.10}$$

Il s'agit d'un problème classique de minimisation au sens des moindres carrés si l'on suppose connus les indices $j(1)$,..., $j(M)$. On obtient le minimum en annulant la dérivée partielle de E par rapport à chaque gain $g\,j(i)$ (i = 1,..., M).

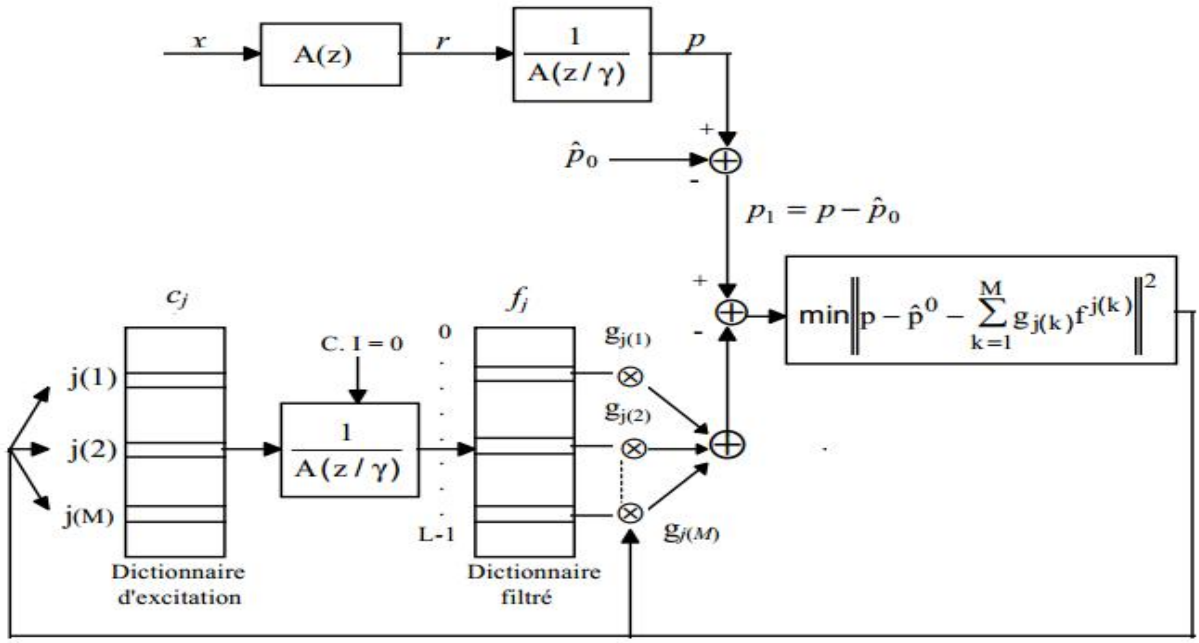


Figure 2.5 : Modélisation du signal perceptuel par M vecteurs du dictionnaire filtré et M gains [16].

$$\frac{\partial E}{\partial g_{j(i)}} = 2\left(p - \sum_{k=1}^{M} g_{j(k)} f_{j(k)}\right) f_{j(i)} = 0 \qquad (2.11)$$

Ainsi, nous obtenons les équations normales suivantes :

$$\sum_{k=1}^{M} g_{j(k)} < f_{j(k)}, f_{j(i)} > = < p, f_{j(i)} >,\ i = 1, \ldots, M, \qquad (2.12)$$

Où<*x*, *y*> désigne le produit scalaire de deux vecteurs *x* et *y*, définie par :

$$< x, y >= \sum_{n=0}^{N'-1} x(n) y(n). \qquad (2.13)$$

2.4.3 Algorithme itératif standard

La complexité des calculs est due essentiellement à la sélection simultanée des vecteurs *fj*(1),..., *fj*(*M*). Une façon de simplifier le traitement est de se limiter à une méthode sous optimale qui consiste en la recherche d'un seul vecteur à la fois. Le calcul des indices et des gains se fait de façon itérative [25].

A chaque itération un vecteur *fj* est choisi et un gain g_j est calculé, de façon à minimiser l'expression suivante :

$$E' = \left\| p - \mathrm{g}_j f_j \right\|^2 \tag{2.14}$$

Ceci revient à annuler la dérivée de E' par rapport à g_j, ce qui donne :

$$\mathrm{g}_j = \frac{< p, f_j >}{< f_j, f_j >} \tag{2.15}$$

En développant E' et en remplaçant g_j par l'expression de ci-dessus, on obtient :

$$E' = \|p\|^2 - \frac{< p, f_j >^2}{< f_j, f_j >} \tag{2.16}$$

La minimisation de la relation (2.16) est équivalente à la maximisation du terme :

$$\frac{< p, f_j >^2}{< f_j, f_j >} \tag{2.17}$$

Ce dernier est toujours positif ou nul. De plus le terme $\|p\|^2$est à priori connu.

A la première itération, on cherche donc l'indice $j(1)$ qui maximise (2.17). On écrira :

$$j(1) = Arg\max_{\mathrm{j}} \frac{<p,f_j>^2}{<f_j,f_j>} \quad , \quad \text{pour j=0,…,L-1} \tag{2.18}$$

Ensuite, on calcule le gain $g_{j(1)}$ correspondant :

$$\mathrm{g}_{j(1)} = \frac{< p, f_{j(1)} >}{< f_{j(1)}, f_{j(1)} >} \tag{2.19}$$

A la $k^{\text{ième}}$ itération, la contribution des $k-1$ premiers vecteurs $f_j(i)$ (i =1,.., $k-1$) est retirée de p :

$$p^k = p - \sum_{i=1}^{k-1} \mathrm{g}_{j(i)} f_{j(i)} \tag{2.20}$$

Et un nouvel indice $j(k)$ et un nouveau gain $g_j(k)$ sont calculés vérifiant :

$$j(k) = Arg\max_{\mathrm{j}} \frac{<p^k,f_j>^2}{<f_j,f_j>} \quad , \quad \text{j=0,…, L-1 et } j \neq j(1) \ldots \neq j(k-1) \tag{2.21}$$

$$\mathrm{g}_{j(k)} = \frac{< p^k, f_{j(k)} >}{< f_{j(k)}, f_{j(k)} >} \tag{2.22}$$

Dans les codeurs CELP classique, le dictionnaire d'excitation est souvent initialisé par des tirages d'une variable aléatoire gaussienne centrée ou construit par apprentissage en utilisant des techniques de quantification vectorielle.

2.5 Choix du dictionnaire d'excitation

Cet algorithme est applicable quel que soit le contenu du dictionnaire d'excitation. Si on choisit C = I où C'est la matrice composée de vecteurs colonnes $\underline{c}^j$ et *I* la matrice identité de dimension N*N, on obtient une excitation "multi-impulsionnelle". Les indices sélectionnés j(k) caractérisent les positions des impulsions choisies et les gains g_k définissent les amplitudes. Dans le codeur CELP classique, on initialise le dictionnaire d'excitation par des tirages d'une variable aléatoire gaussienne centrée ou on le construit par apprentissage en utilisant une variante de l'algorithme de Lloyd-Max. Dans ces deux cas, on peut vérifier que la complexité du traitement est assez importante. Pour réduire cette complexité, on peut chercher à définir un dictionnaire comportant une forte structure par exemple imposer des valeurs ternaires {+1, -1, 0} et choisir des positions régulières dans le dictionnaire pour les valeurs non nulles. C'est la particularité du codeur G.729 qui est appelé pour cette raison le codeur ACELP (Algebric CELP) [24].

2.6 Introduction d'un dictionnaire adaptatif

Pour déterminer les coefficients du filtre A(z), on minimise l'énergie de l'erreur de prédiction [24] :

$$D_1 = \sum_n \left[x(n) - \sum_{i=1}^{p} a_i x(n-i) \right]^2 \qquad (2.23)$$

Par rapport aux *P* paramètres inconnus a_i. Il s'agit d'une prédiction dite à court terme puisque, pour prédire la valeur du signal à l'indice *n*, on utilise les *P* échantillons précédents. Une fois le calcul réalisé par simple résolution du système linéaire obtenu en dérivant D_1 par rapport aux *P* paramètres inconnus, on filtre le signal *x(n)* par le filtre de fonction de transfert *A(z)* d'ordre *P*. On obtient le signal résiduel à court terme *y(n)*. La visualisation de ce signal, spécialement pour des sons voisés, montre que toute la redondance placée dans le signal de parole n'a pas été extraite. Il reste une certaine périodicité comme le montre les tracés de la Figure 2.6. Cette périodicité correspond, physiologiquement, à la période de vibration des cordes vocales. On cherche à caractériser cette information en introduisant deux nouveaux paramètres *b* et *Q* puis en minimisant l'énergie d'une nouvelle erreur de prédiction par rapport à ces deux paramètres inconnus [24].

$$D_2 = \sum_n \left[y(n) - \sum_{i=1}^{p} b y(n-Q) \right]^2 \qquad (2.24)$$

On parle alors de prédiction à long terme. On remarquera que cette minimisation ne peut pas être réalisée comme la précédente puisque, pour D_2, P est fixé, alors que pour D_2, Q est un paramètre à déterminer.

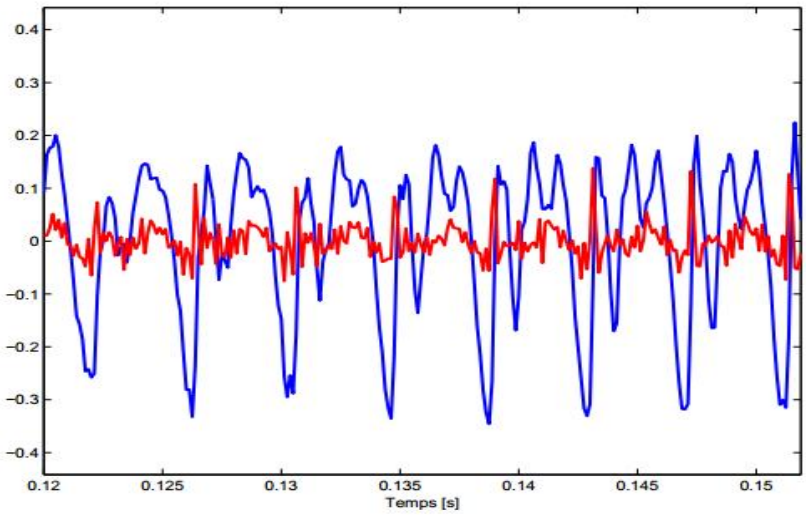


Figure 2.6 : Signal de parole *x(n)* et signal résiduel *y(n)* pour un locuteur féminin.

Minimisation en "boucle ouverte"

Pour calculer les valeurs optimales de b et Q, il suffit de dériver D_2 par rapport à b pour obtenir b optimal comme une fonction de Q, de reporter cette valeur dans D_2 puis de choisir la valeur de Q qui minimise le critère. Ce n'est pas forcément la solution la plus adaptée. On préfère, généralement, une solution en boucle fermée [25].

Minimisation en "boucle fermée"

Introduisons la fonction de transfert

$$B(z) = 1 - bz^{-Q} \tag{2.25}$$

A la synthèse, on utilise le filtre inverse *1/B(z)* qui doit être placé en amont du filtre de fonction de transfert *1/A(z).*

Reprenons le développement présenté précédemment en supposant b et Q prédéterminés. On cherche des vecteurs $\underline{c}^{j(k)}$ dans le dictionnaire d'excitation et des gains g_k de façon à ce que le vecteur $\sum_{k=1}^{K} g_k \underline{c}^{j(k)}$ filtré par le filtre *1/B(z)* puis par le filtre perceptuel $1/A(z/\gamma)$donne le vecteur modélisé $\underline{\hat{p}}$ le plus ressemblant possible au vecteur $\underline{p}$. On a vu que le signal perceptuel modélisé avait pour expression [25] :

$$\hat{p}(n) = \sum_{i=0}^{n} h(i)\,\hat{y}\,(n-i) + \sum_{i=n+1}^{\infty} h(i)\,\hat{y}\,(n-i), \qquad n = 0, \dots, N-1 \tag{2.26}$$

Mais $\hat{y}(n)$ pour n ≥ 0, a priori inconnu, se décompose aussi en une partie inconnue qui ne dépend que des $\underline{c}^{k}$ et g_k et une partie connue

$$\hat{y}(n) = \sum_{k=1}^{K} g_k \underline{c}^{j(k)}\,(n) + b\hat{y}(n-Q) \tag{2.27}$$

Si on admet les hypothèses que les paramètres b et Q du prédicteur à long terme ont été déterminés et que

$$n - Q < 0 \qquad \forall\, n \in 0 \dots N-1$$

C'est-à-dire

$$Q \geq N$$

La valeur du décalage doit donc être supérieure ou égale à la taille de la fenêtre d'analyse.

Finalement le signal perceptuel modélisé s'écrit :

$$\hat{p}(n) = \sum_{k=1}^{K} g_k \sum_{i=0}^{n} h(i) c^{j(k)}\,(n-i) + b\sum_{i=0}^{n} h(i)\,\hat{y} + \sum_{i=n+1}^{\infty} h(i)\,\hat{y}\,(n-i) \tag{2.28}$$

2.7 Conclusion

Dans ce chapitre nous avons étudié en détails le principe du codeur CELP ; nous avons précisé qu'il contient deux dictionnaires à base de quantificateur vectoriel d'ordre 2. Le premier est de nature adaptative ; c'est le dictionnaire prédictif qui modélise le signal d'excitation. Le deuxième est de nature statique ; c'est le dictionnaire d'excitation statique qui quantifie les gains. Les coefficients du filtre de synthèse sont calculés à chaque fenêtre d'analyse par un algorithme de Levinson-Durbin.

Chapitre 3 :

Mise en œuvre du codeur CELP à 8 Kbps et 4.8 Kbps

3.1 Introduction :

La famille des algorithmes de codeur CELP sont l'approche la plus efficace actuellement pour un codage de la parole de haute qualité avec des débits minimaux ; la plupart des recherches connues dans la compression de la parole se concentrent sur ces techniques décodage [1]. Ce chapitre est consacré à l'étude de deux codeurs CELP ; le premier opérant à un débit de 8Kbps et le second à 4.8 Kbps. Notre objectif est de synthétiser la parole avec une bonne perception avec ces deux codeurs afin de comparer la synthèse de la parole à bas débit.

3.2 Description du standard CELP à 8 Kbps *:*

3.2.1 Encodeur CELP à 8 Kbps :

Le principe de l'encodeur G.729 est donné à la figure 3.1. Avant tout traitement le signal de parole d'entrée subit dans une procédure dite de prétraitement une normalisation et un filtrage passe-haut dont la fréquence de coupure du filtre est égale à 140 Hz. En sortie de cette opération de prétraitement, le signal noté s(n), est utilisé comme entrée de tous les blocs successifs du l'encodeur [26].

