

N° d'ordre : 24/2013-M/ELN

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université des Sciences et de la Technologie Houari Boumediene

Faculté d'Electronique et Informatique



MEMOIRE

Présenté pour l'obtention du diplôme de Magister

En: ELECTRONIQUE

Spécialité: Communication parlée

Par : LASKAR Bekkar

Systeme de transmission sécurisé de la parole basé sur la stéganographie

Soutenu publiquement le : 04 /07/2013, devant le jury composé de:

M. H. TEFFAHI

Professeur à l'USTHB

Président

M. M. BOUZID

Maître de Conférences /A, à l'USTHB

Directeur de mémoire

M. A. AMROUCHE

Professeur à l'USTHB

Examineur

M. S. MEKAOUI

Maître de Conférences /A, à l'USTHB

Examineur

Remerciements

Louange à ALLAH le miséricordieux, qui m'a permis d'accomplir ce travail.

Je tiens à remercier sincèrement et profondément mon promoteur Mr. M. BOUZID de sa disponibilité, son encadrement, ses bonnes suggestions et tout l'effort qu'il a consacré à l'amélioration de ce rapport.

J'adresse aussi mes vifs remerciements aux membres du jury Mr H. TEFFAHI, Mr A. AMROUCHE et Mr S. MEKAOUI pour l'honneur qu'ils m'ont accordé en acceptant d'examiner ce travail et de participer au jury.

Mes remerciements vont également à tous les enseignants du laboratoire communication parlée qui ont contribué à notre formation et particulièrement aux membres de l'équipe Analyse et Codage de la parole.

Enfin, je tiens à exprimer toute ma reconnaissance à ma famille: ma mère, mes frère et sœurs, pour l'assistance qu'elle a sue m'apporter pendant toute la durée de ce travail.

Je dédie ce travail

A

Ma chère mère

La mémoire de mon père

Mes frères et sœurs

Toute ma famille et à tous mes proches petit et grand

Tous mes amis et mes camarades de la promotion.

Et à tous ceux qui me sont chers

Table des Matières

	Page
Table des Matières	i
Liste des figures	vi
Liste des tableaux	vii
Liste des abréviations	viii
Liste des notations	x
Introduction générale.....	1
Chapitre 1: Système de transmission sécurisé.....	5
1.1 Introduction.....	5
1.2 Système de communication numérique.....	5
1.3 Système de transmission cryptographique.....	7
1.4 Système de tatouage.....	8
1.4.1 Phase d'insertion de la marque.....	9
1.4.2 Phase d'extraction de la marque	9
1.5Système de transmission stéganographique.....	9
1.6 Stéganalyse.....	11
1.7 Applications de la dissimulation des données.....	11
1.7.1 Protection de copyright.....	11
1.7.2 Les Empreintes.....	12

1.7.3 Protection contre la copie.....	12
1.7.4 Contrôle de diffusion.....	12
1.7.5 Authentification de données.....	12
1.7.6 Indexation.....	12
1.7.7 Sécurité médicale.....	12
1.7.8 Transmission sécurisée de données.....	13
1.8 Système de transmission stéganographique de la parole.....	13
1.9 Classification des systèmes stéganographiques de la parole.....	15
1.9.1 Classification selon la stégo clé.....	15
1.9.1.1 Stéganographie pure	15
1.9.1.2 Stéganographie à clé secrète.....	16
1.9.1.3 Stéganographie à clé publique	16
1.9.2 Classification selon le domaine d'insertion.....	16
1.9.2.1 Domaine temporel.....	17
1.9.2.2 Domaine fréquentiel.....	17
1.9.2.3 Domaine de compression.....	17
1.9.3 Classification selon l'architecture.....	17
1.9.3.1 Dissimulation avant le codage.....	17
1.9.3.2 Dissimulation durant le codage.....	18
1.9.3.3 Dissimulation après le codage.....	18
1.10 Techniques de dissimulation de la parole.....	18
1.10.1 Dissimulation LSB	18
1.10.2 Dissimulation par "Patch Work".....	19
1.10.3 Quantification par modulation d'indice.....	20
1.10.4 Dissimulation par écho.....	20
1.11 Critères d'évaluation d'un système stéganographique.....	21
1.11.1 Sécurité.....	21
1.11.1.1 Indétectabilité	21
1.11.1.2 Imperceptibilité (transparence).....	21
1.11.2 Robustesse.....	21
1.11.3 Capacité.....	22
1.11.4 Le coût.....	22
1.12 Conclusion.....	22

Chapitre 2: Codage de la parole.....	24
2.1 Introduction	24
2.2 Quantification.....	25
2.2.1 Quantificateur Scalaire.....	25
2.2.2 Quantificateur vectoriel.....	26
2.2.2.1 Principe de codage par quantificateur vectoriel.....	27
2.2.2.2 Quantificateur vectoriel divisé.....	28
2.2.2.3 Quantificateur vectoriel multi étages.....	29
2.3 Principe de la prédiction linéaire.....	31
1.4 Les paramètres fréquence spectrale d'immittance.....	32
2.5 Classification des techniques de codage de parole.....	34
2.5.1 Codage de forme d'onde.....	34
2.5.2 Codage hybride.....	34
2.5.3 Codage paramétrique.....	35
2.6 Généralités sur le codeur MELP.....	36
2.7 Codage de la parole en large bande.....	37
2.7.1 Avantages du codage de la parole en large bande.....	38
2.7.2 Applications du codage de la parole en large bande.....	39
2.8 Codeur de parole large bande AMR-WB, norme G.722.2.....	40
2.8.1 Quantification des paramètres ISF.....	41
2.8.2 Analyse en boucle ouverte du délai tonal.....	42
2.8.3 Dictionnaire algébrique.....	42
2.8.3.1 Structure du dictionnaire.....	43
2.8.3.2 Exploration du dictionnaire.....	43
2.9 Conclusion.....	49
Chapitre 3 : Techniques de stéganographie appliquées à la parole.....	50
3.1 Introduction.....	50
3.2 Schémas de dissimulation dans la quantification vectorielle.....	51
3.2.1 Principe de la QIM basée sur la division de dictionnaire QV.....	52
3.2.1.1 Phase de prétraitement.....	53
3.2.1.2 Phase de la dissimulation	53
3.2.1.3 Phase de l'extraction.....	53

3.2.2 Principe de la QIM basée sur la modification des schémas VQ (QIM_MS_VQ)	54
3.3 Dissimulation dans les indices de quantification des paramètres ISF	54
3.3.1 QIM à base de division du dictionnaire de quantification des ISFs.....	55
3.3.1.1 Division du dictionnaire.....	55
3.3.1.2 Procédure de la dissimulation.....	56
3.3.1.3 Procédure de l'extraction.....	58
3.3.1.4 Méthodes de la division du dictionnaire.....	58
3.3.1.5 Débit de la dissimulation	61
3.3.2 QIM basée sur la modification des schémas VQ des ISFs.....	62
3.3.2.1 Procédure de la dissimulation	62
3.3.2.2 Débit de dissimulation	65
3.4 Dissimulation dans l'indice de l'excitation	65
3.4.1 Préparation du message et débit de dissimulation	66
3.4.2 Division du dictionnaire de l'excitation.....	66
3.4.3 Procédure de dissimulation.....	68
3.4.4 Procédure d'extraction	70
3.4.5 Complexité de la méthode de dissimulation.....	70
3.5 Dissimulation dans le délai tonal.....	71
3.6 Génération du message secret.....	73
3.7 Conclusion.....	73
Chapitre 4 : Implémentation de systèmes stéganographiques dans un codeur de parole en large bande.....	75
4.1 Introduction.....	75
4.2 Mesures d'évaluation des performances.....	76
4.2.1 Mesures subjectives.....	77
4.2.2 Mesures objectives.....	78
4.2.1.1 Mesures objectives : performance dans le domaine spectral.....	79
4.2.1.2 Mesures objectives : performance dans le domaine temporel.....	80
4.2.3 Mesures perceptuelles objectives.....	80
4.3 La QIM basée sur la division du dictionnaire des ISFs	81
4.3.1 Performances des systèmes stéganographiques QIM-DD.....	81
4.3.2 Performance du G722.2 modifié par QIM_DD.....	83
4.4 Système QIM-MS-VQ appliqué au quantificateur des ISFs.....	86

4.4.1 Performances des systèmes stéganographiques QIM-MS-VQ	87
4.4.2 Performance du G722.2 modifié par QIM_MS_VQ des ISFs.....	89
4.5 Comparaison entre les méthodes QIM dans ISFs.....	92
4.6 Dissimulation dans l'indice de l'excitation du G.722.2	93
4.6.1 Performance du système de dissimulation dans l'indice d'excitation.....	93
4.7 Dissimulation dans le délai tonal.....	95
4.8 Conclusion.....	98
Conclusion générale.....	99
Bibliographie.....	101

Liste des Figures

Figure	Page
Figure 1.1 : Schéma synoptique du système de transmission classique.....	6
Figure 1.2: Problème d'évasion des prisonniers : message secret.....	10
Figure 1.3 :Exemple d'un système de transmission stéganographique de la parole.....	14
Figure 1.4 Exemple de dissimulation par codage LSB dans une couverture code sur 8 bits.....	19
Figure 1.5 Exemple de dissimulation dans l'écho.....	21
Figure 2.1 : Codage et décodage avec le QV.....	27
Figure 2.2 : Codage et décodage pour 2-SVQ.....	29
Figure 2.3 : Codage et décodage pour 2-MSVQ.....	30
Figure 2.4 : Modèle du filtrage.....	32
Figure 2.7 : Importance de la bande passante à l'indice d'articulation.....	39
Figure 2.8 : Spectre du signal parole en bandes étroite et large.....	39
Figure 3.1 : Principe général de la division du dictionnaire C.....	52
Figure 3.2 : Exemple d'un schéma de dissimulation par QIM.....	53
Figure 3.3 : Exemple d'un schéma d'extraction de données par QIM.....	54
Figure 3.4 : Exemple de division du dictionnaire C de taille L=8.....	56
Figure 3.4 : Procédure de dissimulation en utilisant l'indice du deuxième étage MSVQ.....	63
Figure 4.1 : Formes d'ondes synthétiques avec dissimulation par les systèmes QIM-DD.....	85
Figure 4.2 : Comparaison de l'évolution de la SD entre la division équilibrée et non équilibrée..	86
Figure 4.3 : Formes d'ondes synthétiques avec dissimulation QIM-MS-VQ par DRS et DRT...	91
Figure 4.4 : Comparaison de l'évolution de la SD entre DRS et DRT.....	92
Figure 4.5 : Complexité de la QIM-DD appliquée sur l'indice d'excitation.	95
Figure 4.6 Capacité de la méthode LSB adaptative.....	97

Liste des Tableaux

Tableaux	Page
Tableau 2.1 : Table d'allocation des bits du codeur MELP de 2.4 Kbits/s.....	37
Tableau 2.2 : Allocation des bits pour chaque mode du G722.2.....	41
Tableau 2.3: Quantification de vecteur ISF pour les modes entre 8,85 et 23,85 kbit/s.....	42
Tableau 2.4 : Positions possibles d'impulsions individuelles dans le répertoire algébrique du débit 12,65 kbit/s.....	43
Tableau 3.1: Débits de dissimulation dans chaque indice du deuxième étage S-MSVQ.....	65
Tableau 4.1 : Echelle de variation du score MOS.....	77
Tableau 4.2 : Echelle de variation du score DMOS.....	78
Tableau 4.3: Performance du système QIM-DD conçu par la méthode de division non équilibré.....	82
Tableau 4.4: Performance du système QIM-DD conçu par la méthode de division équilibrée...	82
Tableau 4.5: Performance du système QIM-DD conçu par la méthode de division aléatoire.....	82
Tableau 4.6 : Performance du codeur global G.722.2 avec implémentation du QIM-DD.....	84
Tableau 4.7: Performance du système QIM-MS-VQ conçu par la méthode DRS.....	87
Tableau 4.8: Performance du système QIM-MS-VQ conçu par la méthode DRT.....	88
Tableau 4.9 : Performance du codeur G.722.2 avec implémentation du système QIM-MS-VQ des ISFs.....	90
Tableau 4.10: Performance du G.722.2 avec dissimulation dans l'indice de l'excitation	94
Tableau 4.11: Performance du G.722.2 avec dissimulation dans le délai tonal.....	96

Liste des abbreviations

3GPP	Third Generation Partnership Project.
ACELP	Algebraic Code Excited Linear Prediction.
AMR-WB	Adaptive Multi Rate Wide Band.
AR	Modélisation Auto Régressive.
CELP	Code Excited Linear Prediction (Prédiction linéaire excitée par code).
dB	Décibel.
DoD	Department of Defense (défense américaine)
ETSI	European Telecommunication Standards Institute
EQM	Erreur Quadratique Moyenne.
ISF	Paramètres d'immittance de fréquence.
MELP	Mixed Excitation Linear Prediction (Excitation Mixte Prediction Linéaire)
LP	Linear Prediction (prédiction linéaire).
LPC	Linear Predictive Coding (Codage par prédiction linéar).
LSB	Least significant bits (bits les moins significatifs)
LSF	Line Spectral Frecencies (Fréquences de raies spectrales).
MA	Moyenne mobile (Moving Average).
MICDA	Modulation par impulsions et codage différentiel adaptatif
MOS	Mean Opinion Score (Note Moyenne d'Opinion).
MSVQ	Multistage Vector Quantizer
PESQ	Perceptual Speech Quality Measure.
QIM	Quantizer-index modulation (Quantification par modulation d'indice).
QIM_DD	QIM basée sur la division de dictionnaire
QIM_MS_VQ	QIM basée sur la modification des schémas dequantificateurs Vectoriels
QS	Quantificateur (ou Quantification) Scalaire.
QV	Quantificateur Vectoriel (ou Quantification Vectorielle).

RSB	Rapport Signal sur Bruit.
RSBseg	Rapport Signal sur Bruit segmental.
SD	Spectral Distorsion (Distorsion spectrale).
S-MSVQ	Split Multi Stage Vector Quantization (quantification vectoriel multi étages divisée).
SNR	Signal to Noise Ratio (Rapport signal sur bruit).
SVQ	Split Vector Quantization (Quantificateur vectoriel divisé).
TIMIT	DARPA TIMIT Acoustic-Phonetic Speech Database.
UIT-T	Union Internationale des Télécommunications.
WB-PESQ	Wide Band Perceptual Speech Quality Measure.

Liste des notations

$A(z)$	Filtre inverse avec coefficients non quantifiés
$b(n)$	Signal utilisé pour prérégler les signes lors de la recherche dans le dictionnaire algébrique
$\mathbf{c}_k$	Vecteur d'innovation
$\mathbf{d}$	Corrélation entre le signal cible $x_2(n)$ et la réponse impulsionnelle $h(n)$,
$d(n)$	Cible filtrée en inverse
$d'(n)$	Cible filtrée en inverse avec extension au signe
E	La corrélation dans le numérateur
g_p	Gain du répertoire adaptatif
H	Partie triangulaire inférieure de la matrice convolutionnelle de Töpliz avec diagonale $h(0)$ et diagonales inférieures $h(1), \dots, h(63)$
m_i	Position de la $i^{\text{ème}}$ impulsion.
N_p	Nombre d'impulsions dans l'excitation par le répertoire fixe.
$\mathbf{p}(n)$	Vecteur ISF prédit à la trame n
R	La corrélation dans le numérateur
R_i	Débit partiel.
$R_{hv}(m)$	la corrélation entre la réponse impulsionnelle $h(n)$ et un vecteur $v_n(n)$
$\mathbf{r}(n)$	Vecteur résiduel de prédiction ISF à la trame n
$\hat{\mathbf{r}}(n - 1)$	Vecteur résiduel quantifié à la trame antérieure
r_{LTP}	Signal résiduel après prédiction à long terme
SD	Distorsion spectrale moyenne.
SD_i	Distorsion spectrale d'une trame.
$x_2(n)$	Signal cible pour recherche dans le répertoire algébrique
$\mathbf{z}(n)$	Vecteur ISF à moyenne supprimée à la trame n

Φ	Matrice de corrélations de $h(n)$
$\Phi(i, j)$	Eléments de la matrice symétrique Φ
$\Phi'(i, j)$	Eléments modifiés de la matrice Φ , y compris les informations de signe

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.

