الجمهورية الجزائرية الديمقراطية الشعبية

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DR. YAHIA FARES DE MEDEA

FACULTE DE TECHNOLOGIE

Département du Génie Électrique

Projet de Fin d'Études

Présenté pour l'obtention du Diplôme
de Master en Génie Électrique

Spécialité : Système de télécommunications

THEME

Développement d'une méthode stéganographique basée sur l'interpolation polynomiale des coefficients LSFs sur le codeur G.723.1 de la VoIP

Préparé et présenté par :

- BACHFERRAG Mohamed.
- OULD RAMOUL Mohamed Amine.

Proposé et dirigé par :

- M. KHEDDAR Hamza.

Année Universitaire 2018-2019

REMERCIMENT

En premier lieu, nous remercions Dieu qui nous a procuré ce succès.

Nous tenons à remercier vivement notre encadreur Mr. Kheddar Hamza.

Nos remerciements les plus vifs s'adressent aussi aux messieurs le président et les membres de jury d'avoir accepté d'examiner et d'évaluer notre modeste travail.

Nous exprimons également notre gratitude à tous les professeurs et enseignants qui ont collaboré à notre formation depuis le début de nos études jusqu'à la fin de notre cycle universitaire. Sans omettre bien sûr de remercier profondément tous ceux qui ont contribué de près ou de loin à la réalisation du présent travail

Et enfin, que nos chers parents et familles, et bien avant tout, Trouvent ici l'expression de nos remerciements les plus sincères et les plus profonds en reconnaissance de leurs sacrifices, aides, soutien et encouragement afin de nous assurer cette formation de master dans les meilleures conditions.

DEDICACES

Grace à Dieu que j'ai achevé ce travail que je dédie à :

A mon très cher père, source d'amour. D'affection, de générosité et de sacrifices. Tu étais toujours la près de moi pour me soutenir, m'encourager et me guider avec tes précieux conseils. Que ce travail soit le témoignage des sacrifices que vous n'avez cessé de déployer pour mon éducation et mon instruction. Aucune dédicace ne saurait exprimer l'amour et l'admiration que je porte au grand homme que vous êtes. Puisse Dieu le tout puissant vous préserver et vous accorder santé, longue vie et bonheur.

A ma très chère mère source d'amour et de tendresse qui n'a pas cessé de m'encourager et de prier pour moi. Vous m'avez toujours aidé par vos conseils et vos sacrifices. Puisse Dieu le tout puissant l'accorder meilleure santé et longue vie.

A ma sœur et mes deux frères pour leurs patiences, soutiens et leurs sentiments d'amour aux moments les plus difficiles. Je vous souhaite plein de succès, de joie et de bonheur. Que dieu vous garde et illumine vos chemins.

A mon meilleur amis mon binôme : Hamada, merci pour les très bons moments qu'on a partagés ensemble. Sans oublier mes collègues de la promotion 2019.

A tout ceux qui m'ont aidé, encouragé, apprécié mon effort et crée le milieu favorable, l'ambiance joyeuse et l'atmosphère joviale pour me procurer ce travail.

AMINE

DEDICACES

À MES CHERS PARENTS

Aucune dédicace ne saurait exprimer mon respect, mon amour
éternel et ma considération pour les sacrifices que vous avez consenti pour mon
instruction et mon bien être.
Je vous remercie pour tout le soutien et l'amour que vous me portez
depuis mon enfance et j'espère que votre bénédiction m'accompagne toujours.
Que ce modeste travail soit l'exaucement de vos vœux tant
formulés, le fruit de vos innombrables sacrifices, bien que je ne vous en acquitte jamais assez.
Puisse Dieu, le Très Haut, vous accorder santé, bonheur et longue
Vie et faire en sorte que jamais je ne vous déçoive.

A MES CHERS ET ADORABLE FRERES ET SOEURS

Nassima, la prunelle de mes yeux, Narimene, la douce, au cœur
si grand, Tahar le généreux, Walid mon petit
frère que j'adore, que j'aime profondément.
En témoignage de mon affection fraternelle, de ma profonde
tendresse et reconnaissance, je vous souhaite une vie pleine de bonheur et de succès et que
Dieu, le tout puissant, vous protège et vous garde.

A MON MEILLEUR AMIS MON BNIMONE

Amine, merci pour les très bons moments qu'on a partagés ensemble. Sans oublier mes
collègues de la promotion 2019.

À toutes les personnes qui ont participé a
l'élaboration de ce travail à tous ce que j'ai omis de citer.

MOHAMED

ملخص

تطورت الشبكات الرقمية في خضم السنوات الأخيرة أصبحت حتمية لا يمكن الاستغناء عنها فيما يخص الاتصالات المعاصرة، بث إشارات في هذه الشبكات يحتمل مشاكل عدة والتي لم تحل إلى حد الساعة وكمثال لهذه المشاكل أمن البث، الخصوصية، حالة الشيء المرسول، دقة الإرسال. الهدف من مشروعنا هو دراسة انشاء طريقة اخفاء المعلومات مرتكزة على المشفر G.723.1 على الإشارة الصوتية. هذه الطريقة تستخدم لإخفاء المعلومة في إشارة الكلام، تعرف باسم الستيغانوغرافيا.

الكلمات المفتاحية

كريبتوغرافيا، تشفير، ستيغانوغرافيا، نظام وشم، نظام بث ستيغانوغرافي، غلاف، نطاق التردد، ملف ستيغو، نظام السمع البشري، نوعية إدراك الكلام. معاملات LSF، معلمات LP.

Résumé

Les réseaux numériques ont fortement évolué ces dernières années et sont devenus inévitables pour les communications modernes. La transmission des signaux sur ces réseaux soulève un nombre important de problèmes qui ne sont pas encore tous résolus comme par exemple, la sécurité, la confidentialité, l'intégrité et l'authenticité des signaux audio pendant leur transmission. Le but de notre projet est le développement et l'implémentation d'une méthode stéganographique sur les signaux audios. Cette méthode est implémentée sur le codeur G.723.1 pour cacher des données dans un signale parole, elle est connue sous le nom : stéganographie de la parole.

Mots clés

Stéganographie, système de transmission stéganographique, fichier stego, le système auditif humain, la qualité perceptuelle de la parole, coefficient LSF, paramètre LP.

Abstract

Digital networks have evolved greatly in recent years and have become inevitable for modern communications. The transmission of signals on these networks raises a number of problems which are not yet solved, such as the security, confidentiality, integrity and authenticity of the audio signals during their transmission. The aim of our project is to develop and a steganographic method on audio signals. This method is based on G.723.1 coder which is used to hide data in a speech signal, it is known as: speech steganography.

Keywords

Steganography, steganographic transmission system, stego file, human auditory system, the perceptual quality of speech, LSF coefficients, LP parameters.

Table des matières

Table des matières I

Liste des figures V

Liste des tableaux VI

Liste des abréviations VII

Introduction générale 1

Chapitre I : Codage de la parole

Introduction 3

I.1 Origine des signaux de parole 3

I.1.1 Production et modélisation de la parole 4

I.1.2 Classification des signaux de parole 6

I.1.3 Propriétés du système auditoire humain 7

I.1.3.1 Le masquage 7

I.2 Les différentes étapes suivies dans le processus de codage de la parole 9

I.2.1 Filtrage anti-repliement 10

I.2.2 L'échantillonnage 10

I.3 La quantification 11

I.3.1 Quantification scalaire 11

I.3.1.1 Quantification uniforme 12

I.3.1.2 Quantification non uniforme 12

I.3.2 Quantification vectorielle 12

I.4 Aperçu du codage vocal 13

I.4.1 Structure d'un système de codage de la parole 13

I.4.2 Propriétés désirables d'un codeur de parole 14

I.4.3 Applications des codeurs de la parole 15

I.4.3.1 Le codeur TIA IS54 VSELP 15

I.4.3.2 Le codeur AMR 16

I.5 Classification des codes vocaux 17

I.5.1 Classification par débit binaire 17
I.5.2 Classification par techniques de codage 18
I.5.2.1 Codeurs de forme d'onde 18
I.5.2.2 Codeurs paramétriques 21
I.5.2.3 Codeurs hybrides 23
Conclusion 24

Chapitre II : Le codage AMR

Introduction 25
II.1 Le codeur G.723.1 25
II.1.1 Description générale 25
II.1.1.1 Traitement au niveau de la trame 25
II.1.1.2 Filtre passe-haut 27
II.1.1.3 Analyse LP 28
II.1.1.4 Quantification LSF 30
II.1.1.5 Filtre de pondération perceptive des formants 30
II.1.1.6 Estimation de pitch en boucle ouverte 31
II.1.1.7 Pondération du bruit harmonique 32
II.1.2 Bit-stream 33
II.1.2.1 Mode de multi-impulsion 34
II.1.2.2 Mode d'ACELP 35
II.1.2.3 Mode SID 36
II.2 Le décodeur G.723.1 36
II.2.1 Description générale 36
II.2.2 Génération d'excitation 37
II.2.2.1 Contribution dictionnaire adaptatif 37
II.2.2.2 Excitation multi-impulsion 38
II.2.2.3 Excitation ACELP 39
II.2.2.4 Mode SID et le Mode Nul 39
II.2.2.5 Excitation de coupure 40
II.2.3 Filtre postérieur de pitch 40
II.2.4 Paramètres LP 41
II.2.5 Le filtre postérieur de formants 42
Conclusion 43

Chapitre III : Méthode stéganographique basée sur l'interpolation polynomiale des coefficients LSFs sur le codeur G.723.1 de la VoIP

Introduction 44

III.1 Système de transmission stéganographique 44

III.1.1 La Stéganographie classique 45

III.1.2 La Stéganographie moderne 45

III.1.3 Système de transmission stéganographique de la parole 46

III.1.4 Comparaison entre la stéganographie et le tatouage 47

III.2 Classification des systèmes stéganographiques de la parole 47

III.2.1 Selon la clé 47

III.2.1.1 Stéganographie à clé publique 48

III.2.1.2 Stéganographie à clé secret (privé) 48

III.2.2 Selon les domaines de stéganographie 48

III.2.2.1 Domaine de fréquence 48

III.2.2.2 Domaine temporel 49

III.2.2.3 Domaine de compression 49

III.3 Évaluation de la qualité perceptuelle de la parole 49

III.3.1 Évaluation perceptuelle de la qualité vocale 50

III.4 Description des signaux de parole utilisés dans les tests 51

III.5 Évaluations des codeurs AMR implémentés 51

III.6 Outils de travail 52

III.7 Méthode implémenté 52

III.7.1 Interpolation linéaire 52

III.7.2 Interpolation polynomiale 54

III.7.3 L'intérêt de changement d'interpolation 55

III.7.4 Capacité 56

III.8 Résultat d'implémentation 56

III.8.1 Résultat obtenue avant appliquer l'interpolation polynomiale 56

III.8.2 Résultat obtenue après appliqué interpolation polynomiale 57

III.8.3 Seuil de modification 58

III.8.4 Résultat obtenue après l'application de méthode stéganographique 60

III.8.5 Extraction le message secret à la réception 64

Conclusion..... 65
Conclusion générale 45
Bibliographie..... 45

Liste des figures

Figure I.1: L'appareil Phonatoire. 4
Figure I.2: Modèle simplifié de production de la parole. 5
Figure I.3: Exemple de courbe de masquage associée à une seule tonalité. 8
Figure I.4: Exemple de spectre de signal et de la courbe de masquage associée. 9
Figure I.5: Etapes suivies pendant le processus de codage de la parole. 10
Figure I.6: Modèle d'un quantificateur vectoriel. 13
Figure I.7: Schéma fonctionnel d'un système de codage de la parole. 14
Figure I.8: Codage PCM. 19
Figure I.9: Schéma de principe du codeur DPCM. 20
Figure I.10: Analyse et synthèse de LP. 22
Figure I.11: Le modèle MELP de production de la parole. 23
Figure I.12: Modèle de production de la parole CELP. 24
Figure II.1: Le codeur G.723.1. 27
Figure II.2: Réponse en fréquence du filtre passe-haut. 28
Figure II.3: Fenêtres LP. 29
Figure II.4: Effet de l'ordre du filtre LP. 30
Figure II.5: Réponse du filtre de pondération des formants. 31
Figure II.6: Réponse du filtre de pondération du bruit harmonique (g_{HNW} = 0.2, L= 40 (200 Hz)). 33
Figure II.7: Le décodeur G.723.1. 37
Figure II.8: Génération de signal de parole. 42
Figure II.9: Réponse en fréquence du filtre postérieur de formants. 43
Figure III.1: Types de masquage de données. 45
Figure III.2: Modèle stéganographique audio. 46
Figure III.3: Chiffrement par clé secret. 48
Figure III.4: Schéma synoptique permettant l'estimation la distance perceptuelle PESQ. 50
Figure III.5: Interpolation des LSF. 53
Figure III.6: Exemples de vecteurs LSF interpolés correspondants à huit sous-trames successives. 54
Figure III.7: Interpolation linéaire et polynomiale pour dissimuler un message secret de 8 bits "01010001". 55
Figure III.8: PESQ obtenus après la variation de seuil. 60

Liste des tableaux

Tableau I.1: Classification des codeurs de parole selon le débit binaire. 17
Tableau II.1: Type de Bit-stream. 34
Tableau II.2: Allocation de bits de l'algorithme de codage. 35
Tableau III.1: Comparaison des caractéristiques de la stéganographie et tatouage. 47
Tableau III.2: Limites des évaluations de la qualité de parole selon la recommandation P.862. 51
Tableau III.3: Résultats des tests objectifs de deux codeurs AMR6.4 kbps et 5.6kbps. 52
Tableau III.4: Facteurs de pondération. 53
Tableau III.5: Qualité de parole avant appliquer interpolation polynomiale pour les locuteurs masculins. 57
Tableau III.6: Qualité de parole avant appliquer interpolation polynomiale pour les locutrices féminines. 57
Tableau III.7: Qualité de parole après appliquer interpolation polynomiale pour les locuteurs masculins. 58
Tableau III.8: Qualité de parole après appliquer interpolation polynomiale pour les locutrices féminines. 58
Tableau III.9: Les valeurs de PESQ obtenus à partir de variation de seuil. 59
Tableau III.10: Une trame avant appliquer la méthode de dissimulation. 61
Tableau III.11: Une trame après appliquer la méthode de dissimulation. 62
Tableau III.12: Une trame après appliquer la méthode de décodage au niveau de récepteur. 64

Liste des abréviations

ACB	Adaptive Codebook
ACELP	Algebraic Codebook Excited Linear Prediction
ADM	Adaptive Delta Modulation
ADPCM	Adaptive Differential Pulse Code Modulation
APC	Adaptive Predictive Coding
APCM	Adaptive Pulse Code Modulation
AR	Auto-Régressif
CELP	Code Excited Linear Predictor
CNG	Confort Noise Generation
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DM	Delta Modulation
DPCM	Differential Pulse Code Modulation
DS	Distorsion Spectral
DTMF	Dual Time Multi Frequency
HNW	Harmonic Noise Weighting
LPC	Linear Prediction Coefficients
LSB	Least Significant Byte
MELP	Mixed Excitation Linear Predictive
MOS	Mean Opinion Square
PESQ	Perceptual Evaluation of Speech Quality
QS	Quantization Scalar
QV	Quantization Vector
RSB	Rapport Signal/Bruit
SNR	Signal to Noise Ratio

TDMA	Time Division Multiple Acces
TIMIT	Acoustic Phonetic Speech Data Base
UIT-T	Union Internationale Des Télécommunications
UMTS	Universal Mobile Telecommunications System
VSELP	Vector Sum Excited Linear Prediction
WAV	Waveform Audio File Format
MP3	MPEG1 Audio Layer 3

Introduction générale

L'histoire de l'évolution de l'être humain est intimement liée à l'histoire de la communication. Depuis ses origines, l'humanité a trouvé en la communication un outil qui lui a permis d'évoluer considérablement. De nos jours, le développement et les progrès en technologie font que la communication joue un rôle de plus en plus important. Les communications numériques constituent sans aucun doute l'une des plus importantes révolutions de la science et de la technologie au début de ce troisième millénaire. Les systèmes de transmission de la parole, indifféremment de n'importe quel autre système de communication, ont évolués très rapidement ces dernières décennies. En plus des besoins classiques de communication et de stockage, la sécurité des systèmes de communication revêt une importance capitale. La sécurité est avant tout définie par la résistance d'un système face aux menaces d'un adversaire (ou groupement d'adversaires) cherchant à compromettre ce système.

L'objectif principal de ce projet de master, s'est intéressé particulièrement à la stéganographie qui est l'art de cacher une information privée ou secrète dans un support hôte (fichier texte, image, audio, vidéo...). En effet, les techniques de stéganographie peuvent assurer des communications sécurisées en dissimulant des informations secrètes et les rendre invisibles. En contraste avec la cryptographie traditionnelle qui consiste à masquer le contenu des messages secrets échangés, le but de la stéganographie consiste à cacher non seulement le contenu mais aussi son existence.

Pour atteindre cet objectif nous avons divisé notre travail comme suivant :

Dans Le premier chapitre, nous allons présenter le mécanisme de production de la parole et passé en suite a les différentes étapes suivies dans le processus de codage de la parole.

Dans le deuxième chapitre, nous présenterons une description générale du codeur G723.1, et passé en suite à l'analyse du signal de parole par une analyse par synthèse de mode CELP (Prédiction Linéaire Excitée Par Code) et multi-impulsion, ensuit une description générale du décodeur G723.1.

Dans le troisième chapitre, nous allons présenter le système de transmission stéganographique et leur architecture. Nous définirons brièvement ensuite un autre système de transmission sécurisés. Par la suite, nous allons présenter le système de transmission stéganographique de la parole en décrivant les principaux éléments de ce système et leurs classifications. Après, nous présenterons le principe de notre méthode proposée.

Enfin, nous présenterons les résultats d'implémentation de notre technique stéganographique. Une évaluation de résultat de simulation des techniques implémentées dans le codeur de parole AMR G.723.1 sera présentée.

Nous terminerons par une conclusion générale qui présentera le fruit de ce travail, ainsi que les perspectives envisagées.

Chapitre I

Codage de la parole

Introduction

La parole est le moyen de communication par excellence entre les êtres humains. C'est pourquoi le traitement de la parole est devenu aujourd'hui une composante fondamentale des sciences de l'ingénieur. Le traitement de la parole est divisé en trois disciplines, la reconnaissance, soit du locuteur soit de la parole, la synthèse, et le codage de la parole.

Le codage de la parole est un aspect important des télécommunications modernes. Le codage de la parole est le processus de représentation numérique d'un signal de parole. L'objectif principal du codage de la parole est de représenter le signal de parole avec le plus petit nombre de bits.

Dans le codage de la parole, il est crucial de comprendre la physiologie de la phonation, les propriétés de base du signal parole et sa perception, ainsi que les différentes étapes suivies dans le processus de codage de la parole, plus particulièrement, l'échantillonnage et la quantification scalaire et vectorielle, et les différents types de codage.

I.1 Origine des signaux de parole

La forme d'onde de la parole est une onde de pression sonore provenant de mouvements contrôlés de structures anatomiques constituant la production de la parole humaine.

La parole est essentiellement générée par une onde acoustique émanant des narines et de la bouche lorsque l'air est expulsé des poumons, ce qui perturbe le flux d'air résultant des constrictions à l'intérieur du corps. Il est utile d'interpréter la production de parole en termes de filtrage acoustique. Les trois principales cavités du système de production de la parole sont *nasales, orales et pharyngées* formant le filtre acoustique principal. Le filtre est excité par l'air des poumons et est chargé à sa sortie principale par une impédance de rayonnement associée aux lèvres [1].

Le tractus vocal fait référence aux cavités pharyngées et buccales regroupées. La voie nasale commence au niveau du vélum et se termine aux narines du nez. Lorsque voile du palais est abaissé, le tractus nasal est couplé acoustiquement au tractus vocal afin de produire les sons nasaux de la parole.

Les formes vocales et nasales changent continuellement avec le temps, créant un filtre acoustique avec une réponse fréquentielle variant dans le temps. Comme l'air des poumons

traverse les voies, le spectre de fréquences est déterminé par la sélectivité en fréquence de ces voies. Les fréquences de résonance du tractus vocal sont appelées fréquences de formants ou simplement formants, qui dépendent de la forme et des dimensions du conduit vocal.

À l'intérieur du larynx se trouve l'un des composants les plus importants du système de production de la parole : les cordes vocales. La position des cordons est à la hauteur de la "pomme Adams " la saillie à l'avant du cou de la plupart des hommes adultes. Les cordes vocales sont une paire de bandes élastiques de muscle et de membrane muqueuse qui s'ouvrent et se ferment rapidement lors de la production de la parole. La vitesse d'ouverture et de fermeture des cordons est unique pour chaque individu et définit les caractéristiques et la personnalité de la voix en question.

I.1.1 Production et modélisation de la parole

L'analyse de la parole est une étape indispensable à toute application de synthèse, décodage ou de reconnaissance. Elle repose en général sur un modèle. Il existe de nombreux modèles de parole. On distingue : les modèles articulatoires, les modèles de production, et les modèles phénoménologiques. Dans le processus de codage, on s'intéresse au modèle de production. On y décrit la parole comme un signal produit par un assemblage de générateurs et de filtres numériques (modèle source-filtre). Les paramètres de ces modèles sont ceux des générateurs et filtres qui les constituent. Le modèle Auto-Régressif (AR) en est l'exemple le plus simple.