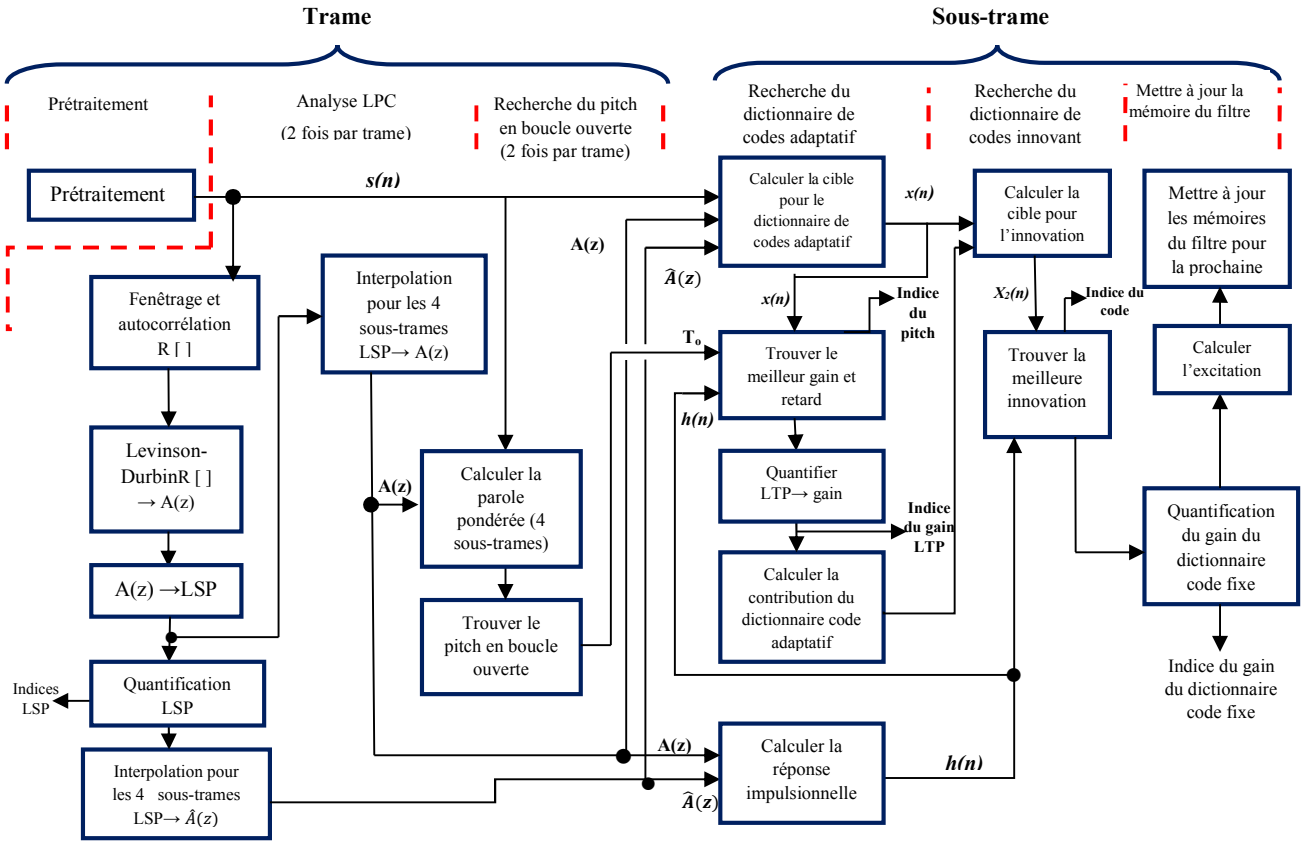


Figure 3.1 : Schéma avec les différents blocs du codeur CELP à 8 Kbps [27].

L'encodeur G.729 opère sur des trames de parole de 20 ms correspondant à 160 échantillons pour une fréquence d'échantillonnage de 8 kHz. Le signal de parole est analysé à chaque trame pour extraire les paramètres du modèle CELP ; coefficients du filtre de Prédiction Linéaire : A(z)=1+$\sum_{k=1}^{10} a_k\, z^{-k}$du dixième ordre (m=10) en utilisant l'algorithme de Levinson-Durbin (Appendice A). La prédiction linéaire est déroulée deux fois par trame utilisant une approche d'autocorrélation, avec des fenêtres non symétriques. Le filtre perceptif est basé sur les coefficients de prédiction $\{a_i\}_{i=1,..,p}$(non quantifiés) et est donné par : $W(z) = A(z/\gamma 1)/A(z/\gamma 2)$, où les valeurs de $\gamma 1$ et$\gamma 2$ sont en fonctions de la forme spectrale du signal d'entrée (spectre plat ou non) [26].

Les coefficients de prédiction sont convertis en fréquences de raies spectrales (LSF) pour des motifs d'interpolation et de Quantification et sont codés sur 18 bits puis quantifiés et interpolés. Les coefficients interpolés, quantifiés et non quantifiés seront reconvertis en coefficient de prédiction linéaire, afin de reconstruire les filtres de synthèse pour chaque sous trame.Les deux groupes de paramètres de prédiction linéaire obtenus seront convertis en paires de ligne de spectre, et jointement quantifiés utilisant la SMQ (*Split Matrix Quantization)* sur 38 bits. Chaque trame est divisée en deux sous trames de 10 ms, soit 80 échantillons chacune. Par la suite, les paramètres d'excitation, tels que les indices ainsi que les gains des dictionnaires fixe et adaptatif, sont estimés sur la base de sous-trames de 80 échantillons, soit de 10 ms [26].

Le signal d'excitation est codé en recherchant dans deux dictionnaires (adaptatif et fixe) les formes d'onde (la séquence d'excitation optimale) qui minimisent et synthétisent l'erreur entre les signaux de parole original et reconstruit suivant une mesure de distorsion en tenant compte d'une pondération perceptive qui améliore la qualité de restitution de la parole. Les formes d'onde étant normalisées, un gain leur est associé de manière à modéliser au mieux les séquences du signal d'excitation. Cette méthode de recherche est dite en boucle fermée ou encore analyse par synthèse.

Le tableau 3.1 donne le schéma d'allocation des bits pour les paramètres issus de l'encodeur G.729. Ces paramètres correspondant aux informations nécessaires à la reconstitution d'une trame vocale de 20 ms. Donc un total de 160 bits est transmis par trame de 20 ms. Il en résulte un débit binaire de 8Kbps [26].

Tableau 3.1 : Allocation des bits du codeur CELP à 8Kbps.

Paramètres	Nombre de bits				
	Trame n°1 de 10 ms		Trame n°2 de 10 ms		Transmission Par Trame
Indice LPC	18		18		36
	Sous-trame1	Sous-trame2	Sous-trame1	Sous -trame2	
Indice du dictionnaire adaptatif (période du pitch)	8	5	8	5	26
Bit de parité pour la période du pitch	1		1		2
Indice du dictionnaire fixe	13	13	13	13	52
Signe de la contribution du dictionnaire fixe	4	4	4	4	16
Gains des contributions fixe et adaptatif	(3+4) 7	(3+4) 7	(3+4) 7	(3+4) 7	(12+16) 28
Total : à 8 Kbps	80		80		**160**

3.2.2 Décodeur CELP à 8 Kbps :

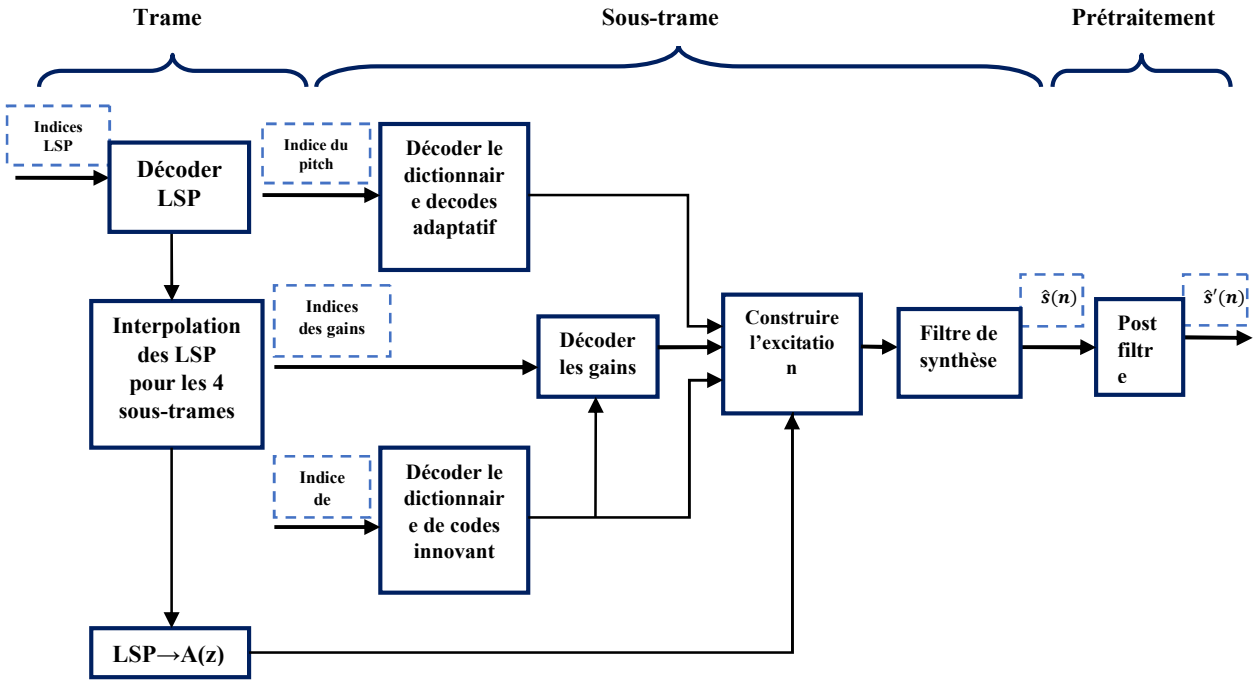


Figure 3.2 : Schéma fonctionnel détaillé du décodeur CELP à 8 Kbps [27].

La fonction du décodeur consiste en décoder les paramètres transmis [26] :

- Paramètres LP
- Vecteur de code du dictionnaire adaptatif
- Gain du dictionnaire de code adaptatif
- Vecteur de code du dictionnaire fixe
- Gain du dictionnaire de code fixe

Ensuite ce décodeur effectue une synthèse pour obtenir la parole reconstruite, qui sera ensuite filtrée et accentuée.

3.2.2.1 Décodage et synthèse de parole :

Décodage des paramètres de prédiction linéaires :

Les indices reçus de la quantification des LSP sont utilisés pour reconstruire les deux vecteurs LSP quantifiés. L'interpolation est appliquée pour retrouver les 4 vecteurs LSP interpolés, correspondants au 4 sous trames. Pour chacune des sous trame, le vecteur LSP interpolé est converti en coefficients de filtre de prédiction linéaire $a_k$qui sont utilisés pour synthétiser la parole de cette sous trame [26].

Pour chacune des sous trames on répète les étapes suivantes :

1) Décodage du vecteur de code du dictionnaire adaptatif :

L'indice de pitch ou le l'indice du dictionnaire de code reçu est utilisé pour trouver la partie entière et la partie fractionnelle de la période fondamentale (durée de pitch). Le vecteur de code du dictionnaire adaptatif $v(n)$ est trouvé par interpolation de l'excitation passée $u(n)$ utilisant le filtre à réponse impulsionnelle finie

2) Décodage du gain du dictionnaire de code adaptatif :

L'indice reçu est utilisé pour trouver le gain quantifié dans une table de quantification.

3) Décodage du vecteur d'innovation du dictionnaire de code

L'indice dictionnaire algébrique est utilisé pour extraire les positions et les amplitudes (signes) des impulsions d'excitation et pour trouver le vecteur de code algébrique $c(n)$.

4) Décodage du gain du dictionnaire de code fixe :

L'indice reçu donne le facteur de correction de gain du dictionnaire de code$\widehat{\gamma}_{gc}$.

Le gain estimé est g'$_c$ trouvé comme suite :

$$g'_c=10^{0.05(\tilde{E}(n)+\bar{E}-E_i)} \qquad (3.1)$$

Le gain quantifié est donné par

$$\hat{g}_c=\widehat{\gamma}_{gc}g'_c \qquad (3.2)$$

5) Calcul de la parole reconstruite :

$$u(n) = \hat{g}_p v(n) + \hat{g}_c c(n). \quad n=0...39 \tag{3.3}$$

Avant la synthèse de la parole, les éléments d'excitation subissent un post traitement. Cela signifie que l'excitation tout entière est modifiée en accentuant la contribution du vecteur de code du dictionnaire adaptatif.

$$\hat{u}(n) = \begin{cases} u(n) + 0.25\beta\, \hat{g}_p\, v(n), & \hat{g}_p > 0.5 \\ u(n), & \hat{g}_p \leq 0.5 \end{cases} \tag{3.4}$$

Le contrôle adaptatif du gain est utilisé pour compenser la différence de gain entre l'excitation non modifié $u(n)$ et celle modifié (accentuée) $\hat{u}(n)$. Le facteur de gain affecté à l'excitation modifiée est donné par :

$$\eta = \begin{cases} \sqrt{\frac{\sum_{n=0}^{39} u^2(n)}{\sum_{n=0}^{39} \hat{u}^2(n)}}, & \hat{g}_p > 0.5 \\ 1.0, & \hat{g}_p \leq 0.5 \end{cases} \tag{3.5}$$

Le signal d'excitation modifié et affecté du gain $\hat{u}'(n)$ est donné par :

$$\hat{u}'(n) = \hat{u}(n)\, \eta \tag{3.6}$$

Le signal de parole reconstruit pour une sous trame de 40 échantillons est donné par :

$$\hat{s}(n) = \hat{u}'(n) - \sum_{i=1}^{10} \hat{a}_i\, \hat{s}(n-i), \; n = 0, \dots, 39 \tag{3.7}$$

Où $\hat{a}_i$représente les coefficients interpolés du filtre de prédiction linéaire.

La parole synthétisée $\hat{s}(n)$ est ensuite filtrée par un filtre adaptatif qui sera décrit dans le paragraphe qui suit.

3.2.2.2 Post traitement :

Le post traitement se compose de deux fonctions : post filtrage adaptatif et amélioration du signal [27].

Post filtrage adaptatif *:*

Le processus du post filtrage se déroule comme suit :

- La parole synthétisée$\hat{s}(n)$ subit le filtrage inverse à travers $\hat{A}(Z/\gamma_n)$ Pour produire un signal résiduel $\hat{r}(n)$.
- Le signal $\hat{r}(n)$ est filtré par le filtre de synthèse $\frac{1}{\hat{A}(Z/\gamma_n)}$.
- Le signal à la sortie du filtre de synthèse $\frac{1}{\hat{A}(Z/\gamma_n)}$ passe au filtre de la compensation du biais $h(z)$ t résultant du post filtrage dans signal de parole $\hat{s}_f(n)$.

Le contrôle adaptatif du gain est utilisé pour compenser la différence du gain entre le signal de la parole synthétisé $\hat{s}(n)$et le signal post filtré $\hat{s}_f(n)$. Le facteur de gain pour la sous trame actuelle est calculer par la formule :

$$\gamma_{sc} = \sqrt{\frac{\sum_{n=0}^{39} \hat{s}^2(n)}{\sum_{n=0}^{39} \hat{s}^2{}_f(n)}} \quad (3.8)$$

Le signal post filtré et augmenté par le facteur du gain est donné par :

$$\hat{s}'(n)=\beta_{sc}(n)\hat{s}_f(n) \quad (3.9)$$

Où $\beta_{sc}(n)$ est mis à jour pour chaque échantillon, et il est donné par :

$$\beta_{sc}(n)=\alpha\beta_{sc}(n-1)+(1-\alpha)\gamma_{sc} \quad (3.10)$$

Où α est le facteur du contrôle adaptatif du gain et sa valeur est 0.9.

Les facteurs de post filtrage adaptatif sont donnés parγ_n=0.7, γ_d=0.75 et

$$\gamma_t=\begin{cases}0.8\ ,\ k_1' > 0\\ 0\ ,\ ailleurs\end{cases} \quad (3.11)$$

3.2.3 Domaines d'application :

- Téléphonie visuelle,
- Communications mobile et sans fil,
- Systèmes de communication par satellite numérique [28].
- Pris en charge sur les passerelles vocales VoIP,
- Adaptateurs VoIP populaires et téléphones IP [29].
- Transmission de la voix sur IP (VoIP) et téléconférence [30].
- TDMA et CDMA téléphones cellulaires,
- Streaming audio internet [31].