La problématique liée à la sécurisation des communications répondait principalement à des besoins militaires et commerciaux. En effet, lorsque deux personnes géographiquement distantes veulent échanger une information confidentielle (comme un ordre martial), plusieurs questions principales se posent :

- 1). Comment être sûr qu'une personne mal intentionnée n'a pas eu accès au message pendant son transfert ?
- 2). Comment être sûr que le message n'a pas été modifié ?
- 3). Comment être sûr que la personne avec laquelle on communique est bien la bonne ?
- 4). Comment s'assurer de la persistance du message dans le canal de transmission utilisé ?
- 5). Comment envoyer un message sans que l'adversaire puisse seulement se rendre compte qu'il y a eu transmission ?

Afin d'illustrer ces différents scénarios, les cryptographes ont inventé un jeu de sécurité impliquant trois personnages fictifs : une personne qui cherche à transmettre une information à une autre personne en essayant de satisfaire un des points cités plus haut et un adversaire qui essaiera de contourner cet objectif. Il existe trois catégories principales de sécurité : la cryptographie permet de répondre aux trois premières questions, le tatouage à la quatrième et la stéganographie à la cinquième.

L'objectif principal de ce projet de magister est de concevoir un système de transmission sécurisée de la parole. Conventionnellement, la conception d'un tel système se basait principalement sur les techniques de cryptographie afin d'assurer une bonne sécurité contre les attaques (passives ou actives) de tout type. A l'heure actuelle, ces techniques s'avèrent insuffisantes surtout avec l'augmentation des divers applications de la communication parlée à travers les réseaux internet,...etc; d'où la nécessité de développer de nouvelles approches plus puissantes pour leur apporter plus de sécurité. Actuellement, on parle de tatouage et de stéganographie qui sont devenus des techniques de base pour assurer la sécurité et la confidentialité des systèmes de transmission.

Dans ce travail, on 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 peu visibles. 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.

Au début, nous avons étudié les techniques steganographiques qui peuvent être adaptées au codage de la parole, comme par exemple la technique de dissimulation par modulation de l'indice de quantification QIM (Quantized Index Modulation) et la technique de dissimulation dans bits les moins significatifs (LSB). Plusieurs systèmes stéganographiques à base des techniques QIM et LSB adaptative ont été mis au point et implémentés dans un système de transmission de la parole. Pour concevoir un tel système, nous avons utilisé le codeur AMR-WB G.722.2 comme codeur de la parole publique (hôte). Le standard ITU-T AMR-WB G.722.2 est un codeur de parole en large bande de type ACELP (Algebraic Code Excited Linear Prediction) adaptatif opérant à des débits multiples compris entre 6,6 kbit/s et 23,85 kbit/s. Il modélise le signal parole par extraction des paramètres pertinents du modèle de prédiction ACELP qui sont ensuite codés et transmis au décodeur. D'autre part, nous utilisons le codeur de parole bas débit MELP (Mixed Excitation Linear Prediction) de 2.4 kbits/s pour générer la parole secrète codée qui sera dissimulée dans la parole publique. Dans notre travail, on s'est intéressé à l'application des techniques de dissimulation de messages

aux blocs de codage d'une partie des paramètres pertinents du G.722.2 qui sont : les paramètres ISF, l'indice d'excitation du dictionnaire algébrique et le délai tonal. Les performances de nos systèmes stéganographiques implémentés dans le codeur G.722.2 seront évaluées selon deux critères : la capacité de dissimulation représentée par le débit de dissimulation de la parole secrète et l'imperceptibilité (transparence) représentée par la qualité perceptuelle de la parole publique synthétisée.

Ce mémoire de magister est composé de quatre chapitres qui sont organisés comme suit :

Dans le chapitre 1, nous définirons brièvement les systèmes de communications numériques, ainsi que les principaux constituants d'un tel système. Nous présenterons ensuite les trois catégories de systèmes de transmission sécurisés : cryptographie, tatouage et stéganographie. Un bref aperçu sur les domaines d'application de la dissimulation des données sera donné. 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. A la fin, nous présenterons les différentes techniques stéganographiques ainsi que les critères d'évaluation de tels systèmes.

Dans le chapitre 2, nous présenterons les méthodes de codage de source par quantification scalaire et vectorielle. Nous allons détailler particulièrement le quantificateur vectoriel divisé SVQ (Split Vector Quantizer) et le quantificateur vectoriel multi-étages MSVQ (Multistage Vector Quantizer). En suite, nous présenterons une méthode d'estimation de l'enveloppe spectrale du signal parole, appelée prédiction linéaire LP. Nous définirons les paramètres spectraux LSF et les paramètres d'immitance ISF qui représentent les coefficients LPC. Par la suite, Nous donnerons une classification des techniques de codage de la parole (forme d'onde, paramétrique et hybride) avec une brève description du codeur de parole MELP. A la fin, nous passerons au domaine de la parole en large bande où les avantages du codage de la parole en large bande seront cités ainsi que les domaines d'utilisation. Une description générale du codeur AMR-WB norme G.722.2 sera donnée.

Dans le chapitre 3, nous détaillerons les techniques de stéganographie que nous avons choisies pour les applications à la parole. Nous avons opté pour ces techniques

de dissimulation vu leurs efficacités et leurs possibilités d'implémentation sur des codeurs de parole comme le codeur AMR-WB G.722.2. Au début, nous présenterons la dissimulation par modulation d'indice de quantification (QIM) ainsi que ces versions dérivées. En suite, nous décrirons la technique de dissimulation LSB adaptative. Une première application de la technique de dissimulation par QIM sera dirigée vers le bloc de quantification des paramètres ISF du codeur G. 722.2. Deux versions de systèmes stéganographiques QIM seront appliqués : QIM basée sur la division de dictionnaires de quantification (QIM-DD) et QIM à base de modification de schéma de quantificateurs vectoriels (appelé système QIM-MS-VQ)). La technique LSB adaptative sera utilisée au niveau du bloc de quantification du délai tonal (pitch).

Dans le chapitre 4, nous présenterons les résultats d'implémentation des différentes techniques stéganographiques détaillées dans le chapitre 3. Pour chaque technique de dissimulation, nous chercherons les meilleurs paramètres d'entrée de notre système stéganographiques (débit de dissimulation, distribution de bits de message, ...). Une évaluation comparative des performances des techniques implémentées dans le codeur de parole AMR-WB G.722.2 sera présentée.

Enfin, nous terminerons par une conclusion générale qui présentera le fruit de ce travail tout en donnant les perspectives futures qui peuvent enrichir ce travail.

Chapitre 1

Système de transmission sécurisé

1.1 Introduction

Les communications numériques ont considérablement évoluées lors des dernières décennies. Cette révolution de la société de l'information est due à la numérisation de l'information. En plus des besoins classiques de communication et de stockage, la sécurité des transmissions est devenue aussi un besoin principal.

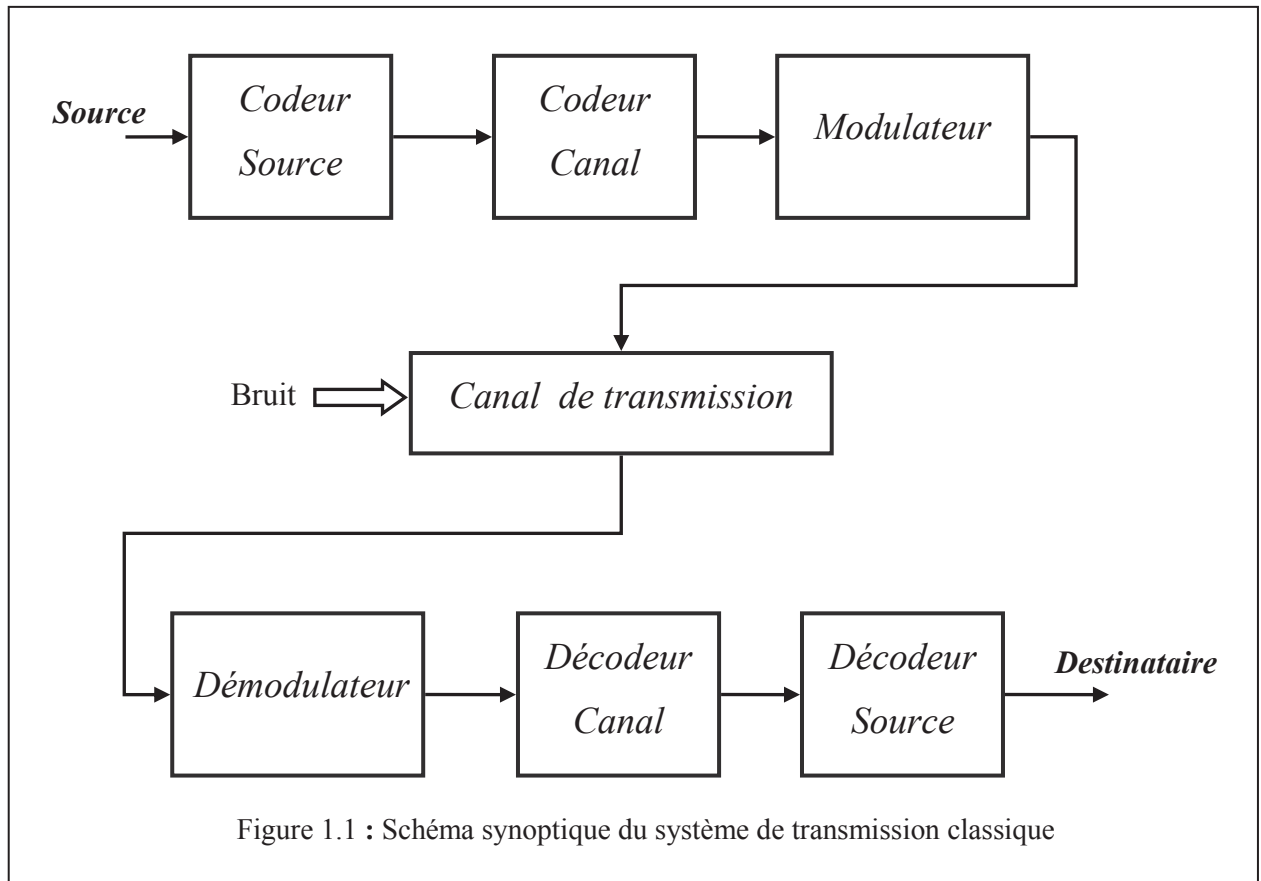
Dans ce chapitre, nous décrirons brièvement les fondements de base des systèmes de transmission en présentant le modèle classique d'un système de communication numérique. Ensuite, nous présenterons les systèmes de transmission sécurisés basés sur la cryptographie, le tatouage et la stéganographie. Nous détaillerons aussi le système destiné à la transmission de la parole.

1.2 Système de communication numérique

Un système de communication numérique assure la transmission de l'information entre une source et un destinataire en utilisant un support physique comme le câble, la fibre optique ou encore la propagation sur un canal radioélectrique. Les signaux transportés peuvent être soit

directement d'origines numériques, comme dans les réseaux de données, soit d'origines analogiques (parole, image...) mais convertis en une forme numérique.

Le schéma synoptique de base d'un système de transmission numérique [1] est donné par la figure 1.1.



Source : tout organe ou dispositif qui produit des signaux numériques constitués de séquences de symboles (bits ou échantillons, par exemple) discrets prenant leurs valeurs dans un alphabet fini. Dans le cas où la source produit des signaux analogiques (audio, vidéo...), ces derniers sont convertis en séquence de digits binaires (bits).

Codeur de source : le codeur de source a pour rôle de transformer le signal issu de la source en une suite (séquence) binaire de "0" et de "1". Il assigne à chaque bloc de source x de k symboles un index (indice) binaire i représenté sur R bits. Cet indice, associé à un représentant de source de k symboles, est généralement de taille fixe.

Codeur de canal : Le codeur de canal peut éventuellement introduire de la redondance dans l'information en vue de la protéger contre le bruit et les perturbations présentes sur le canal de transmission. Il peut détecter et/ou corriger l'information bruitée selon sa capacité.

Modulateur : Le modulateur a pour fonction principale d'assurer l'adaptation du message à la voie de transmission. Le signal issu de l'émetteur est modulé. Ce signal représentatif du message à transmettre doit posséder un spectre compatible avec la bande passante de la voie de transmission et une puissance qui assure au récepteur de travailler dans des bonnes conditions.

Canal de transmission : Le canal de transmission représente le lien physique entre l'émetteur et le récepteur. Il peut être constitué par l'espace libre entre deux antennes, par un câble coaxial ou par une fibre optique. Le canal de transmission est en général une donnée du problème (bruit) et l'on doit s'y adapter.

Destinataire : c'est la personne ou l'appareil qui va exploiter l'information reçue.

En fin le démodulateur, le décodeur de canal et le décodeur de source réalisent respectivement les fonctions inverses du modulateur, du codeur de canal et du codeur de source.

Dans ce système classique, la notion de la sécurité n'existe pas car son but principal est de transférer seulement l'information de la source vers un utilisateur (destinataire) à travers un canal de transmission. Dans notre travail, nous nous sommes intéressés à la sécurité du système de transmission. Ces systèmes sont généralement regroupés en trois catégories : système cryptographique, système de tatouage et système stéganographique.

1.3 Système de transmission cryptographique

La cryptographie (du grec ancien *kryptos*(caché) et *graphein* (écriture)) est la science qui traite de la communication en présence d'adversaires, permettant ainsi de transmettre des données de manière confidentielle [2]. Afin de protéger un message, on lui applique une transformation qui le rend incompréhensible; c'est ce qu'on appelle le chiffrement qui, à partir d'un texte en clair, donne un texte chiffré ou cryptogramme. Inversement, le déchiffrement est l'action qui permet de reconstruire le texte en clair à partir du texte chiffré. Dans la cryptographie moderne, les transformations en question sont des fonctions mathématiques, appelées algorithmes cryptographiques qui dépendent d'un paramètre appelé clé. Selon ce paramètre on distingue deux types de cryptographie: la cryptographie à clés secrètes (symétrique) et la cryptographie à clés publiques (asymétrique).

La cryptographie symétrique [3, 4] également dite à clé secrète (ou encore conventionnelle), est la plus ancienne forme de chiffrement qui regroupe les algorithmes pour lesquels expéditeur et destinataire partagent une même clé secrète.

La cryptographie à clé publique repose sur un schéma asymétrique [5] qui utilise une paire de clés pour le chiffrement: une clé publique qui chiffre les données et une clé privée secrète correspondante qui sera utilisée pour le déchiffrement. Vous publiez largement votre clé publique tout en gardant votre clé privée secrète. Toute personne en possession d'une copie de votre clé publique peut chiffrer des informations que vous seul pourriez lire. Il est mathématiquement impossible de déduire la clé privée de la clé publique. Ainsi, quiconque a une clé publique peut chiffrer des informations mais ne peut pas les déchiffrer. Seule la personne qui a la clé privée correspondante peut déchiffrer les informations.

La cryptographie a été une première proposition pour sécuriser des transferts de documents numériques. Aujourd'hui les algorithmes de chiffrement modernes, avec des clefs de longueurs importantes, permettent d'assurer la confidentialité. Néanmoins, une fois décrypté, le document n'est plus protégé et il peut être distribué ou malhonnêtement modifié. Actuellement, la dissimulation d'information, qui est une libre adaptation de l'expression anglaise "*data hiding*", est couramment utilisée dans la littérature [6, 7, 8]. Le concept de dissimulation d'information est très général. Il s'agit simplement du fait de cacher une information dans un support. Toutefois, selon les objectifs ou les contraintes, il est possible de distinguer plusieurs types et techniques qui peuvent être des solutions aux problèmes et aux inconvénients de la cryptographie [9]. On parle alors du tatouage et de la stéganographie.

1.4 Système de tatouage

Le tatouage des données numériques, qui est une discipline relativement récente, permet une protection des médias (texte, audio, image et vidéo). Par exemple, le tatouage des images d'une vidéo consiste à insérer une marque de façon robuste et imperceptible [10].

L'objectif du tatouage est d'introduire dans les médias une marque invisible appelée *signature* ou *marque*, contenant un code de copyright. Les médias ainsi marqués ou tatoués peuvent alors être distribués; ils porteront toujours la marque de son propriétaire. Ces médias sont susceptibles de subir diverses transformations. Ces transformations peuvent être licites ou illicites, elles ont alors pour but de détruire le marquage. Si elles ne dégradent pas trop la qualité du média, ces modifications ne doivent pas gêner la détection de la marque.

La plupart des algorithmes de tatouage reposent sur le même modèle et diffèrent seulement par la stratégie spécifique à certains niveaux du processus d'insertion ou d'extraction /détection de la marque. Il est possible donc de présenter d'une façon générique un système tatouage.