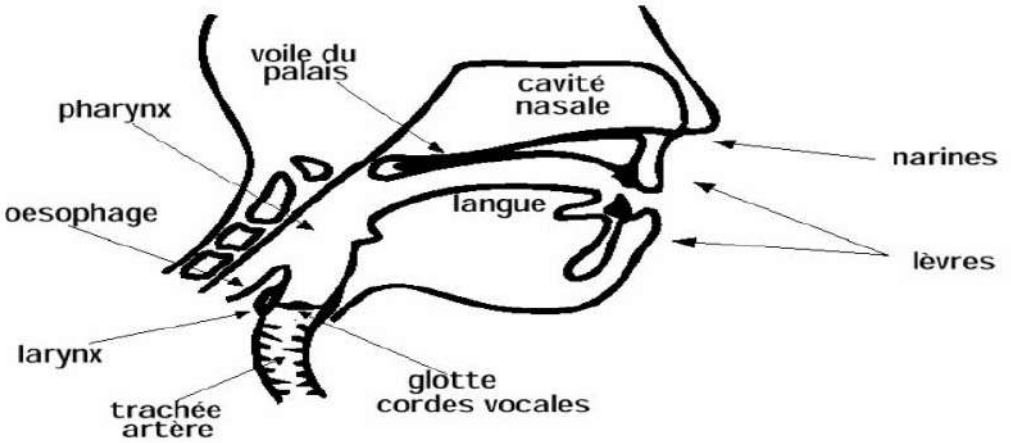


Figure I.1: L'appareil Phonatoire.

Fant proposé en 1960 [2] un modèle de production qui spécifie qu'un signal voisé peut être modélisé par le passage d'un train d'impulsions $u(n)$à travers un filtre numérique récursif de

type tous-pôles. On montre que cette modélisation reste valable dans le cas des sons non voisés, à condition que $u(n)$soit cette fois-ci un bruit blanc. Le modèle final est illustré à la figure I.1. Il est souvent appelé modèle auto-régressif, parce qu'il correspond dans le domaine temporel à une régression linéaire de la forme.

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

Où $u(n)$ et p sont respectivement le signal d'excitation et l'ordre du système. Chaque échantillon est obtenu en ajoutant un terme d'excitation à une prédiction obtenue par combinaison linéaire des p échantillons précédents. Les coefficients du filtre $\{a_i\}$ sont appelés coefficients de prédiction et le modèle AR est souvent appelé modèle de prédiction linéaire LP (Linear Prédiction). Les paramètres du modèle AR sont : la période du train d'impulsions (sons voisés uniquement), la décision sur le son voisé/non voisé, le gain G et les coefficients du filtre $1/A(z)$, appelé filtre de synthèse, comme illustré a la figure I.2 [3].

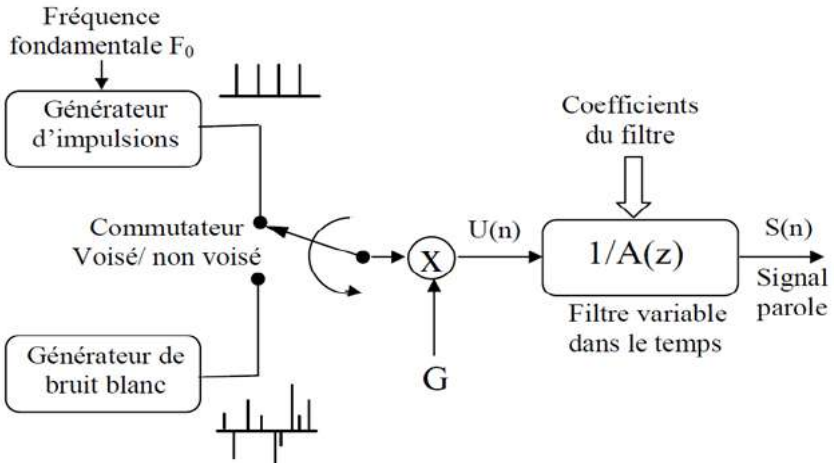


Figure I.2: Modèle simplifié de production de la parole.

Par analogie entre le modèle physique et le modèle mathématique, on peut donner les relations d'équivalences suivantes [3]:

Conduit vocal	⟺	$1/A(Z)$, Le filtre LP
Le flux d'air	⟺	Le signal d'excitation $u(n)$ ou signal résiduel ou encore signal d'erreur de prédiction
Vibration des cordes vocales	⟺	Voisé
Période de vibration des cordes vocales	⟺	$T = 1/F_0$, période du pitch
Fricatives et plosives	⟺	Non voisé/ voisé
Volume d'air	⟺	G, le gain

I.1.2 Classification des signaux de parole

Un signal de parole peut être classé comme étant voisé et non voisé. Les sons vocaux sont générés lorsque les cordes vocales vibrent de telle sorte que le flux d'air provenant des poumons est interrompu périodiquement, créant une séquence d'impulsions pour exciter le tractus vocal. Avec les cordes vocales immobiles, la turbulence créée par le flux d'air traversant un étranglement du conduit vocal génère des sons non voisés. Dans le domaine temporel, le son voisé est caractérisé par une forte périodicité présente dans le signal, avec la fréquence fondamentale appelée fréquence de ton, ou tout simplement pitch. Pour les hommes, le ton varie de 50 à 250 Hz, tandis que pour les femmes, elle se situe généralement entre 120 et 500 Hz et de 200 à 600 Hz chez les enfants. Les sons non voisés, en revanche, ne présentent aucun type de périodicité et sont essentiellement de nature aléatoire.

Pour expérimenter avec le son voisé et le son non voisé et l'implication des cordes vocales, essayez de placer vos doigts sur l'avant de votre cou pendant que vous parlez. Considérez le son « fa » comme dans father « le père ». Commencez par essayer de prononcer pendant quelques secondes le son « f » seul, qu'est une consonne en anglais américaine. Ensuite, prononcez « a » pendant quelques secondes, ce qui est une voyelle. Dans le premier cas, vous ne devriez ressentir aucune vibration à l'avant de votre cou, dans le second cas, une pulsation

est détectée. Parlez plus fort si vous avez du mal à le ressentir. L'oscillation est associée aux activités des cordes vocales et est présente pour la prononciation des voyelles [1].

Il est nécessaire d'indiquer que la classification voisé/non voisé peut ne pas être absolument claire pour tous les trames, car lors des transitions (voisé en non voisé ou inversement), il y aura un caractère aléatoire et une quasi-périodicité difficile à évaluer, strictement voisé ou strictement non voisé.

Pour la plupart des codeurs de parole, le signal est traité trame par trame, une trame comprenant un nombre fini d'échantillons. La longueur de la trame est de telle sorte que les statistiques du signal restent presque constantes dans l'intervalle. Cette longueur est généralement comprise entre 20 et 30 ms, soit 160 et 240 échantillons pour un échantillonnage de 8 kHz.

I.1.3 Propriétés du système auditoire humain

Le fonctionnement du système auditif humain joue un rôle important dans la conception des systèmes de codage de la parole. En comprenant comment les sons sont perçus, les ressources du système de codage peuvent être allouées de la manière la plus efficace, ce qui permet d'améliorer la rentabilité.

I.1.3.1 Le masquage

Le masquage fait référence au phénomène dans lequel un son est rendu inaudible à cause de la présence d'autres sons. La présence d'un ton unique par exemple.

Peut masquer les signaux voisins avec la capacité de masquage inversement proportionnelle à la différence absolue de fréquence. La figure I.3 montre un exemple où une seule tonalité est présente, la tonalité génère une courbe de masquage qui rend tout signal dont une puissance inférieure devient imperceptible. En général, la capacité de masquage augmente avec l'intensité du signal de référence, ou le ton unique dans ce cas [1].

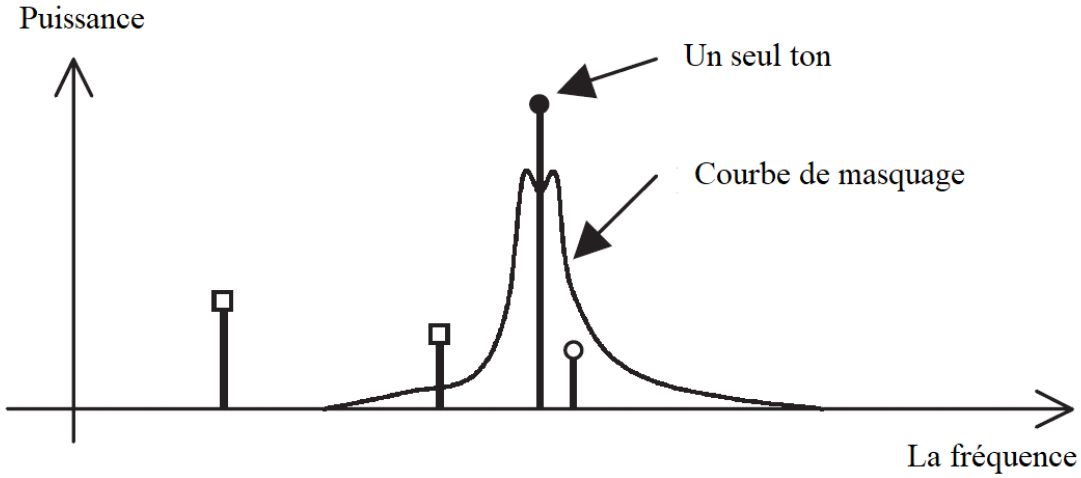


Figure I.3: Exemple de courbe de masquage associée à une seule tonalité.

Les caractéristiques de la courbe de masquage dépendent de chaque individu et peuvent être mesurées dans la pratique en plaçant le sujet dans un environnement de laboratoire et en lui demandant de percevoir un certain son réglé sur certaines valeurs d'amplitude et de fréquence en présence d'un ton de référence.

Le masquage peut être exploré pour les développements de codage de la parole. Par exemple, en analysant le contenu spectral d'un signal, il est possible de localiser les régions de fréquence les plus susceptibles de distorsion. Un exemple est présenté à la figure I.4. Dans ce cas, un spectre typique est représenté, qui consiste en une série de régions de hautes et basses puissances, appelées respectivement pics et vallées. Il existe une courbe de masquage associée qui suit les hauts et les bas du spectre initial. Les signaux dont la puissance est inférieure à la courbe de masquage sont inaudibles. Ainsi, en général, les pics peuvent tolérer plus de distorsion ou de bruit que les vallées.

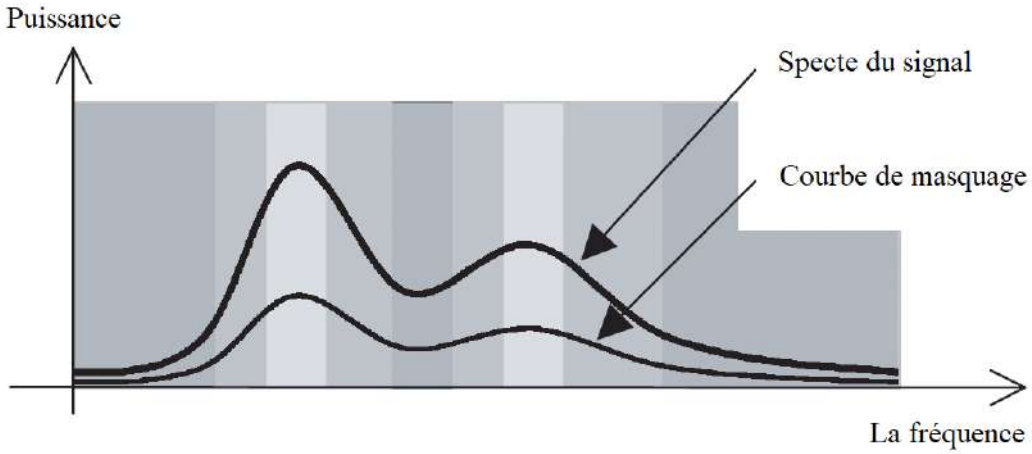


Figure I.4: Exemple de spectre de signal et de la courbe de masquage associée.

Un système de codage bien conçu doit garantir que les vallées sont bien préservées ou relativement exemptes de distorsions, tandis que les pics peuvent tolérer une plus grande quantité de bruit. En suivant ce principe, l'efficacité de l'algorithme de codage est améliorée, ce qui conduit à une qualité de sortie améliorée.

Les codeurs obéissant au principe de CELP s'appuient sur le filtre de pondération perceptuelle pour pondérer le spectre d'erreur lors du codage, la réponse en fréquence du filtre varie dans le temps et dépend du spectre d'origine du signal d'entrée. Le mécanisme est très efficace et est largement appliqué dans la pratique [1].

I.2 Les différentes étapes suivies dans 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, comme illustré a la figure I.5. 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.

.

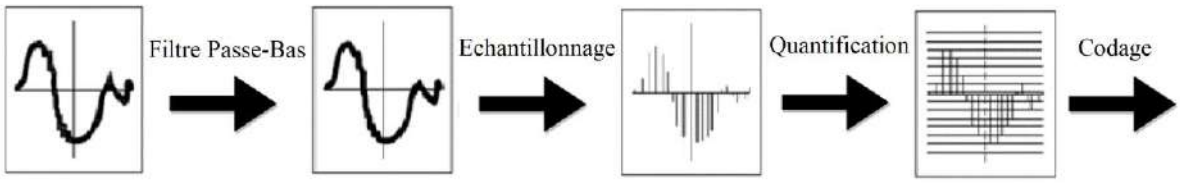


Figure I.5: Etapes suivies pendant le processus de codage de la parole.

I.2.1 Filtrage anti-repliement

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 ». Etapes suivies pendant le processus de codage de la parole.

I.2.2 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 fe. En ce qui concerne le signal vocal, le choix de fe résulte d'un compromis. Son spectre peut s'étendre jusqu'à 12 kHz [4]. Il faut donc en principe choisir une fréquence fe égale à 24 kHz au moins pour satisfaire raisonnablement au théorème de Shannon [5].

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 *fc* est choisie en fonction de la fréquence d'échantillonnage retenue.

Pour la téléphonie, on estime que le signal garde une qualité suffisante lorsque son spectre est limité à 3400 Hz et l'on choisit $fe = 8000\,Hz$. Pour les techniques d'analyse, de synthèse ou de reconnaissance de la parole, la fréquence peut varier de 6000 à 16000 Hz. Par contre pour les signaux audio (parole et musique), on exige une bonne représentation du signal jusqu'à 20 kHz et l'on utilise des fréquences d'échantillonnage de 44.1 ou 48 kHz. Pour les applications multimédia, les fréquences sous-multiples de 44.1 kHz sont de plus en plus utilisées : 22.5 kHz, 11.25 kHz.

I.3 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 exacte 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 [6].

La quantification peut être scalaire ou vectorielle selon que la variable x est à une ou plusieurs dimensions. On peut considérer la quantification scalaire comme étant un cas particulier de la quantification vectorielle lorsque la dimension de la variable x est 1.

I.3.1 Quantification scalaire

La quantification scalaire (QS) attribue à une valeur d'entrée *x* une autre valeur plus proche appartenant à un ensemble fini prédéterminé, ou dictionnaire (codebook) de *N* valeurs [7].

$$C = \{y_k | \mathrm{k} = 1, \dots, \mathrm{N}\} \tag{I. 2}$$

Le quantificateur divise le signal d'entrée en N intervalles I_k , de sorte qu'une entrée *x* appartenant à l'intervalle I_k soit codée avec la sortie y_k.

Pour coder un vecteur *x* de m dimensions par QS, il faut quantifier chaque valeur de ce vecteur x_i indépendamment tel que :

$$Q_i , X_q = y_i, k = Q_i(x_i), \mathrm{i} = 1 \dots \mathrm{m} \tag{I. 3}$$

La quantification scalaire introduit une erreur ou bruit $e = Q(x) - x$, sur l'échantillon d'entrée x. La mesure de distorsion la plus répandue dans la conception QS est l'erreur quadratique entre la valeur originale et la valeur quantifiée :

$$D(x, x_q) = \left| x - x_q \right| 2 = \left| e \right| 2 \tag{I. 4}$$

La performance d'un quantificateur scalaire est souvent évaluée en utilisant la moyenne de l'erreur quadratique :

$$\mathrm{D} = \min E \left[d(x, x_q \right] \tag{I. 5}$$

La quantification scalaire se divise en deux types, la quantification uniforme et la quantification non uniforme.

I.3.1.1 Quantification uniforme

Cette quantification est utilisée dans la conversion analogique-numérique, les intervalles de décision I_k ont la même valeur Δ et les niveaux de sortie y_k sont au milieu de ces intervalles, tels que :

$$\Delta = \frac{x_{max} - x_{min}}{\mathrm{N}} \tag{I.6}$$

$$\mathrm{I_k} = \{\mathrm{x} | \mathrm{x_k} < x \leq \mathrm{x_{k+1}}\} \tag{I.7}$$

$$\mathrm{y_k} = \mathrm{x_{min}} + (\mathrm{k} - 0.5)\Delta, k = 1, \dots, N, \tag{I.8}$$

$$Q(x) = \{y_k | \mathrm{x} \in \mathrm{I_k}\} \tag{I.9}$$

Avec $\mathrm{x_{min}}$ et $\mathrm{x_{max}}$ sont respectivement le minimum et le maximum du signal d'entrée. L'opérationde troncation ou d'arrondie des nombres réels en nombres entiers est un exemple de la quantification uniforme.

I.3.1.2 Quantification non uniforme

Malgré que les quantificateurs uniformes soient faciles à implémenter, ils n'offrent pas une bonne performance, on fait donc recours à la quantification non uniforme, qui consiste à varier les intervalles de décision I_k, en effet, on réduit la largeur de l'intervalle I_k dont les valeurs d'entrée ayant une grande probabilité d'apparition, tandis qu'on augmente celle ayant une faible probabilité d'apparition. La PCM (Pulse Code Modulation) loi A et loi μ sont les plus populaires des quantifications non uniformes.

I.3.2 Quantification vectorielle

La quantification vectorielle (QV) est l'extension de la quantification scalaire à un espace multidimensionnel. La quantification vectorielle attribue un bloc ou un vecteur de valeurs d'entrée à un seul vecteur à partir d'un ensemble fini de vecteurs de sortie. Shannon a démontré que pour un débit donné, le codage de longs bocs d'information donne une très bonne performance de point de vu distorsion. La figure I.6 illustre la structure de base d'un

quantificateur vectoriel. Le quantificateur vectoriel ou l'encodeur, fait correspondre le vecteur d'entrée x à K dimensions à un symbole de canal, ou index i, qui sera transmis sur le canal. L'encodeur partitionne le vecteur d'espace multidimensionnel d'entrée en N régions tel que [3]:

$$P = \{R_1, R_2, \dots, R_3\} \; ou \; R_i = \{x | \mathrm{d(x,y)} <= \mathrm{d}(x, y_j), j \neq i\} \qquad (\mathrm{I}.10)$$

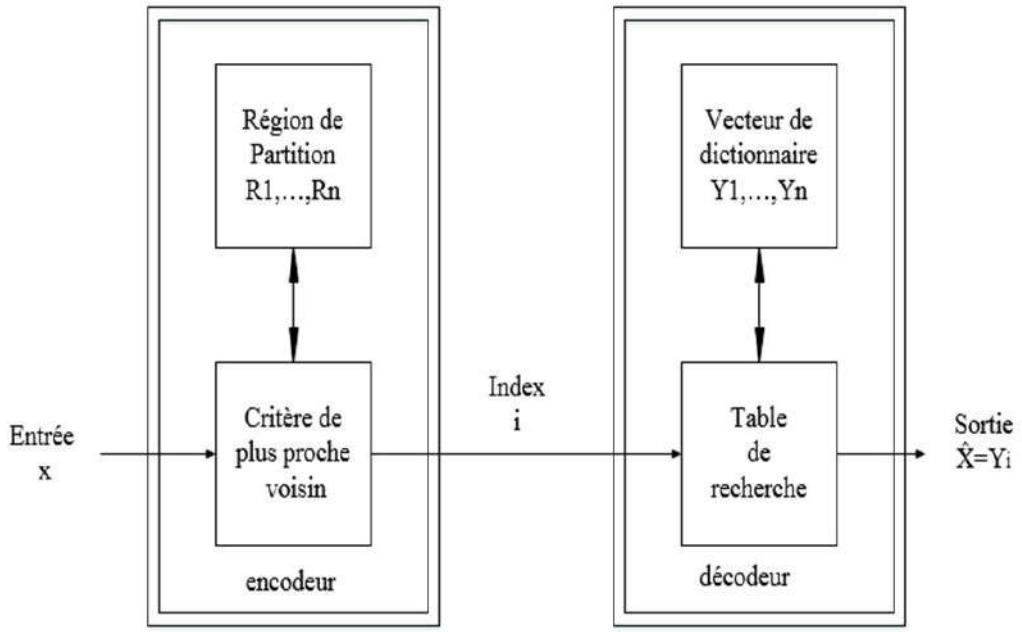


Figure I.6: Modèle d'un quantificateur vectoriel.

Le vecteur y_i est le vecteur de code associé avec la région R_i. L'index est choisi d'une manière à ce que x appartient à la partition Ri, le quantificateur inverse, ou décodeur, doit correspondre le symbole i à un code vecteur approprié x_q=y_i en utilisant une simple procédure de recherche.

I.4 Aperçu du codage vocal

Cette section décrit la structure, les propriétés et les applications de la technologie de codage de la parole.