3.3 Description du standard CELP à 4.8Kbps :

3.3.1 Encodeur CELP à 4.8 Kbps :

Le codeur CELP à 4.8Kbpsutilise une fréquence d'échantillonnage de 8 KHZ et une fenêtre d'analyse de 30 ms divisée en quatre sous-fenêtres de 7.5 ms. Le signal synthétique est obtenu par le passage d'un signal d'excitation à travers un filtre de prédiction linéaire 1/A(z). L'excitation est calculée par sous-fenêtre. Elle est le résultat de l'addition de deux excitations élémentaires :

- La première est un vecteur-code extrait du dictionnaire adaptatif à l'indice i_a et pondéré par un gain g_a.

- La deuxième est un vecteur-code extrait du dictionnaire stochastique à l'indice i_s et pondéré par un gain g_s.

L'analyse du codeur CELP consiste à trouver les paramètres d'excitation (les indices et les gains) de manière à minimiser l'erreur perceptuelle (sortie du filtre perceptuel A(z)/A(z/α) avec α=0,8) entre le signal de parole originale s et le signal synthétisé $\hat{s}$. L'analyse est dite en boucle fermée. Le codeur contient donc à la fois le module d'analyse et le module de synthèse. L'allocation des bits des différents paramètres à transmettre est donnée au tableau 3.2 [18].

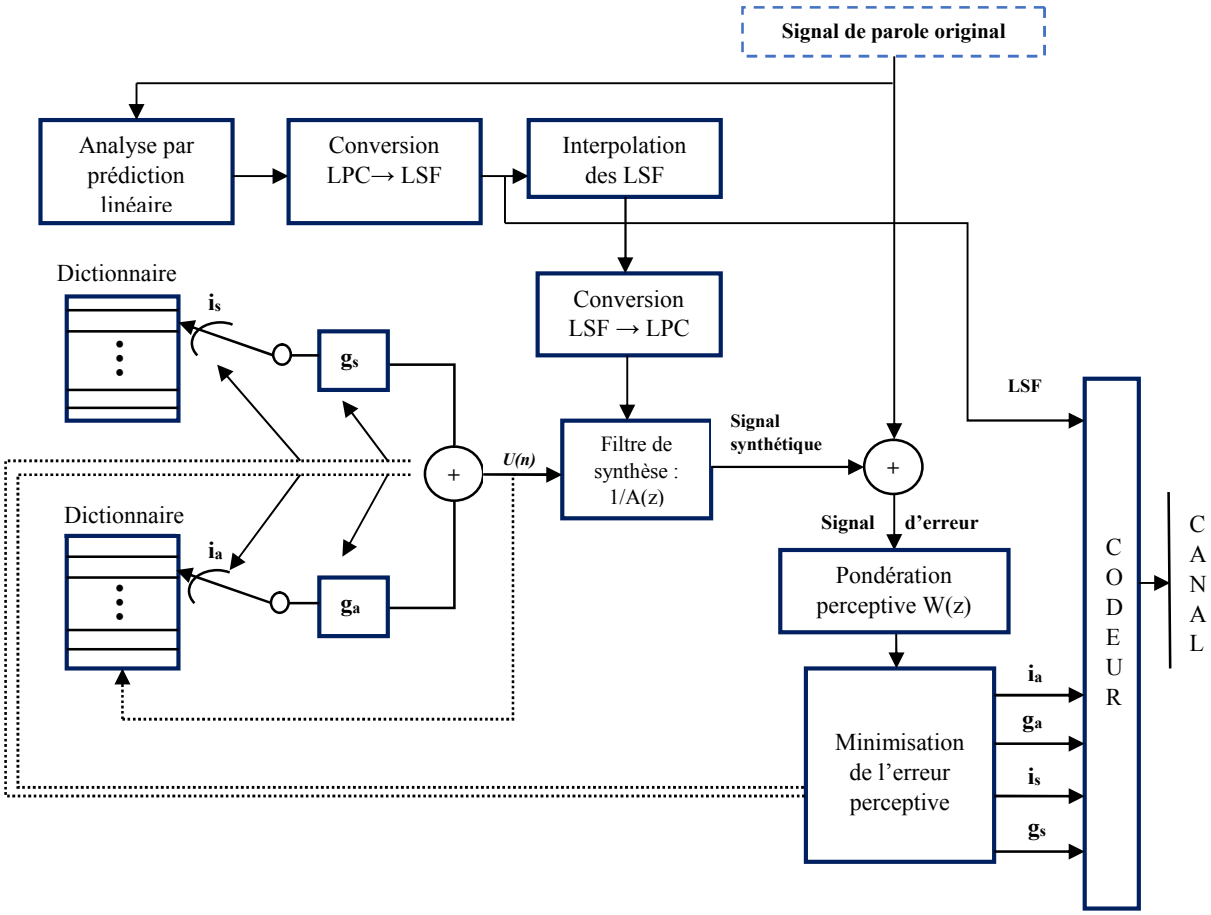


Figure 3.3 : Schéma avec les différents blocs du codeur CELP à 4.8 Kbps [32].

Tableau 3.2 : Allocation des bits du codeur CELP de 4.8Kbps [32, 33].

Paramètres/ trame	**Bits/trame**	**Débits**
Fréquence d'échantillonnage	**8KHz**	
Taille de la trame	**240échantillons (30 ms)**	
Débit en trame	**33.33 trame/ seconde**	
10 LSP	**[3,4,4,4,4,3,3,3,3,3]**	**1.13333**
Pitch (i_a)	**8 + 6 + 8 + 6**	**1.600**
Gain adaptatif (g_a)	**4 x 5**	
Indice d'excitation (i_s)	**4 x 9**	**1.86667**
Gain stochastique (g_s)	**4 x 5**	
Synchronisation	**1**	**0.2**
Bit de correction d'erreur	**4**	
Bit non utilisé	**1**	
Total bit par trame	**144 bits**	
Débit total		**4.8Kbits/s**

3.3.1.1 Analyse par prédiction linéaire :

Elle comporte l'estimation de 10 coefficients de prédiction a_i par la méthode d'autocorrélation sur des fenêtres de 30 ms (pondéré par la fenêtre de Hamming) [34].Les coefficients de prédictions $\{a_k\}_{k=1,\dots,P}$sont convertis aux coefficients fréquences de raies spectrales LSF : $\{w_k\}_{k=1,\dots,P}$. La trame à coder est généralement divisée en 2 à 4 sous-trames [35]. Dans le traitement par sous-trame, une interpolation linéaire est généralement réalisée entre deux ensembles de p coefficients LSF chacun (correspondant à deux trames consécutifs) pour former un ensemble intermédiaire de p coefficients LSF pour chaque sous-trame [36]. L'interpolation linéaire est effectuée pour lisser l'évolution des paramètres LP et réduire ainsi la présence de transitions brutales dues aux changements rapides des paramètres LP entre les trames [18].

3.3.1.2 Recherche dans le dictionnaire adaptatif :

Le dictionnaire adaptatif, modélise la périodicité du signal à long terme due au Pitch. Il est constitué d'une mémoire tampon composée de 256 vecteurs-codes correspondants aux excitations passées [34].Pour chaque trame de N échantillons, l'encodeur transmet p coefficients LSF et N / L fois les paramètres du signal d'excitation :$\{i_a, g_a\}$ et$\{i_s, g_s\}$. Le décodeur peut reconstituer le même signal d'excitation qu'à l'encodeur puisque il dispose des mêmes dictionnaires fixe et adaptatif [18].

3.3.1.3 Recherche dans le dictionnaire stochastique :

Le dictionnaire stochastique est composé de 512 vecteurs codes. Il modélise le signal qui provient du filtre inverse A(z) et auquel la structure périodique due au Pitch est soustraite. Ce signal présente une distribution d'échantillons similaire à celle d'une Gaussienne [34].

C'est pourquoi le dictionnaire stochastique est un dictionnaire entrelacé (déplacement de 2 échantillons) dont les éléments sont quantifiés à 3 niveaux (-1, 0,1) et distribués selon une Gaussienne de moyenne nulle et de variance unitaire [36].

Le codeur transmet, à chaque fenêtre de 30 ms, une trame de 144 bits (tableau 2) à travers le canal de transmission. Après décodage de cette trame, le décodeur CELP reconstruit le signal de parole synthétique en utilisant le même module de synthèse que celui du codeur. C'est à dire par passage de l'excitation à travers le filtre de prédiction linéaire [34].

3.3.2 Décodeur CELP à 4.8 Kbps :

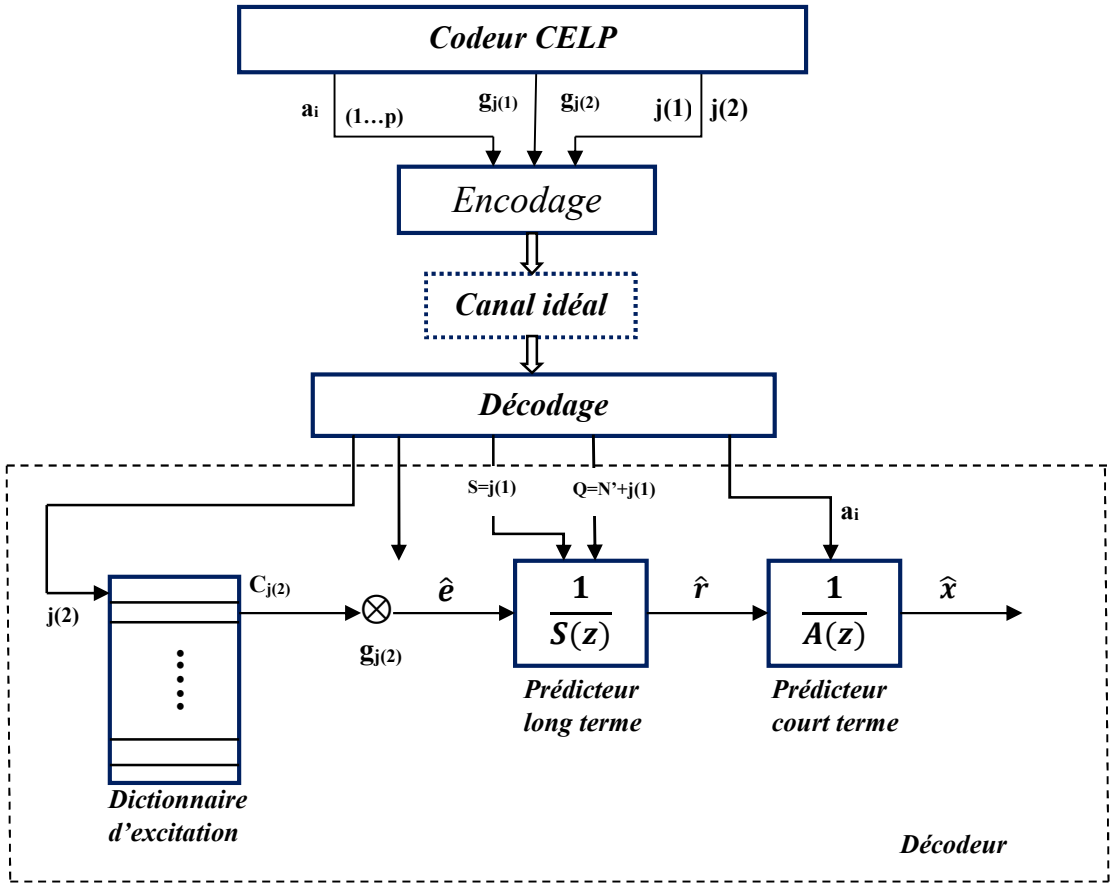


Figure 3.4 : Schéma avec les différents blocs du décodeur CELP à 4.8 Kbps [33].

Les paramètres transmis au canal par le codeur CELP sont : les coefficients LPC, les indices j(1) et j(2) ainsi que les gains correspondants gj(1) et gj(2). Bien sûr, ces paramètres doivent être quantifiés et codés en binaire avant de les transmettre. Nous supposons que le canal de transmission est sans bruit.

Une fois les indices (statistiques et prédictifs) et les gains correspondants ont été reçus et décodés, on peut reconstituer la même excitation $\hat{r}$ qu'au niveau du codeur (figure 3.3). Il suffit pour cela de sélectionner le vecteur-code, d'indice statistique j(2) décodé du dictionnaire d'excitation de le multiplier par le gain correspondant gj(2) et de faire passer le vecteur résultat à travers le prédicteur à long terme de gain s= gj(1) et de retard Q = N' + j(1). En faisant passer l'excitation $\hat{r} = [\hat{r}(0),\ldots, \hat{r}(N'-1)]$ à travers le filtre de synthèse (dont les paramètres LPC a_i sont également transmis par le codeur), on peut construire ainsi le signal synthétisé$\hat{x} = [\hat{x}(0),\ldots, \hat{x}(N'-1)]$. Ce signal doit être le plus ressemblant possible au signal original$x = [x(0),..., x(N'-1)]$ [16].

Le décodeur dispose des mêmes dictionnaires fixe et adaptatif que l'encodeur ; le dictionnaire adaptatif est actualisé de la même manière qu'à l'encodeur. Le décodeur est relativement plus simple à mettre en œuvre [18] :

- les fréquences de raies spectrales (LSF) sont interpolées et reconverties en coefficients de filtre de Prédiction Linéaire pour chaque sous-trame de L échantillons,
- l'excitation est construite par combinaison des contributions adaptatif et fixe,
- le signal de parole est reconstitué par filtrage de l'excitation à travers le filtre de synthèse 1/A(z).

Un bloc de post-traitement est ajouté afin d'améliorer la qualité subjective du signal de parole synthétisé. Ce bloc comprend un post-filtre adaptatif composé de trois filtres en cascade [37] :

- un post-filtre long-terme (utilisé pour améliorer la périodicité du signal) qui nécessite l'estimation de la période du pitch,
- un post-filtre court terme (ou post-filtre de formant) utilisé pour améliorer la structure formantique du signal synthétisé,
- un filtre de compensation de l'inclinaison spectrale dû au filtrage passe bas causé par le post-filtrage court terme ; suivi par une procédure de contrôle adaptatif du gain (utilisé pour compenser la différence du gain entre le signal de parole synthétisé et le signal post-filtré).

3.3.2.1 Domaines d'application :

Ce codeur est utilisé dans plusieurs domaines tel que [36, 38, 39] :

- Les communications militaires sécurisées,
- Unité téléphonique sécurisée de troisième génération (STU 3),
- téléphonie acoustique sous-marine,
- Communications sans fil sécurisées.

3.4 Conclusion :

Dans ce chapitre nous avons présenté en détails le principe du codeur CELP à deux débits respectivement (le standard ACELP à 8Kbps et le FS1016 à 4.8Kbps). Le codeur CELP est basé sur le modèle LPC classique. Le codeur CELP contient deux dictionnaires à base de quantificateur vectoriel d'ordre 2. Le premier est de nature adaptative ; c'est le dictionnaire prédictif et le deuxième est de nature statique ; c'est le dictionnaire d'excitation statique. Les paramètres transmis par ce codeur sont les 10 paramètres LSF, l'indice du pitch, les indices des dictionnaires stochastique et adaptatif ainsi que les gains associés.