1.4.1 Phase d'insertion de la marque

L'insertion de la marque consiste à insérer dans le média original une marque et créer ainsi un nouveau média appelée média tatoué. Un troisième paramètre facultatif peut être ajouté: la clé secrète de marquage qui permet d'assurer un certain niveau de sécurité au processus du tatouage.

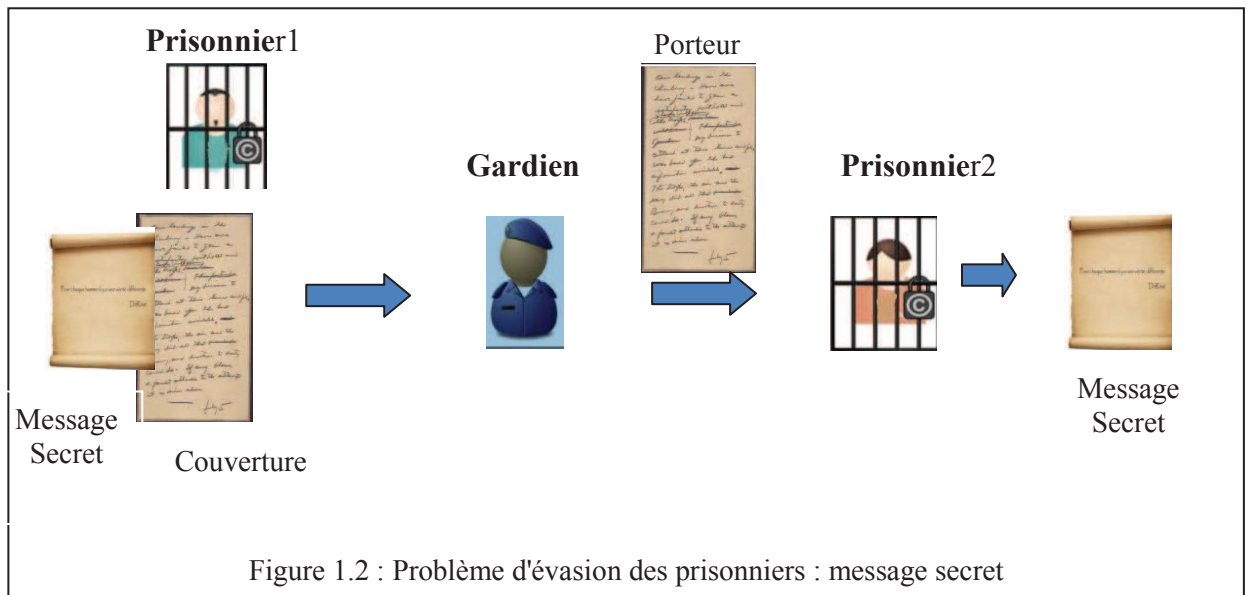
1.4.2 Phase d'extraction de la marque

Selon la conception de l'algorithme de tatouage, on peut avoir besoin du média original lors de cette phase. Dans ce cas, on parle d'un tatouage informé ou non aveugle. Dans le cas contraire, le tatouage est dit non informé ou aveugle. Dans certains cas, l'utilisation d'un tatouage informé permet de déterminer si le média original tatoué a été attaqué. La présence du média original fournit des informations supplémentaires qui peuvent servir pour améliorer l'extraction de la marque.

1.5 Système de transmission stéganographique

L'inventeur du mot stéganographie est "Trithemius" qui est l'auteur de l'une des premières publications sur la cryptographie: Polygraphia & Steganographia [11]. La stéganographie vient du Grec "steganos" (dissimulé), et "graphein" (écriture); ce qui signifie donc "écriture dissimulée". En effet, la stéganographie est l'art de la dissimulation d'un message privé ou secret dans un support [12]. Ce message doit être secret ainsi que son existence.

Un exemple d'application de la stéganographie est présenté dans la figure 1.2. Il s'agit du problème d'évasion des prisonniers [13]. 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 jugé 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 l'utilisation de la cryptographie impossible.



Histoire de la stéganographie :

Bien que ce qui nous intéresse ici soit en rapport avec les données numérique, cependant il peut être intéressant de revenir un peu en arrière. La stéganographie connaît ses premières prémices à l'antiquité. A cette époque, les messages étaient écrits sur le crâne rasé d'un esclave. Une fois que ses cheveux avaient repoussés, il pouvait transmettre le message sans se faire inquiéter.

D'autres méthodes bien connues sont apparues par la suite. Ainsi, l'encre invisible eue grand succès il y a quelques décennies. Le principe était d'écrire le message secret avec une encre invisible (jus de citron, urine, etc...) puis d'écrire ensuite avec une encre bien visible un message au contenu anodin. Une fois arrivé à destination, un traitement particulier permettait de recouvrir le message. Ces encres furent utilisées durant la guerre de sécession (1775-1783) pour transmettre des messages entre George Washington et des agents tels que Benjamin Tallmadge.

De nos jours, la stéganographie a pris de l'ampleur dans des supports bien particuliers, les supports numériques. Que ce soit des fichiers parole, vidéo ou image, ils représentent des supports privilégiés pour la transmission d'information. Avec la généralisation d'Internet et les réseaux GSM, le volume de données qu'il représente est énorme. Cela représente autant de support possible à la stéganographie.

1.6 Stéganalyse

L'effort pour détecter la présence de la stéganographie est appelé stéganalyse. La stéganalyse est supposée comme contrôleur du canal de transmission qui surveille les matières suspectes [14]. Contrairement à la cryptanalyse, dont le but est de récupérer les données cryptées, la stéganalyse n'a pas comme objectif de retrouver les données dissimulées par un algorithme stéganographique. Elle consiste uniquement en la détection de contenu stéganographie, ce qui n'est déjà pas une mince affaire.

Plusieurs catégories de méthodes peuvent être utilisées dans cette optique. La plus simple consiste en la détection de signature d'un logiciel donné. Certains programmes laissent derrière eux, de manière intentionnelle ou par erreur de conception, des artefacts permettant de caractériser leur passage. Cette façon de faire possède l'avantage de fournir directement le nom de l'application utilisée pour la dissimulation de donnée. Par contre, chaque application doit faire l'objet d'une analyse spécifique.

1.7 Applications de la dissimulation des données

Parmi les applications de la dissimulation des données on peut trouver :

1.7.1 Protection de copyright

La dissimulation de données offre une alternative intéressante pour protéger les données, même lorsque celles-ci sont diffusées. La protection des droits d'auteur représente, quant à elle, l'application la plus courante aujourd'hui. L'objectif est d'insérer une information dans un support (image, vidéo ou parole), typiquement le copyright du propriétaire, afin de prévenir toute revendication frauduleuse de propriété. Cette signature ne doit être connue que de la personne ou de l'organisme de tatouage [10, 11]. Elle dépend donc d'une clé secrète qui permet son insertion et sa détection. Cette application nécessite la mise en place d'un algorithme de tatouage d'un niveau de robustesse très élevé.

1.7.2 Les Empreintes

Cette application est utilisée pour tracer les copies illégales de media (suivi des pirates). Ce type d'application engendre un marquage unique pour chaque copie distribuée (typiquement un numéro de série). Cependant, la distribution de copies composées de différentes marques peut engendrer des problèmes de collusions. Ainsi, les marques utilisées devront satisfaire à un critère de sécurisation de collusions.

1.7.3 Protection contre la copie

Un souhait des distributeurs de multimédia est l'existence d'un moyen de protection contre la copie afin d'interdire une circulation de médias illégaux. Cependant, il est possible d'utiliser des marques spécifiant le statut de la copie de la donnée. Un exemple pratique est le système DVD dans lequel les données contiennent des informations de marquage [11]. En général, il existe deux types de DVD : les conformes et les non conformes. Les lecteurs de DVD dits "non conformes" sont basés sur une norme qui ne souscrit pas au traité de protection des DVD. Un DVD sur lequel s'exercent les droits d'auteur, peut être lu et décrypté par un lecteur conforme et ne peut pas être lu par un lecteur non conforme. Un problème survient lorsque le DVD crypté est attaqué et qu'il devient possible de se procurer une copie décryptée de ce DVD. L'utilisation du tatouage permet de combler cette faille en insérant des informations au sein du flux MPEG4. Les lecteurs non conformes sont capables de lire seulement les DVD illégaux et les lecteurs conformes, les DVD légaux.

1.7.4 Contrôle de diffusion

On peut insérer une marque dans une publicité afin d'en contrôler la diffusion. Cela peut également servir à réaliser une audiométrie.

1.7.5 Authentification de données

L'objectif est de détecter toutes modifications éventuelles des données afin de pouvoir certifier si celles-ci ont été modifiées ou non.

1.7.6 Indexation

Le domaine de l'indexation consiste à classer de manière automatique des couvertures selon leurs contenus. Il permet de faciliter une recherche dans une base de données [15]. La dissimulation d'un document permet ainsi d'insérer une information (contenant peu de bits) décrivant le contenu de la couverture. Cela permet d'insérer un pointeur vers une description plus complète.

1.7.7 Sécurité médicale

Il s'agit d'insérer un "identifiant" confidentiel assurant le lien entre le patient et la radio afin d'éviter toute confusion.

1.7.8 Transmission sécurisée de données

Dans ce contexte applicatif de transmission de données, la dissimulation doit satisfaire à un ensemble de contraintes spécifiques. En effet, il faut obtenir un débit de transmission le plus élevé possible tout en maintenant l'imperceptibilité de l'information cachée [16, 17]. D'autre part, la fiabilité de la transmission doit être la meilleure possible avec un coût de calcul réduit permettant l'application en temps-réel. Ces enjeux spécifiques établissent explicitement la distinction entre tatouage pour la protection de la propriété intellectuelle et la stéganographie pour la transmission de données auquel nous nous intéressons. Ils définissent également les critères de performances d'un système de transmission de données basé sur la stéganographie.

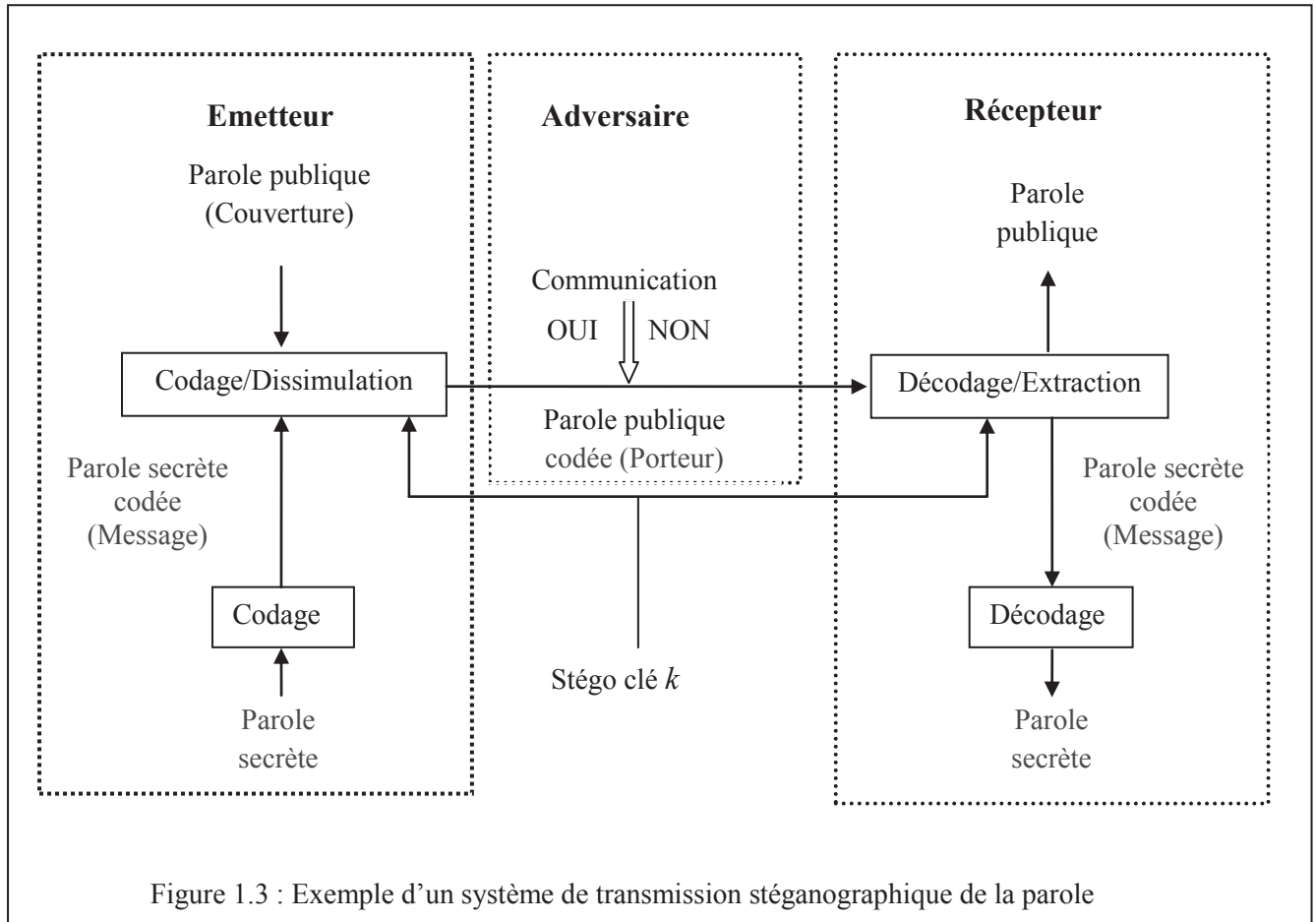
1.8 Système de transmission stéganographique de la parole

Un système stéganographique appliqué à la parole n'est pas différent de n'importe quel autre système de transmission stéganographique. La figure 1.3 montre qu'il y a principalement deux éléments dans un tel système [17]. D'un côté, un processus de dissimulation d'un message dans une couverture et de l'autre côté un processus d'extraction (recouvrement) de ce message. Cependant, ce système a quelques particularités à cause de la nature du message caché et de la couverture (parole secrète et parole publique). Ces deux derniers ne se transmettent jamais dans leurs états bruts [18], d'où le passage par une procédure de codage est obligatoire. Cette procédure permet de générer un flux binaire comprimé nommé parole codée.

La figure 1.3 présente les éléments particuliers aux techniques de stéganographie et d'insertion de données cachées (Data hiding), qui sont :

La couverture: est le fichier "host" original qui va être utilisé pour couvrir ou cacher une information. Plusieurs supports numériques peuvent être utilisés comme médium de couverture comme par exemple les protocoles de communication [19, 20] et les cellules de mémoire [21]. Cependant, le support numérique le plus employé est le fichier sous toutes ses formes, particulièrement celui sous la forme de données multimédias (documents texte, son, image, vidéo). Dans nos applications, cette couverture est la parole publique codée. Généralement, toutes les couvertures (parole publique codée) sont regroupées dans un ensemble X^* .

Le message : est l'information à cacher. Dans un système stéganographique de la parole, ce message est représenté par la parole secrète codée. On peut définir un ensemble M qui contient tous les messages possibles.



Le porteur : est le fichier qui porte un message encapsulé. Dans notre cas le porteur est la parole publique codée. On remarque que la couverture et le porteur ont la même nature, donc les porteurs et les couvertures peuvent être regroupées dans le même ensemble X^* .

Clé secrète k :

La clé secrète ou stégo clé est utilisée en conjonction avec un algorithme de dissimulation publique connu. Elle peut contrôler, par exemple, les endroits où la couverture est modifiée ou peut être utilisée pour générer d'autres primitives qui entrent dans le processus de la dissimulation [11]. Dans ce qui suit, l'ensemble qui contient toutes les clés possibles sera noté par K . Le message secret peut être crypté avant de passer à la procédure de dissimulation. Dans ce cas, il y'aura aussi une clé de chiffrement. Par conséquent, la stégo clé et la clé de

chiffrement pourraient être dérivées d'une clé maîtresse. Cependant, il est plus simple de considérer les deux clés séparément.

Dissimulation : $M \times X^* \times K \rightarrow X^*$

La dissimulation peut être considérée comme une fonction ayant comme entrées : le message secret $m \in M$, une couverture $x^{(0)} \in X^*$ et la clé secrète $k \in K$ [14]. La sortie de cette fonction est le porteur $x^{(m)}$ (résultat de la dissimulation du message m dans la couverture $x^{(0)}$) qui est pratiquement indiscernable de la couverture $x^{(0)}$. A cause de la particularité du système stéganographique de la parole, la procédure de dissimulation est liée toujours avec le codage de la parole publique. Dans certains cas, ils sont séparables comme dans le travail de Zhijun Wu [22]. Mais dans le cas général, les deux procédures (codage et dissimulation) sont effectuées au même temps.