I.4.1 Structure d'un système de codage de la parole

La figure I.7 montre le schéma fonctionnel d'un système de codage de la parole. Le signal vocal analogique en temps continu provenant d'une source donnée est numérisé par une connexion standard de filtre (élimine les aliasings), d'échantillonneur (conversion en temps

discret), et convertisseur analogique / numérique (on suppose une quantification uniforme). La sortie est un signal de parole à temps discret dont les valeurs d'échantillonnage sont également discrétisées. Ce signal est appelé la parole numérique.

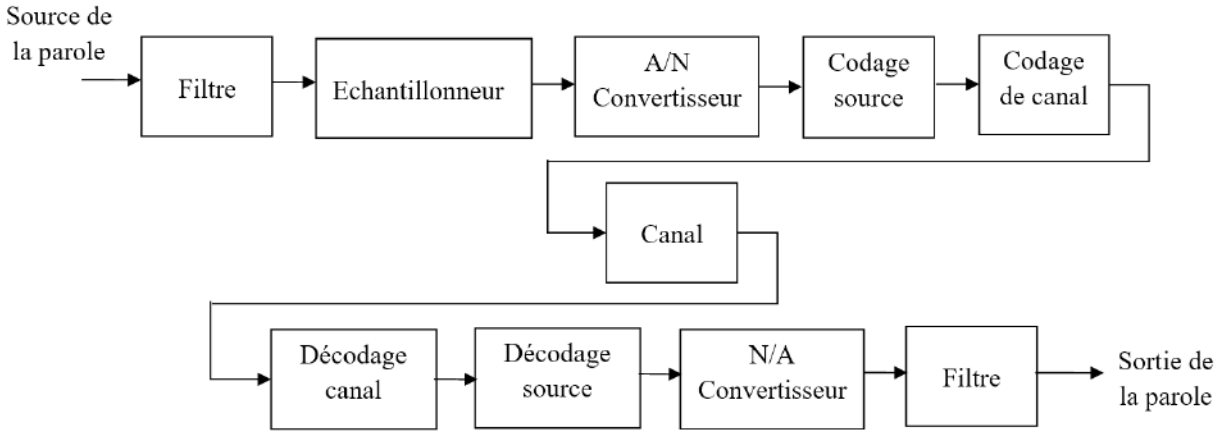


Figure I.7: Schéma fonctionnel d'un système de codage de la parole.

I.4.2 Propriétés désirables d'un codeur de parole

L'objectif principal du codage de la parole est soit de maximiser la qualité perçue à un moment donné, ou pour minimiser le débit binaire pour une qualité de perception particulière.

Le débit binaire approprié auquel la parole doit être transmise ou stockée dépend sur les coûts de transmission ou de stockage, le coût de codage (compression) du signal de parole numérique et les exigences de qualité de la parole. Dans presque tous les codeurs de parole, le signal reconstruit diffère de celui d'origine. Le débit binaire est réduit en représentante signal de parole (ou les paramètres d'un modèle de production de parole) avec une précision réduite et en supprimant la redondance inhérente du signal [1].

- Faible débit binaire : Plus le débit binaire du flux de bits codé est faible, le moins de bande passante requise pour la transmission, conduisant à un système plus efficace. Cette exigence est en conflit constant avec d'autres bonnes propriétés du système, telles que la qualité de la parole.

- Qualité vocale élevée : La parole décodée doit avoir une qualité acceptable pour l'application cible. Il y a plusieurs dimensions dans la perception de la qualité, y compris l'intelligibilité, le naturel, l'agréable et la reconnaissabilité du locuteur.

- Robustesse à travers différentes langues de locuteurs : La technique sous-jacente du codeur de parole devrait être suffisamment générale pour modéliser correctement différents locuteurs et différentes langues (hommes adultes, femmes adultes et enfants). Notez que ce n'est pas une tâche facile, car chaque signal vocal à ses caractéristiques uniques.

- Robustesse en présence d'erreurs de canal : Ceci est crucial pour les systèmes de communication numériques où les erreurs de canal auront un impact négatif sur la qualité de la parole.

- Bonne performance sur les signaux non-voisés (c'est-à-dire la signalisation téléphonique) : Dans un système de télécommunications typique, d'autres signaux peuvent être présents en plus de la parole. Des tonalités de signalisation telles que la DTMF (Dual Time Multi Frequency) dans la composition au clavier et la musique sont souvent rencontrés.

- Petite taille de la mémoire et faible complexité du calcul : Pour le codeur de la parole pour être réalisable, les coûts associés à sa mise en œuvre doivent être faibles, celles-ci incluent la quantité de mémoire nécessaire pour prendre en charge son fonctionnement, ainsi que la demande en calcul.

- Faible délai de codage : Dans le processus de codage et de décodage de la parole, un retard est inévitablement introduit, qui correspond au décalage temporel entre la parole d'entrée du codeur par rapport à la parole de sortie du décodeur. Un délai excessif crée des problèmes avec les conversations bilatérales en temps réel.

I.4.3 Applications des codeurs de la parole

Le codage de la parole joue un rôle important dans la technologie moderne basée sur la voix, en particulier pour la communication vocale numérique, où la qualité et la complexité ont un impact direct sur la qualité marchande et le coût des produits ou services sous-jacents. Il existe de nombreuses normes de codage de la parole conçues pour répondre aux besoins d'une application donnée. Quelques exemples sont les suivants [1].

I.4.3.1 Le codeur TIA IS54 VSELP

Est dérivé des principes d'un codeur CELP standard, en fait-il s'agit d'un codeur CELP doté d'une structure de dictionnaire particulière et dont les coûts de calcul sont réduits. Il existe au moins trois codeurs standard basés sur les mêmes principes de VSELP (Vector Sum

Excited Linear Prediction), tous développés à l'origine par Motorola Inc. Ces normes sont les suivantes :

- TIA IS54 : Normalisé par la TIA en 1989 pour la téléphonie cellulaire numérique à accès Multiple Par Répartition Dans Le Temps (TDMA) en Amérique du Nord. Il fait partie de la norme intérimaire 54 (IS54) et fonctionne à 7,95 kbps.

- ETSI GSM 6.20 : Ceci a été créé afin de doubler la capacité du système cellulaire GSM. Le codeur fonctionne à 5,6 kbps.

- RCR STD 27B : Ce système a été normalisé par le Centre de recherche et de développement pour les systèmes radio (RCR) pour le service de téléphonie cellulaire numérique TDMA au Japon.

I.4.3.2 Le codeur AMR

Ce codeur a été normalisé en 1999 dans le cadre du système de télécommunications mobiles universelles (UMTS) [8] liés au projet de partenariat de troisième génération (3GPP). Le codeur AMR (Adaptive Multi-Rate) appartient à la famille de codeurs appelée multi mode contrôlée par réseau. Définition précise du codage multi mode. Contrôle de réseau signifie que le codeur répond à un signal de contrôle externe pour commuter le débit de données sur un ensemble de débits prédéterminé. Le signal de commande est supposé être généré à distance, généralement en réponse à des niveaux de trafic dans le réseau ou à des demandes d'informations de signalisation.

Le codeur AMR comprend une famille de codeurs à taux fixe, chacun ayant un taux différent. Il y a un total de huit codeurs et tous sont basés sur l'algorithme ACELP (Algebraic Codebook Excited Linear Prediction).

Les débits binaires disponibles sont 4.75, 5.15, 5.90, 6.70, 7.40, 7.95, 10.2 et 12.2 kbps. Au débit le plus élevé de 12,2 kbps, l'algorithme est identique à GSM EFR. À des débits binaires inférieurs, les bits disponibles diminuent, et différents schémas sont utilisés pour allouer les bits à divers paramètres du modèle ACELP.

Plus récemment, avec la croissance explosive d'Internet, le marché potentiel du protocole de voix sur Internet (voix sur IP ou VoIP) a poussé de nombreuses entreprises à développer des produits et des services autour du concept. Contrairement à la téléphonie classique, internet achemine le trafic vocal sous forme de paquets de données sur un réseau de données à

commutation de paquets au lieu d'un flux synchrone de données binaires. Pour les clients résidentiels, un avantage majeur de la téléphonie internet est la réduction des factures pour les appels vocaux longue distance. Pour les entreprises, la VoIP permet l'intégration des données et de la voix dans un seul réseau, ce qui se traduit par des économies de coûts et une efficacité d'administration considérables.

I.5 Classification des codes vocaux

La classification des codeurs de parole modernes n'est pas simple et prête souvent à confusion, en raison du manque de séparation claire entre les différentes approches. Cette section présente certains critères de classification existants. Les lecteurs doivent garder à l'esprit qu'il s'agit d'un domaine en constante évolution et que de nouvelles classes de codeurs seront créées au fur et à mesure de l'introduction de techniques alternatives.

I.5.1 Classification par débit binaire

Tous les codeurs de parole sont conçus pour réduire le débit de référence de 128 kbps vers les valeurs les plus basses. En fonction du débit du flux binaire codé, il est courant de classer les codeurs de parole conformément au tableau I.1. Différentes techniques de codage conduisent à des débits binaires différents. Une méthode donnée fonctionne bien dans une certaine plage de débits binaires, mais la qualité de la parole tombera radicalement si elle est abaissée en dessous d'un certain seuil. Le débit binaire minimal que les codeurs de parole obtiendront est limité par le contenu informationnel du signal de parole. À en juger par le débit de messages récupérables d'un point de vue linguistique pour des signaux de parole typiques, il est raisonnable de dire que le minimum se situe aux alentours de 100 bps. Les codeurs actuels peuvent produire une bonne qualité à 2 Kbits/s et plus, ce qui laisse à penser que de nombreuses améliorations sont possibles [1].

Tableau I.1: Classification des codeurs de parole selon le débit binaire.

Catégorie	**Plage de débit**
Débit binaire élevé	> 15 kbps
Moyen débit binaire	5 à 15 kbps
Faible débit binaire	2 à 5 kbps
Très faible débit binaire	< 2kbps

I.5.2 Classification par techniques de codage

I.5.2.1 Codeurs de forme d'onde

Une tentative est faite pour conserver la forme d'origine de la forme d'onde du signal, et les codeurs résultants peuvent donc généralement être appliqués à n'importe quelle source de signal. Ces codeurs conviennent mieux au codage à haut débit, car les performances chutent avec la diminution du débit binaire. En pratique, ces codeurs fonctionnent mieux à un débit de 32 kbps et plus.

Le rapport signal sur bruit (SNR)

Peut être utilisé pour mesurer la qualité des codeurs de forme d'onde. Étant donné la parole original $x[n]$et la version synthétique $y[n]$, le RSB est défini par :

$$RSB = 10\log_{10}\left(\frac{\sum_n x[n]^2}{\sum_n (x[n]-y[n])^2}\right) \quad \text{(I. 11)}$$

Avec la plage de l'indice de temps *n* couvrant l'intervalle de mesure.

I.5.2.1.1 Codeurs dans le domaine temporel

Les codeurs de formes d'ondes dans le domaine temporel 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 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). Dans ce qui suit, on décrit brièvement quelques schémas importants de codage dans le domaine temporel.

Le Codeur PCM

C'est le plus simple type de codage de formes d'ondes. C'est essentiellement 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 (figure I.8), deux variantes (incluses dans la norme CCITT G.711) [9] 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.

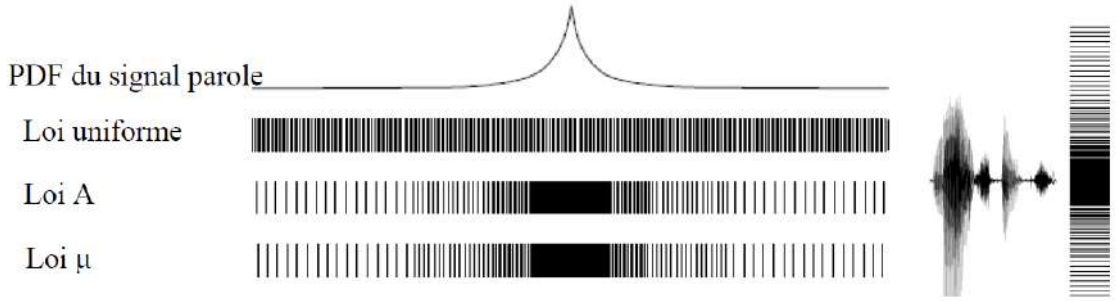


Figure I.8: Codage PCM.

Loi μ : L'équation est modifiée comme suit :

$$f(x) = \frac{(\mathrm{Log}\left(1 + \frac{\mathrm{m}|\mathrm{x}|}{\mathrm{x_{max}}}\right) * \mathrm{sgn\ x})}{\mathrm{Log}(1+\mathrm{m})} \qquad pour \qquad \frac{|x|}{x_{max}} \leq 1 \qquad (\mathrm{I.12})$$

On obtient une compression d'autant plus grande que m est grand. Typiquement, les valeurs de m peuvent varier entre 100 et 500 normalement, le plus utilisé est $\mu = 255$ $(28\ bits - 1)$.

Loi A : Comme l'équation précédente, on a celle de loi A :

$$F(x) = \frac{(\mathrm{A}\ |\mathrm{x}| * \mathrm{sgn\ x})}{(1+\ln \mathrm{A})} \qquad pour \qquad \frac{|x|}{x_{max}} \leq \frac{1}{A}$$

$$F(x) = \frac{\left(1 + \ln\left(\mathrm{A}\frac{|\mathrm{X}|}{Xmax}\right)\right) * sgn\ x}{(1+\ln \mathrm{A})} \qquad pour \qquad 1 \leq \frac{|x|}{x_{max}} \leq 1 \qquad (\mathrm{I.13})$$

La norme européenne pour la téléphonie est :

- Fréquence d'échantillonnage 8 kHz.
- Loi de compression : A= 87.56.
- Filtre de garde, bande passante de 100 à 3400 Hz.
- Débit binaire de 64 kbps.
- Quantification sur 256 niveaux (8 bits).

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 (figure I.9).

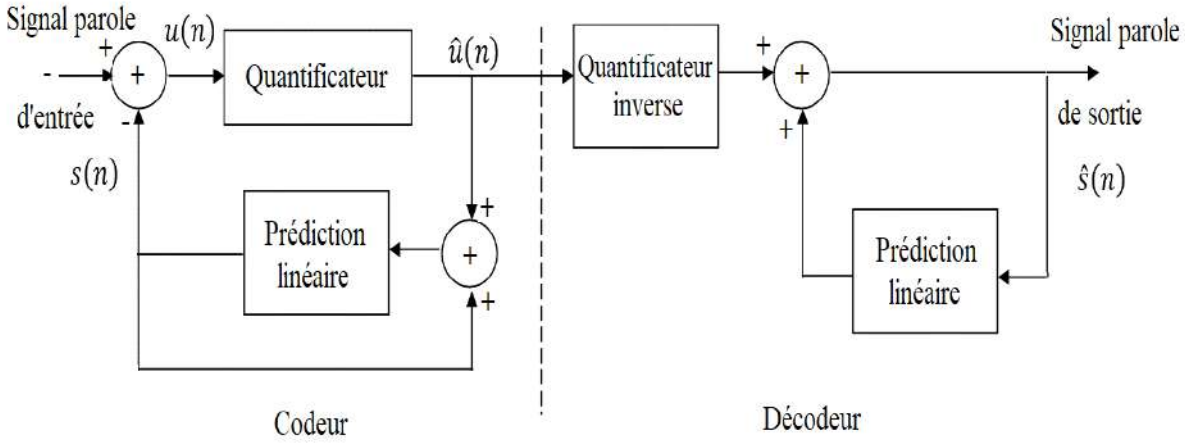


Figure I.9: Schéma de principe du codeur DPCM.

Une version améliorée de la DPCM est l'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 16, 24, 32 et 40 Kbits/s [10]. La complexité de l'ADPCM est légèrement faible.

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 (Discrète Fourier Transform) ou DCT (Discrète Cosine Transform) peuvent être utilisées.

I.5.2.2 Codeurs paramétriques

Dans le cadre des codeurs paramétriques, le signal de parole est supposé être généré à partir d'un modèle contrôlé par certains paramètres. Au cours du codage, les paramètres du modèle sont estimés à partir du signal de parole d'entrée, avec les paramètres transmis en tant que débit binaire codé. Ce type de codeur ne tente pas de préserver la forme originale de la forme d'onde, et donc le SNR (Signal to Noise Ratio) est une mesure de qualité inutile. La qualité perceptuelle de la parole décodée est directement liée à la précision et à la sophistication du modèle sous-jacent. En raison de cette limitation, le codeur est spécifique au signal et présente des performances médiocres pour les signaux non voisé [1].

Il existe plusieurs modèles proposés dans la littérature. Le plus réussi, cependant, est basé sur la prédiction linéaire. Dans cette approche, le mécanisme de production de parole humaine est résumé à l'aide d'un filtre variant dans le temps, les coefficients du filtre étant déterminés à l'aide de la procédure d'analyse par prédiction linéaire.

Cette classe de codeurs fonctionne bien pour un débit binaire bas. L'augmentation du débit binaire ne se traduit normalement pas par une meilleure qualité, car elle est limitée par le modèle choisi. Le débit binaire typique est compris entre 2 et 5 kbps. Les exemples de codeurs de cette classe comprennent le codage par prédiction linéaire (LPC) et la prédiction linéaire à excitation mixte (MELP).

I.5.2.2.1 Codage prédictif linéaire

Le LPC (Codage Prédictif Linéaire) est l'une des techniques de codage de la parole les plus courantes. Le LPC exploite les redondances d'un signal de parole en le modélisant sous la forme d'un filtre linéaire, excité par un signal appelé signal d'excitation. Le signal d'excitation est aussi appelé le signal résiduel. Les codeurs de parole traitent un groupe d’échantillons particulier, appelé trame ou un segment. Le codeur de parole trouve les coefficients de filtre et le signal d'excitation pour chaque trame. Les coefficients de filtre sont calculés de manière à minimiser l'énergie à la sortie du filtre pour cette trame. Ce filtre s'appelle un filtre d'analyse LP. Le signal de la parole est d'abord filtré à travers le filtre d'analyse LP. Le signal résultant est appelé signal résiduel pour cette trame particulière. En réalité, pour le décodeur, l'inverse du filtre d'analyse LP sert de filtre de synthèse LP, tandis que le signal résiduel agit comme

signal d'excitation pour le filtre de synthèse LP. L'ensemble du processus est illustré à la figure I.10.

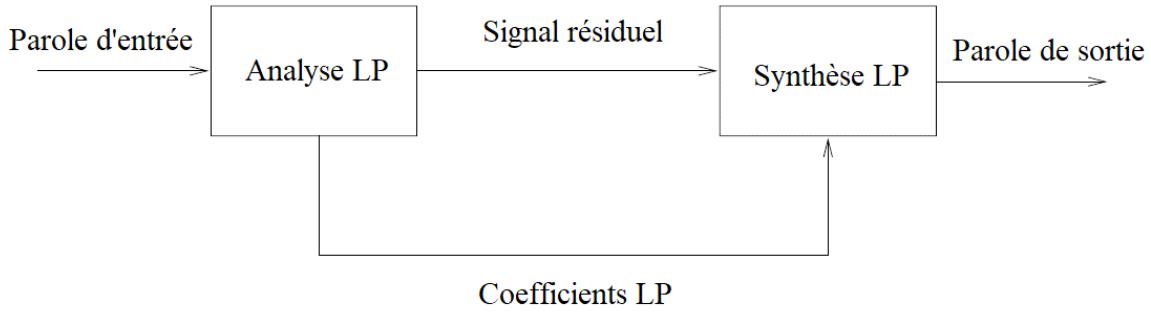


Figure I.10: Analyse et synthèse de LP.

I.5.2.2.2 Le modèle de production de la parole MELP

Un schéma de principe du modèle MELP de production de la parole est présenté à la figure I.11, qui tente d'améliorer le modèle LPC existant. Nous pouvons voir que le modèle MELP est plus complexe. Cependant, les deux modèles partagent certaines similitudes fondamentales, comme le fait que les deux s'appuient sur un filtre de synthèse pour traiter un signal d'excitation afin de générer la parole synthétique. Le décodeur MELP utilise une technique d'interpolation sophistiquée pour lisser les transitions entre les trames. Les principales améliorations du modèle MELP par rapport au modèle LPC sont soulignées ci-dessous [1].

- Une gigue de période générée de manière aléatoire est utilisée pour perturber la valeur de la période de hauteur afin de générer un train d'impulsions apériodique.

L'une des limitations fondamentales de LPC est la classification stricte d'une trame de parole en deux classes : voisé et non voisé. Le codeur MELP étend le nombre de classes en trois catégories : non voisé, voisé et voisé nerveux. Ce dernier état correspond au cas où l'excitation serait apériodique mais pas complètement aléatoire, ce qui est souvent le cas dans les transitions d'harmonisation. Cet État exprimé par une gigue est contrôlé dans le modèle MELP par le paramètre de gigue de pitch et est essentiellement un nombre aléatoire. Expérimentalement, il a été constaté qu'une gigue de période uniformément répartie jusqu'à 25% de la période de tangage produisait de bons résultats. Les tonalités courtes isolées, souvent rencontrées dans les paroles codées en LPC en raison d'une classification erronée de l'état des voix, sont réduites au minimum.