Chapitre 4 :

Implémentation et évaluation des codeurs CELP à 8Kbps et 4.8Kbps

4.1 Introduction :

La manière la plus fiable et la plus performante d'évaluer la qualité vocale des transmissions téléphoniques est de demander directement l'avis aux utilisateurs. Cependant, cette méthode est très coûteuse en temps de réalisation et en nombre de personnes à interroger. Des instruments de mesure ont été développés afin d'estimer automatiquement la qualité vocale perçue par un grand nombre d'usagers. Les instruments les plus performants sont normalisés à l'UIT (Union Internationale des Télécommunications) comme les modèles PESQ qui représentent une étape importante dans l'évolution de la technologie d'évaluation de la qualité [40, 41].

Notre objectif est de synthétiser la parole avec une bonne perception afin de comparer la synthèse de la parole à bas débit. Pour cela, nous avons utilisé les deux codeurs qui fonctionnent à 8 Kbps pour l'ACELP et FS1016 à 4.8 Kbps pour le CELP mis en œuvre dans le chapitre précédant.

Ce chapitre est divisé en deux parties essentielles. Dans la première partie on a simulé les deux codeurs CELP à 8 Kbps et 4.8 Kbps en utilisant le langage C (Builder C++ 5.0) pour la partie programmation et Matlab pour les représentations. La deuxième partie est consacrée pour la phase d'évaluation ; notre choix s'est porté sur les mesures PESQ comme mesures perceptuelles en raison de leur bonne corrélation avec les tests subjectifs.

Une comparaison de performance entre les deux codeurs CELP est effectuée avec la mesure objective PESQ en utilisant des échantillons extraits des bases de données de trois langues différentes (arabe, français et anglais). Ces bases de données vocales varient pour des voix masculine et féminine.

4.2 Présentation du logiciel d'évaluation PESQ :

Le PESQ (Évaluation perceptuelle de la qualité vocale) est un critère objectif d'évaluation de la qualité. Ce critère a un caractère perceptuel justement parce qu'il est fondé sur des notions psycho-acoustiques pour simuler notre perception vis-à-vis du signal de parole [13].

La plus ancienne méthode subjective acceptée internationalement était le MOS (Mean Opinion Score). Le MOS dépend normalement de demander aux utilisateurs de noter le système en testant de nombreux appels. Cela rend les méthodes objectives plus couramment utilisées. L'algorithme le plus largement utilisé est l'évaluation perceptuelle de la qualité de la parole. Le processus clé de PESQ, est la transformation des signaux originaux et dégradés en représentations psychophysiques proches de ceux des signaux auditifs du système auditif humain pour obtenir une écoute prédite à sens unique. Il est normalisé en tant que recommandation UIT-T P.862. Le PESQ compare signal d'entrée avec signal de sortie qui est passé par un système de communication et évalue l'équivalent MOS (score d'opinion moyen) [42].

Le score PESQ est produit sur une échelle similaire au MOS, avec des valeurs situées entre 0,5 et 4,5. Les valeurs habituelles sont entre 1,0 et 4,5, pour les scores MOS obtenus dans des expériences subjectives sur la qualité de l'écoute [41].

La relation entre les scores PESQ et la qualité audio est la suivante [41] :

- Des scores PESQ entre 3 et 4,5 désignent une qualité perçue acceptable (avec 3,8 comme seuil de la qualité dans les systèmes téléphoniques traditionnels) on va se référer à ce niveau comme qualité « très bonne » ;
- Des valeurs entre 2 et 3 indiquent qu'un effort est nécessaire pour la compréhension du parler on va se référer à ceci comme qualité « basse » ;
- Scores inférieurs à 2 signifient que la dégradation a rendu la communication très difficile ou même impossible, par conséquent la qualité est « inacceptable ».

- ***Algorithme PESQ :***

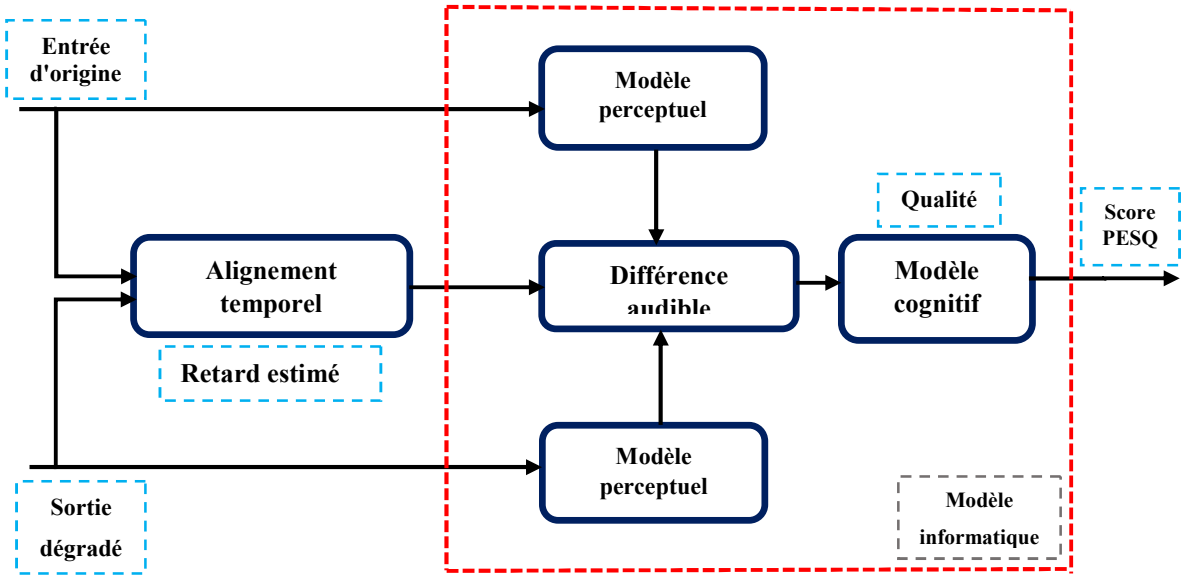


Figure 4.1 : le diagramme de base de l'algorithme PESQ

La première étape de l'algorithme PESQ consiste à utiliser une série de délais entre le signal d'origine et le signal dégradé pour obtenir le délai réel entre les deux. Cette étape s'appelle l'algorithme d'alignement temporel. L'algorithme peut gérer les changements de délai pendant le silence et la parole active. Sur la base de l'ensemble des retards détectés, le PESQ compare le signal d'origine avec le signal différé à l'aide d'un modèle informatique, comme illustré sur la figure 4.1 [42].

Le modèle informatique remplace le sujet (humain). Ce modèle est constitué de deux modèles. Le premier est le modèle de perception responsable de l'extraction des paramètres de la parole et le second est le « modèle cognitif » qui rend le jugement réel. Ces modèles permettent de comparer l'entrée et la sortie du périphérique testé. Le modèle PESQ transforme le signal dégradé et le signal d'origine correspondant en représentations internes, puis utilise leur différence pour calculer une note de qualité d'écoute [42].

4.3 Description de corpus de parole utilisé :

Les signaux de test ont été pris à partir de plusieurs bases de données dont la fréquence d'échantillonnage est de 8 KHz ; des échantillons extrait des bases de données de test TIMIT (Texas Instruments- Massachusetts Institute of Technology) et PAPE (Phrases Arabes Phonétiquement Equilibrées) respectivement pour les langues anglais et arabe. Pour la langue française des phrases ont étés prises d'une ancienne base de données mais restent toujours valables pour des tests de simulation. Les phrases sont prononcées par des hommes et des femmes.

La stratégie de test peut être résumée comme suit : Afin de tester ces codeurs nous avons utilisé un corpus de 3 langues : arabe, français et anglais qui est composé de phrases phonétiquement équilibrées. Les 20 premières phrases [enregistrés par plusieurs personnes] pour la langue arabe sont prononcées par 10 locuteurs féminins et 10 locuteurs masculins, pour la langue française on a pris 2 phrases prononcées respectivement par un locuteur masculin et un autre féminin. Enfin, les phrases pour la langue anglaise sont prononcées par 5 locuteurs masculins et 5 locuteurs féminins.

4.4 Résultats de simulation :

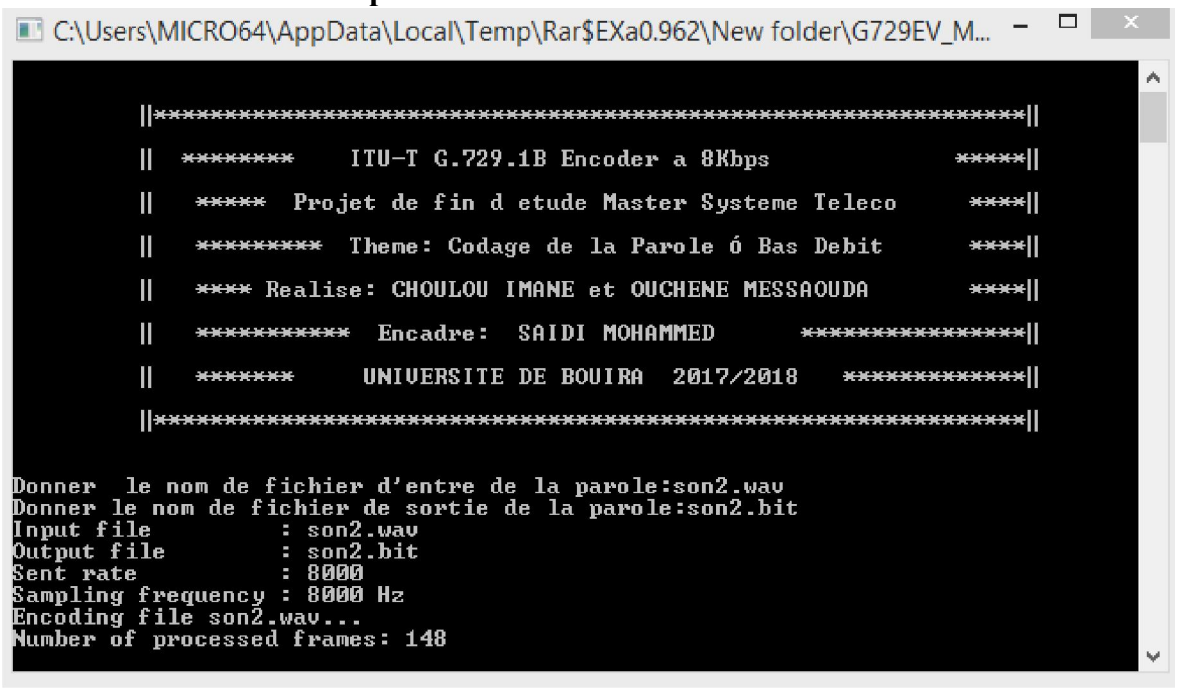
Nous commençons de premier principe par le Codec CELP à 8 Kbps suivi du deuxième à 4.8 Kbps, Pour ce faire nous avons simulé le corpus cité précédemment. Pour illustrer ce travail nous allons présenter la simulation de deux phrases prononcées par un locuteur et une locutrice pour les deux langues (Arabe et Français).

4.4.1 Codec CELP à 8 Kbps :

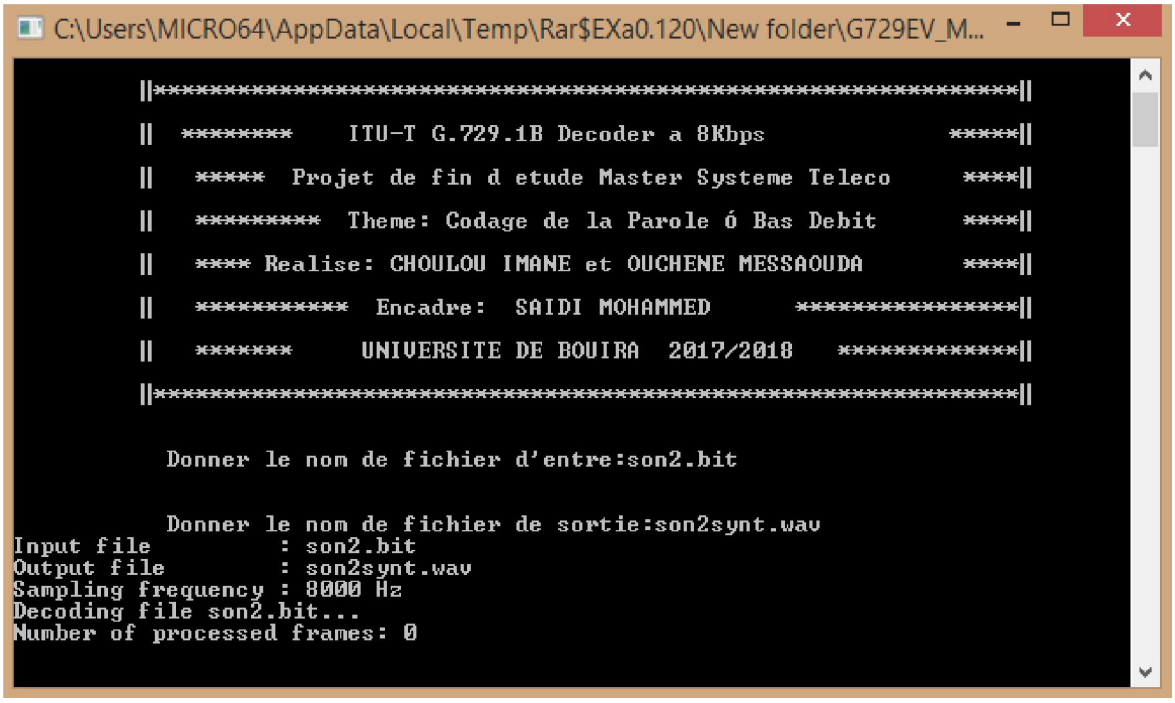
Les figures en dessous montrent le déroulement de la simulation pour ce codec réalisé avec le langage C. Suivi des figures représentant les signaux (original, synthétique et résiduel) sous Matlab.

En premier lieu, on a introduit un fichier son de type (.**wav**) enregistré dans la base de données cité précédemment. Pour avoir en sortie un flux binaire qui sera ensuite pris comme entrée du décodeur afin de restituer notre signal synthétique.

- **Codeur CELP à 8 Kbps :**



- **Décodeur CELP à 8 Kbps :**



1)- Langue Arabe :

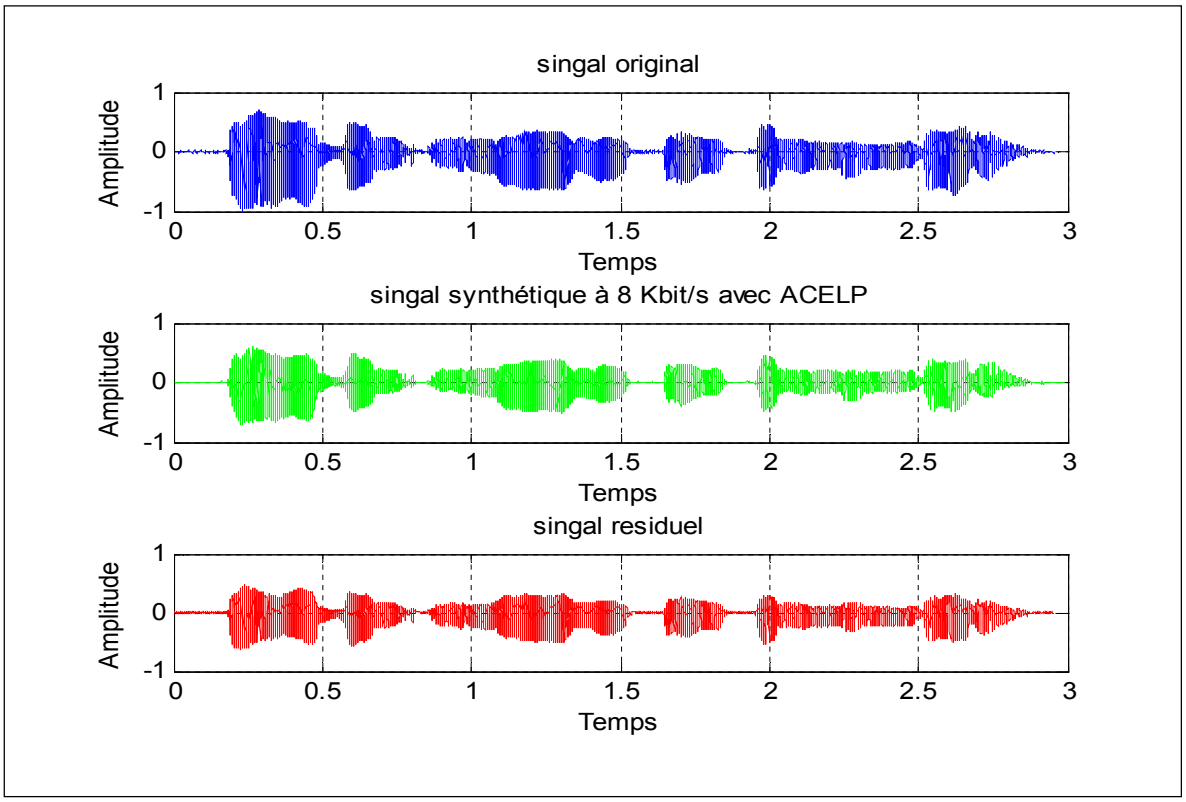


Figure 4.2 : Phrase prononcée par un locuteur " صعد الإمام فوق المنبر ".