Extraction : $X^* \times K \rightarrow M$

L'extraction est la fonction qui utilise la clé secrète k et le porteur $x^{(m)}$ (la parole publique codée avec dissimulation de message m) pour récupérer le message secret m (parole secrète).

1.9 Classification des systèmes stéganographiques de la parole

Plusieurs classifications des systèmes stéganographiques peuvent être trouvées dans la littérature. Ces classifications dépendent des caractéristiques de ces systèmes (comme la stégo clé), du domaine d'insertion des messages [14, 23] de l'architecture des systèmes [16, 22] ou du domaine d'application [24].

1.9.1 Classification selon la stégo clé

Il y a principalement trois types de techniques stéganographiques selon le type de la stégo clé: stéganographie pure, stéganographie à clés secrètes et stéganographie à clés publique. Cette dernière est basée sur les principes de cryptographie à clés publique.

1.9.1.1 Stéganographie pure

La stéganographie pure est un système dans lequel le secret de dissimulation des données ne réside que dans l'algorithme utilisé à cet effet. Donc, il n'exige aucun échange de clés. C'est-à-dire que le système stéganographique de la parole n'utilise que la parole publique (couverture) et la parole secrète (message) dans la dissimulation et dans l'extraction, il utilise

seulement la couverture. La découverte de l'algorithme rompt la dissimulation de la communication. Ceci revient à mettre en place de la "sécurité par l'obscurité". Les deux fonctions de la dissimulation et de l'extraction peuvent être redéfinies comme :

Dissimulation : $M \times X^* \rightarrow X^*$

La dissimulation dans ce cas est la fonction ayant comme entrées : la parole secrète (message secret) $m \in M$ et la parole publique (couverture) $x^{(0)} \in X^*$. La sortie de cette fonction est le porteur $x^{(m)}$ qui est pratiquement indiscernable de la couverture $x^{(0)}$.

Extraction: $X^* \rightarrow M$

L'extraction est la fonction qui utilise le porteur $x^{(m)}$ pour récupérer le message secret m .

1.9.1.2 Stéganographie à clé secrète

La stéganographie à clé secrète est similaire à la cryptographie symétrique. L'échange de données confidentielles nécessite, au préalable, l'échange d'une clé secrète (stégo clé) que l'on ne partagera qu'avec le récepteur autorisé. Il est donc nécessaire d'avoir un canal sécurisé, afin d'être certain que cette dernière ne soit pas compromise. Cette clé aura une influence sur la manière de dissimuler la parole secrète, donc elle est générée en exploitant quelques caractéristiques de la parole publique et du codage. Ce système est le même système présenté à la figure.1.3.

1.9.1.3 Stéganographie à clé publique

La stéganographie à clé publique est similaire à la cryptographie asymétrique. La personne voulant envoyer des données à un autre interlocuteur, sans éveiller de soupçons, utilisera la clé publique de ce dernier. La clé publique étant à priori connue de tout le monde, il n'y aura pas besoin d'échange préalable "sécurisé". La personne recevant ce message sera la seule à pouvoir en extraire son contenu à l'aide de sa clé privée.

1.9.2 Classification selon le domaine d'insertion

Les techniques de la stéganographie de cette classe sont liées aux différents espaces de représentations appelés domaines d'insertion. Chaque domaine dispose de divers schémas d'insertion.

1.9.2.1 Domaine temporel

Les méthodes d'insertion typique dans le domaine du temps modifient directement les valeurs des échantillons de la couverture (parole publique). Dans l'ensemble, des petits changements dans les intensités ou les amplitudes des échantillons causent des différences perceptuelles très faibles pour la couverture.

Ce sont des méthodes simples et peu coûteuses en temps de calcul. Le problème majeur de ces méthodes est le type du codeur public qui doit coder directement la forme d'onde du signal, ce qui n'est pas toujours le cas.

1.9.2.2 Domaine fréquentiel

Dans le domaine fréquentiel, la modification est effectuée sur la présentation fréquentielle de la couverture. Les méthodes dans ce domaine doivent prendre en considération les caractéristiques du codeur (bande passante, type de codeur,...).

1.9.2.3 Domaine de compression

Les données comprimées contiennent beaucoup plus d'informations que le domaine brut. Une petite modification de ces données peut conduire à une détérioration considérable de la qualité du porteur décodée.

Ainsi, les méthodes pour la dissimulation du message doivent être bien adaptées pour les données comprimées, de sorte que la dégradation de la qualité perceptuelle du porteur soit minimale. Dans la parole, la technique de compression la plus utilisée est la quantification vectorielle [25]. Donc, toutes les méthodes d'insertion doivent être adaptées avec cette technique.

1.9.3 Classification selon l'architecture

Dans cette section, nous donnons la classification des systèmes stéganographiques selon l'architecture du bloc codage/dissimulation. Il existe trois architectures possibles : dissimulation avant le codage, durant le codage et après le codage.

1.9.3.1 Dissimulation avant le codage

Dans cette approche, le message est dissimulé (caché) directement sur les échantillons du signal parole publique [22, 26]. Cette approche, basée sur le domaine temporel et la mise en œuvre des algorithmes, est très simple du point de vue implémentation.

L'application de cette approche nécessite quelques conditions liées toujours avec les propriétés du codeur utilisé pour la parole publique. Généralement, ce codeur doit être de type forme d'onde; c'est la raison pour laquelle cette approche n'est pas vraiment utilisée dans les travaux de recherche récents.

1.9.3.2Dissimulation durant le codage

Il y'a beaucoup de méthodes utilisant l'approche de dissimulation durant le codage dans le domaine compressé [8, 16, 22]. Ces méthodes ont été largement utilisées puisqu'elles offrent la possibilité d'obtenir une meilleure adaptation entre la procédure de dissimulation et le codage qui donne une dégradation minimale de la parole publique.

L'application de cette approche nécessite une modification de la procédure de codage qui n'est pas toujours facile à réaliser. En effet, cette modification nécessite une connaissance spécifique du codeur de la parole publique utilisé, pas seulement son type. Néanmoins, cette approche offre généralement de meilleurs résultats [16] selon les critères que nous allons détailler dans le paragraphe 1.11.

1.9.3.3Dissimulation après le codage

Dans cette approche, le message est dissimulé directement sur la parole publique codée [27, 28]. Cette approche est utilisable pour tous les types de codeur de parole et spécialement pour les codeurs GSM à bas débit. Généralement, cette approche se base sur les propriétés du signal de parole publique comme l'énergie par exemple, mais dans certain cas elle se base aussi sur les propriétés de la parole publique codée [29].

1.10 Techniques de dissimulation de la parole

Les techniques de dissimulation ont été développées principalement pour les systèmes de tatouage. Leurs applications pour des systèmes de communication parlée sécurisés ont commencées un peu plus tard. En exploitant les propriétés des codeurs de la parole, plusieurs techniques de dissimulation ont été adaptées pour ces systèmes au cours des dernières années.

1.10.1 Dissimulation LSB

L'une des premières techniques de dissimulation de l'information est le codage/remplacement du bit le moins significatif LSB (Least Significant Bit) [30, 15]. Une simple implémentation de cette technique consiste à remplacer le bit le moins significatif du signal couverture par un

Au niveau de l'extraction, la même clé secrète est utilisée pour choisir les mêmes deux ensembles de données. Ensuite, la différence de l'espérance de ces deux ensembles de données est calculée. Si elle est égale à $2d$, le signal reçu est modifié par la procédure de dissimulation présenté par l'équation 1.1.

1.10.3 Quantification par modulation d'indice

Les travaux de Chen et Wornell en 2001 marquent le début des travaux de recherche sur les fonctions d'insertion par quantification [15, 22]. Ils ont proposé un nouveau système de dissimulation basé sur la quantification par modulation d'indice QIM (Quantized Index Modulation).

Le principe de la QIM peut être résumé comme suit : la fonction d'insertion correspond à la quantification d'une couverture x par un quantificateur $Q_m(x)$ dépendant du message m à émettre :

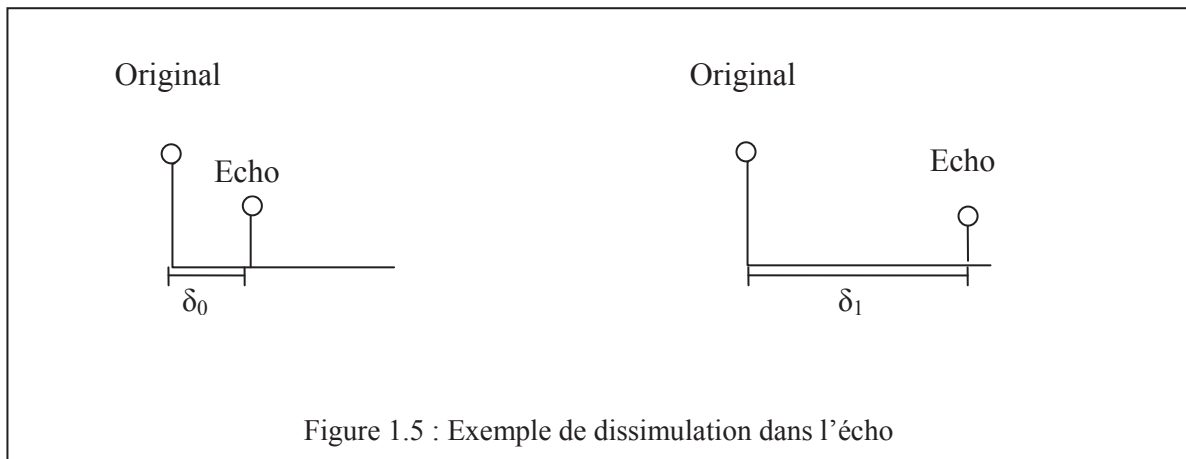
$$f(x, m) = Q_m(x) \quad (1.2)$$

Cette technique utilise la quantification scalaire ou vectorielle dans la procédure de dissimulation et de l'extraction. Elle nécessite quelques connaissances sur ces deux types de quantification. Donc, nous allons détailler cette technique dans le chapitre 3.

1.10.4 Dissimulation par écho

Cette technique dissimule des données dans un signal porteur audio par l'introduction d'un écho court. Le retard (offset) entre le signal original et l'écho est tellement petit que l'écho est perçu comme une résonance ajoutée [30]. La variation dans ce décalage est utilisée pour dissimuler le message. La dissimulation de message est effectuée par la manipulation de trois paramètres du signal écho: le taux d'amplitude initial, le décalage (retard) et le taux de décroissance de l'amplitude (lorsque on augmente le décalage entre le signal original et l'écho on doit diminuer l'amplitude de l'écho) de sorte que l'écho reste toujours pas audible [23]. Par exemple, si on veut dissimuler un seul bit (un message de 0 ou 1), le décalage est illustré dans la figure 1.5.

Dans la procédure d'extraction, on compare le décalage entre l'écho et le signal original. Si le décalage est égale à δ_0 , le bit secret est 0; par contre, s'il est égale à δ_1 , le bit secret est alors 1.



1.11 Critères d'évaluation d'un système stéganographique

Les méthodes d'insertion de données à cacher nécessitent différentes propriétés selon leurs domaines d'application et leurs finalités.

1.11.1 Sécurité

La sécurité est utilisée pour exprimer la façon dont le message caché peut être protégé [14]. Généralement, la sécurité d'un système stéganographique est représentée par deux paramètres qui sont l'indéfectabilité et l'imperceptibilité.

1.11.1.1 Indéfectabilité

L'indéfectabilité est la capacité d'un message caché à ne pas être détecté par des analyses statistiques. Cette condition est indispensable pour les systèmes de communication secrète basés sur la stéganographie.

1.11.1.2 Imperceptibilité (transparence)

Les données cachées doivent être entièrement imperceptibles par le système auditif humain (SAH). Ainsi, l'opération d'insertion ne doit pas détériorer le porteur d'une façon perceptible. L'imperceptibilité est une propriété fortement liée au tatouage invisible et à la stéganographie.

1.11.2 Robustesse

La robustesse est l'aptitude à préserver les données cachées face aux attaques. Cette demande est fortement liée à la plupart des types de tatouage, particulièrement ceux pour la protection des droits d'auteurs.

1.11.3 Capacité

La capacité est la quantité d'information maximale que le fichier couverture peut dissimuler. Elle est généralement mesurée en bits. Dans le contexte de marquage pour la protection des droits d'auteurs, la capacité n'est pas primordiale. L'insertion d'un numéro d'identification codé sur 64 bits suffit dans la plupart des applications. Cependant, cette propriété est très importante pour la stéganographie.

1.11.4 Le coût

Le coût représente le temps consommé par la procédure de dissimulation (insertion) et la détection/extraction [31]. L'importance de ce temps dépend de l'application considérée. Par exemple, il est très important dans les systèmes de transmission des données cachées mais moins important dans les systèmes de tatouage.

1.12 Conclusion

Dans ce chapitre, nous avons présenté brièvement trois schémas de systèmes de transmission sécurisés. Le premier est le système cryptographique qui est le système le plus ancien parmi eux. Il est basé sur le chiffrement des messages afin de les rendre illisibles. Le deuxième schéma est celui d'un système de tatouage. L'idée de base d'un tel système est d'insérer une marque dans une couverture pour assurer sa protection. Enfin, le troisième est le schéma d'un système stéganographique qui se base aussi sur le principe de dissimulation de données dans une couverture. Cependant, il existe quelques différences importantes entre les deux derniers systèmes. Par exemple, dans le système stéganographique le message dissimulé ne doit jamais être détecté; par contre, dans le système de tatouage, il est possible de détecter la marque mais il est impossible de la supprimer.

Dans les systèmes de transmission de la parole sécurisés à base de stéganographie, le codage a une importance capitale, comme nous l'avons mentionné précédemment. Il est donc indispensable de connaître le type de codage de la parole utilisé. Le codage de la parole est un domaine très vaste qui se base sur les techniques utilisées par les codeurs de parole normalisés par plusieurs organisations internationales comme l'ITU-T (international telecommunication union), l'ETSI (European Telecommunication Standards Institute),...etc.

Dans le chapitre suivant, nous détaillerons quelques techniques de codage de la parole comme la quantification qui sera utilisée dans les techniques de dissimulation que nous avons

conçues. Nous donnerons aussi une brève description de deux codeurs de parole normalisés : le AMR-WB (Adaptive multi-Rate Wideband) norme G 722.2 et le codeur de parole bas débit MELP (Mixed Excitation Linear Predictive). En effet, ces deux standards seront utilisés ultérieurement dans l'implémentation de nos systèmes de codage de la parole sécurisés à base de stéganographie.

Chapitre 2

Codage de la parole

2.1 Introduction

Dans les systèmes de communication actuels, le signal de parole est représenté dans un format numérique binaire. Pour des applications de transmission, il est souvent souhaitable que le signal soit représenté par le moins de bits possible afin de diminuer les exigences en termes de bande passante, de taille mémoire, de puissance et de stockage.

Dans ce chapitre, nous décrivons les méthodes de codage de source par quantification scalaire et vectorielle. Nous détaillerons particulièrement le quantificateur vectoriel divisé SVQ (Split Vector Quantizer) et le quantificateur vectoriel multi-étages MSVQ (MultistageVector Quantizer). Ensuite, nous passerons dans le domaine de la parole où nous présenterons d'abord le principe de la prédiction linéaire et les paramètres fréquence spectrale d'immittance (ISF). Nous donnerons aussi une classification des techniques de codage de la parole (forme d'onde, paramétrique et hybride) et une brève description du codeur de parole MELP. Après, nous donnerons une description du domaine de la parole en large bande (fréquence d'échantillonnage, bande passante, avantage du codage en large bande et domaine d'utilisation). Nous terminerons par une description générale du codeur AMR-WB norme

G722.2 en se basant sur les méthodes de quantification utilisées et la procédure de recherche (ACELP) du vecteur d'excitation.

Nous nous sommes intéressés particulièrement à ces codeurs de parole car ces derniers seront utilisés ultérieurement dans la conception d'un système de transmission de la parole sécurisé. En effet, la parole publique (couverture) sera codée par le codeur AMR-WB G722.2 et la parole secrète par le codeur MELP.