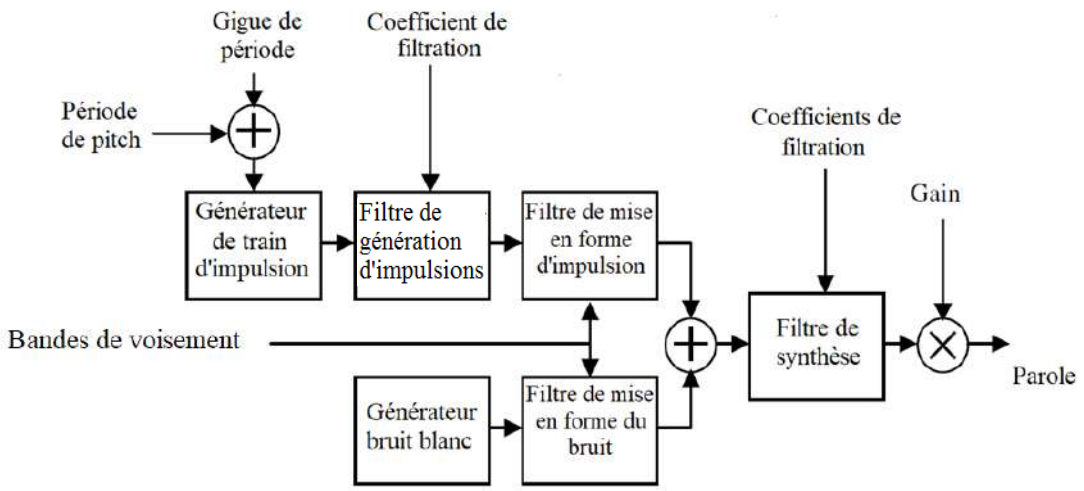


Figure I.11: Le modèle MELP de production de la parole.

I.5.2.3 Codeurs hybrides

Comme son nom l'indique, un codeur hybride combine la force d'un codeur de forme d'onde avec celle d'un codeur paramétrique. À la manière d'un codeur paramétrique, il repose sur un modèle de production de parole, lors du codage, les paramètres du modèle sont localisés. Des paramètres supplémentaires du modèle sont optimisés de telle sorte que la parole décodée soit aussi proche que possible de la forme d'onde d'origine, avec la proximité souvent mesurée par un signal d'erreur pondéré de manière perceptuelle. Comme dans les codeurs de forme d'onde, une tentative est faite pour faire correspondre le signal d'origine avec le signal décodé dans le domaine temporel.

D'un point de vue technique, la différence entre un codeur hybride et un codeur paramétrique est que la première tente à quantifier ou de représenter le signal d'excitation au modèle de production de parole, qui est transmis dans le cadre du flux binaire. Ce dernier produit toutefois un faible débit en éliminant toutes les informations détaillées du signal d'excitation, seuls les paramètres grossiers sont extraits.

Un codeur hybride a tendance à se comporter comme un codeur de forme d'onde pour un débit binaire élevé et comme un codeur paramétrique à faible débit, avec une qualité moyenne à bonne pour un débit moyen [1].

I.5.2.3.1 Le modèle de production de la parole CELP

Le codeur CELP s’appuie sur les modèles de prédiction linéaire à court et à long terme, la figure I.12 montre le schéma fonctionnel du modèle de production de parole, où une séquence

d'excitation est extraite du dictionnaire à travers un index. L'excitation extraite est mise à l'échelle jusqu'au niveau approprié et filtrée par la connexion en cascade du filtre de synthèse de pitch et du filtre de synthèse de formants, pour obtenir la parole synthétique. Le filtre de synthèse de pitch crée une périodicité dans le signal associée à la fréquence fondamentale de pitch, et le filtre de synthèse de formants génère l'enveloppe spectrale.

Le dictionnaire peut être fixe ou adaptatif et peut contenir des impulsions déterministes ou du bruit aléatoire. Pour des raisons de simplicité, nous supposons pour l'instant que le dictionnaire est corrigé et contient des échantillons de bruit blanc. Ainsi, l'indice d'excitation sélectionne une séquence de bruit blanc dans le dictionnaire en tant qu'entrée pour la connexion en cascade des filtres de synthèse. Le modèle de production de parole sur lequel s'appuie le codeur CELP consiste simplement en une source de bruit blanc excitant les filtres de synthèse [1].

Les notes suivantes peuvent être soulignés lorsqu'on comparé aux autres codeurs.

- Une classification stricte voisé / non voisé, comme pour le codeur LPC est éliminée.
- Les informations de phase partielle du signal d'origine sont préservées.

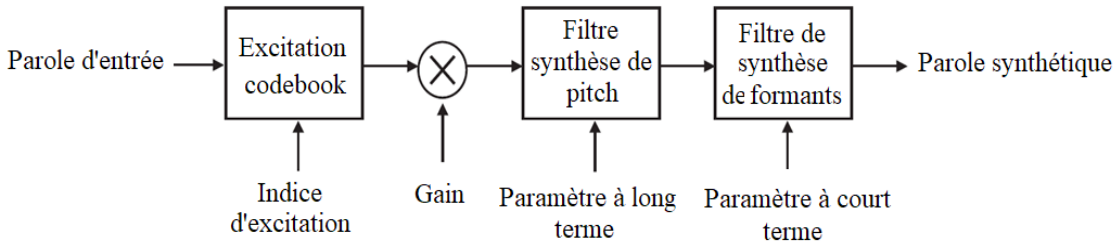


Figure I.12: Modèle de production de la parole CELP.

Conclusion

Afin de bien saisir et de pouvoir traiter le codage de la parole, il est indispensable de connaître le signal parole, ses caractéristiques et sa production, c'est pourquoi nous avons commencé ce chapitre par une description du signal de parole.

Nous avons ensuite, présenté la structure, les propriétés et les applications de la technologie de codage de la parole, et les différents types de codeurs de parole.

Chapitre II

Le codage AMR

Introduction

La recommandation UIT-T G.723.1, appelée aussi AMR (Adaptive Multi-Rate audio codec), spécifie une représentation codée qui peut être utilisée pour compresser le signal vocal ou un autre signal audio de services multimédias à débit binaire moyen. Lors de la conception de ce codeur, la principale application qui étudiait est la téléphonie qui fonctionne à débit binaire moyen, dans le cadre de la série des normes du type H.324 et SIP.

Le codeur G.723.1 possède deux débits binaires associés : 5,3 kbit/s et 6,3 kbit/s. Le débit supérieur correspond à une meilleure qualité. Le débit inférieur assure un niveau de qualité satisfaisant et offre aux concepteurs de systèmes un degré supplémentaire de flexibilité. Ces deux débits sont des caractéristiques obligatoires du codeur et du décodeur. Il est possible de passer de débit à l'autre à chaque limite de trame (toutes les 30 ms). Il est également possible de choisir le fonctionnement en débit variable au moyen d'une émission discontinue avec remplissage de bruit pendant les intervalles de silence.

Ce codeur a été optimisé afin de représenter la parole avec une qualité élevée aux débits cités par avant, avec un faible degré de complexité. Les signaux acheminant des fréquences musicales et autres signaux audio ne sont pas restitués aussi fidèlement que les signaux vocaux mais ils peuvent être compressés et décompresser au moyen de ce codeur.

II.1 Le codeur G.723.1

II.1.1 Description générale

Ce codeur est conçu de façon à fonctionner avec un signal numérique que l'on obtient en filtrant d'abord la bande téléphonique (Recommandation G.712) [11] de l'entrée analogique, puis en l'échantillonnant à 8000 Hz puis en le convertissant linéairement en MIC de 16 échantillons binaires pour l'entrée du codeur. Il convient d'effectuer la conversion inverse vers le mode analogique par des moyens similaires à la sortie du décodeur. D'autres caractéristiques d'entrée/sortie, comme celles qui sont spécifiées par la Recommandation G.711 pour les données MIC à 64 kbit/s [12].

II.1.1.1 Traitement au niveau de la trame

Le codeur est fondé sur les principes du codage prédictif linéaire (LPC) par analyse et synthèse, en vue de minimiser un signal d'erreur pondéré par une courbe de perception. Le codeur travaille sur des blocs (trames) contenant chacun 240 échantillons, ce qui correspond à une longueur de trame de 30 ms à 8 kHz. Chaque bloc passe d'abord par un filtre passe-haut

précède afin d'éliminer la composante apériodique. Il est ensuite subdivisé en quatre sous-trames de 60 échantillons chacune [13].

Les opérations au niveau de trame dans le codeur se composent des étapes suivantes (figure II.1) [13]:

- Le filtrage passe-haut du signal d'entrée.
- Former un signal étendu (filtré passe-haut) constitué de trois parties : Des échantillons de trame précèdent, des échantillons de trame courante, et des échantillons de trame suivant. Les échantillons de trame courante sont divisés en 4 sous-trames.
- Analyse de prédiction linéaire, se fait sur chaque sous-trame : Cela crée quatre ensembles de coefficients LP (Linear Prédiction), un pour chaque sous-trame.
- Les coefficients LP pour la dernière sous-trame (sous-trame 3) sont quantifiés dans le domaine LSF (Line Spectral Frequency).
- Les coefficients LP quantifiés sont interpolés linéairement dans le domaine LSF à l'aide des coefficients LP quantifiés de la trame précédente. Cela crée quatre ensembles de coefficients LP (quantifiés), un pour chaque sous-trame, les derniers de ces ensembles étant les coefficients LP quantifiés pour la dernière sous trame. Ces coefficients quantifiés sont utilisés pour le filtre de synthèse.
- Les coefficients non quantifiés LP (4 ensembles) sont utilisés pour former un filtre de pondération perceptive des formants. Ce filtre est utilisé pour pondérer le signal d’erreur.
- Le signal d'entrée (après filtrage passe-haut) est traité avec filtre de pondération perceptive des formants.

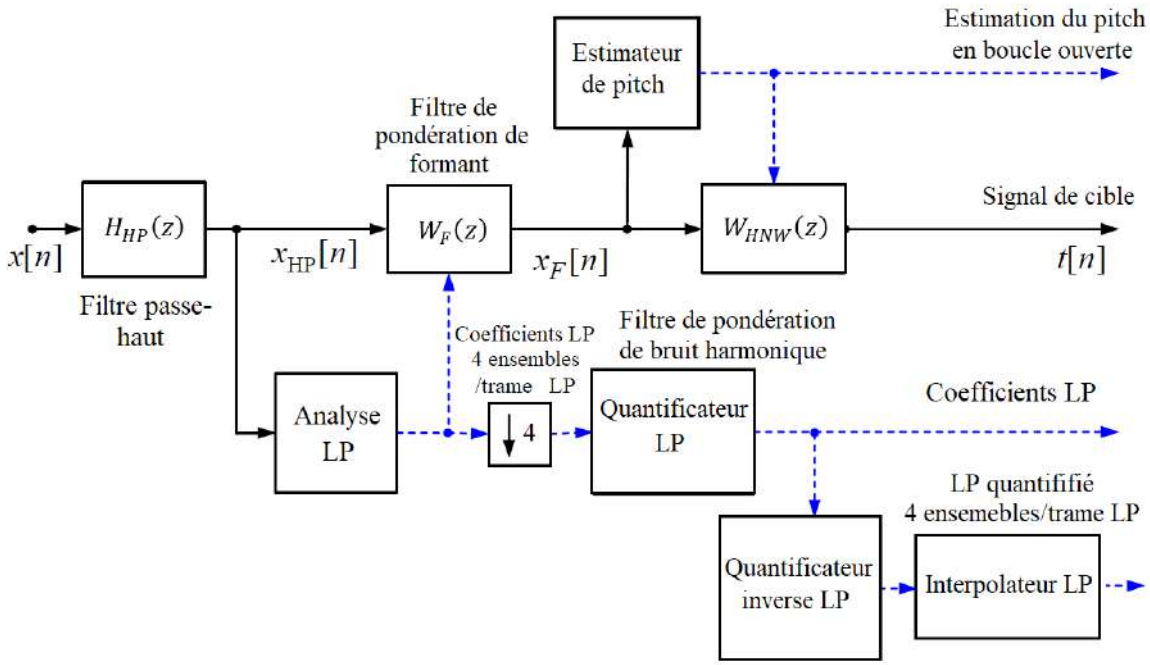


Figure II.1: Le codeur G.723.1.

- La sortie du filtre de pondération des formants est utilisée pour former une première estimation du retard de pitch. On appelle cela l'estimation de pitch en boucle ouverte. Cette estimation est basée sur deux sous-trames à la fois, ce qui donne deux estimations de pitch en boucle ouverte par trame, l'un pour les sous-trames 0 et 1, et un autre pour les sous-trames 2 et 3.

- L'estimation de pitch en boucle ouverte est utilisée pour générer un deuxième filtre de pondération, le filtre de pondération du bruit harmonique, qui permet de suivre les pics harmoniques lors de la parole voisée.

- Le signal d'entrée, traité par la combinaison de filtre passe-haut, le filtre de pondération des formants et le filtre de pondération de bruit harmonique forme le soi-disant signal de cible.

II.1.1.2 Filtre passe-haut

Le filtre passe-haut utilisé pour éliminer les composantes apériodiques. La fonction de transfert de ce filtre est la suivante [14]:

$$H(z) = \frac{1 - \mathrm{z}^{-1}}{1 - \frac{127}{128}\mathrm{z}^{-1}} \qquad \text{(II. 1)}$$

Où $a = \frac{127}{128}$. La réponse en fréquence du filtre est représentée à la figure II.2.

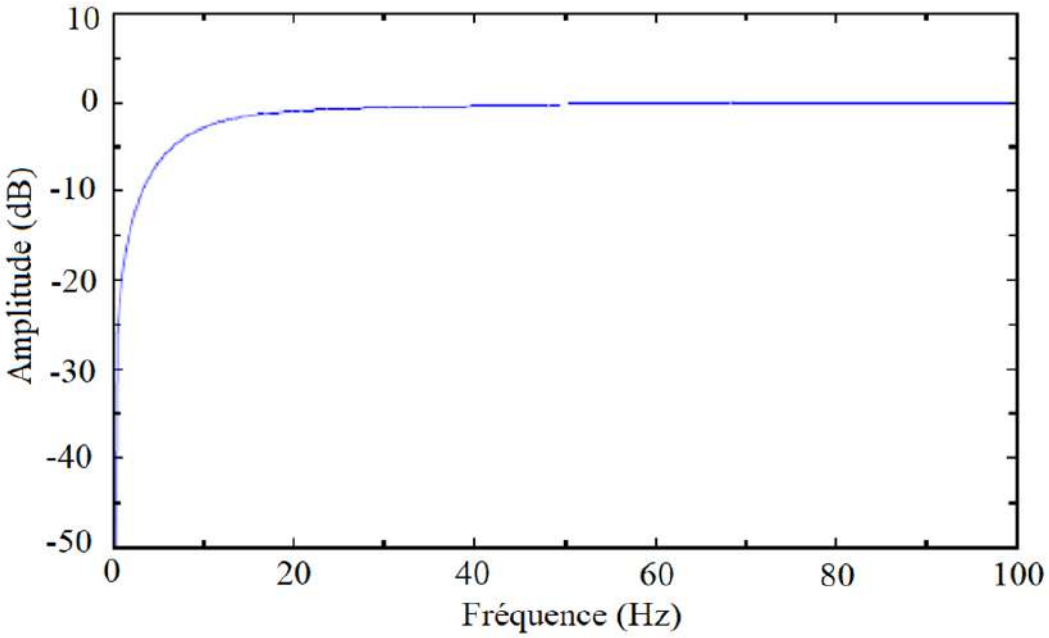


Figure II.2: Réponse en fréquence du filtre passe-haut.

II.1.1.3 Analyse LP

L'analyse de prédiction linéaire (LP), opère sur les signaux filtrés passe-haut. L'analyse LP est réalisée pour chaque sous-trame. Une fenêtre de Hamming de 180 échantillons est appliquée pour chaque sous-trame. La fenêtre est centrée sur une sous-trame et se prolonge donc sur chaque côté de la sous-trame (60 échantillons en arrière, 60 échantillons au cours de la sous-trame, et les 60 échantillons à l'avance). Le recouvrement en arrière sur la trame est de 60 échantillons pour accommoder l'étendue de la fenêtre arrière lors du traitement de la première sous-trame. Le recouvrement en avant sur la trame est de 60 échantillons pour accommoder l'étendue à la fenêtre avant lors du traitement de la dernière sous-trame. Les positions des fenêtres pour les sous-trames sont illustrées à la figure II.3 [14].

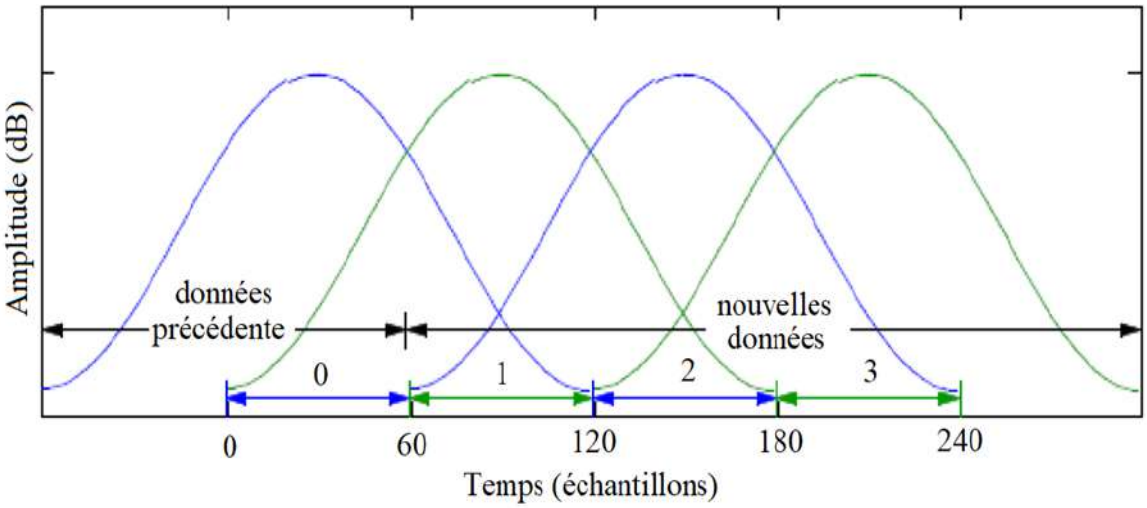


Figure II.3: Fenêtres LP.

Les positions des fenêtres pour les sous-trames montrent que le traitement nécessite 120 échantillons de signal passés. Les nouveaux 240 échantillons pour une trame sont ajoutés pour donner les 360 échantillons nécessaires. Après analyse LP, les 120 premiers échantillons deviennent la mémoire pour la trame suivante.

L'analyse LPC est effectuée sur le signal $x[n]$ de la manière suivante. L'analyse prédictive linéaire (LP) du 10 ème ordre est utilisé. Pour chaque sous-trame, une fenêtre de 180 échantillons est centrée sur la sous-trame. Une fenêtre de Hamming est appliquée à ces échantillons pour calculer 11 coefficients d'autocorrélation. Les coefficients prédictifs (LPC) sont calculés à l'aide de la récursivité conventionnelle de Levinson-Durbin [15]. Les LPCs sont utilisés pour construire le filtre de pondération perceptuelle à court terme. Le filtre de synthèse LPC est défini comme suit :

$$A_i(z) = \frac{1}{1 - \sum_{j=1}^{10} a_{ij}\, z^{-j}} \qquad (II.2)$$

Où, i est l'index de sous-trame.

L'analyse de prédiction linéaire fournit des informations sur le spectre de fréquence du signal. Comme le montre la figure II.4, tout en augmentant l'ordre du filtre LP, l'enveloppe spectrale se rapproche du spectre d'origine.

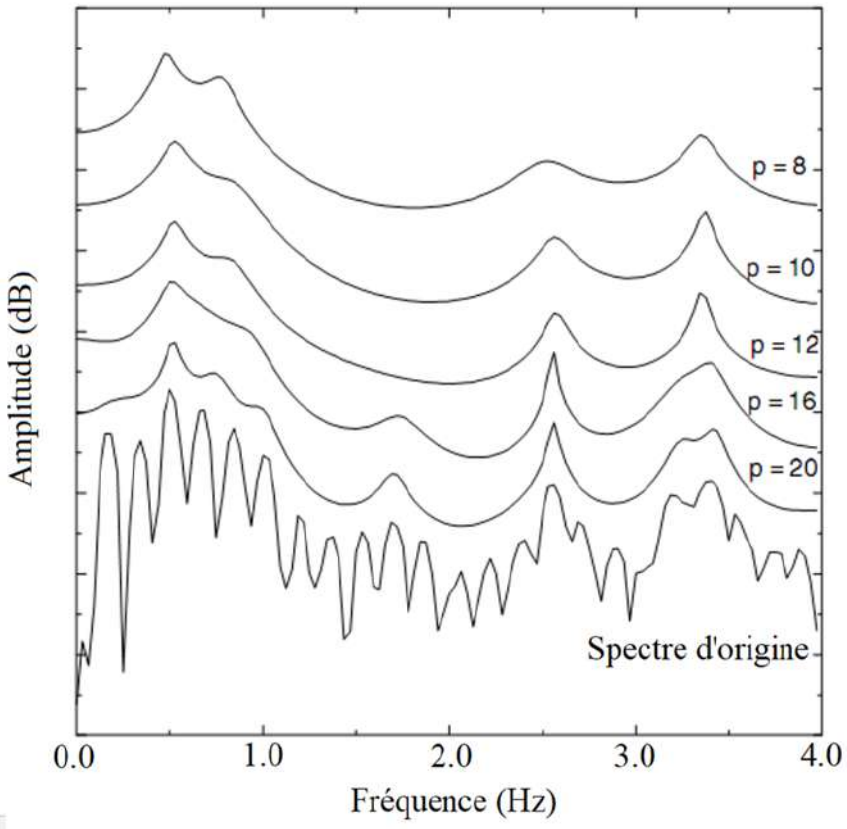


Figure II.4: Effet de l'ordre du filtre LP.