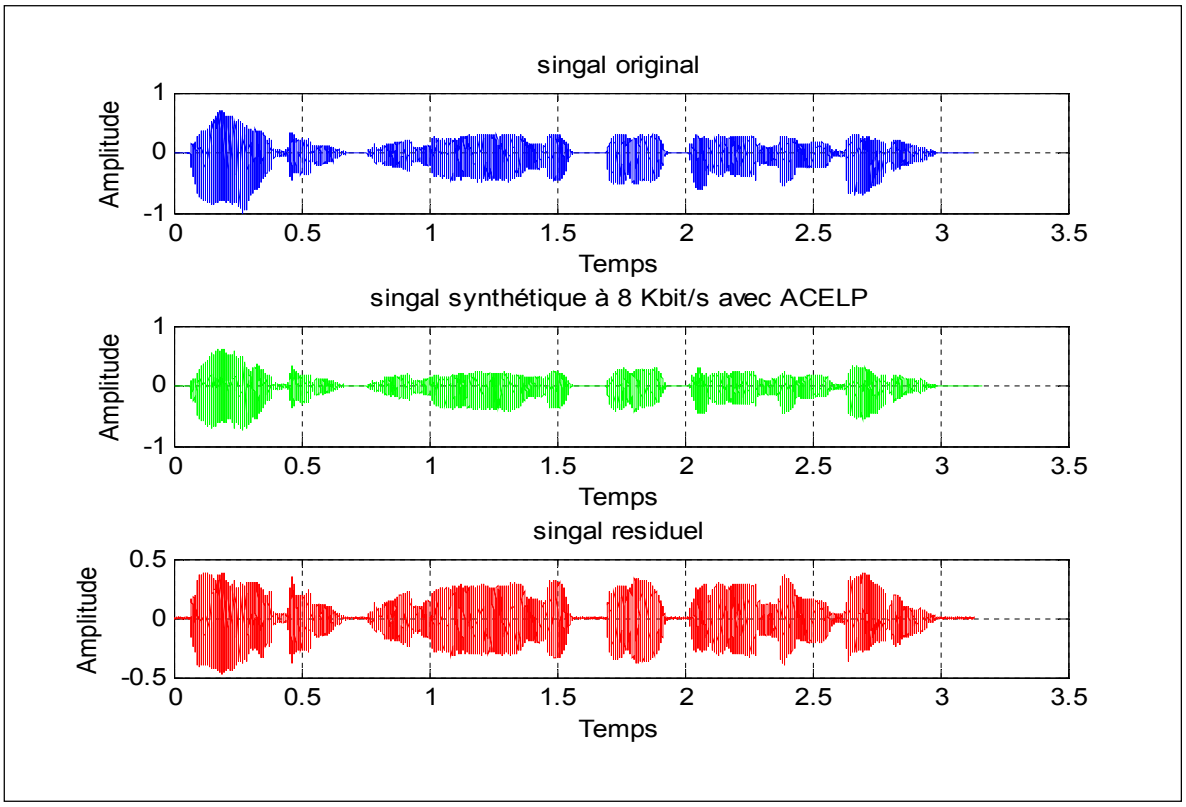


Figure 4.3 : Phrase prononcée par une locutrice " صعد الإمام فوق المنبر ".

2)- Langue Française :

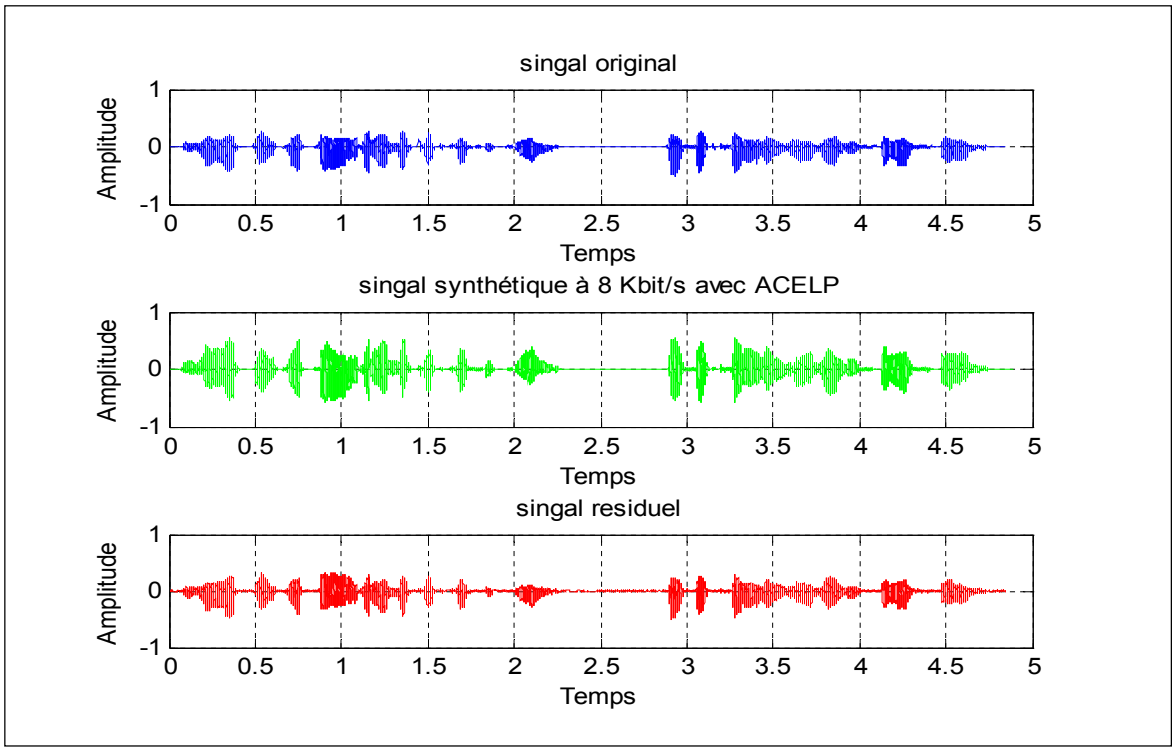


Figure 4.4 : Phrase prononcée par un locuteur "Je ne peux atteindre les bocaux de confiture dans cette crémerie on vend du fromage fort".

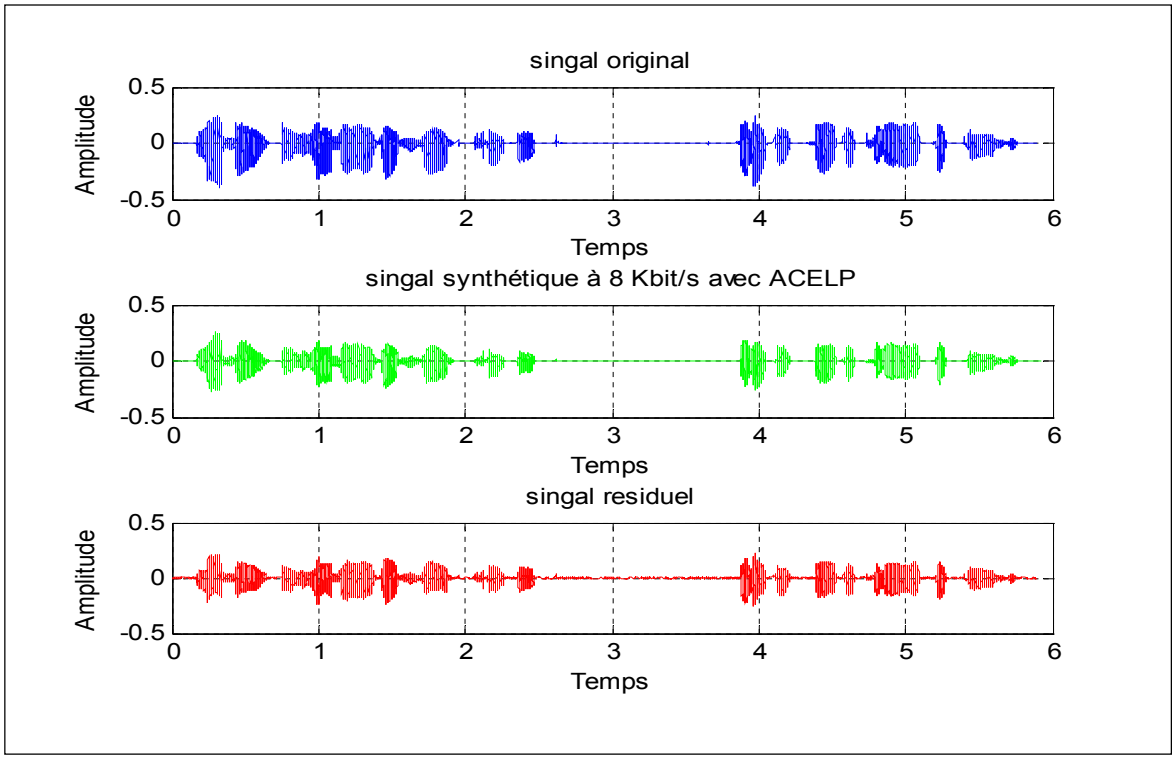
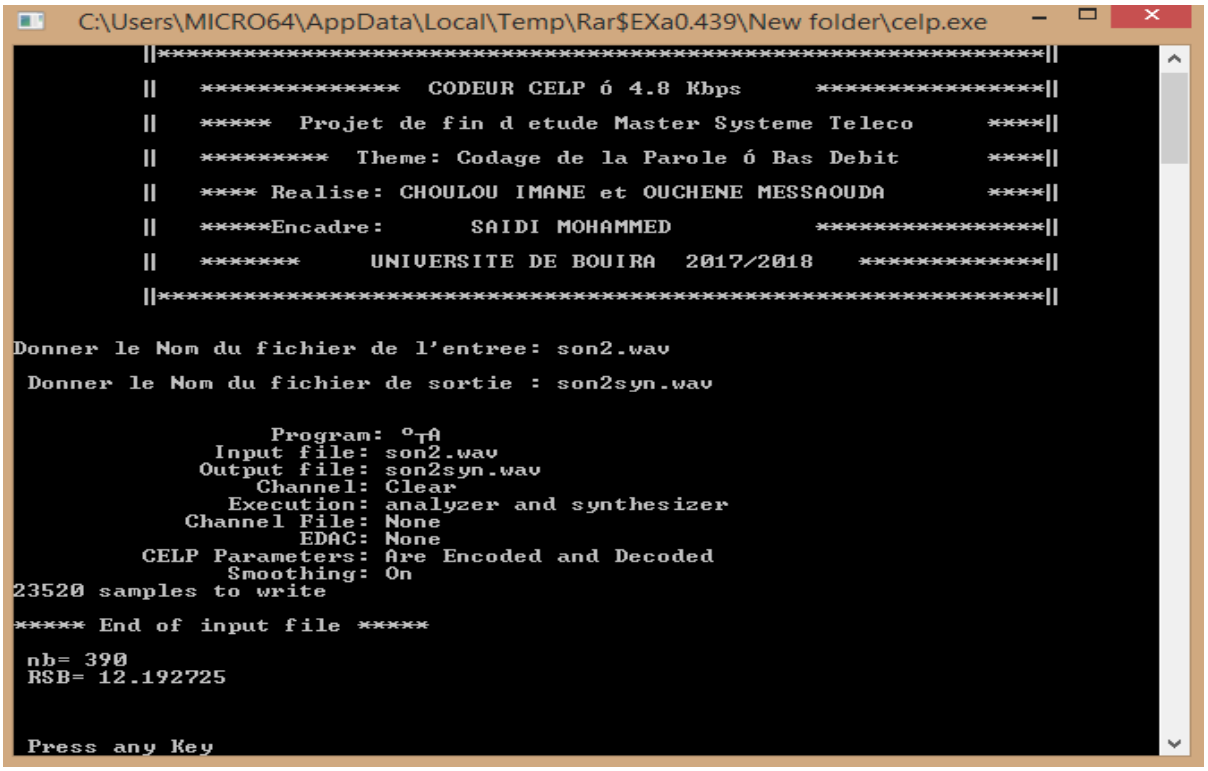


Figure 4.5 : Phrase prononcée par une locutrice "La bas il y a de mauvaises vagues très hautes c'est la question que tout le monde se pose".

4.4.2 Codec CELP à 4.8 Kbps :

Contrairement au Codec précédant, les opérations de codage et décodage sont faites en même temps.