2.2 Quantification

La quantification [1, 32] revient à approximer la valeur d'un signal original par la valeur la plus proche nommée valeur reproduite. Plus le nombre de valeurs possibles est élevés, plus la valeur reproduite sera proche de la valeur originale. Le signal quantifié sera donc plus fidèle à l'original. Les méthodes connues de quantification peuvent être classées en deux principales catégories :

- La quantification scalaire QS.
- La quantification vectorielle QV.

2.2.1 Quantificateur Scalaire

La quantification scalaire QS [33] consiste à coder des échantillons (un par un) qui sont représentés par des valeurs. Autrement dit, la QS est une application qui associe à chaque valeur x de l'ensemble des réels $\mathfrak{R}$ une valeur y_i d'un ensemble fini $Y \subset \mathfrak{R}$, appelé alphabet. Soit Q un quantificateur scalaire, X la base d'apprentissage et Y l'alphabet de taille L .

Alors : $Q : \mathfrak{R} \rightarrow Y$

$$x \rightarrow y_i \quad (2.1)$$

Avec $Y = \{y_1, y_2, y_3, \dots, y_L\}$.

Les L valeurs de sortie y_i sont nommés niveaux de sortie (de quantification) ou représentants.

Le quantificateur scalaire QS partitionne l'ensemble X en L classes R_i et associe à chaque classe d'indice i un représentant y_i .

$$D'où : R_i = \{x \in \mathfrak{R} : Q(x) = y_i\} \quad (2.2)$$

Où $Q(x) = y_i$ pour tout $x \in R_i = [x_i, x_{i+1}]$, avec x_i nommé niveau de discision.

On détermine un QS par:

- Les L classes R_i ou les $L + 1$ niveaux de discision x_i .
- Les L représentants y_i .

Principe de codage par quantificateur scalaire

Le codeur est muni d'un alphabet Y indexé qui est construit par le quantificateur scalaire. Son principe est de calculer la distance entre l'échantillon x et tous les représentants y_i , puis transmettre l'indice i du représentant le plus proche. Le fait de transmettre l'indice et non pas la valeur réelle x de l'entrée provoque une compression. D'où un gain de débit (ou d'espace mémoire). A la réception, le décodeur reçoit l'indice i et produit en sortie le représentant y_i d'indice i . L'alphabet du décodeur est identique à celui du codeur.

Le changement de l'échantillon original x par un représentant y_i provoque une erreur e .

$$e = Q(x) - x \quad (2.3)$$

Il existe deux types d'erreurs de quantification :

- Le bruit granulaire : apparait si $x \in [x_0, x_L]$.
- Le bruit de surcharge : apparait dans le cas où x est hors de l'intervalle $[x_0, x_L]$.

Ce type de codage utilise en général un nombre L de puissance 2 puisque la transmission des indices se fait par bit logique ($L = 2^R$). On définit R comme le débit de transmission en bits par échantillon bpe .

2.2.2 Quantificateur vectoriel

La quantification vectorielle QV [33] est une simple généralisation du cas scalaire, où l'on regroupe plusieurs échantillons dans un vecteur et on cherche à quantifier l'ensemble. Elle permet de prendre en compte la corrélation entre les échantillons. De plus, il a été montré par Shannon [1, 34] que plus on regroupe des échantillons plus les performances deviennent meilleures.

Autrement dit, la QV est une application qui associe à chaque vecteur x de l'ensemble $\mathfrak{R}^k$ un vecteur y_i d'un ensemble fini $Y \subset \mathfrak{R}^k$, appelé dictionnaire.

Soit Q un quantificateur vectoriel (VQ), X la séquence d'apprentissage et Y le dictionnaire.

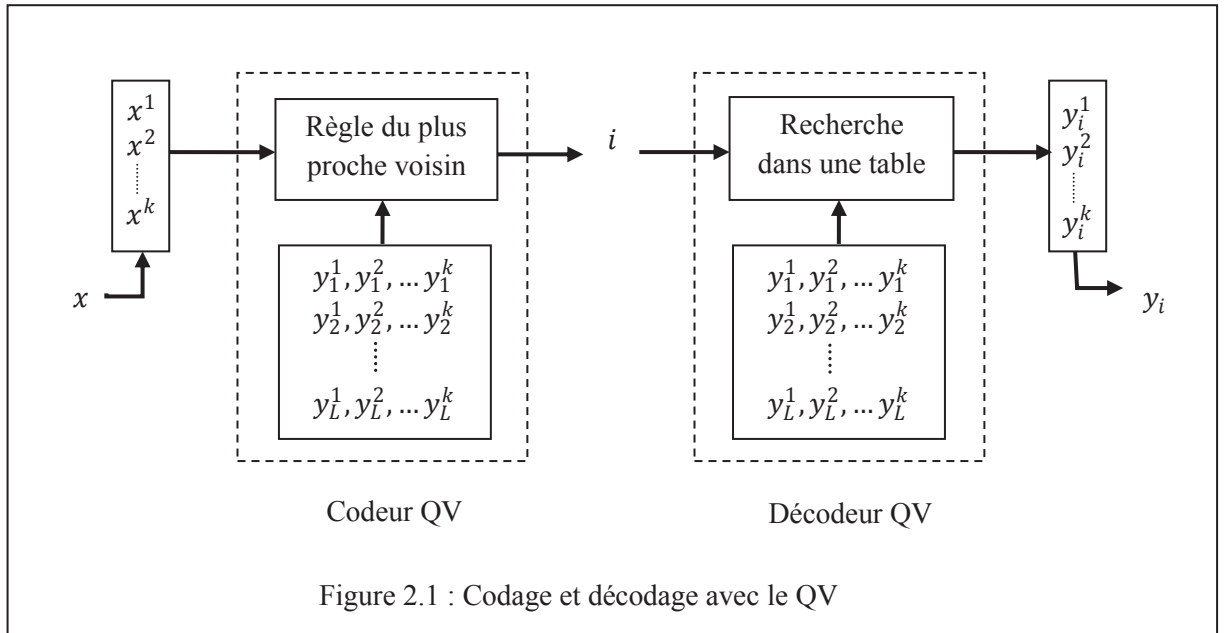
Alors : $Q: \mathfrak{R}^k \rightarrow Y$

$$x \rightarrow y_i \quad (2.4)$$

Avec $Y = \{y_1, y_2, y_3, \dots, y_L\}$ ($L = 2^b$ où b est le débit en bits par vecteur $b = k.R$).

2.2.2.1 Principe de codage par quantificateur vectoriel

Le codeur est muni d'un dictionnaire indexé Y qui est construit par le quantificateur vectoriel QV (figure 2.1). L'idée de base est de coder le vecteur x par le représentant y_i puis transmettre l'indice du représentant. Le fait de transmettre l'indice et non pas tout le vecteur de dimension k provoque une compression, d'où un gain en débit. A son tour, le décodeur reçoit l'indice (supposons la transmission est idéal) et produit en sortie le représentant y_i qui correspond à l'indice reçu i .



De nombreux schémas de quantificateurs vectoriels structurés ont été proposés dans la littérature. Les principaux sont les suivants. Dans le quantificateur vectoriel à étages multiples MSVQ [35], on met en cascade des quantificateurs vectoriels de niveaux moindres. Les dictionnaires sont de tailles réduites et il en va de même en ce qui concerne le temps de calcul et le coût en mémoire. Le quantificateur vectoriel divisé SVQ [36] (Split Vector Quantizer)

est composé de N quantificateurs vectoriels classiques de tailles et de dimensions plus petites. On décompose le vecteur d'entrée en N sous vecteurs et chaque sous vecteurs sera quantifié indépendamment des autres. Le VQ à divisions commutées SSVQ [37] (Switched Split Vector Quantizer) est un autre type de quantificateur vectoriel structuré. Ce quantificateur décompose la base d'apprentissage en m classes, puis il quantifie chaque classe avec un SVQ correspondant. Dans le quantificateur vectoriel multi étages divisé S-MSVQ (Split Multistage Vector Quantizer) [38], le vecteur d'entrée est divisé en N sous vecteurs où chaque sous vecteurs sera quantifié indépendamment des autres en utilisant un MSVQ.

2.2.2.2 Quantificateur vectoriel divisé

Le quantificateur vectoriel divisé SVQ [36] est composé de N quantificateurs vectoriels classiques de tailles et de dimensions plus petites. On décompose le vecteur d'entrée x en N sous vecteurs ou chaque sous vecteurs sera quantifié indépendamment des autres.

L'idée de base du SVQ consiste à partitionner les vecteurs x de dimension k de la base d'apprentissage en N sous vecteurs (ou parties) de dimension k_i (avec $\sum_{i=1}^N k_i = k$). En suite, chaque partie sera quantifiée indépendamment des autres. Cette technique nous donne N dictionnaires de tailles $L_i = 2^{R_i \times k_i}$ (où $L = \prod_{i=1}^N L_i$ et R_i est le débit partiel en *bpe*). D'où, la possibilité d'allouer des débits différents $b_i = R_i \times k_i$ pour chaque partie. Donc selon l'importance, on donne plus de précision de quantification pour une partie choisi et moins pour une autre. On note cette quantification par N -SVQ où N représente le nombre de partie (ou division). Si le nombre de partie N est égal à la dimension k des vecteurs, on obtient un quantificateur scalaire; par contre si $N = 1$ alors c'est un quantificateur vectoriel conventionnel.

Le schéma bloc suivant (figure 2.2) montre une quantification 2-SVQ d'une séquence X composée de vecteurs x de dimension k . Les vecteurs x sont décomposés en deux sous vecteurs x_a de dimension m et x_b de dimension $k - m$. Chaque partie est quantifiée respectivement avec un débit $b_1 = R_1 \times m$ bits par sous vecteur et $b_2 = R_2 \times (k - m)$ bits par sous vecteur. Par conséquent, on obtient deux dictionnaires Y_a de taille $L_1 = 2^{b_1}$ et Y_b de taille $L_2 = 2^{b_2}$.

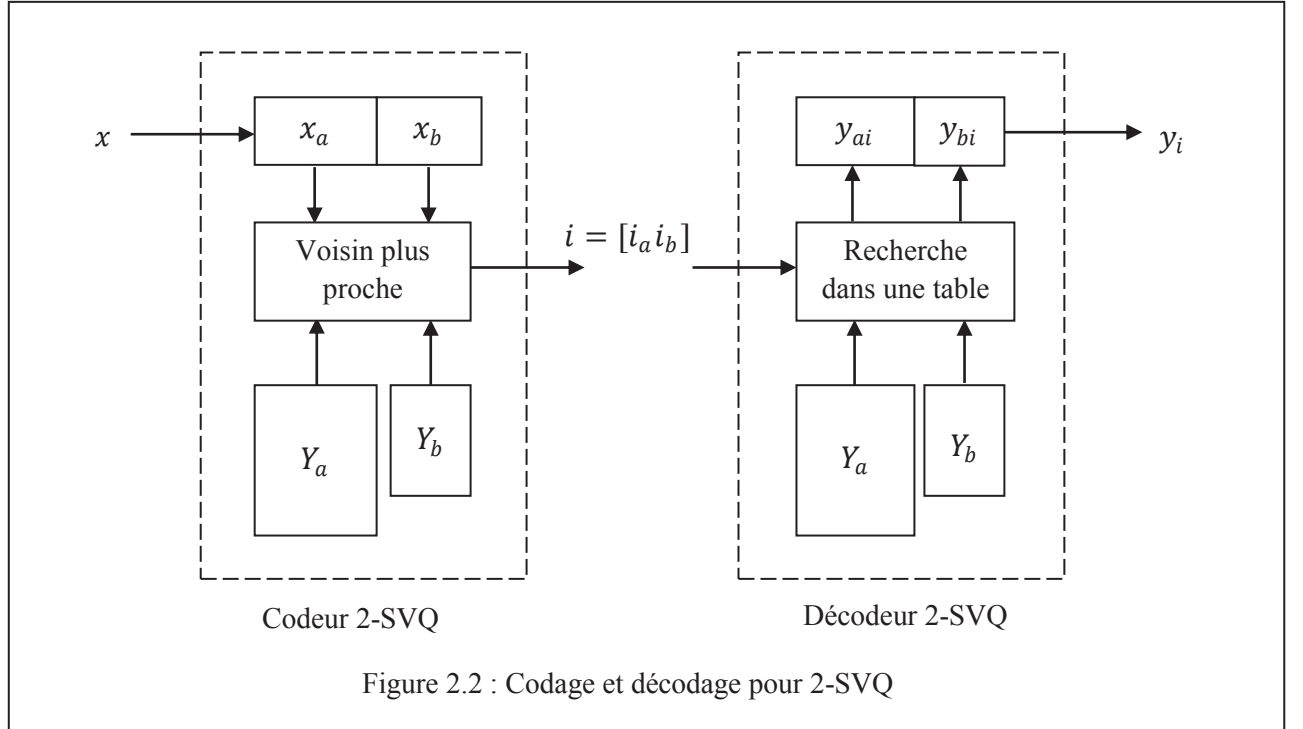
Soit Q le quantificateur 2-SVQ précédant, x un vecteur de dimension k , Y_a et Y_b les dictionnaires construit par ce quantificateur.

Alors : $Q: \mathcal{R}^k \rightarrow Y_a \cup Y_b$

$$x \rightarrow y_i \quad (2.5)$$

Avec $x = [x_a x_b]$ et $y_i = [y_{ai} y_{bi}]$ ($y_{ai} \in Y_a, y_{bi} \in Y_b$).

D'où



En se basant sur l'exemple précédent du 2-SVQ, le principe du codage et décodage par le SVQ peut être résumé comme suit (figure 2.2): l'encodeur doit trouver d'abord l'indice i_a du représentant le plus proche de x_a dans le dictionnaire Y_a . Comme tâche séparée et indépendante, il doit chercher ensuite l'indice i_b du représentant le plus proche de x_b dans le dictionnaire Y_b . L'indice i à transmettre est la concaténation des deux indices i_a et i_b .

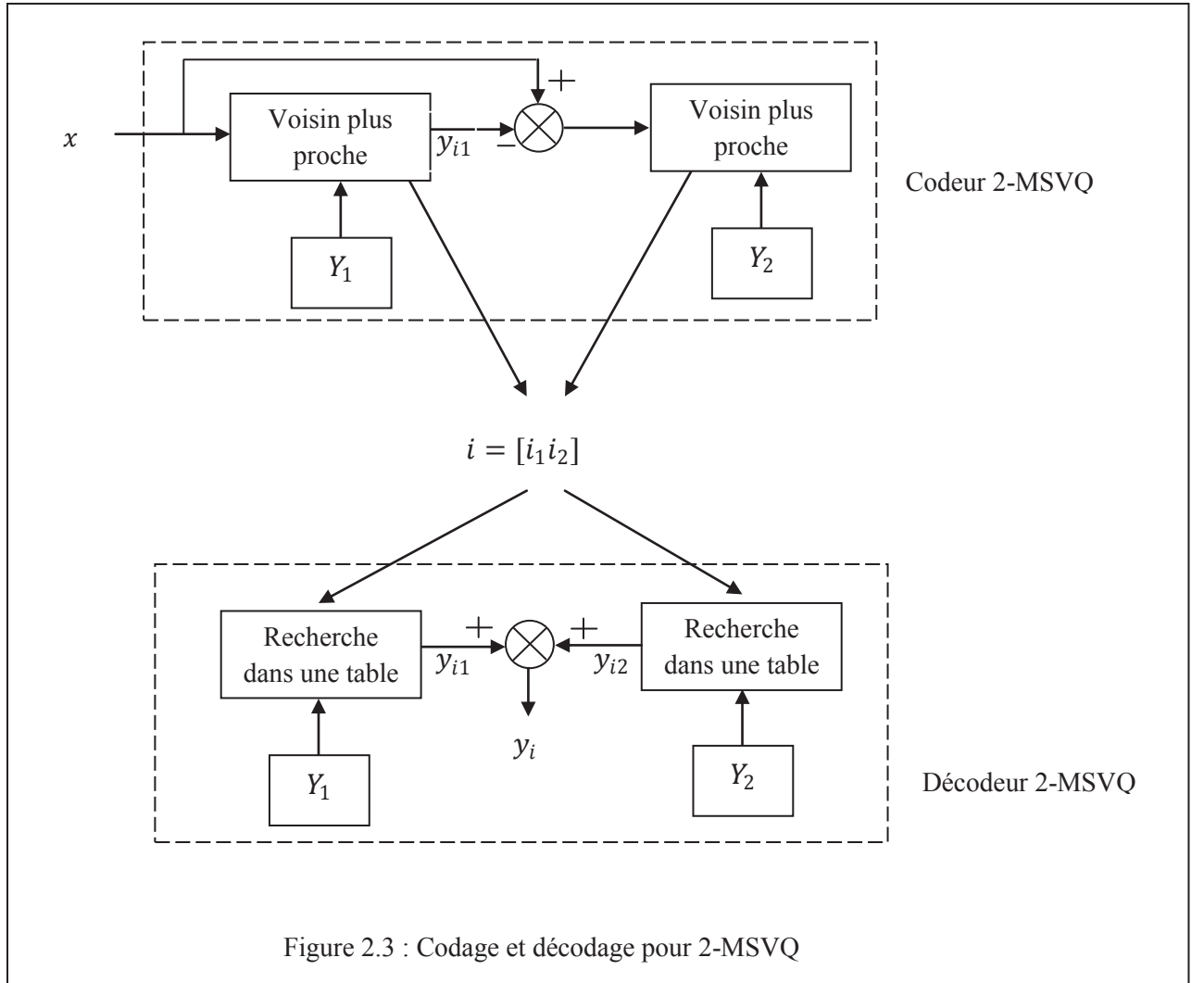
A la réception, le décodeur reçoit l'indice i et procède à la construction du représentant $y_i = [y_{ai} y_{bi}]$ où le sous vecteur y_{ai} est le représentant d'indice i_a et le sous vecteur y_{bi} est le représentant d'indice i_b dans les deux dictionnaires Y_a et Y_b respectivement.