II.1.1.4 Quantification LSF

La quantification des paramètres LP est effectuée dans le domaine fréquentiel pour avoir les paramètres LSF (Line Spectral Frequency). Un ensemble de paramètres LP par trame (correspondant à la dernière sous-trame dans la trame) est quantifié après le passage à des étapes suivantes [13]:

- Les coefficients LP sont convertis en paramètres LSF.
- La quantification du vecteur LSF.
- La quantification des sous-vecteurs LSF.

II.1.1.5 Filtre de pondération perceptive des formants

Dans le cadre du processus de former un signal synthétique, la parole filtrée par le filtre passe haut est passée à travers un filtre de pondération perceptive des formants.

Ce filtre à la fonction de transfert suivante :

$$W_i(z) = \frac{1 - \sum_{j=1}^{10} a_{ij} z^{-j} y_1^j}{1 - \sum_{j=1}^{10} a_{ij} z^{-j} y_2^j}, \qquad 0 \leq i \leq 3 \qquad \text{(II. 3)}$$

Où, $y_1 = 0{,}9$ et $y_2 = 0{,}5$.

La trame du signal vocal d'entrée, est ensuite subdivisée en quatre sous-trames dont chacune passe par le filtre de pondération ce qui donne le signal vocal de sortie pondéré.

L'effet du filtre de pondération de formants est d'atténuer l'importance de ces régions du spectre où le spectre présente des pics de LP et d'insister sur les régions entre les pics [13]. L'idée est que les pics (formants) du spectre LP auront tendance à masquer le bruit à ces fréquences, tandis que le bruit dans les vallées est plus audible. Un exemple de réponse du filtre de pondération est présenté à la figure II.5.

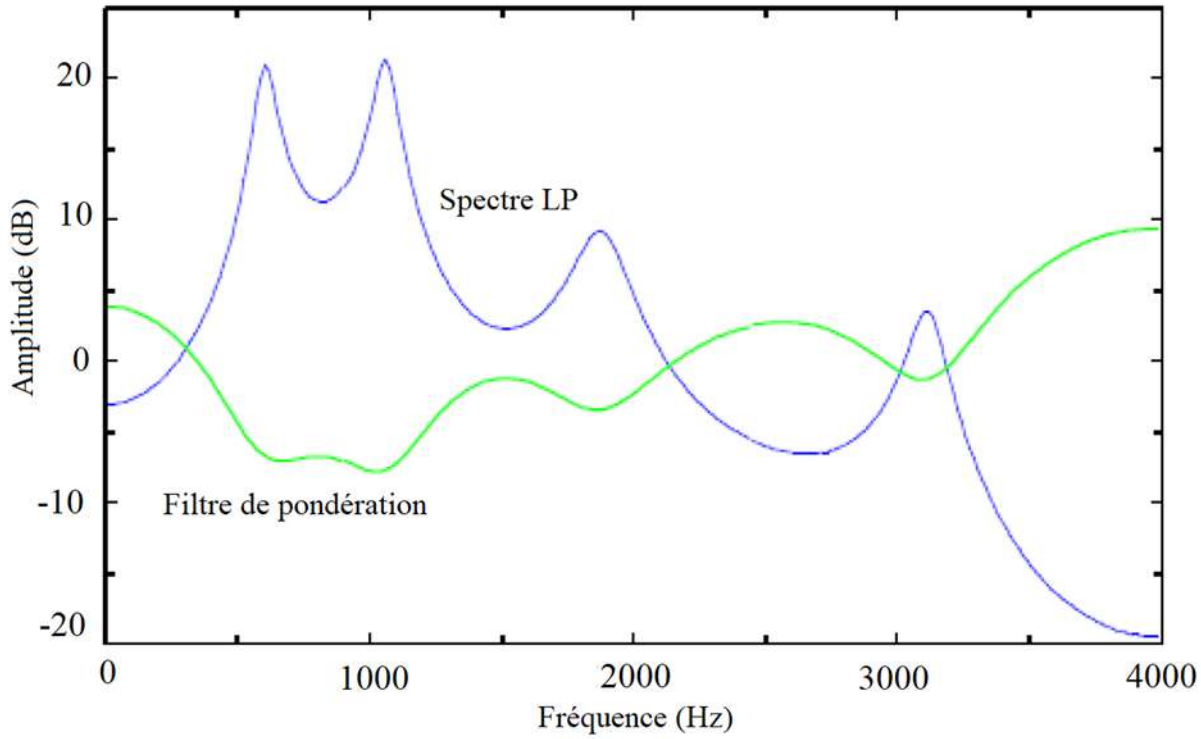


Figure II.5: Réponse du filtre de pondération des formants.

II.1.1.6 Estimation de pitch en boucle ouverte

L'estimation du pitch en boucle ouverte trouve le retard et les valeurs de pitch et de gain qui minimisent l'erreur de prédiction quadratique moyenne. Le pitch en boucle ouverte est déterminé à partir de la sortie du filtre de pondération perceptuelle. Pour déterminer la période

fondamentale, on fait appel à une méthode de maximisation du critère d'intercorrélation, $C_{OL}(j)$, exprimé comme suit :

$$C_{OL}(j) = \frac{(\sum_{n=0}^{119} f[n]f[n-j])^2}{\sum_{n=0}^{119} f[n-j]f[n-j]}, \qquad 18 \leq j \leq 142 \qquad \text{(II. 4)}$$

L'index j qui maximise le critère d'intercorrélation, $C_{OL}(j)$, est choisi comme estimation de la période fondamentale en boucle ouverte pour les deux sous-trames appropriées. Lorsque l'on recherche le meilleur index, on donne une certaine préférence aux périodes fondamentales de plus faible grandeur, afin d'éviter de choisir des multiples de pitch. Les valeurs maximales du critère $C_{OL}(j)$ sont recherchées en commençant par $j = 18$. Chaque valeur maximale de $C_{OL}(j)$ trouvée est comparée au meilleur maximum déjà obtenu $C_{OL}(j¢)$. Si la différence entre les index j et $j¢$ est inférieure à 18 et si $C_{OL}(j)$>$C_{OL}(j¢)$, on choisit le nouveau maximum. Si la différence entre les index est supérieure ou égale à 18, on ne choisit le nouveau maximum que si $C_{OL}(j¢)$ est supérieur à $C_{OL}(j)$ de 1,25 dB.

II.1.1.7 Pondération du bruit harmonique

Un autre composant du filtrage perceptuel global pour former le signal synthétique est un filtre de pondération de bruit harmonique (HNW). Afin d'améliorer la qualité des signaux vocaux codés, on construit un filtre de mise en forme du bruit harmonique. Exprimé comme suit :

$$y[n] = x[n] - g_{HNW}x[n-L] \qquad \text{(II. 5)}$$

Ceci est une version à gain réduit d'une prédiction de pitch. La réponse du filtre est :

$$W_{HNW}(z) = 1 - g_{HNW}z^{-L} \qquad \text{(II. 6)}$$

L'entrée de ce filtre est un signal pondéré en formants et la sortie est le signal synthétique. Le retard est choisi en cherchant autour de la valeur de pitch en boucle ouverte [14]. La valeur HNW est trouvée pour chaque sous-trame. Le choix du retard est régi par les mêmes équations que pour la recherche de pitch en boucle ouverte, mais un ensemble séparé de retards et de coefficients est déterminé pour chaque sous-trame. Cela signifie que la somme pour la détermination des valeurs de corrélation est plus grande que la longueur de la sous-trame de 60 échantillons.

Enfin, le filtre HNW n'est utilisé que si le gain de prédiction est suffisamment élevé. Le gain de prédiction est le rapport entre l'énergie d'entrée et l'énergie de sortie du filtre HNW (équation II.7).

$$P_G = \frac{R[0,0]}{\varepsilon_{opt}}$$

$$P_G = \frac{1}{1 - \frac{R[0,\ L_0]}{R[0,0]R[\ L_0,\ L_0]}} \qquad \text{(II. 7)}$$

Le gain de prédicteur optimal est :

$$g_{opt} = \frac{R[0,L]}{R[L,L]} \qquad \text{(II. 8)}$$

Un exemple de réponse HNW est présenté à la figure II.6.

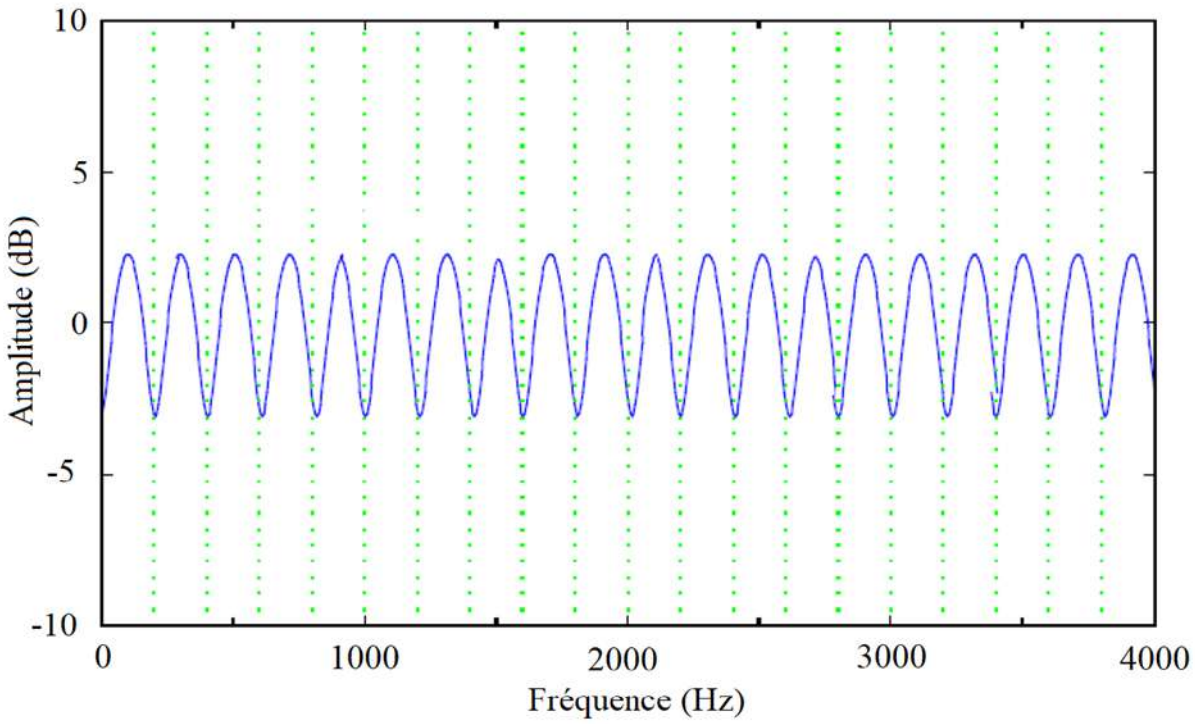


Figure II.6: Réponse du filtre de pondération du bruit harmonique (g_{HNW} =0.2, L= 40 (200 Hz)).

II.1.2 Bit-stream

Le codeur génère un flux de bits, stockés sous forme de données binaires dans un fichier. Chaque trame de données est précédée par un code à deux bits, indiquant le type de trame.

Pour les différents types de trames, le nombre de bits de données change. Le nombre de données dans une trame de données est la suivante (y compris l'indicateur de trame de type 2 bits) [13].

- Mode multi-impulsion : Le mode multi-impulsion utilise 24 octets (192 bits) par trame.
- Mode ACELP : Le mode ACELP utilise 20 octets (160 bits) par trame.
- Mode SID : Une trame de descripteur d'insertion de silence utilise 4 octets (32 bits) par trame dans le fichier de flux binaire.
- Mode Nul : Une trame nul utilise 1 octet (dont 2 bits sont utilisés) par trame dans le fichier de flux binaire.

Le nombre de bits de données (après l'indicateur de type de trame à 2 bits) est résumé dans le tableau II.1.

Tableau II.1: Type de Bit-stream.

Code de trame	**Type de trame**	**Nombre de bits de données**
00	Multipulse	192
11	ACELP	160
10	SID	32
11	Null	2

Il y a des points communs pour les modes. Les trois premiers modes utilisent le même codage de paramètre LP. Multi-impulsion et ACELP utilisent le même codage adaptatif, dictionnaire de codes et le même codage de gain combiné.

II.1.2.1 Mode de multi-impulsion

Les bits de données pour le mode multi-impulsion sont les suivants [13]:

- Paramètres LP : Au total, 24 bits pour les paramètres LSF.
- Retard de dictionnaire adaptatif : Codé avec [7, 2, 7, 2] bits par sous-trame, au total 18 bits.

- Codes de gain combiné : 12 bits par sous-trame, au total 48 bits.
- Codes de grille multi-impulsion : 1 bit par sous-trame, au total 4 bits.
- Bit réservé : Un bit inutilisé.
- Codes de position multi-impulsion : Au total 73 bits.
- Signes multi-impulsion : [6, 5, 6, 5] bits par sous-trame, au total 22 bits.

L'affectation des bits de l'algorithme de codage est présentée dans le tableau II.2.

Tableau II.2: Allocation de bits de l'algorithme de codage.

Paramètres codés	**S.trame 0**	**S.trame 1**	**S.trame 2**	**S.trame 3**	**Total**
Indices LPC					24
Décalage de code adaptatif	7	2	7	2	18
Gains combinés	12	12	12	12	48
Positions d'impulsion	20	18	20	18	73
Signes d'impulsion	6	5	6	5	22
Indice de grille	1	1	1	1	4
Total					189

II.1.2.2 Mode d'ACELP

Les bits de données pour le mode ACELP sont les suivantes [13]:

- Paramètres LP : Au total, 24 bits pour les paramètres LSF.
- Retard de dictionnaire adaptatif : Codé avec [7, 2, 7, 2] bits par sous-trame, au total 18 bits.

- Codes de gain combiné : 12 bits par sous-trame, au total 48 bits.
- Codes de grille ACELP : 1 bit par sous-trame, au total 4 bits.
- Codes de position des impulsions ACELP : 12 bits (3 bits pour chacune des pistes 4) par sous-trame, au total48 bits.
- Signes d'impulsions ACELP : 4 bits (un bit par piste) par sous-trame, au total 16 bits.

II.1.2.3 Mode SID

Les bits de données pour le mode de SID sont les suivants [13]:

- Paramètres LP : Au total, 24 bits pour les paramètres LSF.
- Code de gain SID : 6 bits par trame.

II.2 Le décodeur G.723.1

II.2.1 Description générale

Comme le codeur, le décodeur (figure II.7) fonctionne trame par trame. Il décode d'abord les index de codage LPC quantifiés, puis construit le filtre de synthèse des coefficients LPC. Pour chaque sous-trame, on décode aussi bien les séquences d'excitation par un dictionnaire adaptatif et les séquences d'excitation par un dictionnaire fixe, avant de les injecter dans le filtre de synthèse. Le filtre postérieur (Post-filter) adaptatif se compose d'un filtre postérieur de formants et d'un filtre postérieur fondamental anticipée/différée. Le signal d'excitation envoyé dans le filtre postérieur fondamental, qui à son tour traverse le filtre de synthèse dont la sortie va vers le filtre postérieur de formants. Un module de normalisation du gain maintient l'énergie du signal au niveau d'entrée du filtre postérieur de formants [12].

Le décodeur ne fait pas la recherche, est donc beaucoup moins simple a implémenté que le codeur [13].

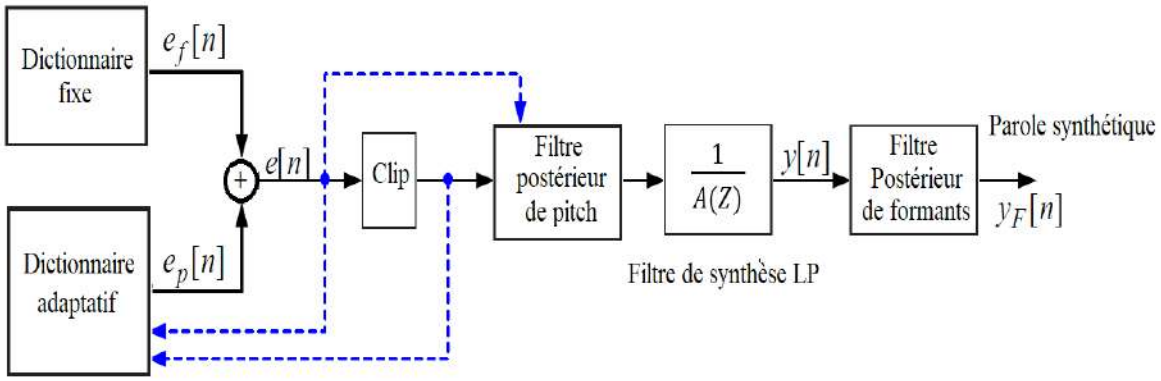


Figure II.7: Le décodeur G.723.1.

Le décodeur prend en charge cinq modes : mode multi-impulsion, mode ACELP, mode SID (silence de descripteur d'insertion), mode de trame nul (silence), mode de perte de paquets.

Considérez les quatre premiers modes pour l'instant. La parole active est codée avec multi-impulsion ou ACELP. Trames SID suivent toujours une trame vocale active. Une trame SID met à jour les paramètres de LP pour une trame « silence » et définit le niveau de bruit de confort (CNG) pour cette trame. Trames nulles génèrent un bruit de confort en utilisant les paramètres envoyés au cours de la trame SID précédente.

II.2.2 Génération d'excitation

L'excitation a deux composantes, l'une du dictionnaire adaptatif et un du dictionnaire de dictionnaire fixe.

II.2.2.1 Contribution dictionnaire adaptatif

Pour les modes de parole actifs (multi-impulsion et ACELP), le traitement des retards et des gains du dictionnaire adaptatif est le même pour les modes de bruit de confort, la contribution ACB (Adaptive Codebook) n'est pas transmise et n'est pas utilisée pour générer une contribution de pitch. Au lieu de cela, il est utilisé pour ajouter un composant aléatoire afin d'accroître la richesse de l'excitation [13]. Pour les modes de parole actifs, la contribution ACB est décodée comme suit :

- Décoder les retards ACB : Il s'agit des décalages absolus pour les sous-trames 0 et 2 et des retards relatifs pour la sous-trame 1 et 3. Notez que les 4 plus grands retards des

sous-trames 0 et 2 sont des valeurs de retard « interdites ». La présence de ces valeurs déclenche le mode de perte de paquets.

- Le codage de gain d'impulsions de 24 niveaux est extrait du codage de gain combiné.
- Pour le codage multi-impulsion : Le codage de coefficient ACB peut prendre 85 ou 170 valeurs. Le dictionnaire de codage à 85 éléments est utilisé pour les décalages de pitch courts (moins de 58 échantillons).
- Pour le codage ACELP : Le codage de coefficient ACB est toujours pris de la table 170 éléments. Le codage de gain combiné laisse certains codes de gain combiné utilisé. Si ceux-ci apparaissent, une perte de paquets est déclarée.

La contribution ACB est déterminée en prenant l'excitation passée, en répétant et en décalant si nécessaire, et en filtrant à l'aide des coefficients du filtre ACB vectoriel. Le filtre de pitch est un filtre IIR multi-tap de la forme :

$$e_p[n] = \sum_{\mathrm{k}=K_l}^{K_u} b_k \, \tilde{e}[\mathrm{n} - k - L] \qquad \text{(II. 9)}$$

Où $\tilde{e}$ [n] est une version à répétition de pitch de l'excitation passée (contenant à la fois les contributions de l'ACB et celles du livre de codes fixe).

$$\tilde{e}[\mathrm{n}] = \pi \begin{cases} e[\mathrm{n}], & \mathrm{n} < 0, \\ e[mod(n, L) - L] & \mathrm{n} \geq 0, \end{cases} \qquad \text{(II. 10)}$$

II.2.2.2 Excitation multi-impulsion

La contribution est déterminée comme multi-impulsion pour chaque sous-trame, comme suit :

- Le bit de grille d'impulsion sélectionne la grille d'impulsions (échantillons impairs ou même dans une sous-trame).
- Les codes de position d'impulsion sélectionner les positions d'impulsions (6 positions de sous-trames 0 et 2, et 5 pour les sous-trames 1 et 3).
- Définir les amplitudes d'impulsion en même temps que les signes d'impulsions dans les positions d'impulsions pour former l'excitation.

- La répétition n'est pas une option pour les sous-trames 0 et 1, si le retard de pitch pour 0 sous-trame est inférieur à 58. De même, pour les sous-trames, 2 et 3, la répétition de pitch peut être utilisée si le retard de pitch pour la sous-trame 2 est inférieur à 58. Si la répétition de pitch est activée et le bit de répétition de pitch est réglé.

La contribution multipulse à l'excitation est [13]:

$$\tilde{e}_f\,[n] = \sum_{\mathrm{k}=1}^{N_f} g_k\,\delta[\mathrm{n} - m_k] \qquad (\mathrm{II}.11)$$

Où les amplitudes des impulsions sont les mêmes, mais les signes peuvent différer. Si le bit de répétition du pitch doit être appliqué, la contribution de l'excitation répétée du pitch est :

$$\tilde{e}_f\,[n] = \sum_{\mathrm{k}=0}^{K} e_f\,[\mathrm{n} - kL], \qquad 0 \leq \mathrm{n} \leq \mathrm{N} - 1 \qquad (\mathrm{II}.12)$$

Où la limite supérieure est $K = [(N - 1) / L]$.