1)- Langue Arabe :

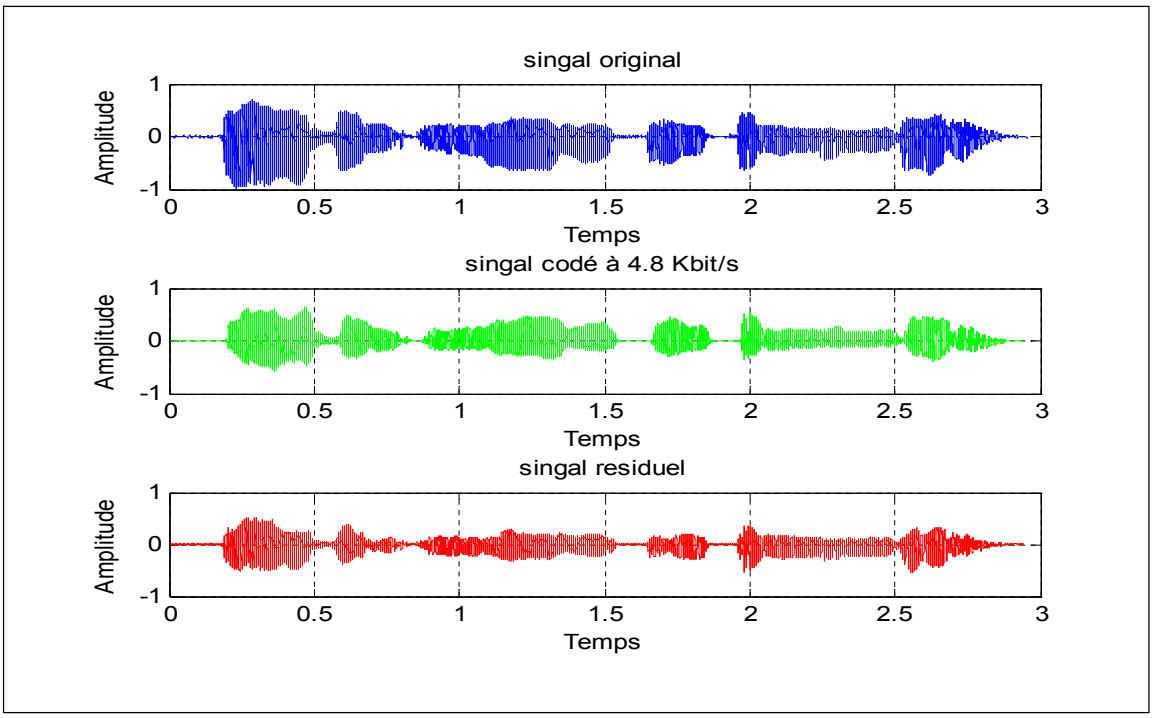


Figure 4.6 : Phrase prononcée par un locuteur"صعد الإمام فوق المنبر "

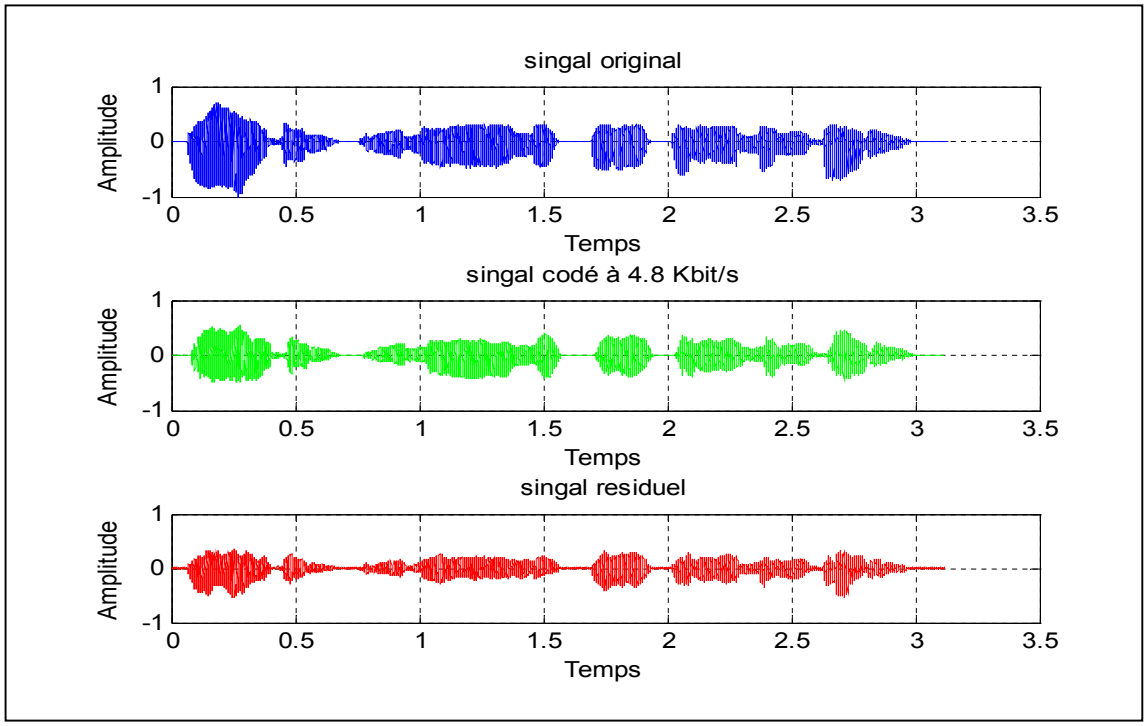


Figure 4.7 : Phrase prononcée par une locutrice " صعد الإمام فوق المنبر "

2)- Langue Française :

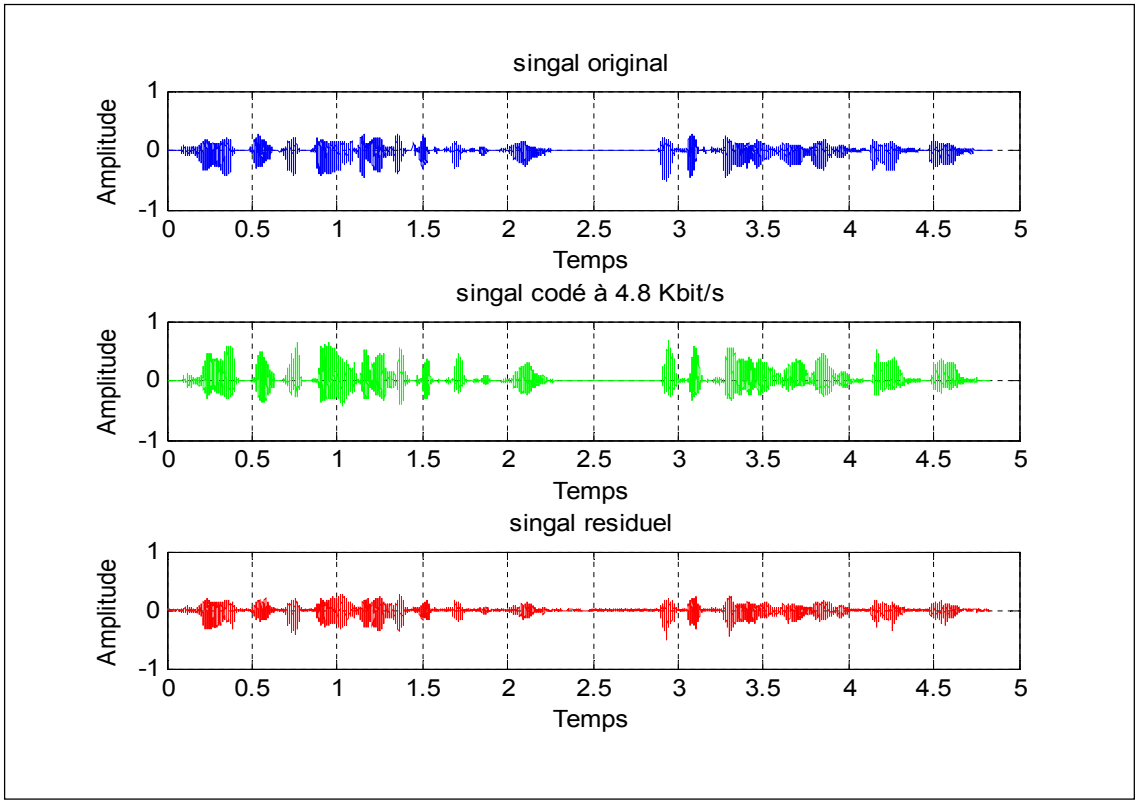


Figure 4.8 : Phrase prononcée par un locuteur "Je ne peux atteindre les bocaux de confiture dans cette crémerie on vend du fromage fort".

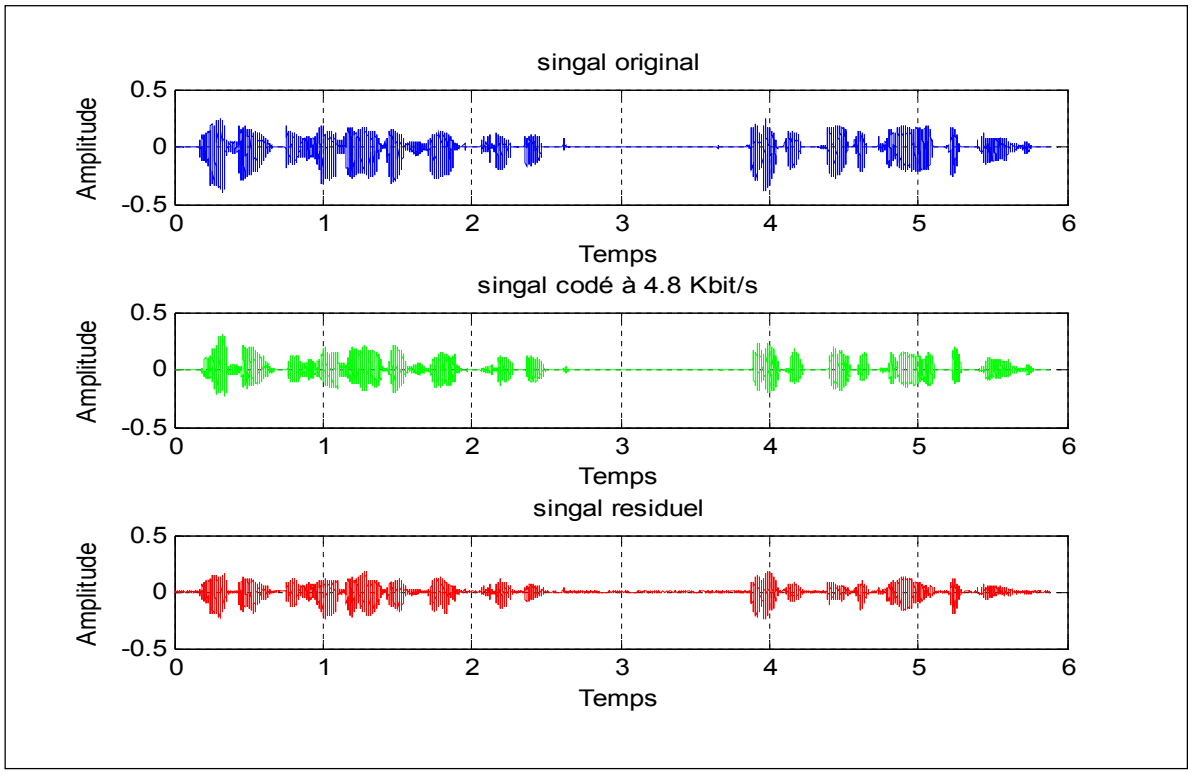
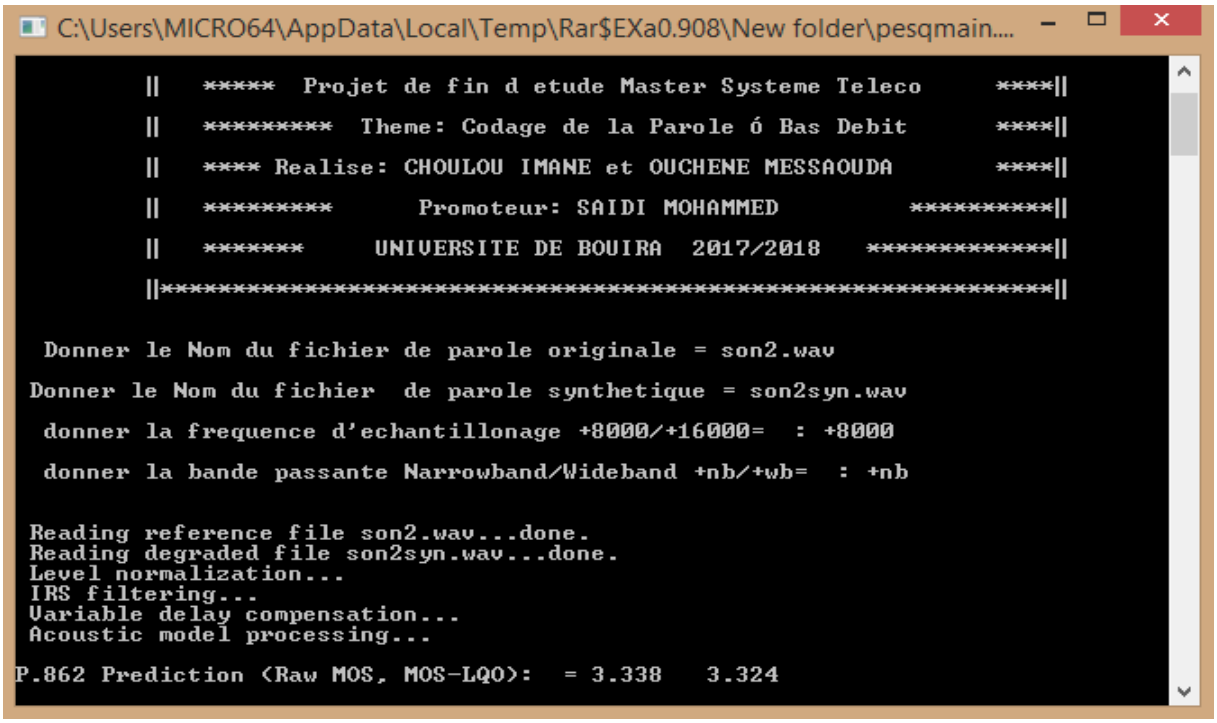


Figure 4.9 : Phrase prononcée par une locutrice "Là-bas il y a de mauvaises vagues très hautes c'est la question que tout le monde se pose".

4.5 Evaluation objective des Résultats :

Plusieurs simulations, ont été réalisées pour évaluer les performances de ces deux codeurs fonctionnant à 8 Kbps et 4.8 Kbps. Le but est de chiffrer la qualité perceptuelle de nos codeurs et d'évaluer leurs performances.

La figure ci-dessous présente l'exécution sous le logiciel PESQ. On a inséré un signal d'entrée original de type (.wav) et le signal synthétisé avec l'un des deux codecs (8 Kbps ou 4.8 Kbps) pour les comparer et avoir la note de qualité qui détermine la différence entre ces deux signaux. Sachant que la fréquence d'échantillonnage introduite est de 8 KHz (la bande étroite).



Les trois tableaux ci-dessous résument la moyenne du score PESQ pour les bases de données utilisées. Les mesures de ces notes ont été faites suivant cette exécution.

Tableau 4.1 : Scores PESQ pour la langue arabe.

PESQ CELP à 8 Kbps		**PESQ CELP à 4.8 Kbps**	
Masculin	3.970	Masculin	3.380
Féminin	3.794	Féminin	3.043

Tableau 4.2 : Scores PESQ pour la langue française.

PESQ CELP à 8 Kbps		**PESQ CELP à 4.8 Kbps**	
Masculin	3.556	Masculin	3.070
Féminin	3.263	Féminin	3.044

Tableau 4.3 : Scores PESQ pour la langue anglaise.

PESQ CELP à 8Kbps		**PESQ CELP à 4.8Kbps**	
Masculin	3.970	Masculin	3.173
Féminin	3.794	Féminin	3.036

D'après les résultats obtenus, nous constatons que :

- Le score PESQ obtenu est meilleur pour les locuteurs masculins que les locutrices féminines pour les deux codeurs ACELP et FS-1016 ;
- La qualité synthétique du codeur ACELP est meilleure que celle du FS-1016 ;
- Pour les deux codeurs, l'intelligibilité synthétique est assurée par ces deux codeurs avec un niveau assez bon.