2.2.2.3 Quantificateur vectoriel multi étages

La quantification MSVQ à m étages [36] consiste à utiliser une partie du débit total de la quantification pour construire un dictionnaire de taille réduite avec le QV. Ensuite, utiliser

une autre partie du débit pour quantifier l'erreur causée par le premier quantificateur (1^{er} étage). Et ainsi de suite, en utilisant chaque fois une partie du débit jusqu'à avoir m étages.

Cette technique nous donne m dictionnaires de tailles $L_i = 2^{R_i \times k}$ (où $L = \prod_{i=1}^N L_i$ et R_i est le débit partiel en *bpe*). D'où, la possibilité d'allouer des débits différents pour chaque étage $b_i = R_i \times k$. On note cette quantification par m -MSVQ où m représente le nombre d'étage. Si le nombre d'étage m égale à 1, alors on obtient un quantificateur vectoriel classique.



Le principe du codage et décodage par le MSVQ peut être résumé comme suit (En se base sur l'exemple du 2-MSVQ présenté dans la figure 2.3) : l'encodeur doit trouver d'abord l'indice i_1 du vecteur-code y_{i1} le plus proche au vecteur x dans le dictionnaire Y_1 du 1^{er} étage. Ensuite, dans le dictionnaire Y_2 de l'étage 2, il doit chercher l'indice i_2 du vecteur-code y_{i2} le plus proche au vecteur résiduel de la quantification du 1^{er} étage $x - y_{i1}$. L'indice i à transmettre est la concaténation des deux indices i_1 et i_2 .

A la réception, le décodeur reçoit l'indice i et procède à la construction du vecteur de reproduction $y_i = y_{i1} + y_{i2}$, où le vecteur y_{i1} est le vecteur-code d'indice i_1 et le vecteur y_{i2} est le vecteur-code d'indice i_2 dans les deux dictionnaires Y_1 et Y_2 respectivement.

2.3 Principe de la prédiction linéaire

Dans la plupart des applications du codage de la parole, l'analyse par prédiction linéaire LP (Linear prediction) est largement utilisée pour représenter l'enveloppe spectrale à court-terme, d'un signal de parole. La LP est une technique d'estimation qui suppose qu'un échantillon futur $s(n)$ d'un signal (de parole par exemple) peut être prédit approximativement par une combinaison linéaire d'un certain nombre (P) d'échantillons le précédant immédiatement [18].

Si l'on désigne par $\hat{s}(n)$ la valeur estimée de $s(n)$, celle-ci s'écrit :

$$\hat{s}(n) = - \sum_{i=1}^p a_i s(n-i) \quad (2.7)$$

Où $a_i \{1 \leq i \leq p\}$ représentent les coefficients de la prédiction linéaire (LPC) et p l'ordre de la prédiction.

L'erreur $e(n)$ de la prédiction s'écrit sous cette forme :

$$e(n) = s(n) - \hat{s}(n) = s(n) + \sum_{i=1}^p a_i s(n-i) \quad (2.8)$$

En appliquant la transformée en Z à l'équation précédente, on obtient :

$$E(z) = S(z) \sum_{i=1}^p a_i z^{-i} + S(z) \quad (2.9)$$

Ce qui donne le modèle d'analyse LP suivant :

$$E(z) = S(z) \cdot A(z) \quad (2.10)$$

Avec :

$$A(z) = 1 + \sum_{i=1}^p a_i z^{-i} \quad (2.11)$$

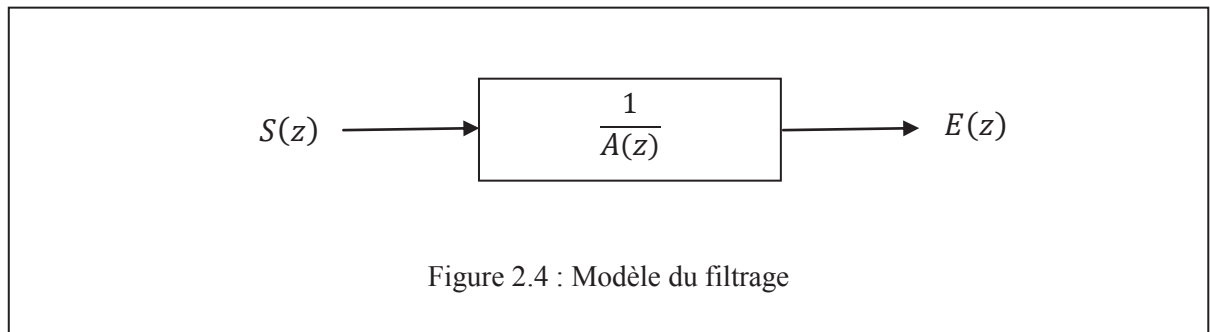
$A(z)$ est appelé filtre d'ordre P d'analyse LP ou encore "filtre inverse".

Inversement, le modèle de synthèse par LP est régi par l'équation suivante :

$$S(z) = \frac{1}{A(z)} \cdot E(z) \quad (2.12)$$

Où $1/A(z)$ est connu par "filtre direct " ou "filtre de synthèse".

Il s'agit donc, de considérer le signal (de parole synthétique) $S(z)$ comme étant la sortie d'un filtre ne comprenant que des pôles. C'est-à-dire engendré par un modèle autorégressif (AR) comme il est montré ci-dessous.



On peut ainsi interpréter la synthèse par LP comme un processus de filtrage autorégressif sur le signal d'erreur de prédiction (c.-à-d., signal d'excitation $E(z)$). Le calcul des coefficients LPC (LPCs) de ce filtre est une étape importante dans la modélisation autorégressive. Le problème de la détermination optimale des coefficients a_i du filtre de synthèse est résolu par la minimisation de l'erreur de prédiction quadratique moyenne. Tout cela a conduit à mettre en œuvre des algorithmes de résolution plus ou moins simples (Algorithme de Levinson-Durbin, de Burg,...) que l'on utilise selon différentes hypothèses d'analyse et qui sont généralement bien documentés [36, 39]

1.4 Les paramètres fréquence spectrale d'immitance

Les paramètres fréquence spectrale d'immitance ISF [40] sont des représentations mathématiques des coefficients LPC. Ces représentations offrent des avantages importants pour le codage de la parole (stabilité du filtre de synthèse et quantification efficace).

Les paramètres ISF (ISFs) sont obtenues à partir du filtre d'analyse LP d'ordre p défini par l'équation 2.11. La dimension k des vecteurs ISFs correspond à l'ordre p utilisé.

Les coefficients LPC sont convertis en ISPs, vu l'efficacité de ces derniers dans la quantification. Les paires ISPs sont définis comme les racines des deux polynômes suivants :

$$F_1(z) = A(z) - z^{-p}A(z^{-1}) \quad (2.13)$$

et

$$F_2(z) = A(z) + z^{-p}A(z^{-1}) \quad (2.14)$$

Le polynôme $F_1(z)$ présente deux solutions toujours fixe ($z = -1$ et $z = 1$). Donc, il est divisé sur $(1 - z^{-2})$ pour supprimer ces deux solutions. Pour un ordre de prédiction $p=16$, les polynômes F_1 et F_2 peuvent s'écrire comme suit :

$$F_1(z) = \sum_{i=0}^{14} f_1(i).z^{-i} \quad (2.15)$$

et

$$F_2(z) = \sum_{i=0}^{16} f_2(i).z^{-i} \quad (2.16)$$

Les coefficients $f_1(i)$ et $f_2(i)$ sont respectivement symétriques. C'est à dire que $f_1(i) = f_1(14 - i)$ et $f_2(i) = f_2(16 - i)$ pour $i = \{0, 1, \dots, 7\}$.

Soient q_i les racines de $F_1(z)$ et $F_2(z)$ qu'on appelle paires spectrales d'immittance (ISP). Les paramètres ISF sont une présentation des ISP dans le domaine des cosinus. D'où, les fréquences spectrales d'immittance ISFs sont données par la relation suivante :

$$\begin{cases} f_i = \frac{f_s}{2\pi} \cos^{-1}(q_i) & i = 0, 2, \dots, 14 \\ f_i = \frac{f_s}{4\pi} \cos^{-1}(q_i) & i = 15 \end{cases} \quad (2.17)$$

Dans la pratique, on travail avec des paramètres ISF normalisés avec la fréquence d'échantillonnage f_s . Ils répondent à la relation d'ordonnancement suivante :

$$0 < w_0 < w_1 < \dots < w_{14} < 0.5 \quad (2.18)$$

2.5 Classification des techniques de codage de parole

Le but de toute technique de codage de la parole est de réduire la quantité de données binaires à transmettre et son attribut fondamental est la qualité de la parole restituée. Typiquement, plus le débit de codage est faible, moins la qualité du signal restitué est bonne. Les différentes normes couvrent diverses applications avec des débits variant de 128 Kbits/s jusqu'à moins de 600 bits/s. Les algorithmes de codage de la parole utilisés dans les normes peuvent être groupés en 3 grandes catégories : le codage de formes d'onde, le codage paramétrique et le codage hybride.

2.5.1 Codage de forme d'onde

Les codeurs temporels codent directement la forme d'onde du signal, sans connaissance a priori de ses caractéristiques. Ce genre de codage est utilisé dans plusieurs standards destinés principalement aux réseaux téléphoniques filaires. Il conduit à un débit de codage élevé, typiquement compris entre 16 et 64 Kbits/s et il permet de préserver une bonne qualité perceptuelle pour le signal vocal. La complexité de ces codeurs est minimale ainsi que le délai de codage [18].

Le codage porte directement sur les échantillons de la forme d'onde du signal de parole qui sont quantifiés et transmis. Puisque le signal variant peu d'un échantillon à un autre, un codage différentiel est alors introduit afin de réduire le débit. Dans ce cas, c'est la différence entre la valeur d'un échantillon et sa valeur prédite qui est quantifiée puis transmise. La prédiction est réalisée par une combinaison linéaire à partir des valeurs précédemment quantifiées.

2.5.2 Codage hybride

Les codeurs dits hybrides codent l'aspect temporel du signal tout en tirant profit d'un modèle de production de la parole et de certains aspects de la perception auditive. Le modèle de production de la parole sépare la source d'excitation du canal vocal, qui peut être quasi-périodique pour les sons voisés ou de type bruit pour les sons fricatifs ou plosifs. La forme du conduit vocal détermine ses fréquences de résonance et l'enveloppe spectrale (formants) du signal de parole. Le signal de parole est modélisé par un modèle " source-filtre ". Dans ce genre de modèle, le signal de parole est la sortie d'un filtre tout-pôle appelé filtre de synthèse, dont la fonction de transfert représente l'enveloppe spectrale. Ce filtre est excité par une

entrée dont les caractéristiques (en particulier la fréquence fondamentale) déterminent la structure fine du spectre.

Les approches les plus classiques utilisent la prédiction linéaire. Le signal est découpé en trames, typiquement d'une durée variant de 10-30 ms, sur lesquelles il est supposé être stationnaire. Dans chaque trame, une analyse spectrale par prédiction linéaire à court-terme est appliquée conduisant à une série de coefficients de prédiction linéaire (LPC Linear Prediction Coefficients). Les LPCs constituent les coefficients du dénominateur du filtre de synthèse tout-pôle $A(z)$ dont la fonction de transfert fréquentielle $1/A(z)$ modélise la densité spectrale de puissance du signal.

Les codeurs hybrides sont pour la plupart dérivés de l'algorithme de codage CELP (Code Excited Linear Prediction) introduit par Schroeder et Atal [41]. Cet algorithme et ses dérivés sont à la base de la majorité des standards de codeurs de parole utilisés dans les réseaux de téléphonie mobiles et les communications vocales sur Internet en voix sur IP.

De nombreux travaux de développement des algorithmes de type CELP ont porté sur le contenu et la structure des dictionnaires de façon à améliorer la qualité du codeur mais aussi à réduire la complexité de calcul.

Parmi les variantes de l'algorithme CELP, on peut citer le codeur ACELP Algebraic CELP [42] qui utilise un dictionnaire stochastique algébrique. Un autre algorithme utilisant un dictionnaire algébrique est le CS-ACELP Conjugate Structure ACELP [43] qui est utilisé par les codeurs multi canaux comme le codeur G729 ou l'AMR Adaptive Multi-Rate.

2.5.3 Codage paramétrique

Les codeurs paramétriques modélisent le signal vocal à l'aide d'un modèle paramétrique propre aux signaux de parole. Le codeur extrait les paramètres du modèle et les code sur un nombre limité de bits. Le décodeur utilise ces paramètres pour alimenter le modèle et synthétiser le signal de parole. Ces codeurs permettent d'obtenir de faibles débits de codage, typiquement inférieurs à 4 Kbits/s, tout en conservant une bonne qualité de parole même si elle est généralement inférieure à celle des codeurs hybrides. Les algorithmes de codage paramétriques sont destinés à des applications de sécurité et ils sont souvent développés et standardisés par des organismes militaires, comme le DoD ou l'OTAN [18].

Les codeurs de parole utilisant le codage paramétrique diffèrent par le type de modèle utilisé et par la façon de coder les paramètres. Deux types de modèles sont couramment utilisés : le modèle " source-filtre " décrit précédemment et les modèles sinusoïdaux STC (Sinusoidal Transform Coder). Ces codeurs calculent différents types de paramètres, en particulier des paramètres spectraux qui représentent l'enveloppe spectrale du signal. Les paramètres spectraux sont généralement obtenus par prédiction linéaire mais on utilise aussi parfois une transformée de Fourier à court terme. D'autres paramètres comme le voisement ou les niveaux de voisement, la fréquence fondamentale, l'énergie (souvent représentée par un gain), les valeurs de quelques harmoniques sont aussi utilisés.

Les codeurs source-filtre utilisent différentes approches pour construire le signal d'excitation. On peut distinguer les codeurs à deux états d'excitation (LPC classique), les codeurs à excitation multi-bande MBE (Multi Band Excited), les codeurs à excitation mixte MELP (Mixed Excitation Linear Prediction), les codeurs HSX (Harmonic Stochastic Excitation) et les codeurs à interpolation de forme d'onde prototypes WI (Waveform Interpolation)

2.6 Généralités sur le codeur MELP

Le codeur de parole LPC à excitation mixte (MELP) est conçu pour surmonter certaines limitations du LPC10 conventionnel. Il utilise un modèle de production de la parole qui contient des paramètres additionnels pour contenir la dynamique fondamentale du signal avec une précision améliorée. L'idée essentielle est la génération d'un signal d'excitation mélangé comme entrée du filtre de synthèse [44, 45].

L'amélioration apportée par le MELP se situe principalement au niveau de l'étage d'excitation où elle devient "mixte", c'est à dire qu'elle peut prendre plusieurs formes (impulsion, bruit, régime transitoire). Les exigences analogiques recommandées pour le codeur MELP sont pour une bande passante s'étendant de 100 Hz à 3800 Hz, bien que celui-ci opère sur des signaux à bande limitée plus faible.