II.2.2.3 Excitation ACELP

La contribution ACELP est déterminée sous-trame par sous-trame, comme suit [13]:

- Le bit de grille d'impulsion sélectionne la grille d'impulsions (échantillons impairs ou même dans une sous-trame).
- Placer une impulsion de chacune des 4 pistes dans la position déterminée par les codes de position ACELP.
- Définir les amplitudes d'impulsion.
- Régler le signe de chaque impulsion en fonction des bits de signe de mode ACELP.
- Répéter une version réduite de l'excitation par impulsions. Le retard de répétition est déterminé à partir du retard de pitch pour une sous-trame, avec un décalage évalué à partir d'une table adressée par le code de coefficient ACB. Si le retard de répétition est inférieur à 58, l'excitation de l'impulsion est décalée de ce retard et mis à l'échelle par un gain déterminé à partir du code de coefficient ACB.

II.2.2.4 Mode SID et le Mode Nul

Les modes SID et Nul génèrent un bruit de confort. Le mode SID définit les paramètres LP et le gain à utiliser pour la génération de bruit de confort. Les paramètres transmis au cours d'une trame SID sont les paramètres LP et une valeur de gain. Aucune information n'a été

transmise pendant une trame nulle. Au cours d'une trame vide, les informations de la trame SID précédente est utilisée [13].

Le niveau d'excitation est commandé par deux variables d'état, le gain de SID et le gain CNG. Pour la première trame SID suivant une trame active, le niveau de bruit de confort est initialisé à la valeur de gain SID reçu.

II.2.2.5 Excitation de coupure

L'excitation est formée par la somme de la contribution ACB et de la contribution de dictionnaire fixe.

$$e\,[n] = e_p[n] + e_f[n] \qquad (\text{II.}13)$$

Lorsque l'excitation est formée sous-trame par sous-trame, l'ACB utilise les valeurs passées pour créer la contribution du pitch à l'excitation.

L'excitation est coupée une fois par trame. Cela signifie qu'en cours de traitement d'une trame, certains échantillons de l'excitation passée sont coupées et d'autres ne sont pas coupées. Par exemple, lors du traitement de la deuxième sous-trame dans une trame, l'excitation passée immédiate de 60 échantillons (une sous-trame) n'est pas coupée, mais des échantillons plus en arrière sont coupés. Cela affecte les calculs de ACB [13].

II.2.3 Filtre postérieur de pitch

Au cours de trames actives, l'excitation coupée est traitée par un filtre postérieur de pitch. Cependant c'est l'excitation non coupée qui est utilisée pour déterminer les paramètres de filtre postérieur de pitch. Le filtre postérieur de pitch n'est pas utilisé pour les modes CNG. La sortie du filtre postérieur de pitch pilote le filtre de synthèse LP.

Si les deux filtres postérieurs de pitch et le filtre de synthèse LP étaient invariants dans le temps, l'ordre des filtres n'aurait pas d'importance. Cependant, il semble avantageux de placer le filtre de pitch en avant du filtre de synthèse LP lorsque les paramètres du filtre de pitch (le retard et le coefficient) changent, ce filtre tend à lisser les transitions [13].

Le filtre postérieur de pitch utilise une formulation similaire à celle du filtre de pondération du bruit harmonique rencontré dans le codeur. Cependant, nous voulons maintenant des creux aux harmoniques de pitch, le filtre HNW à des creux aux harmoniques de pitch. Ce changement est accompli en changeant le signe avant le gain dans le filtre. Le filtre postérieur de pitch est de la forme :

$$y[n] = g_E\big(x[n] + g_{PPf}[n + \mathrm{L}]\big) \quad \text{(II. 14)}$$

Où g_E est un gain global choisi pour rendre l'énergie du signal de sortie égale à l'énergie du signal d'entrée.

II.2.4 Paramètres LP

Pour les modes actifs (Multi-impulsion et ACELP) et le mode de SID, les paramètres LP sont envoyés dans le flux binaire, ceux-ci sont traités comme suit :

- Décoder les valeurs LSF quantifiées du flux binaire.
- Corriger les valeurs LSF mal ordonnées ou proches les unes des autres.
- Interpoler les coefficients LSF quantifiées à partir des coefficients LSF quantifiées de la dernière trame pour créer un vecteur de coefficient LSF quantifiées pour chaque sous-trame.
- Convertir les valeurs LSF interpolées en valeurs de coefficient LP correspondantes.

Le signal d'excitation est filtré avec le filtre tous-pôles spécifiés par les paramètres LP. Les coefficients sont mis à jour sous-trame par sous-trame. Un exemple de génération de signal vocal synthétique est présenté à la figure II.8 [14].

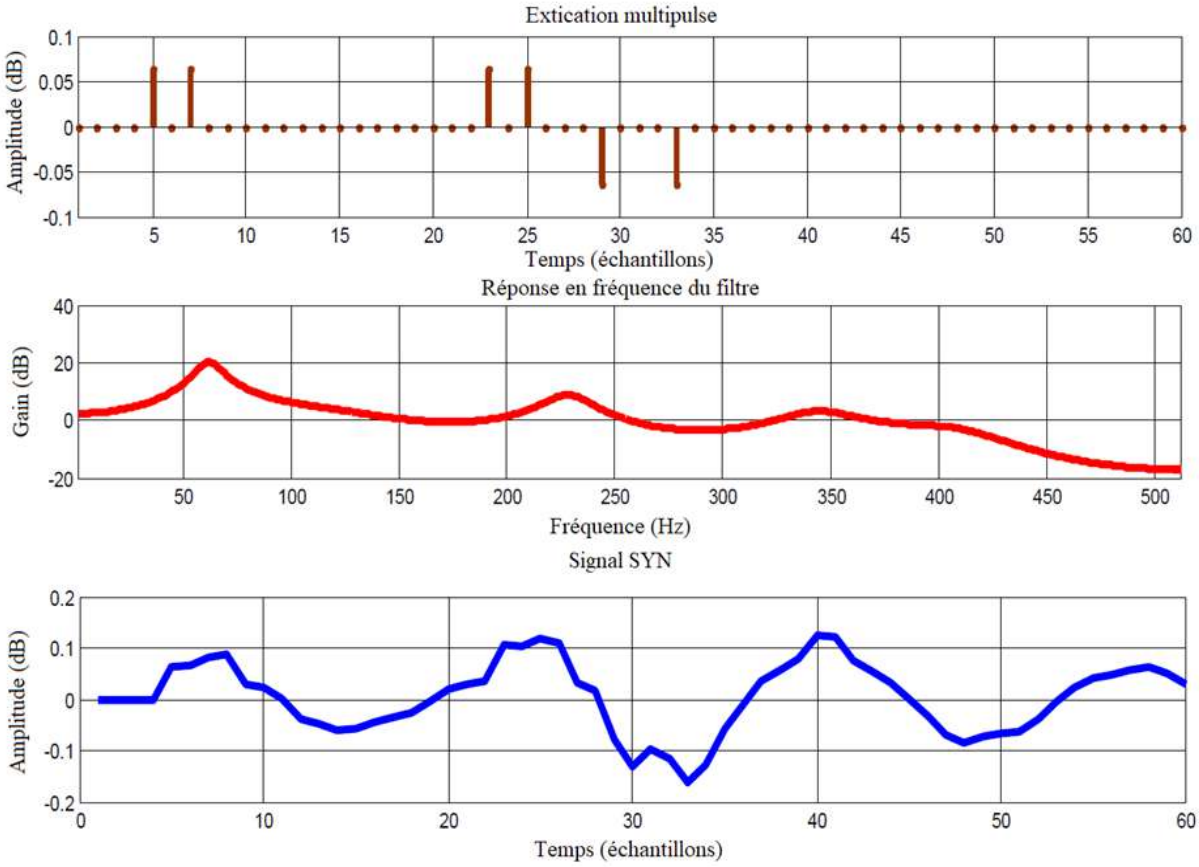


Figure II.8: Génération de signal de parole.

II.2.5 Le filtre postérieur de formants

Le filtre postérieur de formants est appliqué pour tous les modes après le filtre de synthèse LP. Le format de filtre postérieur est mis à jour chaque sous-trame. Exprimé comme suit :

$$F_f(z) = \frac{A(y_n z)}{A(y_d z)} \left(1 - a r_y z^{-1}\right) \qquad (II.15)$$

Où, les facteurs d'expansion de la bande passante sont $y_n = 0{,}65$ et $y_d = 0{,}75$. Le deuxième terme du filtre est une compensation adaptative « d'inclinaison » qui dépend de la corrélation sur un retard de 1. Le paramètre est 0,25 [14].

Un exemple de filtre postérieur du formant est illustré à la figure II.9. Le spectre LP est identique à celui utilisé dans l'exemple d'un filtre de pondération de formants (cf. figure II.5). Pour ce graphe, l'autocorrélation en une période a été prise à partir des valeurs de corrélation utilisées pour générer les coefficients LP. Cet exemple a une réponse assez basse LP, mais le filtre postérieur est assez plat. La ligne pointillée de la figure II.9 montre le filtre postérieur sans inclinaison de compensation.

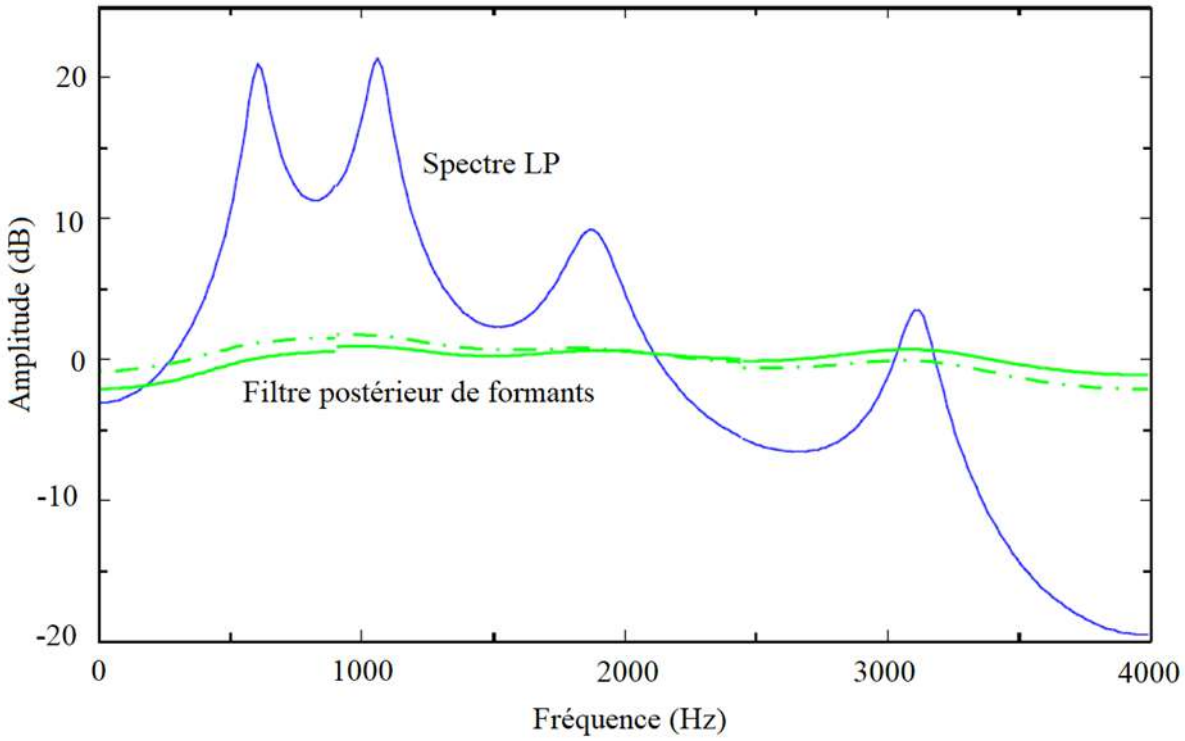


Figure II.9: Réponse en fréquence du filtre postérieur de formants.

Conclusion

Dans ce chapitre nous avons présenté une description générale du codeur G.723.1 à base de CELP. Afin de bien pouvoir comprendre ce codeur, une explication détaillée sur le traitement au niveau de la trame. Et afin de mieux expliquer les différentes étapes d'estimation de coefficients de prédiction linéaire (LPC) qui seront optimisées et quantifiées. Puis, nous avons parlé sur le filtre de pondération des formants qui permet d'atténuer des régions des pics. Cependant, il est clair qu'en présence de tous ces paramètres optimisés au niveau du codeur, une fois reçus au niveau du décodeur, la reproduction du signal de parole paraît la plus simple.

Chapitre III

Méthode stéganographique basée sur l'interpolation polynomiale des coefficients LSFs sur le codeur G.723.1 de la VoIP

Introduction

Depuis la naissance des grandes civilisations, le problème d'échange des données secrètes est toujours existé. La cryptographie offre un moyen efficace pour protéger les données secrètes en les rendant inintelligibles aux personnes non autorisées qui peuvent détruire la communication entre les deux parties. Une communication contenant un message secret entre deux personnes doit sembler une autre chose aux yeux / oreilles de la personne qui contrôle le canal.

Aujourd'hui, la stéganographie est une méthode principale utilisée pour protéger les données, pour envoyer de manière fiable des informations secrètes, et pour cacher un message secret au sein d'un médium hôte (couverture), de telle sorte que le médium résultant (stego) semble inaffecté (dissimulation indétectable) par l'insertion du message secret. Autrement dit, l'objectif est de rendre difficile, ou impossible, la distinction entre un médium original et un médium modifié par l'insertion d'un message secret.

III.1 Système de transmission stéganographique

La stéganographie est l'art de la dissimulation de communications. Contrairement à la cryptographie, la stéganographie pour objectif de sécuriser une communication, par d'en cacher l'existence même. Les deux disciplines ont donc chacune leur propre domaine de compétence. Dans certaines situations, la faite même de vouloir transmettre des données de manière chiffrée sera jugé comme suspect. De la même manière, de plus en plus de pays mettent en place de forte restriction concernant la longueur des clés cryptographiques, ainsi que la cryptographie en elle-même. Le regain d'intérêt actuel pour la stéganographie provient de ces restrictions imposées à la cryptographie. D'une manière générale, la stéganographie arrive en renforcement au chiffrement de données. Pour permettre de garantir une confidentialité maximum, les données sont tout d'abord chiffrées avant d'être dissimulées à l'aide d'un processus stéganographique [16]. La figure III.1 résume les différents types de masquage de données :

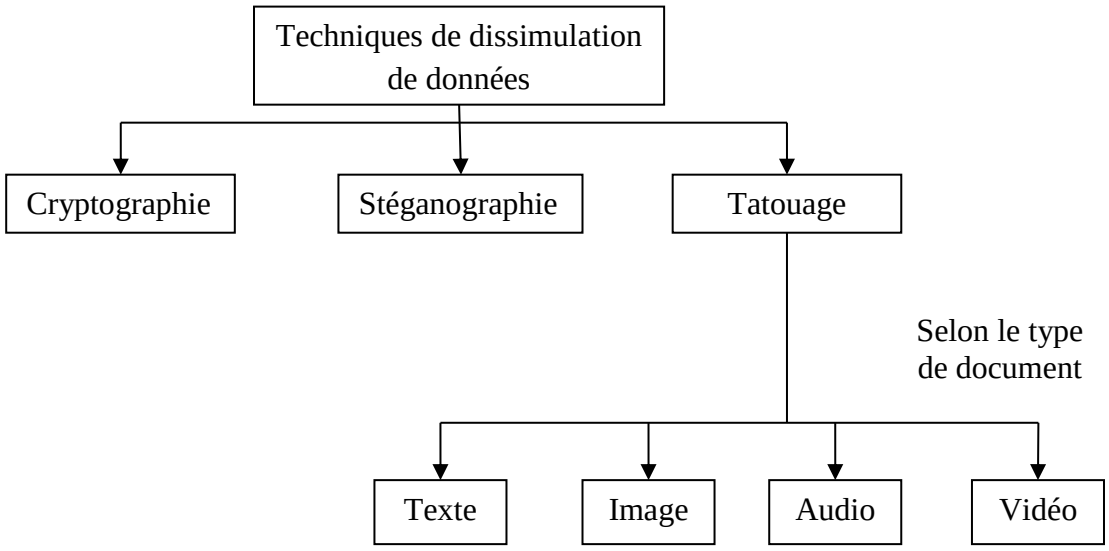


Figure III.1: Types de masquage de données.

III.1.1 La Stéganographie classique

La stéganographie classique ne s'intéresse généralement qu'à la récupération de l'information, la couverture est utilisée uniquement pour la transporter. Cette couverture devrait être un fichier dont la transmission ne pose pas de suspicion. Il existe un cas particulier de stéganographie où la couverture et les informations cachées sont importantes. C'est le cas lorsque l'incorporation d'informations servant de signature d'auteur protège une propriété numérique. C'est généralement appelé tatouage [17].

III.1.2 La Stéganographie moderne

La stéganographie dite moderne, c'est-à-dire adaptée aux données numériques, est relativement nouvelle. En pleine expansion, elle suit depuis le milieu des années 90 un essor corrélé celui d'internet. La première étape est bien évidemment de définir précisément l'objet que l'on va étudier. Le lecteur pourra se référer à d'excellents ouvrages traitant par la dissimulation d'information. Bien que la communauté des stéganographie se soit constituée dans les années 90, le socle de la stéganographie moderne en définissant la notion de canal caché. Pour illustrer son propos, il reprend le problème de deux prisonniers. Deux prisonniers souhaitent établir un plan d'évasion. Pour ce faire, ils ont la possibilité de se transmettre des messages. Cependant, ces messages transitent à travers le gardien, qui a donc accès au contenu. Tout contenu juger inapproprié sera détruit, et pourra engendrer une lourde sanction. Dans cette optique, le contenu doit donc paraître anodin aux yeux du gardien. Cette situation

rend la cryptographie inefficace, car un contenu indéchiffrable attirera l'attention du gardien [16]. La stéganographie a donc comme objectif de pouvoir entretenir des communications sécurisées, sans pour autant attirer l'attention d'autrui.

III.1.3 Système de transmission stéganographique de la parole

Le modèle de base de la stéganographie audio se compose de couverture (fichier audio), message et mot de passe. Le porteur est également connu sous le nom de fichier de couverture, qui cache les informations secrètes [18].

Fondamentalement, le modèle de base de la stéganographie audio est montré par la figure III.2. Le message est les données que l'expéditeur souhaite le garder confidentiel. Le message peut être texte simple, image, audio …etc.

Le mot de passe est connu sous le nom de clé stego, qui garantit que seul le destinataire à la clé de décodage correspondante qui sera capable d'extraire le message secret d'un fichier de couverture. Le fichier de couverture avec les informations secrètes est connu comme un stego-fichier.

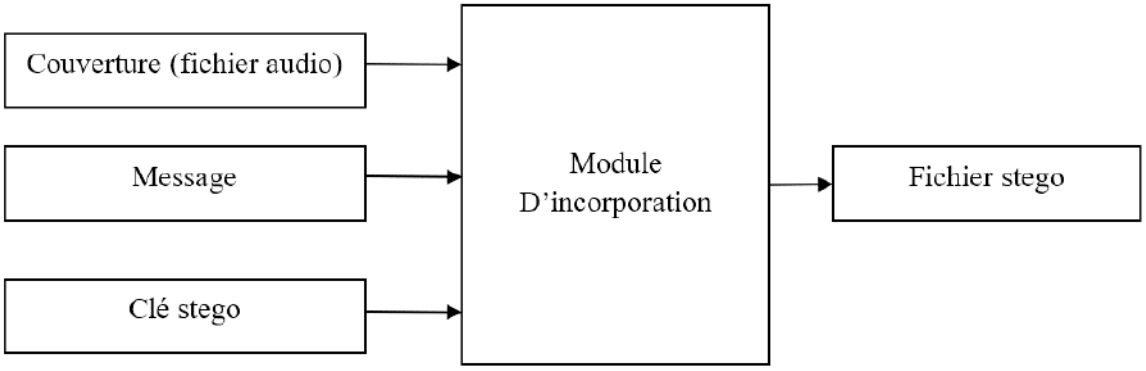


Figure III.2: Modèle stéganographique audio.

Le processus de dissimulation de l'information consiste à suivre deux étapes :

- Identification des bits redondants dans un fichier de couverture. Les bits redondants sont les bits qu'ils peuvent modifiés sans corrompre la qualité ou détruire l'intégrité du fichier de couverture.
- Pour incorporer les informations secrètes dans le fichier de couverture, les bits redondants dans le fichier de couverture sont remplacés par les bits de l'information secrète.

III.1.4 Comparaison entre la stéganographie et le tatouage

La stéganographique et le tatouage peuvent paraître semblables car ils ont la même exigence de distinction (de l'existence du message) ; pourtant, leur différence est fondamentale : la stéganographie recherche une distinction statistique, alors que le tatouage recherche une distinction *perceptuelle*. Le tableau III.1. Montre la différence entre le stéganographique et le tatouage [19].

Tableau III.1: Comparaison des caractéristiques de la stéganographie et tatouage.