4.5.1 Comparaison entre les deux codeurs :

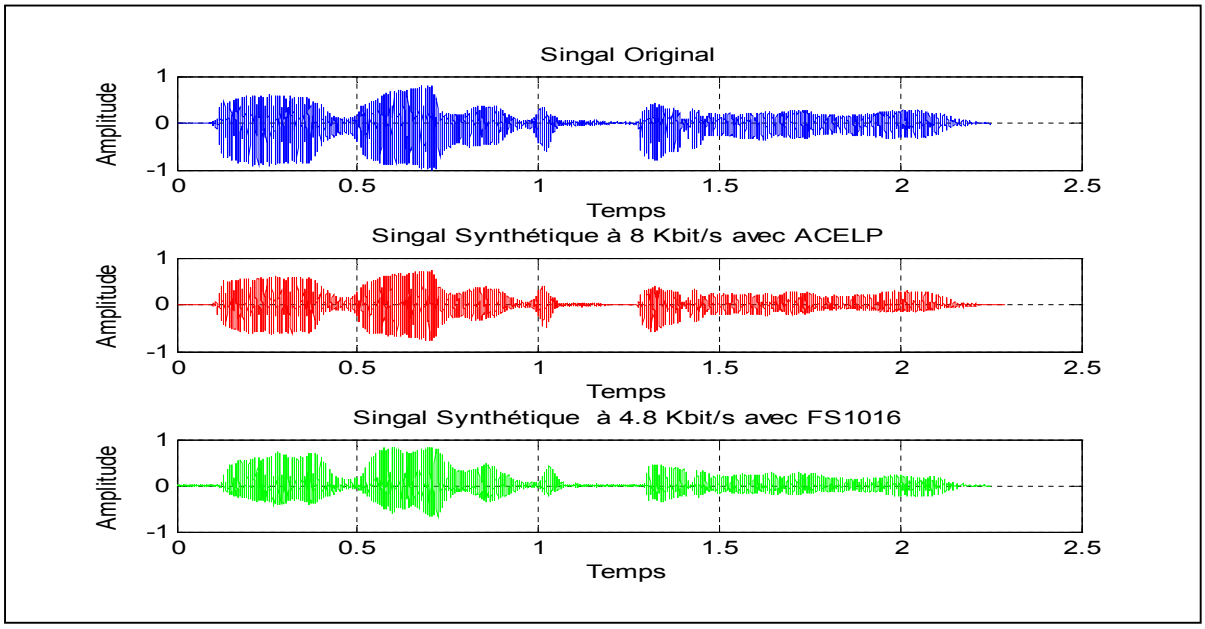


Figure 4.10 : Phrase prononcée par un locuteur « آذاه زحف رمله »

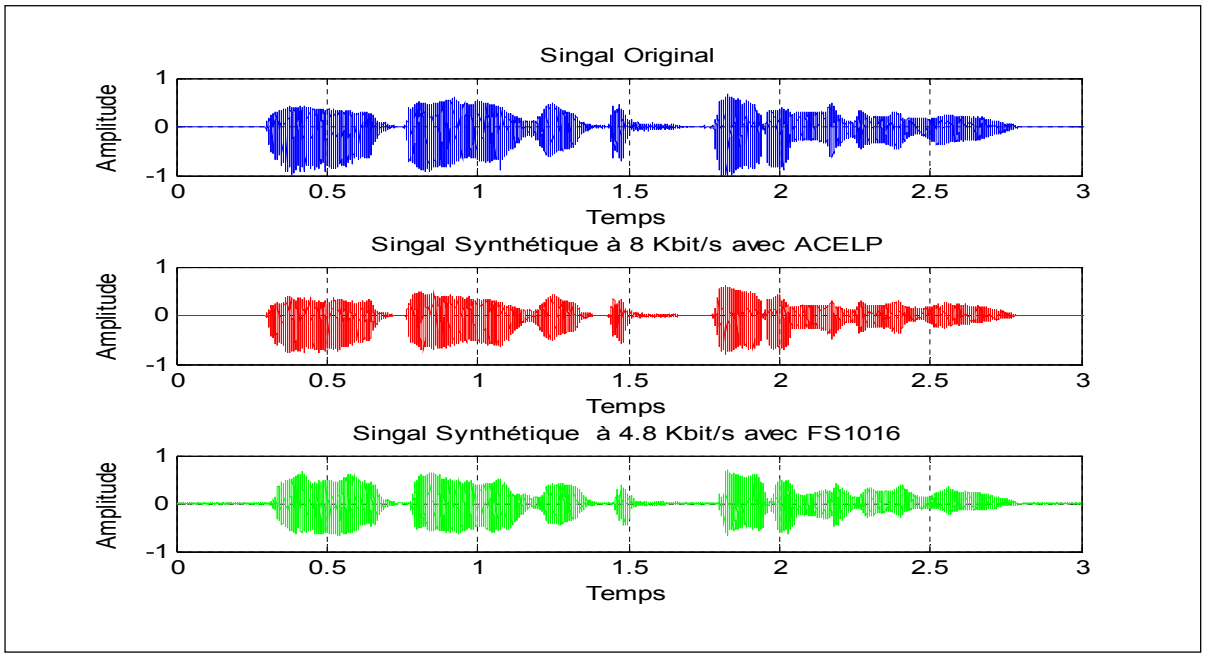
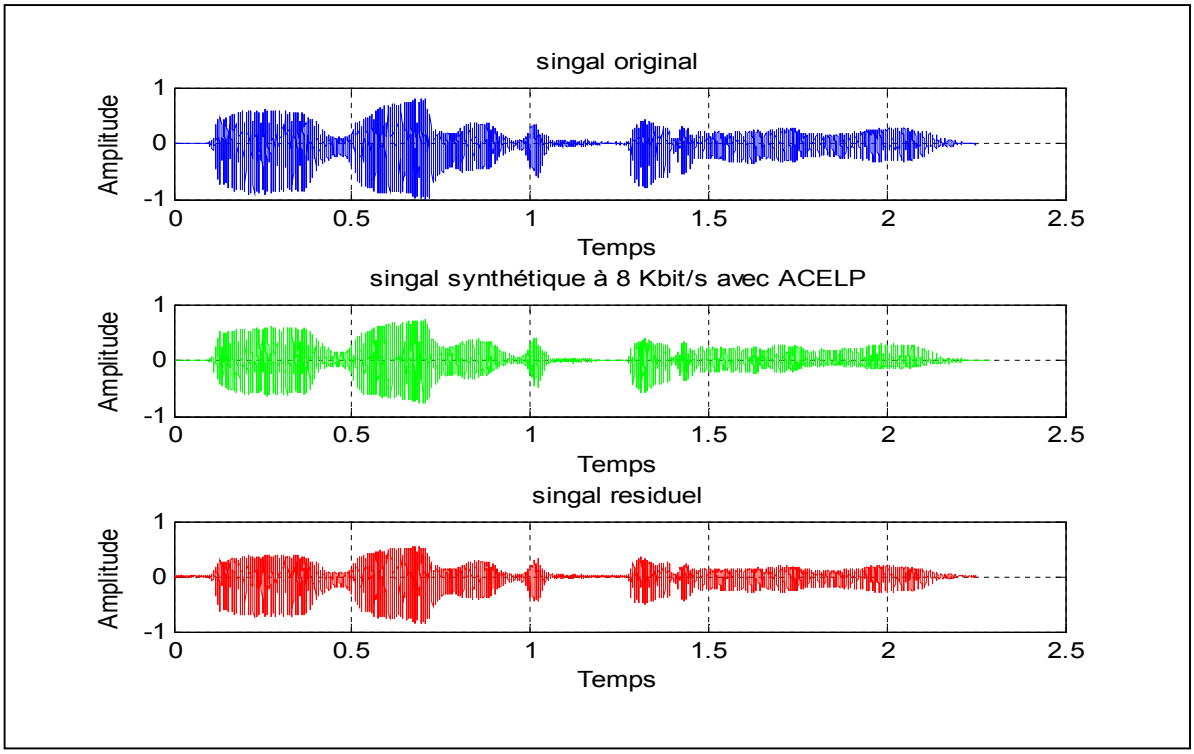
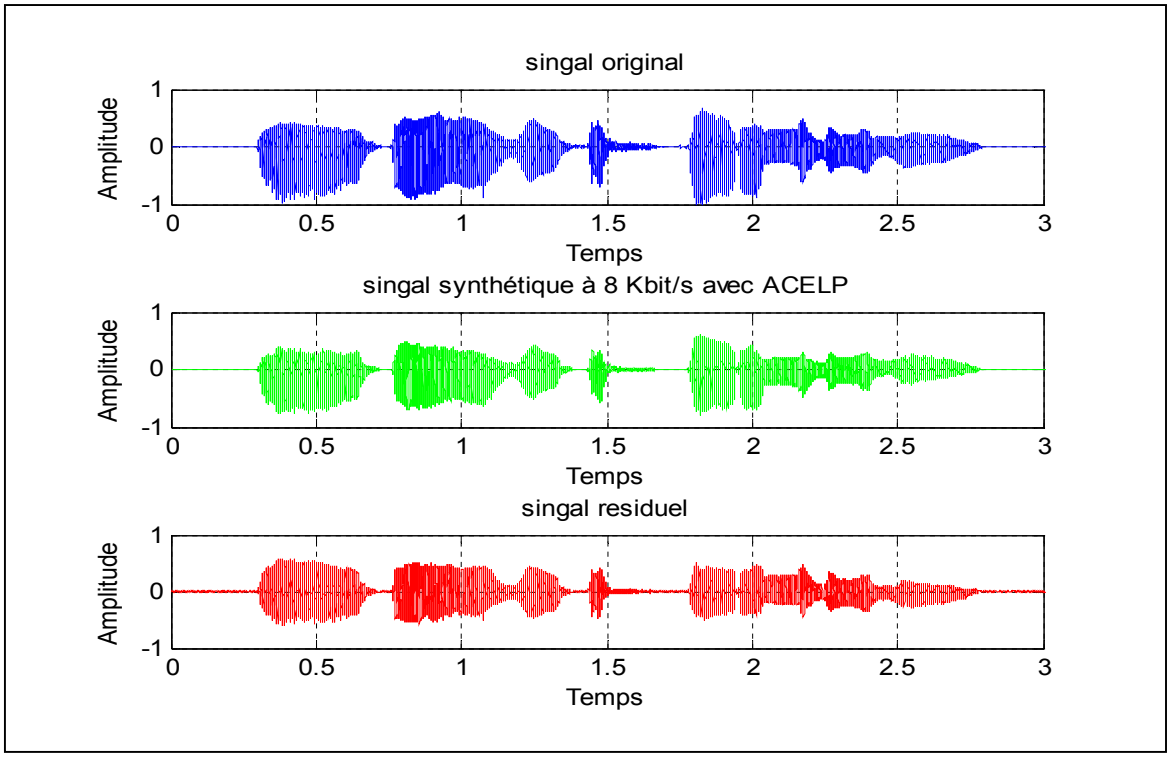


Figure 4.11 : Phrase prononcée par une locutrice" آذاه زحف رمله "

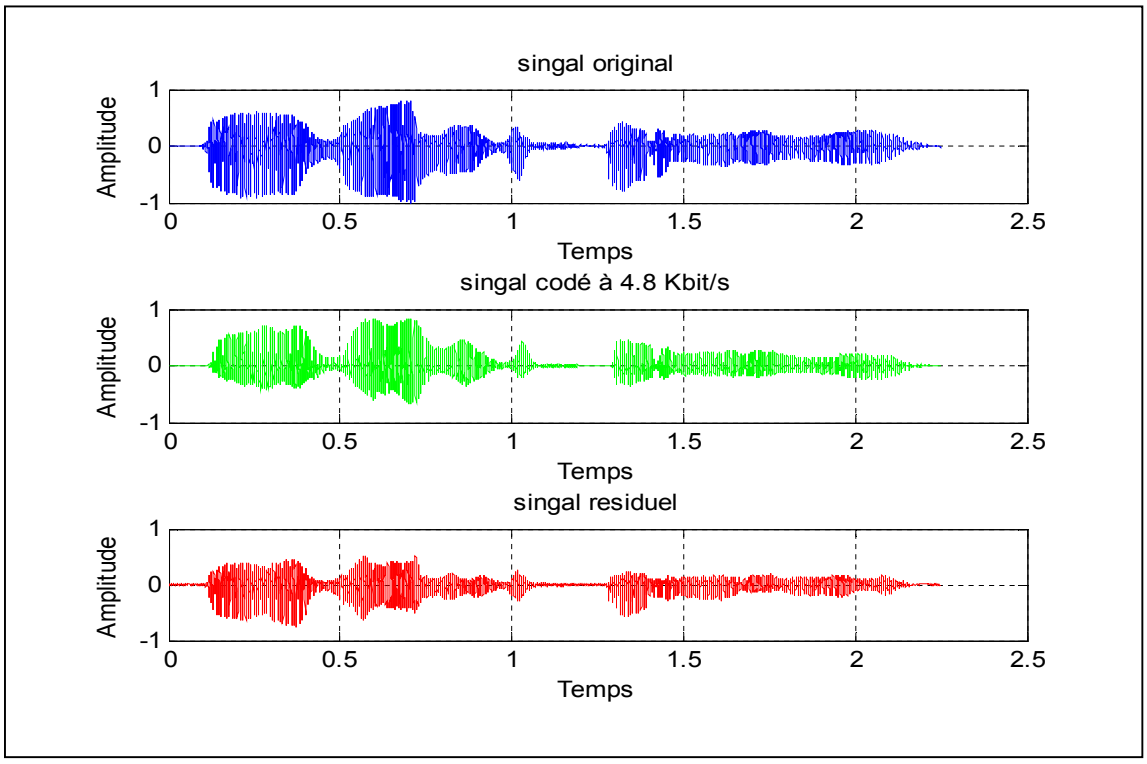
- ***Phrase prononcée par un locuteur " آذاه زحف رمله " à 8 Kbps :***



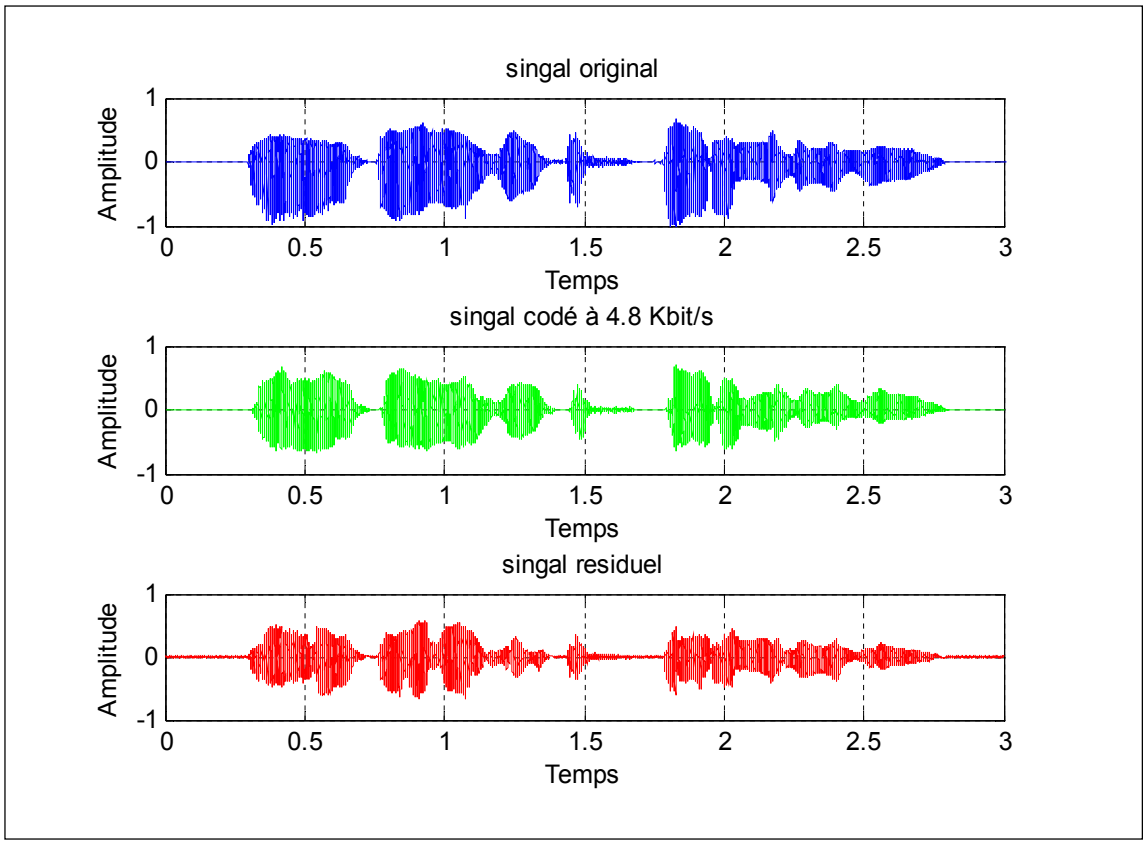
- ***Phrase prononcée par une locutrice "آذاه زحف رمله " à 8 Kbps :***



- ***Phrase prononcée par un locuteur " آذاه زحف رمله " à 4.8 kbps :***



- ***Phrase prononcée par une locutrice " آذاه زحف رمله " à 4.8 kbps :***



- ***Comparaison des résultats obtenus :***

D'après les graphes obtenus après les opérations de codage et décodage avec les standards ACELP et CELP respectivement à 8 Kbps et 4.8 Kbps on peut dire que :

- Les signaux résiduel et synthétique ont une grande ressemblance ; alors que le premier signal est la différence entre le signal d'entrée et le signal synthétique. Donc on peut conclure que la forme d'onde n'est pas considérée comme critère pour juger la qualité de la parole dans ce codage à bas débit.
- D'après les signaux synthétiques obtenus avec nos tests de simulation sur des fichiers audio de langues différentes ainsi que les bons scores PESQ des tableaux précédents, on peut dire que ce type de codec a la capacité de fonctionner sur une variété de langues et des locuteurs de sexe différents.
- L'analyse des tableaux des notes obtenus avec l'évaluation PESQ et la comparaison entre les signaux synthétiques à 8 Kbps et 4.8 Kbps permet de dire que la réduction du débit influe sur la qualité du signal décodé. Bien que l'ACELP à 8 Kbps soit plus performant, le CELP à 4.8 Kbps maintient une bonne note de qualité perceptuelle.