Le codeur MELP version 1.2 (MELP V1.2) représente la nouvelle norme du DOD à 2.4 Kbits/s. Il remédie donc aux problèmes de la LPC-10 en utilisant une représentation plus souple de l'excitation. Le mélange de la composante impulsionnelle et de la composante de bruit dans l'excitation dépend de la bande de fréquence (entre 0-4000 Hz). La composante impulsionnelle peut être périodique ou apériodique. La dernière joue un rôle dans la

modélisation des consonnes plosives, non-voisées et des zones de transition entre sons.

Tableau 2.1: Table d'allocation des bits du codeur MELP de 2.4 Kbits/s

Paramètres	MELP à 2.4 Kbits/s	
Fréquence d'échantillonnage	8kHz	
Taille de la trame	180 échantillons (22.5 ms)	
Nombre de trame	44,44 trames/seconde	
Mode de voisement	Voisé	Non voisé
10 LSFs	25	25
Pitch	7	7
10 Amplitudes de Fourier	8	0
5 Bandes de voisement	4	0
2 Gains	8	8
Flag	1	0
Protection	0	13
Synchronisation	1	1
Total de bits par trame	54 bits	
Débit total	54×44,44 = 2400 Bits/s	

2.7 Codage de la parole en large bande

À l'ère des réseaux de communication mobiles, de la Voix sur IP (VoIP) et de mains-libres interfaces vocales, de nouvelles possibilités ainsi que de nouveaux défis se posent. Bien que les services vocaux traditionnels soient limités à la gamme de fréquence étroite en dessous de 4 kHz, les nouvelles technologies permettent la transition vers une bande passante supérieure en assurant une qualité supérieure de la parole transmise. D'autre part, les systèmes de communication vocale sont de plus en plus utilisés dans des conditions acoustiques défavorables, comme les environnements bruyant et réverbérant. Par conséquent, de nouveaux systèmes de codage de la parole en large bande et de nouvelles méthodes d'évaluation de la qualité sont apparus [39]

La transmission en large bande correspond à l'élargissement de la bande passante utilisée pour la transmission du signal de parole. En effet, la bande passante habituellement utilisée en téléphonie s'étale de 200 Hz à 3400 Hz, or la bande large permet la transmission de la parole de 50 Hz jusqu'à 7000 Hz. Le codage de la parole en large bande n'est pas un

domaine d'étude très récent. Le premier codeur utilisant la bande large, le G.722, fut développé dans les années 1980 pour être utilisé sur le réseau téléphonique [25]. Mais le codage de la parole en large bande entraîne des techniques différentes de celles utilisées dans le codage en bande étroite. En effet, il y a une plus forte dynamique spectrale pour la parole en bande large. Plusieurs études de développement et de standardisation ont permis de créer quelques codeurs de meilleure qualité et moins coûteux en débit.

2.7.1 Avantages du codage de la parole en large bande

L'utilisation d'une faible bande passante limitée au-dessous de 4 kHz (bande étroite) impose certaines limites de la qualité de la communication, mais l'intelligibilité est suffisante [39]. La transmission de la parole en large bande jusqu'à 7 kHz est utilisée dans un nombre croissant de services. Cet élargissement de la bande passante améliore l'intelligibilité de la parole en donnant un son plus naturel. En effet, la plupart de l'énergie dans les signaux de parole est présente en dessous de 7 kHz, mais le codage en large bande nécessite généralement de nouvelles fonctions qui doivent être prises en compte.

L'évaluation de la qualité auditive de la parole résultante suit les mêmes lignes que pour les signaux à bande étroite, mais les résultats nécessitent une nouvelle interprétation. Il existe un paramètre très important pour l'évaluation de la qualité de parole en fonction de la bande passante utilisée. Il s'agit de **l'indice d'articulation (IA)** qui représente l'information importante de la parole située dans une bande passante ($0 \leq IA \leq 1$) [39]. L'indice de l'articulation est calculé par la formule suivante :

$$IA = -c \log_{10}(1 - s) = -c \log_{10}(e) \quad (2.19)$$

Où e représente la probabilité de l'erreur de compréhension, s représente la probabilité de la compréhension correcte et $c = 0.55$ qui est déterminée expérimentalement. Pour une bande passante de 8 KHz, $IA = 1$.

La figure 2.7 représente les valeurs de l'indice d'articulations IA en fonction de la bande passante. Cette figure montre en toute évidence que la bande en dessous 3,4 KHz ajoute environ 18% et cela en dessous de 300Hz environ 4% de l'information importante. Et l'information située dans la bande en dessous 7 KHz et en dessous de 50 Hz est pratiquement négligeable. Les pertes de l'information due au filtre de la bande étroite sont plus importantes aux sons non voisés par rapport aux sons voisés comme le montre ci bien la figure 2.8. Cela explique pourquoi l'élargissement de la bande est important, mais l'intelligibilité moyenne

n'augmente pas beaucoup. Cependant, l'utilisation de la large bande renforce le naturel dans la parole [39]. Dans un test d'écoute formel de cette fonction, une augmentation de 4 points a été trouvée, de 2 à 6 sur une échelle de 10 points pour un passage de la bande étroite à la large bande.

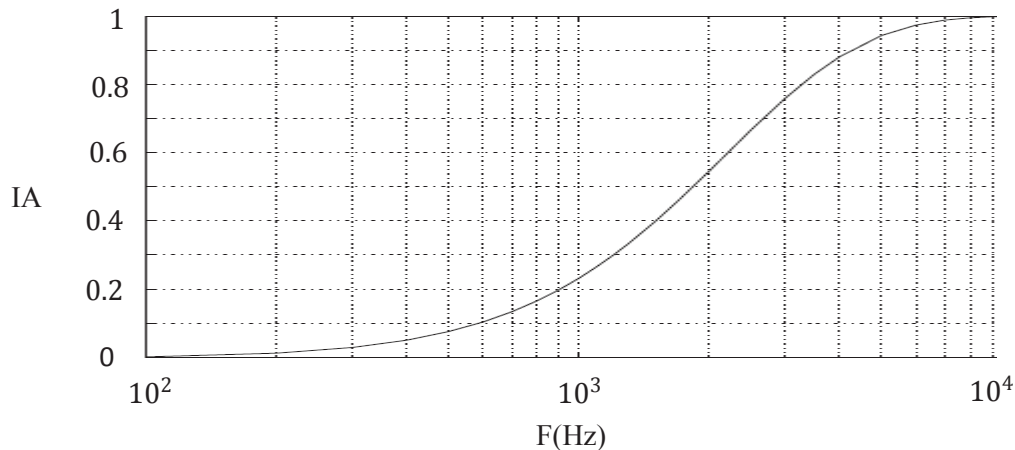


Figure 2.7 Importance de la bande passante à l'indice d'articulation

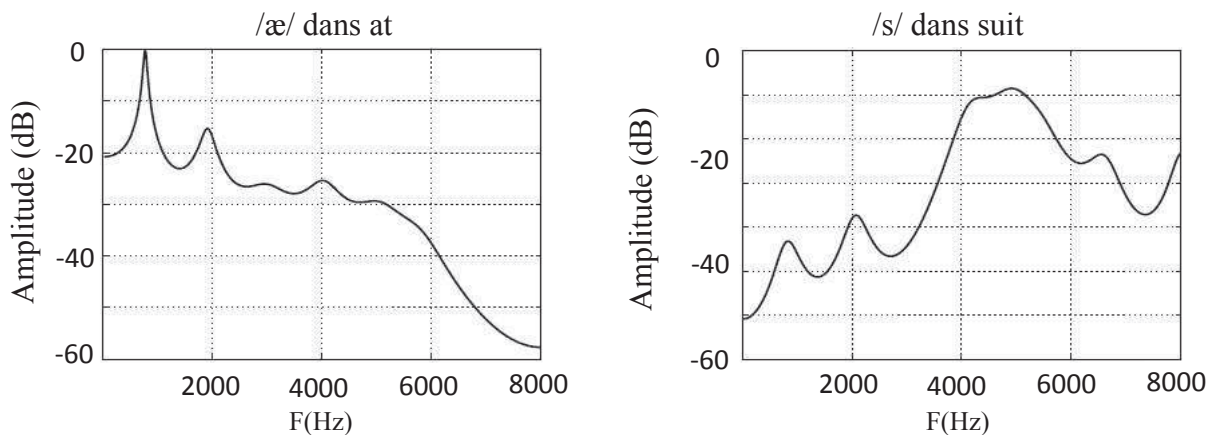


Figure 2.8 : Spectre du signal parole en bandes étroite et large

2.7.2 Applications du codage de la parole en large bande

La qualité et le naturel de la parole améliorés dans la bande large sont des éléments clés qui se dégagent dans certains services actuels de télécommunication, tels que les téléconférences audio et audiovisuelles, l'intrasite LAN VoIP d'entreprise à grande vitesse et aussi dans la future génération des services comme la messagerie multimédia [46]. Ils peuvent également améliorer les applications internet telles que streaming, la radio diffusion, la messagerie instantané et les multimédia en temps réel [25].

2.8 Codeur de parole large bande AMR-WB, norme G.722.2

Le premier codeur de parole large bande à 64 kbit/s, dit G.722, a été normalisé par l'ITU-T en 1986. Ce codeur est basé sur un codage en deux sous-bandes contenant un codeur MICDA (norme G.721) dans chaque sous-bande. Pour pouvoir transmettre simultanément des données sur le même canal, un débit réduit à 56 kbit/s est également possible. Un codeur G.722.1 a été normalisé en 1999 pour des débits de 24 et 32 Kbits/s.

Dans notre travail, nous avons utilisé une normalisation récente ITU-T de 2002. Il s'agit du codeur G.722.2, qui a été au préalable accepté par l'organisme de normalisation 3GPP sous la dénomination codeur large bande à débit Multiple Adaptatif AMR-WB (Adaptive Multi-rate WideBand) [38, 47].

Le codeur AMR-WB utilise la technique ACELP (Algebraic Code Excited Linear Prediction) qui s'appuie sur une modélisation du système de production de la parole. Il est également doté de mécanismes de transmission discontinue DTX permettant d'optimiser la consommation de la ressource radio en ne transmettant pas de signal lors des périodes de non-activité vocale. Pour cela, l'encodeur a un détecteur d'activité vocale VAD (Voice Activity Detection) qui discrimine les instants de parole de ceux de silence ou de bruit. Au décodeur, un générateur de bruit de confort CNG (Comfort Noise Generator) régénère un signal le plus proche possible du bruit original.

Le codeur AMR-WB est destiné à traiter des signaux de parole d'une largeur de bande de 7 kHz. La fréquence d'échantillonnage des signaux de parole d'entrée est donc de 16 kHz. Chaque échantillon est codé sur 16 bits. On obtient donc à l'entrée 256 Kbits/s. Ce codeur fonctionne à une multitude de débits binaires (neuf débits de transmission ou choix 23.85, 23.05, 19.85, 18.25, 15.85, 14.25, 12.65, 8.85 et 6.60 kbit/s). Le débit peut être modifié à toute limite de trame de 20 ms.

Dans chaque trame de 20 ms, 132, 177, 253, 285, 317, 365, 397, 461 ou 477 bits sont produits, correspondant à un débit de 6,60, 8,85, 12,65, 14,25, 15,85, 18,25, 19,85, 23,05 ou 23,85 kbit/s respectivement. Une répartition des bits plus détaillée entre les paramètres de transmission du codeur AMR-WB est présentée dans le tableau 2.2.

Une étape très importante dans notre travail consistait à étudier de près les techniques de quantification et de codage de ces paramètres. L'extraction de ces paramètres est bien détaillée dans [38].

Tableau 2.2 : Allocation des bits pour chaque mode du G722.2

Mode	Débit kbits/s	VAD	ISFs	Filtrage LTP	Délai tonal	Code algébrique	Gain de répertoire	Energie HB	Total bits/trame
0	6.6	1	36	-	23	48	24	-	132
1	8.85	1	46	-	26	80	24	-	177
2	12.65	1	46	4	30	144	28	-	253
3	14.25	1	46	4	30	176	28	-	285
4	15.85	1	46	4	30	208	28	-	317
5	18.25	1	46	4	30	256	28	-	365
6	19.85	1	46	4	30	288	28	-	397
7	23.05	1	46	4	30	352	28	-	461
8	23.85	1	46	4	30	352	28	16	477

Les applications prévues pour la présente recommandation sont les suivantes: voix sur IP (VoIP) et Internet, communications mobiles, Réseau Téléphonique Public Commuté RTPC, téléphonie en large bande sur les Réseaux Numériques à Intégration de Services RNIS, GSM, les communications mobiles de troisième génération, vidéophonie et téléconférence.

2.8.1 Quantification des paramètres ISF

Une prédiction du premier ordre à moyenne mobile (MA) est appliquée et le vecteur de fréquence ISF résiduel est quantifié par une combinaison de quantification vectorielle divisée SVQ et d'une quantification vectorielle en plusieurs phases après division (SMSVQ). La prédiction et la quantification sont effectuées comme suit :

Soit $z(n)$ le vecteur ISF à moyenne supprimée à la trame n . Le vecteur résiduel de prédiction $r(n)$ est donné par:

$$r(n) = z(n) - p(n) \quad (2.20)$$

Où $p(n)$ est le vecteur ISF prédit à la trame n . Une prédiction du premier ordre à moyenne mobile (MA) est utilisée, dans laquelle:

$$p(n) = \frac{1}{3} \hat{r}(n-1) \quad (2.21)$$

Où $\hat{r}(n-1)$ est le vecteur résiduel quantifié à la trame précédente.

Le vecteur ISF résiduel r subit une quantification vectorielle multi étages après subdivision (SMSVQ). Le vecteur est subdivisé en 2 sous-vecteurs $r_1(n)$ et $r_2(n)$ de dimensions 9 et 7, respectivement. Ces 2 sous-vecteurs sont quantifiés en deux étages.

Tableau 2.3: Quantification de vecteur ISF pour les modes entre 8,85 et 23,85 kbit/s

Vecteur ISF non quantifié d'une longueur de 16 éléments				
Etage 1 ($r_1, 0 - 8$) 8 bits			Etage 1 ($r_2, 9 - 16$) 8 bits	
Etage 2 $(r_1^{(2)}, 0 - 2)$ 6 bits	Etage 2 $(r_2^{(2)}, 3 - 5)$ 7 bits	Etage 2 $(r_3^{(2)}, 6 - 8)$ 7 bits	Etage 2 $(r_4^{(2)}, 9 - 11)$ 5 bits	Etage 2 $(r_5^{(2)}, 12 - 16)$ 5 bits

Dans la première phase, le vecteur $r_1(n)$ est quantifié avec 8 bits et le vecteur $r_2(n)$ avec 8 bits. Pour les modes à 8,85, 12,65, 14,25, 15,85, 18,25, 19,85, 23,05 ou 23,85 Kbits/s, les vecteurs d'erreur de quantification $r_i^{(2)} = r - \hat{r}_i$, $i = 1, 2$ sont respectivement subdivisés dans la phase suivante en 3 et 2 sous-vecteurs. Ces sous-vecteurs sont quantifiés au moyen des débits décrits dans le tableau 2.3.

2.8.2 Analyse en boucle ouverte du délai tonal

Selon le mode, l'analyse du délai tonal en boucle ouverte est effectuée une fois par trame (de 20 ms chacune) ou deux fois par trame (de 10 ms chacune) afin de trouver deux estimations du délai tonal dans chaque trame. Cela est effectué afin de simplifier l'analyse tonale et de limiter la recherche en boucle fermée du délai tonal à un petit nombre de délais autour des délais en boucle ouverte estimés.

2.8.3 Dictionnaire algébrique

Le dictionnaire algébrique est un dictionnaire fixe dans lequel un code algébrique sert à construire les vecteurs d'excitation (vecteurs d'innovation). L'excitation contient un petit nombre d'impulsions non nulles avec des ensembles entrelacés de positions possibles. Les amplitudes et les positions des impulsions du vecteur-code d'excitation peuvent être déduites de son indice par l'application d'une règle exigeant une mémorisation physique minimale ou nulle, contrairement aux dictionnaires stochastiques dans lesquels le trajet entre l'indice et le vecteur-code associé implique des tables d'exploration.

2.8.3.1 Structure du dictionnaire

La structure du dictionnaire est fondée sur un modèle de permutation entrelacée d'impulsion isolée (ISPP, interleaved single-pulse permutation). Les 64 positions dans le vecteur-code sont divisées en 4 pistes de positions entrelacées, avec 16 positions dans chaque piste. Les différents dictionnaires aux différents débits sont construits par insertion d'un certain nombre d'impulsions signées dans les pistes (de 1 à 6 impulsions par piste). L'indice de dictionnaire, ou mot-code, représente les positions et les signes d'impulsion dans chaque piste. Ainsi, aucune mémoire de dictionnaire n'est requise, étant donné que le vecteur d'excitation dans le décodeur peut être construit au moyen des informations contenues dans l'indice lui-même (pas de tables de consultation). Soit par exemple la structure de dictionnaire du débit 12,65 Kbits/s présentée dans le tableau 2.4.