		Tatouage	**Stéganographie**
Le but		Protéger la couverture	Protéger l'information secrète
Les caractéristiques	Sécurité	Invisibilité ou visibilité perceptuelle selon les besoins	Les informations incorporées sont « Invisibles » à un spectateur inconscient
	Robustesse	Robustesse contre la falsification ou l'enlèvement	Robustesse contre la détection
	Les attaques	Ne doit pas conduire à la perte de tatouage	Peut entraîner la perte de données cachées
	Type de couverture	Fichiers numériques - audio, vidéo, texte ou images	Tout service, protocole, fichier, environnement utilisant le numérique représentation des données

III.2 Classification des systèmes stéganographiques de la parole

On trouve dans la parole une certaine classification des systèmes stéganographiques comme suivant :

III.2.1 Selon la clé

Selon la clé de cryptage on peut distinguer deux types de stéganographie :

III.2.1.1 Stéganographie à clé publique

Chaque entité possède 2 clés distinctes (une privée S_k, une publiqueP_k) avec impossibilité de déduire la clé privée à partir de la clé publique. De ce fait, il est possible de distribuer librement cette dernière.

III.2.1.2 Stéganographie à clé secret (privé)

Est une ancienne, simple et plus rapide que la clé publique, elle se base sur les formules arithmétiques et décalage circulaire entre les caractères [20].

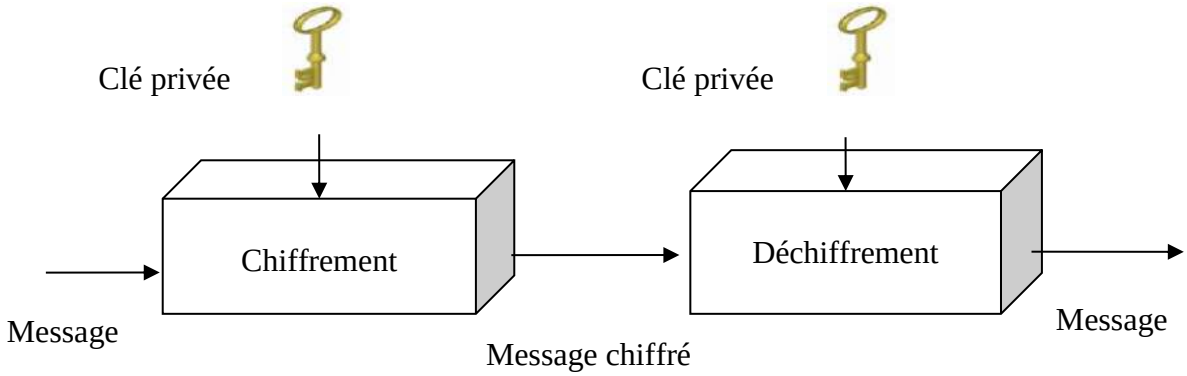


Figure III.3: Chiffrement par clé secret.

III.2.2 Selon les domaines de stéganographie

Il existe trois types de domaines stéganographiques :

III.2.2.1 Domaine de fréquence

Dans le domaine fréquentiel, il existe certaines techniques utilisées dans la stéganographie audio, comme l'insertion de la tonalité, le codage de phase et les techniques de spectre étalé. Dans les techniques d'insertion de tonalité, les tonalités de faible puissance sont masquées en présence d'une plus forte. C'est une méthode robuste mais à faible capacité d'intégration. Dans la méthode de codage de phase, les bits de message secret sont intégrés comme déphasage dans le spectre de phase de l'audio de couverture originale. Cette méthode a une bonne robustesse, mais a également une faible capacité et ne survit pas au filtrage passe-bas. Les techniques de propagation du spectre calculent le seuil de masquage de fréquence en utilisant un modèle psychoacoustique, où le signal de données est propagé par un code de séquence M et le signal de propagation est intégré dans l'audio sous le seuil de masquage de fréquence. Cette méthode augmente la transparence, mais elle occupe une large bande passante [21].

III.2.2.2 Domaine temporel

Sous le domaine temporel, les techniques incluent la technique de bit LSB (Least Significant Byte), le codage de parité et les techniques de dissimulation d'écho. La méthode LSB modifie le bit le plus bas du support de couverture pour intégrer le message secret. C'est l'une des méthodes les plus simples, mais il peut être facilement attaqué. Dans la technique de codage de parité, la parité de chaque groupe d'échantillons est calculée, si elle ne correspond pas au bit du message, le bit le plus bas de l'un des échantillons individuels du groupe est modifié pour rendre le bit de parité égal à bit de message. Cette technique est également facile à extraire. Dans la méthode de dissimulation d'écho, un petit écho est introduit dans des parties de l'audio de la couverture avec une manipulation de trois paramètres : taux de décroissance, décalage et amplitude pour rendre l'écho inaudible, mais la capacité d'inclusion ici est faible [21].

III.2.2.3 Domaine de compression

La compression est une opération où l'information peut être réduite pour diminuer le débit et le temps de transmission. Ces informations peuvent être une image, texte, audio, data, vidéo, etc. Il existe deux types de techniques de compression qui sont les suivantes :

a. Compression sans perte : Elle réduit la quantité de données source transmises de telle sorte que lorsque des données compressées sont décompressées, aucune altération se produite.

b. Compression avec perte : Elle ne récupère pas une copie exacte des informations après la phase de décompression. Ici, certaines informations sont perdues après la décompression. MP3 (MPEG1 Audio Layer 3) est un exemple de compression avec perte utilisé pour compresser le flux continu de son en petits fichiers [22].

III.3 Évaluation de la qualité perceptuelle de la parole

La qualité vocale peut être mesurée en utilisant des méthodes subjectives ou objectives. Les mesures subjectives utilisent souvent le MOS (Mean Opinion Square). Cependant cette procédure est lente et difficile à mettre en œuvre. Ainsi, les notes des participants pour une condition de test donnée sont moyennées pour obtenir la note moyenne d'opinion, qui permet d'exprimer l'effet subjectif sur l'évaluation de la qualité vocale. De plus, la perception de la qualité vocale dépend du contexte et de l'environnement dans lesquels est placée la personne qui juge. De même, l'environnement (bruit, informations visuelles ou sonores supplémentaires) influence le jugement de la qualité. Ainsi, les conditions à tester sont

définies en fonction de l'objectif visé. Le participant est amené à évoluer dans un ou plusieurs contextes (écoute, locution et conversation).

III.3.1 Évaluation perceptuelle de la qualité vocale

La méthode PESQ (Perceptual Evaluation Of Speech Quality) est utilisée pour la prédiction de la qualité subjective pour la téléphonie et pour les codeurs vocaux. Elle est destinée à évaluer l'influence de certains facteurs tels que la perte de paquets, le délai variable et les distorsions dues aux erreurs de canal et qui sont mal évaluées par les méthodes classiques RSB (Rapport Signal/Bruit), RSB segmental, DS (Distorsion Spectral), etc.

La méthode PESQ est conçue pour comparer une version de référence (originale) à celle obtenue par synthèse à partir de cette référence, c'est le cas par exemple des dégradations que subit cette référence après transmission ou après une compression. L'idée de base du PESQ est de transformer les formes d'onde en une représentation perceptuelle, similaire à la représentation des paramètres utilisés par les vocodeurs à bande étroite. En d'autres termes, la différence entre les paramètres de ces représentations est évaluée à l'aide d'un modèle cognitif en vue d'estimer la distance de perception entre les deux signaux : le signal original et le signal synthétique (décodé) (figure III.4). Le tableau III.2 donne un aperçu des limites des évaluations de la qualité selon la recommandation P.862 [23]:

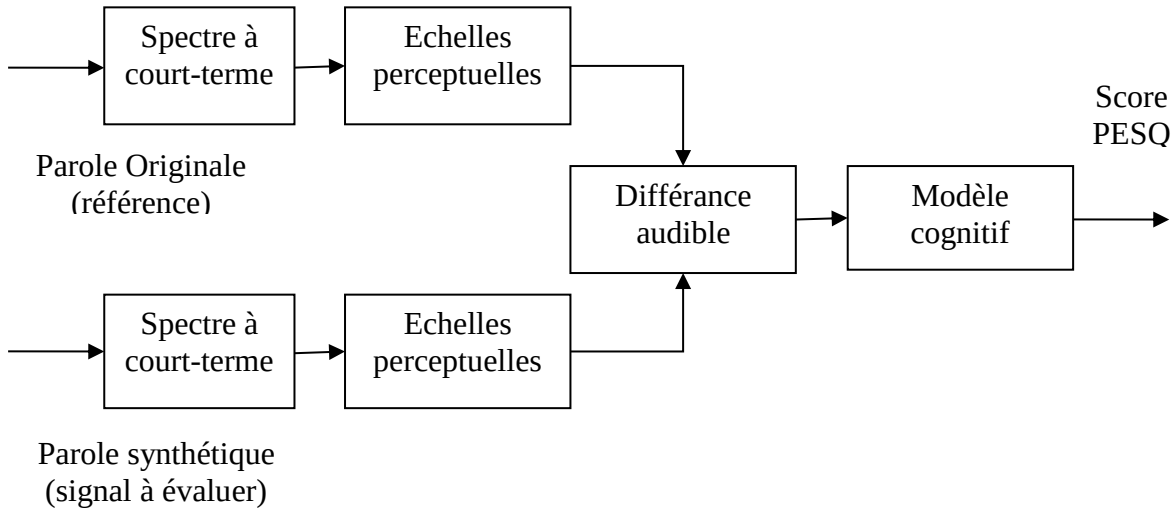


Figure III.4: Schéma synoptique permettant l'estimation la distance perceptuelle PESQ.

Tableau III.2: Limites des évaluations de la qualité de parole selon la recommandation P.862.

Valeur du PESQ	**Évaluation de qualité de parole**
$3 \leq PESQ \leq 4$	Très acceptable
$2.5 \leq PESQ < 3$	Acceptable
$2 \leq PESQ < 2.5$	Basse
$PESQ < 2$	Inacceptable (l'intelligibilité est perdue)

III.4 Description des signaux de parole utilisés dans les tests

Pour tester et valider nos méthodes, nous avons utilisé un matériau linguistique, ce dernier est composé de phrases en anglais TIMIT (Acoustic-Phonetic Speech Data Base) [24] pris de la base de données internationale TIMIT. Ce corpus contient un total de 20 phrases, soit 10 phrases prononcées par 10 locutrices féminins et les autres phrases prononcées 10 locuteurs masculins. La fréquence d'échantillonnage du signal parole des fichiers était de 16kHz, nous avons dû effectuer un sous-échantillonnage de toute la base de données à 8 kHz, pour se mettre dans les conditions de la téléphonie. Pour la langue anglaise, nous avons utilisé des différentes phrases célèbres, phonétiquement équilibrées : « She had your dark suit in greasy wash water all year ».

III.5 Évaluations des codeurs AMR implémentés

L'utilisation de codeurs pour transmettre de la voix sur un canal de communication induit une baisse de la qualité perçue. Cette baisse est due au mécanisme de compression des données utilisées. Par conséquent, pour chaque codeur, il y a un score PESQ maximal qui peut être obtenu. Il est évident que lorsque des dégradations apparaissent dans le réseau, ses performances doivent seulement être estimées par rapport à ce score maximum. Nous avons donc évalué les performances des codeurs AMR fonctionnant à 6.3 kbps et 5.3 kbps respectivement pour les locuteurs masculins et les locutrices. Nous résumons les résultats obtenus dans le tableau III.3.

Tableau III.3: Résultats des tests objectifs de deux codeurs AMR6.4 kbps et 5.6kbps.

	AMR à 6.3 kbps Multipulse (PESQ)	**AMR à 5.3 kbps ACELP (PESQ)**
Locuteurs masculins	3.783	3.781
Locutrices féminines	3.550	3.545

D'après ces résultats, nous constatons que les scores PESQ obtenus sont meilleurs pour les locuteurs masculins que pour les locutrices, dans le cas des deux codeurs AMR. La perte de la qualité observée pour le AMR à 5.3 kbps comparativement au AMR à 6.3 kbps était prévue, vue les taux de compression utilisé dans les deux cas [23].

III.6 Outils de travail

Matlab (Matrix Laboratory) est le logiciel utilisé pour l'implémentation de notre algorithme de compression, il présente plusieurs avantages pour le traitement du signal, il a été choisi pour sa simplicité et sa puissance. Aussi, il existe déjà un code source de codeur G.723.1 en langage Matlab.

Vu le nombre de fonctions offertes et facilitant la manipulation de signal de parole, nous avons décidé de se servir de Matlab. En fait, il est très aisé de lire, afficher, conversion et l'interpolation des paramètres LP et LSF sous Matlab.

III.7 Méthode implémenté

Dans nos applications, nous utilisons le codeur de parole AMR G.723.1à 6.3 kbps et à 5.3 kbps (cf. chapitre II) pour générer le message secret à dissimuler. Ce message sera réorganisé selon les caractéristiques de la méthode de dissimulation (débit de dissimulation, nombre de bits à dissimuler à la fois, etc.).

III.7.1 Interpolation linéaire

Les coefficients LP sont convertis en paramètres LSF. Cela se fait en recherchant des racines entre des valeurs discrètes [25], puis en utilisant une interpolation linéaire entre de deux ensembles de coefficients LSF (calculés pour deux trames d'analyse successives n et $(n+1)$ pour former un ensemble intermédiaire pour chacune des 4 sous-trames de la trame à coder (figure III.5). L'interpolation linéaire utilise la pondération du tableau III.4 [26].

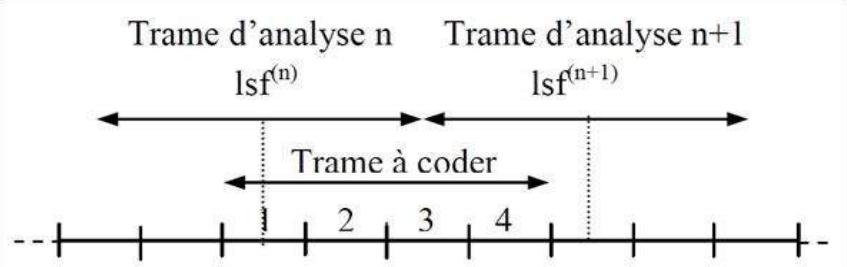


Figure III.5: Interpolation des LP dans le domaine LSF.

Tableau III.4: Facteurs de pondération.

Numéro sous-trame	**Lsf$^{(n)}$**	**Lsf$^{(n+1)}$**
1	1	0
2	0.75	0.25
3	0.50	0.50
4	0.25	0.75

La trame d'analyse désigne la trame de parole (de 30 ms) qui est pondérée par la fenêtre de Hamming et à partir de laquelle les coefficients LSF sont calculés. La trame à coder désigne la trame de parole à partir de laquelle le codage AMR est effectué. Notons que le centre de la trame d'analyse est aligné avec le centre de la première sous-trame de la trame à coder, ce qui implique que les coefficients LSF calculés à partir de cette trame d'analyse sont ceux de la première sous-trame, et les coefficients des autres sous-trames (numéro 2, 3 et 4) sont obtenus par interpolation entre les coefficients LSF calculés à partir de la trame d'analyse courante et de la trame d'analyse suivante [26]. La figure III.6 montre les deux premières trames à coder. Chaque trame est divisée en 4 sous-trames et chaque sous-trame est représentée par son propre ensemble de coefficients LSF calculé selon l'équation III.1. Notons que dans cette figure, le symbole "X" représente les coefficients LSF originaux et le symbole "O" représente les coefficients LSF interpolés.

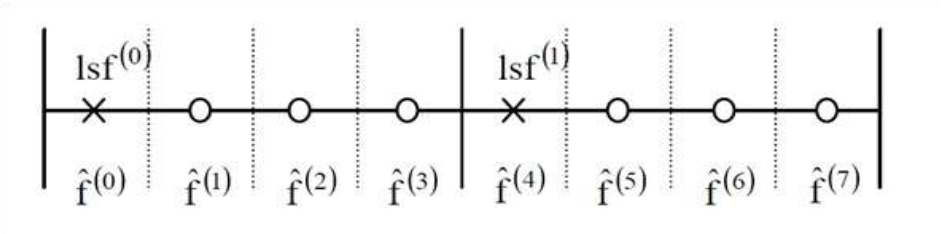


Figure III.6: Exemples de vecteurs LSF interpolés correspondants à huit sous-trames successives.

La fonction d'interpolation linéaire est démontrée comme suit :

$$p(x) = (1 - a) * lsfp + a * lsf \qquad \text{(III.1)}$$

III.7.2 Interpolation polynomiale

Les méthodes d'interpolation des coefficients LSF les plus récentes n'utilisent que l'interpolation linéaire. Cependant, l'application de différentes méthodes d'interpolation dédiées à la modification du signal de parole [27].

Dans cette étude d'interpolation des coefficients LSF, nous avons utilisé une interpolation polynomiale, que nous avons appliquée sur le codeur G.723.1, tout en gardant l'interpolation originale.

L'interpolation polynomiale est une technique d'interpolation d'un ensemble de données ou d'une fonction par un polynôme. En d'autres termes, étant donné un ensemble de points (obtenu, par exemple, à la suite d'une expérience), on cherche un polynôme qui passe par tous ces points, et éventuellement vérifie d'autres conditions, de degré si possible le plus bas.

Il existe un et un seul polynôme P_n de degré inférieur ou égal à *n* vérifiant (équation III.2).

$$P_n(x_i) = \mathrm{f}\,(x_i) \qquad \forall i = 0, \dots, n \qquad \text{(III. 2)}$$

Le polynôme s'écrit :

$$P_n(\mathrm{x}) = \sum_{i=0}^{n} \mathrm{f}\,(x_i) L_i(\mathrm{x})$$

$$L_n(\mathrm{x}) = \prod_{j=1, j\neq i}^{n+1} \frac{x - x_j}{x_i - x_j} \qquad \text{(III. 3)}$$

Le polynôme P_n est appelé polynôme d'interpolation de Lagrange de la fonction f aux points $x_0, x_1, \dots, x_n$.

Les polynômes $L_i(x)$ sont appelés polynômes de base de Lagrange associes à ces n + 1 points $(x_0,\ldots,x_n)$, et ils sont tous de degré *n*.

III.7.3 L'intérêt de changement d'interpolation

Le changement d'interpolation utiliser dans notre étude donne plusieurs avantages qui facilitent le travail aux paramètres LSF, parmi ces avantages :

- Pour avoir la flexibilité de modification des LSF interpolé.
- Garder l'interpolation linéaire pour les utilisateurs qui n'ont pas utilisé la stéganographie.
- Avoir plus des coefficients d'interpolation, donc avoir plus de difficulté de régénérer le polynôme d'interpolation ce qui rendre l'extraction de message secret difficile, donc on a plus de sécurité.

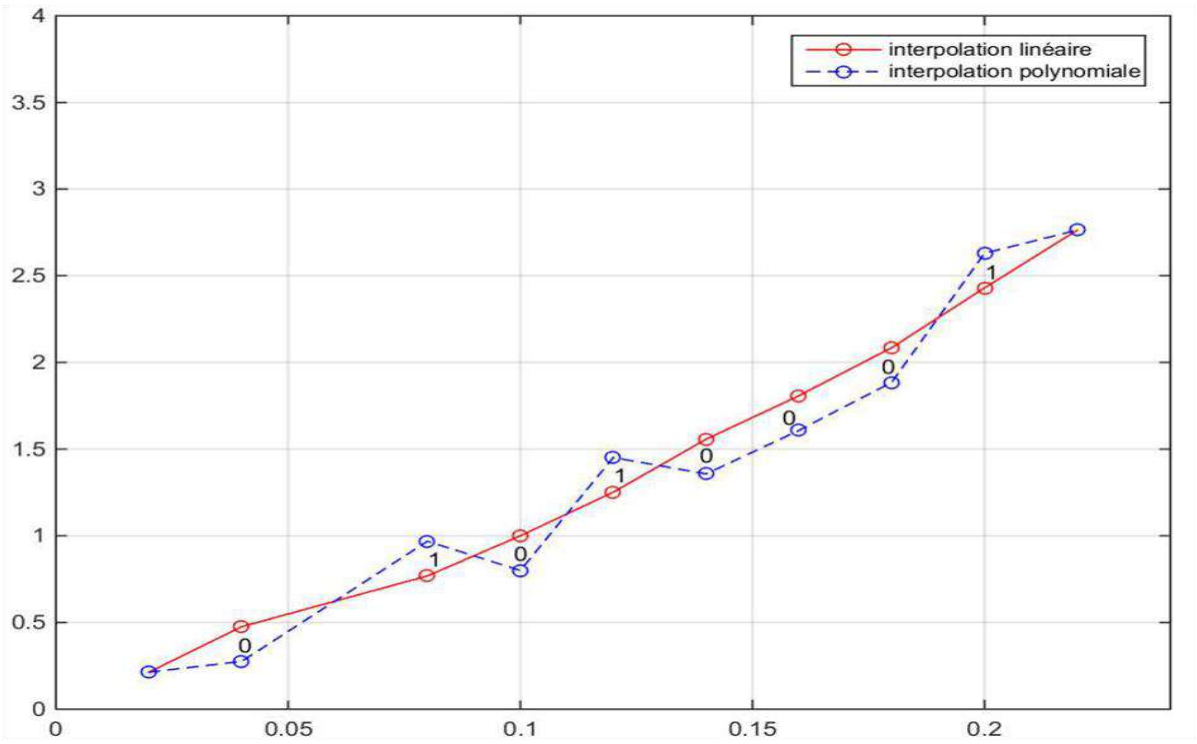


Figure III.7: Interpolation linéaire et polynomiale pour dissimuler un message secret de 8 bits "01010001".

On ne prend pas la première et la dernière des 10 LSFs pour ne pas obtenir une distorsion du signal ainsi pour récupérer notre signal à la réception parce que la première et la derrière valeur se sont des repères (figure III.7).