4.6 Conclusion :

Dans ce chapitre nous avons présenté une évaluation de deux codeurs ACELP et CELP. Les codeurs de type CELP sont des systèmes de codage mixtes qui utilisent à la fois des représentations paramétriques et temporelles du signal de parole. Notre objectif était focalisé envers l'étude des codeurs CELP à 8 et 4.8 Kbps, et afin de tester et comparer leurs performances pour un ensemble de langues, nous avons effectué une implémentation software que nous avons testé avec une base de données phonétiquement équilibré.

Etant donné que les deux codeurs appartiennent à la famille de codeurs hybrides qui préservent la forme d'onde, on se référence sur un fort critère d'évaluation qui est le PESQ. Les différents résultats d'évaluation objective avec le PESQ ont montré que les deux codeurs appartiennent à la bonne catégorie de qualité perceptuelle mais le G.729 à 8 Kbps offre de meilleurs résultats que le FS1016 à 4.8 Kbps.

Conclusion générale

Conclusion générale

Les problèmes critiques qui constituent des contraintes dans les communications sans fil, sont la bande passante, la mémoire de stockage et l'alimentation. L'objectif dans le codage de la parole est associé à la réduction des informations supplémentaires présentes dans le signal afin de représenter le signal vocal avec un nombre restreint de bits tout en mettant à jour sa qualité perceptuelle. Pour cette raison, le codage de la parole à bas débit est et sera la question de recherche la plus importante.

Afin de réaliser un codage performant permettant de préserver la qualité, il est nécessaire de prendre en considération certains points essentiels tels que : la complexité des algorithmes de codage, le débit binaire et l'intelligibilité de la parole.

Dans ce travail nous avons implémenté deux standards de codage de la parole à bas débit, le premier codeur est le G.729 (ACELP) à 8Kbps et le second le FS1016 (CELP) à 4.8 Kbps dans le but de comparer la synthèse de la parole obtenue par ces deux types codeurs. Nous avons simulé ces deux codeurs avec le langage C (Builder C++ 5.0) pour la partie programmation et Matlab pour les représentations. Comme nous nous sommes intéressés à la comparaison entre les deux codeurs ACELP et CELP, on a choisi de se référer à un fort critère d'évaluation objective qui est le «PESQ » afin de juger les performances de chacun d'eux.

Les résultats obtenus montrent que les deux codeurs appartiennent à la bonne catégorie, avec des scores PESQ désignant une qualité perçue acceptable, mais la comparaison entre les signaux synthétique à 8 Kbps et 4.8 Kbps nous permet de dire que la réduction du débit à impacter la qualité du signal de sortie. D'où on conclut que le codeur ACELP à 8 Kbps est le meilleur.

Avant de terminer cette conclusion, on peut dire que ce travail nous a apporté d'immenses intérêts tant sur le plan théorique que expérimental. En effet, de plus qu'il nous a permis de nous approfondir dans la simulation et la programmation, nous avons été amenés, par ce travail, à nous instruire dans un domaine clé dans les télécommunications et les applications multimédias, c'est le codage et le traitement de la parole.

Nous avons, aussi pu nous familiariser avec un domaine d'actualité c'est le codage de parole à bas débit, nous avons vu les différentes classes des codeurs de parole et en détails le codeur CELP, les Mesures d'évaluation des performances dans le codage de la parole. Notamment, on a mis en œuvre les codecs ACELP et CELP opérant respectivement à des débits de 8 et

4.8 Kbps. Enfin, en implémentant ces codeurs on a pu les évaluer avec la mesure PESQ en utilisant un corpus de parole de langues différentes.

Perspectives

Comme perspectives, l'ensemble des codeurs présenté a été programmé en langage haut niveau C/C++, nous souhaitons que cette contribution serve de plate-forme pour de futurs travaux dans le domaine de communication numérique. Particulièrement, pour ceux qui auront objectif d'intégrer le système de codage sur une carte DSP en vue d'une évaluation en temps réel de toutes leurs performances.

Bibliographie

[1] : J. Hernández, «Algorithmes d'acquisition, compression et restitution de la parole à vitesse variable. Etude et mise en place», Projet de fin d'études, Spécialité : Informatique, École Nationale Supérieure de l'Électronique et de ses Applications (ENSEA) Cergy-Pontoise, Paris, Avril 1995.

[2] : R. Boite, H. Bourlard, T. Dutoit, J. Hancq & H. Leich, « *Traitement de la parole*», 1ère édition, Presses Polytechniques et Universitaires Romandes, pp.4-6, Janvier 2000.

[3] : https://www.speex.org/docs/manual/speex-manual/node4.html, (consulté le 7 septembre 2018).

[4] : M. Medjber, « Amélioration du standard G.729 -8Kb/s par la méthode de modification de l'échelle temporelle (WSOLA) », Mémoire de Magister, Option : signal et communications, ENP, 2007.

[5] : A. Le Guyader, P. Philippe & J. B. Rault, « *Synthèse des normes de codage de la parole et du son (UIT-T,* ETSI *ET ISO/MPEG)* », Vol. 55, issue n° 9-10, pp 425–441, 2000.

[6]: Mohammad M. A. Khan, « Coding of Excitation Signals In a Waveform Interpolation Speech Coder », Thesis of Master of Engineering, Department of Electrical & Computer Engineering, McGill University Montreal, Canada, July 2001.

[7]: ITU-T Recommendation G.711, « Pulse code modulation (PCM) of voice frequencies », November1988.

[8]: ITU-T Recommendation G.726, « 40, 32, 24, 16 Kbit/s adaptive differential pulse code modulation (ADPCM) », 1990.

[9]: ITU-T Recommendation G.722, «7 kHz audio-coding within 64 Kbit/s », 1990.

[10]: ITU-T Recommendation G.723.1, « Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 Kbit/s », 1996.

[11] : M. Ladj & M. Larbi, « Extraction des formes d'ondes caractéristiques dans le codeur de la parole par interpolation de formes d'ondes », Projet de fin d'études, ENP, Alger, 2006.

[12] : A. Khelfa, D. Chikouche, K. Rouabah & N. Amardjia, "*Contribution à l'amélioration des signaux parole synthétisés par LPC en utilisant l'information de phase*", Département d'Electronique, Université de Sétif, pp.165-168, Algérie.

[13] : A. Amehraye, « Débruitage perceptuel de la parole », Thèse de doctorat, Option : Traitement du Signal et Télécommunications, Ecole Nationale Supérieure des Télécommunications de Bretagne, Mai 2009.

[14] : A. Ben Aicha, « Réduction du bruit musical et évaluation de la qualité des signaux débruités par approches perceptuelles », Thèse de doctorat, Option : technologie de l'information et de la communication, École Supérieure des Communications de Tunis, Février 2010.

[15] : M. Jelinek, « Modélisation spectrale et compression de la parole à bas débit », Thèse de doctorat, Université de Sherbrooke, Spécialité : génie électrique, Sherbrooke (Québec), Canada, Octobre 1998.

[16] : M. Bouzid, « Codage Conjoint de Source et de Canal pour des Transmissions par Canaux Bruités », Thèse de doctorat en électronique, Université des Sciences et de la Technologie Houari Boumediene (USTHB), Spécialité : Communication Parlée, Alger, Algérie, 2006.

[17] : P. Jardin, « Codage de source », ESIEE : Signaux et Télécommunications, 03 octobre 2008.

[18] : M. Djamah, « Codage échelonnable à granularité fine de la parole utilisant la quantification vectorielle arborescente », Thèse de Doctorat en Télécommunications, INRS (Centre Énergie Matériaux Télécommunications), Université du Québec.

[19] : M. de Meuleneire, « Codage imbriqué pour la parole à 8-32 Kbit/s combinant techniques CELP, ondelettes et extension de bande » ; Thèse de doctorat, École Nationale Supérieure des Télécommunications de Bretagne, 2007.

[20] : A. Guerid & A. Houacine, « CELP Coder Modification for the Voice Conversion »; International Journal of Signal Processing Systems, Vol. 4, No. 2, pp. 124-127, April 2016.

[21]: W. Li, A. Sridhar & T. Teng; « Comparison of Speech Coding Algorithms: ADPCM, CELP and VSELP»; Project for EE 6390, University of Texas at Dallas, fall 1999.

[22] :A. Isar, A. Cubiţchi & M. Naforniţă ; « Algorithmes et techniques de compression », Programme d'action de soutien à la formation et à la recherche 2000-PAS-32, Edition Politehnic Orizonturi, 2002.

[23] : M. D. Kwong, « Transmission efficace en temps réel de la voix sur réseaux ad hoc sans fil », Thèse de doctorat, Spécialité : Génie électrique, Université de Sherbrooke, Québec, Canada, Juillet 2008.

[24] : N. Moreau, « Outils pour la compression. Application à la compression des signaux audio », Télécom Paris Tech, 26 novembre 2007.

[25] : P. Dymarski, N. Moreau & W. Vos, « Détermination et Codage de L'excitation dans un Codeur CELP », Treizième Colloque Gretsi – Juan- Les- Pins, pp. 725-728, 16 au 25 Septembre 1991.

[26] : G. Baudoin, « Codage de la Parole à Bas et Très bas débit Transformation de la voix », Mémoire d'habilitation à diriger des recherches, Université Mame la Vallée, 2000

[27] : J. Černockŷ & V. Hubeika, « Speech Coding II », Department of Computer Graphics and Multimedia (DCGM), Brno University of technology, Czech Republic.

[28]: http://www.gaoresearch.com/products/speechsoftware/other/g729.php, (consulté le 23 juillet 2018).

[29] : http://what-when-how.com/voip/g-729-family-of-low-bit-rate-codecs-voip/, (consulté le 24 juillet 2018).

[30] : https://www.vocal.com/speech-coders/g-729/, (consulté le 23 juillet 2018).

[31]: A. Spanias, T.Painter & V. Atti, «Audio signal processing and coding», John Wiley & Sons, Inc., September 2006.

[32] : A. BOUKHARI, « Quantification des coefficients LSF. Application au codage CELP », Mémoire de Magister, Université des Sciences et de Technologie Houari Boumediene (USTHB), 2005.

[33] : M. Saidi, L. Falek & B. Boudraa, « Codage par prédiction linéaire (LPC) à bas et très bas Débit », Centre de Recherche Scientifique et Technique pour le Développement de la Langue Arabe (CRSTDLA), Laboratoire de Communication Parlée et de Traitement du Signal (LCPTS), USTHB ; Alger, Algérie.

[34]: J. Makhoul, « Linear Prediction: A Tutorial Review », Proc. of the IEEE, Vol. 63, No. 4, April 1975.

[35] : M. Djamah, M. Boudraa, B. Boudraa & M. Bouzid ; « Un logiciel de codage de la parole basé sur le FS1016 », Laboratoire Communication Parlée et Traitement de Signal (LCPTS), Faculté Génie-Electrique, USTHB, Alger, Algérie, Juin 2002.

[36]: R. Fenichel & D. Bodson, « Details to assist in implementation of Federal Standard 1016 CELP», Technical Information Bulletin 92-1, National Communication system, January 1992.

[37]: W. C. Chu, «Speech Coding Algorithms: Foundation and Evolution of Standardized Coders», John Wiley & Sons, Inc., New Jersey, 2003.

[38] : A. Goalic, « Traitements temps réel en codage source et canal pour des Communications hertziennes et acoustiques sous-marines », Habilitation à Diriger des Recherches, Université de Bretagne Occidentale, Bretagne.

[39]: L. Sun et al., « Guide to Voice and Video over IP, Computer Communications and Networks », DOI10.1007/978-1-4471-4905-7_2, Springer-Verlag, London, 2013.

[40] : A. Leman, « Diagnostic et évaluation automatique de la qualité vocale à partir d'indicateurs hybrides », Thèse de doctorat, Spécialité : Acoustique, Ecole doctorale des Sciences pour l'Ingénieur de Lyon, juin 2011.

[41] : A. Cheradid, « Étude des approches adaptatives liées à la QoE dans le cadre des applications de Téléphonie sur IP », Mémoire de Magistère en Informatique, Option : Technologies de l'Information et de la Communication, Université Kasdi Merbah, Ouargla, 2014.

[42] : M. Nasief, N. Messiha & H. Mansour, «The Effect of the Spoken Language on the Linear Prediction Vector Quantization Distortion for Linear Prediction Coders», Faculty of engineering, Cairo, Egypt, January 2013.

Résumé :

Le codage de parole à bas débit est un axe de recherche très important dans le domaine de compression de la parole qui tend à réduire le débit d'information à condition que le signal ne soit pas dégradé et que le coût du traitement reste raisonnable. Les codeurs par prédiction linéaire excitée par code "CELP" sont classés parmi les meilleurs codeurs de parole à bas débit, qui offrent de bonnes performances à un débit aussi faible que 4.8Kbps.

Ce manuscrit aborde la mise en œuvre de deux codeurs CELP ; le G.729 à 8 Kbps et le standard FS1016 à 4.8 Kbps. Ces codeurs sont évalués en termes de qualité pour différents types de langues et de locuteurs en utilisant le PESQ.

Mots clés : CELP, G.729, Codage de la parole, Codage à bas débit.

Abstract:

Critical problems that constitute constraints in the transmission of speech are bandwidth and storage memory. Low bit rate speech coding is a very important search axis in the speech compression domain, which tends to reduce the information rate provided so that the signal is not degraded and the cost of the processing remains reasonable. "CELP" code-excited linear prediction coders are among the best low-bit speech coders, delivering good performance at as low as 4.8 kbps.

This manuscript addresses the implementation of two CELP coders, the G.729 at 8 kbps and the FS1016 standard at 4.8 kbps. These coders are evaluated in terms of quality for different types of languages and speakers using the PESQ.

Key words: CELP, G.729, speech coding, low bit rate coding.