III.7.4 Capacité

Dans le contexte de la stéganographie, où la seule contrainte est celle de l'indétectabilité, définissent la capacité comme le nombre maximum de bits d'information que l'on pourrait cacher dans un fichier-couverture, tout en gardant la probabilité de détection par l'adversaire est négligeable.

Rappelons que chaque trame est subdivisée en quatre sous-trames. Chaque sous-trame est représentée par 10 coefficients où un nombre maximum de 8 bits peut être dissimulé par chaque sous-trame. On a donc une possibilité de dissimuler jusqu'à 32 bits par chaque trame. Autrement dit, notre méthode proposée offre un canal caché est donnée :

$$capacité\ de\ canal\ caché = nbit \times R \div n\ total\ des\ coefficient \qquad \text{(III. 4)}$$

Un canal caché de 1.05 kbps ($32 \times 6.3 \div 192 = 1.05$) pour le mode 6.3 kbps et offre aussi un canal caché de 1.06 kbps ($32 \times 5.3 \div 160 = 1.06$) pour le mode 5.3 kbps.

III.8 Résultat d'implémentation

Pour les simulations suivantes, nous utilisons toujours le G.723.1 en mode 6.3 kbps et 5.3 kbps sont utilisés pour interpoler les paramètres LSF.

La base de données utilisée est composée de 20 séquences de parole de locuteurs masculins et féminins extraites de la base de données TIMIT, les tests reliés à chaque étude seront représentés sous forme des tableaux.

III.8.1 Résultat obtenue avant appliquer l'interpolation polynomiale

Pour cette méthode on a codé chaque fichier WAV (Waveform Audio File Format) par le codeur G.723.1 qui utilise l'interpolation linéaire, le tableau III.5 et tableau III.6 résume les différentes valeurs de PESQ pour locuteurs masculins et les locutrices féminins.

Tableau III.5: Qualité de parole avant appliquer interpolation polynomiale pour les locuteurs masculins.

Fichier	PESQ
1	3.783
2	3.755
3	3.746
4	3.732
5	3.728
6	3.720
7	3.715
8	3.709
9	3.706
10	3.704

Tableau III.6: Qualité de parole avant appliquer interpolation polynomiale pour les locutrices féminines.

Fichier	PESQ
11	3.696
12	3.616
13	3.569
14	3.565
15	3.550
16	3.545
17	3.541
18	3.535
19	3.530
20	3.527

III.8.2 Résultat obtenue après appliqué interpolation polynomiale

Dans cette méthode on a codé les mêmes fichiers WAV sur lesquels nous avons travaillé auparavant après appliquer interpolation polynomiale le tableau III.7 et tableau III.8 résume les différentes valeurs de PESQ.

Tableau III.7: Qualité de parole après appliquer interpolation polynomiale pour les locuteurs masculins.

Fichier	PESQ
1	3.783
2	3.755
3	3.746
4	3.732
5	3.728
6	3.720
7	3.715
8	3.709
9	3.706
10	3.704

Tableau III.8: Qualité de parole après appliquer interpolation polynomiale pour les locutrices féminines.

Fichier	PESQ
11	3.696
12	3.616
13	3.569
14	3.565
15	3.550
16	3.545
17	3.541
18	3.535
19	3.530
20	3.527

Selon les valeurs de PESQ obtenues pour chaque méthode avant et après l'interpolation polynomiale, on a constaté que l'interpolation polynomiale peut nous donne un résultat identique (LSF identique) comparativement à l'interpolation linéaire avant de dissimuler un message secret.

III.8.3 Seuil de modification

Pour cette manipulation on va prend un échantillon de locuteur parmi les 20 locuteurs que nous avons choisi, à chaque fois on a ajouté une valeur de tel sort nous obtiendrons une incrémentation logarithmique des LSF. Ensuite nous avons appliqué notre méthode utilisée.

Le tableau III.9 résume les différentes valeurs de PESQ obtenues pour chaque opération d'addition.

Tableau III.9: Les valeurs de PESQ obtenus à partir de variation de seuil.

Seuil	**PESQ**
0.01	3.751
0.02	3.741
0.04	3.670
0.08	3.442
0.1	3.421
0.12	3.262
0.14	2.966
0.16	2.903
0.18	2.836
0.2	2.766
0.22	2.571
0.24	2.617
0.26	2.222
0.28	2.246
0.3	1.872
0.32	1.961
0.34	1.545
0.36	1.503
0.38	1.797
0.4	1.517
0.42	1.536
0.44	1.283
0.46	1.342
0.48	1.181
0.5	0.899
0.52	0.183
0.54	0.185
0.56	0.435
0.58	0.590
0.6	1.604
0.61	1.599

D’après ces résultats, on remarque que la variation de la valeur de PESQ est liée aux valeurs de seuil modifié dans chaque case présentée dans le tableau, le meilleur résultat de seuil a obtenu lorsqu’on avoir une qualité de parole acceptable. Par exemple, l’utilisation

d'une valeur de seuil minimal égale à 0.12 donne une note PESQ moyen égale à 3.262 ce qui représente le résultat acceptable le plus bas dans le cas de codage moyenne débit.

Si nous prenons une valeur de seuil supérieure à 0.12 nous aurons une valeur de PESQ inférieure à 3.2 ce qui signifie que nous allons travailler avec le résultat équivalent de codeur à bas débit tel que le codage MELP. La figure III.8 présente les valeurs de PESQ obtenus à partir de variation de seuil.

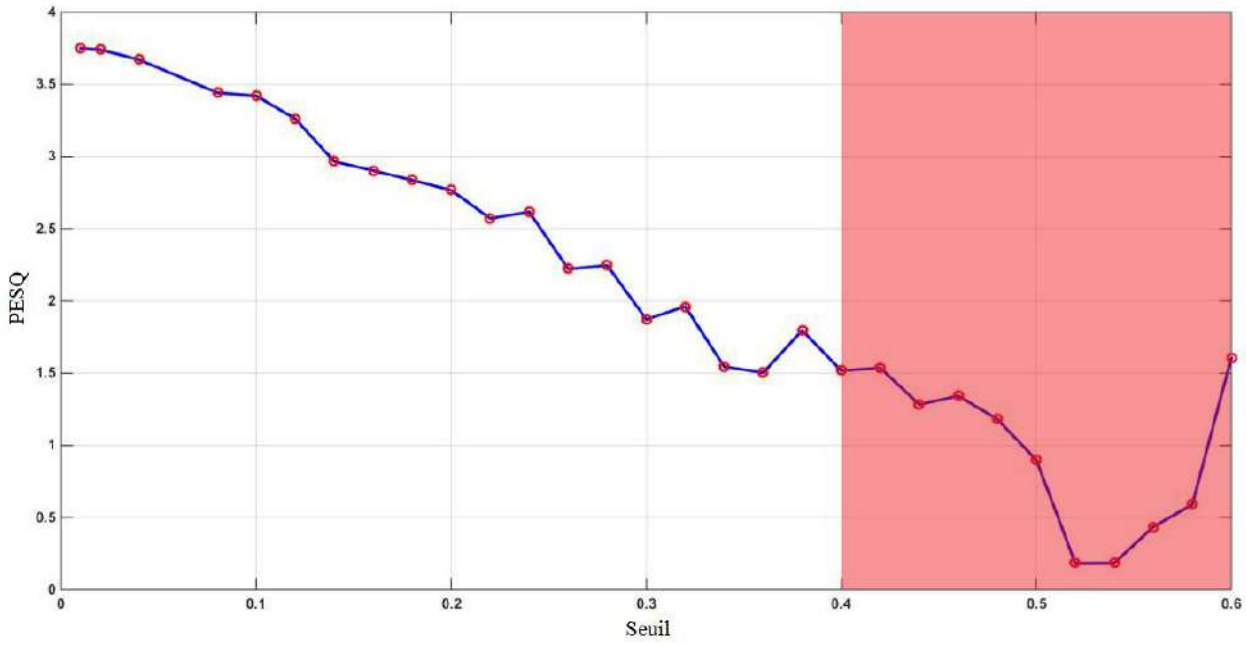


Figure III.8: PESQ obtenus après la variation de seuil.

La variation de PESQ soit décroissant jusqu'à la valeur de seuil 0.4, après cette valeur la variation est aléatoire, ayant toujours $PESQ \leq 1.5$ (zone rouge) comme illustré à la figure III.8.

III.8.4 Résultat obtenue après l'application de méthode stéganographique

Notre but dans cette manipulation est de masquer parfaitement le message secret dans le signal de base. Alors on utilise le changement des coefficients LSF pour transmit et dissimuler le message secret.

On prend un exemple de la dissimulation d'un message dans une trame. Le tableau III.10 montre la trame avant la transmission d'un message secret.

Tableau III.10: les valeurs des LSF avant appliquer la méthode de dissimulation.

LSF S.trame 1	**LSF S.trame 2**	**LSF S.trame 3**	**LSF S.trame 4**
0.2136	0.2071	0.2006	0.1941
0.4760	0.4542	0.4324	0.4107
0.7689	0.7822	0.7956	0.8089
1.0002	1.0103	1.0204	1.0304
1.2517	1.2620	1.2722	1.2825
1.5572	1.5565	1.5558	1.5551
1.8081	1.8079	1.8077	1.8075
2.0848	2.0936	2.1024	2.1111
2.4296	2.4363	2.4429	2.4495
2.7634	2.7670	2.7707	2.7743

Dans un simple scénario de dissimulation d'information, nous avons fait un essai pour transmise un ensemble de bit « 0 1 0 1 0 0 0 1 » dans notre signal de base. L'algorithme lire le premier élément de vecteur, si l'élément est un bit 0 alors le l'algorithme de notre méthode développée soustraire une valeur de seuil égale à 0.01 du deuxième coefficient LSF au niveau de la sous trame.

La fonction suivante démontrée comment transmit un bit 0 :

$$\mathrm{LSF} = \mathrm{LSF} - \mathrm{seuil} \qquad (\mathrm{III}.5)$$

Si le l'élément est un bit 1 alors l'algorithme ajoute une valeur de seuil au LSF. La fonction suivante démontrée comment transmit un bit 1:

$$\mathrm{LSF} = \mathrm{LSF} + \mathrm{seuil} \qquad (\mathrm{III}.6)$$

Le tableau III.11 montre le changement des coefficients LSF obtenue au niveau de la trame après l'application de la méthode de dissimulation.

Donc le codeur restera appliquer ces étapes jusqu'au dernier élément du vecteur qui on veut transmit.

Tableau III.11: les valeurs des LSF après appliquer la méthode de dissimulation.

LSF S.trame 1	**LSF S.trame 2**	**LSF S.trame 3**	**LSF S.trame 4**
0.2136	0.2071	0.2006	0.1941
0.4660	0.4442	0.4224	0.4007
0.7789	0.7922	0.8056	0.8189
0.9902	1.0003	1.0104	1.0204
1.2617	1.2720	1.2822	1.2925
1.5472	1.5465	1.5458	1.5451
1.7981	1.7979	1.7977	1.7975
2.0748	2.0836	2.0924	2.1011
2.4396	2.4463	2.4529	2.4595
2.7634	2.7670	2.7707	2.7743

L'algorithme III.1 récapitule les étapes de la méthode stéganographie proposée.

Algorithme III.1: Algorithme général de méthode stéganographique appliqué.

Données :

- $N = longueur\ (LSF), x_inter_poly = \{1,2,3,4,5,6,7,8,9,10\}$
- $n = longueur\ (x)\ \ ,\ p = \{0,0,0,0,0,0,0,0,0,0\}$
- $y = \{0,0,0,0,0,0,0,0,0,0\}$
- $Seuil = 0.01, msg = \{0\ 1\ 0\ 1\ 0\ 0\ 0\ 1\}, q = 1, e = \{2,3,\ldots,9\}$

Etape 1 :

- Calculer les paramètres LSF par l'interpolation linéaire.
- Prendre les paramètres LSF (1) et LSF(10) obtenue comme des données (entrer) pour l'interpolation polynomiale.

Etape 2 :

- $k = 1 : n.$
- Calcule de polynôme de base de Lagrange L.
- Calcule de polynôme d'interpolation de Lagrange P.
- On obtient les paramètres LSF calculé par l'interpolation polynomiale.

Etape 3 :

- Ajouter ou soustraire la valeur de seuil à partir du LSF(2) jusqu'à LSF (9) au niveau de chaque sous trame.
- **Si msg**(q)=0

  LSF (*e*, i) =LSF-SEUIL

  Sinon

  LSF (*e*, i) =LSF+SEUIL

  FinSi

- $e = e + 1$ et répéter la condition jusqu'à la fin de message secret.
- Obtenant les nouveaux paramètres LSF. Puis converti en coefficients LP.
- Transmission des LP.

III.8.5 Extraction le message secret à la réception

La procédure de l'extraction au niveau de récepteur est très simple. Elle nécessite la connaissance, les méthodes d'interpolation utilisées (linéaire et polynomiale). Pour extraire le bit dissimulé dans un support hôte reçus, il suffit de connaitre la valeur de seuil de modification dans la stégo-clé. Les étapes de l'algorithme III.1 sont alors appliquées.

Alors on va faire la différence entre la trame obtenue par l'interpolation linéaire et la trame obtenue par l'interpolation polynomiale, si on obtient une valeur négative cela se signifier le bit caché est un 1, ou si on obtient une valeur positive cela se signifier le bit caché est un 0, le tableau III.12 montre l'extraction de message dissimulé au niveau de récepteur.

Tableau III.12: Une trame après appliquer la méthode de décodage au niveau de récepteur.

S. trame 1	**Msg 1**	**S.t rame 2**	**Msg 2**	**S. trame 3**	**Msg 3**	**S. trame 4**	**Msg 4**
0	-	0	-	0	-	0	-
0.0100	0	0.0100	0	0.0100	0	0.0100	0
-0.0100	1	-0.0100	1	-0.0100	1	-0.0100	1
0.0100	0	0.0100	0	0.0100	0	0.0100	0
-0.0100	1	-0.0100	1	-0.0100	1	-0.0100	1
0.0100	0	0.0100	0	0.0100	0	0.0100	0
0.0100	0	0.0100	0	0.0100	0	0.0100	0
0.0100	0	0.0100	0	0.0100	0	0.0100	0
-0.0100	1	-0.0100	1	-0.0100	1	-0.0100	1
0	-	0	-	0	-	0	-

A partir de ces résultats de simulation du tableau III.11, on constate que notre méthode de décodage assure la récupération de message « 0 1 0 1 0 0 0 1 » sur chaque sous-trames et aussi fiable, ainsi d'après cette étude on peut utiliser cette méthode comme une méthode de stéganographie et on peut le considérer comme étant une méthode de tatouage parce que la transparence de couverture est préservé, donc on peut considérer que notre message est un signature qui protège note couverture, ainsi, nous donne la possibilité de vérifier la copyright de fichier de la parole.

Conclusion

Dans ce chapitre, nous avons présenté une méthode de dissimulation dont le but d'incorporer un message secret dans un signal audio où le récepteur est toujours capable de récupérer le message secret de parole envoyé intelligiblement. Nous avons implémenté cette méthode dans un codeur de parole en large bande de type AMR G.723.1. Les signaux de parole secrète codés ont été générés par un codeur de parole à moyen débit fonctionnant à 6.3 kbps et ACELP à 5.3 kbps.

Conclusion générale

Dans le premier chapitre, en premier lieu, nous avons abordé le processus de production de la parole. Son principe, et les différentes étapes de codage de parole avec l'analyse par synthèse et les différents codeurs de la parole et leurs applications.

Le deuxième chapitre est réservé au codage AMR G.723.1 Et afin de mieux expliquer les différentes étapes d'estimation de coefficients de prédiction linéaire (LPC) qui seront optimisés et quantifiés. Puis, nous avons parlé sur le filtre de pondération des formants ainsi que les défirent étape de décodage.

Notre objectif principal était de concevoir un système de transmission de la parole sécurisé à base de la technique de stéganographie. Pour atteindre cet objectif, nous avons faire différentes modifications et améliorations ont été proposées et réalisées. Plus particulièrement, la modification des coefficients LPC. Nous avons d'abord étudié et testé le codeur de parole AMR G.723.1 fonctionnant avec deux débits binaires associés : 5,3 kbps et 6,3 kbps, Ce codeur a été utilisé par la suite pour concevoir notre système de codage de la parole sécurisé. Par la suite, nous avons étudié l'impact d'interpolation polynomial des coefficients LPC par la variation de seuil de modification sur le codeur G.723.1 de la VoIP.

Nous avons introduit les concepts de base nécessaires pour la compréhension des recherches menées dans ce mémoire. Nous avons gardé la même qualité perceptuelle de la parole lors d'implémentation de la méthode d'interpolation développée pour la transmission d'un message, ce qui donne une capacité de dissimulation de 32 bits par chaque trame, on conclut que notre méthode offre un canal caché de 1.05 kbps et canal caché de 1.06 kbps pour les deux modes 6.3 kbps et 5.3 kbps. D'après notre étude on peut utiliser cette méthode comme une méthode de stéganographie et on peut le considérer comme étant une méthode de tatouage.

Bibliographie

[1] CHU, Wai C, «Speech coding algorithms: Foundation and Evolution of Standardized Coders A John Wiley & Sons. Inc., Publication (03),» 2003.

[2] HARRAG, Abdelghani, «Extraction des données d'une base: Application à l'extraction des traits du locuteur,» Thèse de doctorat, 2018.

[3] MEDJBER, Mustapha, «Amélioration du standard G. 729-8Kb/s par la méthode de modification de l'échelle temporelle (WSOLA).,» Thèse de doctorat. Ecole Nationale Polytechnique, 2007.

[4] DUTOIT, Thierry, «Introduction au traitement automatique de la parole. Notes de cours, Faculté Polytechnique de Mons,» 2000.

[5] SHANNON, Claude Elwood, «A mathematical theory of communication. Bell system technical journal,» 1948, pp. 379-423 vol. 27, no 3.

[6] LINGUISTIC DATA CONSORTIUM, et al. The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus, «NIST Speech CD,» 1990, pp. 1-1.1.

[7] LOO, James Hung Yan, «Intraframe and interframe coding of speech spectral parameters,» McGill University, 1996.

[8] HOLMA, Harri et TOSKALA, Antti, «WCDMA for UMTS: Radio access for third generation mobile communications. john wiley & sons,» 2005.

[9] ITU-T Recommendation G.711, «Pulse code modulation (PCM) of voice frequencies,» Nov 1988 .

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

[11] RAAKE, Alexander, «Qualité perçue de parole transmise par voie téléphonique large-bande,» 2005.

[12] Recommandation UIT-T G.723.1, «Codeur de signaux vocaux à double débit pour communications multimédia acheminées à 5,3 kbit/s et à 6,3 kbit/s,» chez *Aspects généraux des systémes*.

[13] KABAL, Peter, «ITU-T G. 723.1 speech coder: a Matlab implementation,» McGill Univ, 2004.

[14] ERKAN, Hasan Hüseyin, «Speech processing for voice over IP applications,» Thèse de doctorat. DEÜ Fen Bilimleri Enstitüsü, 2011.

[15] KAZAMARANDE, Evariste et COMON, Pierre, «Stabilité numérique de l'algorithme de Levinson. ESAIM: Mathematical Modelling and Numerical Analysis,» 1995, pp. 123-170 vol. 29, no 2.

[16] Medeni, M. Ould, «Application des codes correcteurs d'erreurs en stéganographie,» thèse de doctorat université Mohammed v-Agdal, Juillet 2012.

[17] Annual Review of Network Management and Security, Volume 1.

[18] KOUIDER, Sarra, «Insertion adaptative en stéganographie: application aux images numériques dans le domaine spatial,» Thèse de doctorat. Université Montpellier II-Sciences et Techniques du Languedoc., 2013.

[19] Elz˙ bieta ZieliN´ ska, Wojciech Mazurczyk, and Krzysztof Szczypiorski , «Trends in Steganography,» March 2014 , p. 89.

[20] S. KHELIDJ, R. HAMADOU, «La stéganographie par la méthode de codage d'Huffman et de modification de direction,» Mémoire de master faculté des sciences et de la technologie département du génie électrique et de l'informatique, université Dr. yahia, 2014.

[21] El-Khamy, Said E., «A security enhanced robust audio steganography algorithm for image hiding using sample comparison in discrete wavelet transform domain and RSA encryption,» Springer DOI 10.1007/s11042-016-4113-8, November 2016.

[22] S.Satapathy, A.Joshi N.Modi, N. Pathak, «advances in intelligent systems and computing,» chez *Proceedings of international conference on ICT for sustainable*

development ICT4SD Volume 1., 2015.

[23] H.KHEDDAR, B.BOUDRAA, «Etude comparative de méthodes de correction de pertes de paquets en transmission de la voix sur IP (VoIP),» Mémoire magister laboratoire LCPT, 2012.

[24] ZUE, Victor, SENEFF, Stephanie, et GLASS, James, «Speech database development at MIT: TIMIT and beyond. Speech communication, vol. 9, no 4,» 1990, pp. 351-356.

[25] KABAL, P, «ITU-T G. 723.1 Speech Coder. A Matlab Implementation,» Department of Electrical & Computer Engineering McGill University, 2009.

[26] Islam, tamanna, «Interpolation of Linear Prediction Coefficients for Speech Coding.MasterofEngineering,» 2000.

[27] Erkelens J. S., Broersen P M. T, «PC Interpolation by Approximation of the Sample Autocorrelation Function IEEE Trans. On Audio and Speech Processing. – Vol. 6, No. 6,» 1998, p. 569–573.