Tableau 2.4 : Positions possibles d'impulsions individuelles dans le répertoire algébrique du débit 12,65 kbit/s

Piste	Impulsion	Positions
1	$i_0 i_4$	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
2	$i_1 i_5$	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
3	$i_2 i_6$	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
4	$i_3 i_7$	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

2.8.3.2 Exploration du dictionnaire

Le dictionnaire algébrique est exploré par minimisation de l'erreur quadratique moyenne entre le signal vocal d'entrée pondéré et le signal vocal de synthèse pondéré. Le signal cible utilisé dans la recherche tonale en boucle fermée est mis à jour par soustraction de la contribution du dictionnaire adaptatif. Soit :

$$x_2(n) = x(n) - g_p y(n) \quad n = 0 \dots 63 \quad (2.22)$$

Où $y(n) = v(n) * h(n)$ est le vecteur filtré du dictionnaire adaptatif et g_p est le gain non quantifié du répertoire adaptatif. La matrice H est définie comme étant la partie triangulaire inférieure de la matrice convolutionnelle de Toeplitz avec diagonale $h(0)$ et diagonales inférieures $h(1), \dots, h(63)$, et $d = H^T x_2$ est la corrélation entre le signal cible $x_2(n)$ et la réponse

impulsionnelle $h(n)$ (également appelé "vecteur cible filtre en inverse") et $\Phi = H^t H$ est la matrice de corrélation de $h(n)$. Les éléments du vecteur d sont calculés par :

$$d(n) = \sum_{i=n}^{63} x_2(i)h(i-n) \quad n = 0 \dots 63 \quad (2.23)$$

et les éléments de la matrice symétrique Φ sont calculés par:

$$\Phi(i, j) = \sum_{n=j}^{63} h(i-n)h(j-n) \quad i = 0 \dots 63 \quad j = i \dots 63 \quad (2.24)$$

Si c_k est le vecteur-code algébrique à l'indice k , le dictionnaire algébrique est exploré par maximisation du critère de recherche:

$$Q_k = \frac{(x_2^t H c_k)^2}{c_k^t H^t H c_k} = \frac{(d^t c_k)^2}{c_k^t \Phi c_k} = \frac{(R_k)^2}{E_k} \quad (2.25)$$

Le vecteur d et la matrice Φ sont habituellement calculés avant la recherche dans le dictionnaire algébrique.

La structure algébrique du dictionnaire autorise des procédures de recherche très rapides étant donné que le vecteur d'excitation c_k contient seulement quelques impulsions non nulles. La corrélation dans le numérateur de l'équation (2.25) est donné par:

$$R = \sum_{i=0}^{N_p-1} a_i d(m_i) \quad (2.26)$$

où m_i est la position de la $i^{ème}$ impulsion, a_i est son amplitude et N_p est le nombre d'impulsions. L'énergie dans le dénominateur de l'équation (2.25) est donnée par:

$$E = \sum_{i=0}^{N_p-1} \Phi(m_i, m_i) + 2 \sum_{i=0}^{N_p-2} \sum_{j=i+1}^{N_p-1} a_i a_j \Phi(m_i, m_j) \quad (2.27)$$

Afin de simplifier la procédure de recherche, les amplitudes d'impulsion sont prédéterminées sur la base d'un certain signal de référence $b(n)$. Dans cette méthode d'amplitude d'impulsion sélectionnée par le signal, le signe d'une impulsion à la position i est

rendu égal au signe du signal de référence à cette position. Ici, le signal de référence $b(n)$ est donné par:

$$b(n) = \sqrt{\frac{E_d}{E_r}} r_{LTP} + \alpha d(n) \quad (2.28)$$

Où $E_d = \int d^2 d$ est l'énergie du signal $d(n)$ et $E_r = \mathbf{r}_{LTP}^t \mathbf{r}_{LTP}$ est l'énergie du signal $r_{LTP}(n)$ qui est le signal résiduel après prédiction à long terme. Le facteur de normalisation commande le degré de dépendance du signal de référence de $d(n)$, et il est diminué lorsque le débit est augmenté. Ici, $\alpha = 2$ pour les modes 6,6 et 8,85; $\alpha = 1$ pour les modes 12,65, 14,25, et 15,85; $\alpha = 0,8$ pour 18,25; $\alpha = 0,75$ pour 19,85; et $\alpha = 0,5$ pour 23,05 et 23,85.

Afin de simplifier la recherche, le signal $d(n)$ et la matrice Φ sont modifiés de façon à incorporer les signes présélectionnés. Soit $s_b(n)$ le vecteur contenant le signe de $b(n)$. Le signal modifié $d'(n)$ est donné par :

$$d'(n) = s_b(n)d(n) \quad n = 0, \dots, 63 \quad (2.29)$$

et la matrice d'autocorrélation Φ' est donnée par:

$$\Phi'(i, j) = s_b(i)s_b(j)\Phi(i, j) \quad i = 0, \dots, 63 \quad j = i, \dots, 63 \quad (2.30)$$

La corrélation au numérateur du critère de recherche Q_k est donc donnée par:

$$R = \sum_{i=0}^{N_p-1} d'(m_i) \quad (2.31)$$

et l'énergie au dénominateur du critère de recherche Q_k est donnée par:

$$E = \sum_{i=0}^{N_p-1} \Phi'(m_i, m_i) + 2 \sum_{i=0}^{N_p-2} \sum_{j=i+1}^{N_p-1} \Phi'(m_i, m_j) \quad (2.32)$$

Afin de réduire la complexité de la recherche, une procédure rapide de recherche dite *procédure de recherche arborescente en profondeur* est d'abord utilisée, dans laquelle les positions d'impulsion sont déterminées N_m impulsions à la fois. Plus précisément, les N_p impulsions disponibles sont réparties en M sous-ensembles non vides de N_m impulsions chacun, tels que $N_1 + N_2 + \dots + N_m + \dots + N_M = N_p$. Un choix particulier de positions pour

les $J = N_1 + N_2 + \dots + N_{m-1}$ premières impulsions considérées est appelé *trajet de niveau-m* ou *trajet de longueur J*. Le critère de base pour un trajet de J positions d'impulsion est le rapport $Q_k(J)$ lorsque seules les J impulsions correspondantes sont considérés.

Une forme spéciale de la procédure de recherche arborescente en profondeur est utilisée ici, dans laquelle deux impulsions sont recherchées à la fois, soit, $N_m = 2$. Ces 2 impulsions appartiennent à deux pistes consécutives. Par ailleurs, au lieu de supposer que la matrice Φ est précalculée et mémorisée, ce qui nécessite une mémoire de $N \times N$ mots ($64 \times 64 = 4k$ mots), une méthode utilisant efficacement la mémoire est utilisée afin de réduire la mémoire requise. Dans cette méthode, la procédure de recherche est effectuée de telle façon que seule une partie des éléments requis de la matrice de corrélation est précalculée et mémorisée. Cette partie correspond à la corrélation de la réponse impulsionnelle correspondant à la position potentielle d'impulsions dans des pistes consécutives, ainsi qu'aux corrélations correspondant à (i, j) , $j = 0, \dots, N-1$ (qui sont les éléments de la diagonale principale de matrice Φ).

Les procédures de recherche pour tous les modes de débit sont similaires. Deux impulsions sont recherchées à la fois et ces deux impulsions correspondent toujours à des pistes consécutives. C'est-à-dire que les deux impulsions recherchées sont dans les pistes T_0-T_1 , T_1-T_2 , T_2-T_3 ou T_3-T_0 . Avant de chercher les positions, le signe d'une impulsion à une position possible n est considéré comme étant le signe de $b(n)$ à cette position. Ensuite le signal modifié $d'(n)$ est calculé comme décrit ci-dessus par y insertion des signes prédéterminés.

Pour les deux premières impulsions (1^{er} niveau de l'arborescence), la corrélation au numérateur du critère de recherche est donnée par:

$$R = d'(m_0) + d'(m_1) \quad (2.33)$$

et l'énergie au dénominateur du critère de recherche Q_k est donnée par:

$$E = \Phi'(m_0, m_0) + \Phi'(m_1, m_1) + 2 \Phi'(m_0, m_1) \quad (2.34)$$

où les corrélations $\Phi'(m_i, m_j)$ ont été modifiées de façon à inclure les signes resélectionnés aux positions m_i et m_j . Pour les niveaux subséquents, le numérateur et le dénominateur sont mis à jour par adjonction de la contribution de deux nouvelles impulsions. Dans l'hypothèse

où deux nouvelles impulsions sont recherchées à un certain niveau d'arborescence avec les positions m_k et m_{k+1} de deux pistes consécutives, la valeur mise à jour de R est donnée par:

$$R = R + d'(m_k) + d'(m_{k+1}) \quad (2.35)$$

et l'énergie mise à jour est donnée par:

$$E = E + \Phi'(m_k, m_k) + \Phi'(m_{k+1}, m_{k+1}) + 2\Phi'(m_k, m_{k+1}) + 2R_{hv}(m_k) + 2R_{hv}(m_{k+1}) \quad (2.36)$$

Où $R_{hv}(m)$ est la corrélation entre la réponse impulsionnelle $h(n)$ et un vecteur $v_n(n)$ contenant l'adjonction de versions retardées de réponse impulsionnelle aux positions déjà déterminées. Soit :

$$v_n(n) = \sum_{i=0}^{k-1} h(n - m_i) \quad (2.37)$$

et:

$$R_{hv}(m) = \sum_{n=m}^{N-1} h(n)v_n(n - m) \quad (2.38)$$

La recherche dans le mode 12,65 kbit/s sera décrite à titre d'exemple. Dans ce mode, 2 impulsions sont placées dans chaque piste, ce qui donne un total de 8 impulsions par sous-trame de longueur 64. Deux impulsions sont explorées à la fois, et ces deux impulsions correspondent toujours à des pistes consécutives. C'est-à-dire que les deux impulsions recherchées sont dans les pistes T_0-T_1 , T_1-T_2 , T_2-T_3 , ou T_3-T_0 . L'arbre possède 4 niveaux dans ce cas. Au premier niveau, l'impulsion P_0 est attribuée à la piste T_0 et l'impulsion P_1 à la piste T_1 . A ce niveau, aucune recherche n'est effectuée et les deux positions d'impulsion sont mises au maximum de $b(n)$ dans chaque piste.

Au deuxième niveau, l'impulsion P_2 est attribuée à la piste T_2 et l'impulsion P_3 à la piste T_3 . Quatre positions pour l'impulsion P_2 sont essayées en fonction des 16 positions de l'impulsion P_3 . Les 4 positions essayées de P_2 sont déterminées sur la base des maxima de $b(n)$ dans les pistes. Au troisième niveau, l'impulsion P_4 est attribuée à la piste T_1 et l'impulsion P_5 à la piste T_2 . Huit positions pour l'impulsion P_4 sont essayées en fonction des 16 positions de l'impulsion P_5 . Comme au niveau de recherche précédent, les 8 positions essayées de P_4 sont déterminées sur la base des maxima de $b(n)$ dans la piste. Au quatrième niveau, l'impulsion

Algorithme 2.1: Procédure de recherche des impulsions en mode 12.65 Kbits/s**Données :**

$i = 0, k = 0, nb_itérations = 4, Q_p = 0, nb_impulsions = 8,$
 $ordre = \{0,1,2,3,1,2,3,0\}, nb = \{4,8,8\}, vec.$

Etape 1:

- Calculer les vecteurs d', Φ' et b .
- Trouver le vecteur **Pos_max** qui contient les positions correspondantes au maximum du vecteur b dans chaque piste.

Etape 2:

- $k=0$.
- Sélectionner les deux positions correspondantes au maximum de b dans les pistes $ordre[0]$ et $ordre[1]$.

$$m_0 = Pos_max[ordre[0]] \quad m_1 = Pos_max[ordre[1]]$$

- Calculer l'énergie E et la corrélation R .
- Incrémenter k par 2.

Etape 3:

- Chercher les deux meilleurs impulsions m_k et m_{k+1} en optimisant les positions correspondantes au $nb[i]$ plus grandes valeurs de b dans la piste $ordre[k]$ et les 16 position de la piste $ordre[k+1]$.
- Mise à jour l'énergie E et la corrélation R .
- Incrémenter k par 2.

Si ($k < nb_impulsions$) : réitérer l'**Etape 3**.

Etape 4:

- Calculer $Q = R^2/E$
- **Si** ($Q > Q_p$) :
 - $Q_p = Q$.
 - Sauvegarder les impulsions $\{m_0 \dots m_7\}$ dans le vecteur vec .

FinSi

- Incrémenter i par 1.
- Décalage à droite cyclique des deux sous-vecteurs $\{ordre[0], \dots, ordre[3]\}$ et $\{ordre[4], \dots, ordre[7]\}$.
- **Si** ($i < nb_itérations$) : réitérer l'**Etape 2**, l'**Etape 3**, et l'**Etape 4**.

Etape 5:

- Créer l'indice du vecteur d'excitation à partir du vecteur vec .

P_6 est attribuée à la piste T_3 et l'impulsion P_7 à la piste T_0 . Huit positions pour l'impulsion P_6 sont essayées en fonction des 16 positions de l'impulsion P_7 . Ainsi le nombre total de combinaisons essayées est $4 \times 16 + 8 \times 16 + 8 \times 16 = 320$. L'ensemble du processus est répété 4 fois (4 itérations) par attribution des impulsions à différentes pistes. Par exemple, dans la 2^{ème} itération, les impulsions P_0 à P_7 sont attribuées aux pistes $T_1, T_2, T_3, T_0, T_2, T_3, T_0$, et T_1 , respectivement. Ainsi le nombre total de combinaisons de positions essayées est de $4 \times 320 = 1280$. Cette procédure de recherche est résumée dans l'algorithme 2.1.

2.9 Conclusion

L'étude des codeurs et des différentes techniques de codage, qui a été présentée dans ce chapitre, a permis de dresser une solution aux problèmes liées au choix du codeur qui sera utilisé dans notre travail pour la dissimulation de données et celui du codage des messages à dissimuler.

Le codeur MELP conçu par le DOD donne une bonne qualité de la parole à très bas débit (2.4 kbits/s). Ces diverses propriétés font de lui un très bon candidat pour être utilisé comme codeur de la parole secrète. D'autre part, le codeur AMR-WB G.722.2 utilise plusieurs techniques de codage comme la quantification vectorielle en plusieurs phases après subdivision (SMSVQ) et le codage ACELP. Les caractéristiques de ce codeur offre plusieurs possibilité de la méthode de la dissimulation d'un message secret.

Dans le chapitre suivant, on s'intéressera à la mise en œuvre de méthodes de dissimulation applicables sur le codeur AMR-WB G.722.2 en utilisant le codeur MELP pour coder la parole secrète à cacher.

Chapitre 3

Techniques de stéganographie appliquées à la parole

3.1 Introduction

Plusieurs techniques de stéganographie peuvent être utilisées pour concevoir des systèmes de transmission de la parole sécurisés. Ces techniques se différencient selon leurs propriétés, la qualité de la parole publique reconstituée et les paramètres de transmission utilisés pour dissimuler les messages secrets.

Dans ce chapitre, nous décrivons en détails les techniques de stéganographie que nous avons utilisées dans la simulation de nos systèmes de transmission sécurisés. Au début, nous nous sommes intéressés à une technique appartenant à la catégorie des schémas de dissimulation

appliqués aux systèmes de quantification vectorielle (QV). Il s'agit de la dissimulation par modulation d'indice de quantification QIM (Quantization Index Modulation) ainsi que ces versions dérivées. Nous avons opté pour cette technique de dissimulation vu son efficacité et sa possibilité d'implémentation sur des codeurs de parole comme le codeur AMR-WB G.722.2 qui utilise la QV dans le codage de plusieurs paramètres de transmission (les ISFs et l'excitation du filtre de synthèse LP, ... etc).

Dans nos implémentations de systèmes de codage de la parole sécurisés, nous nous sommes intéressés à deux versions de la technique QIM: la QIM basée sur la division de dictionnaires QV (QIM-DD) et la QIM basée sur la modification du schéma du quantificateur vectoriel (QIM-MS-VQ). Une première implémentation consistait à appliquer les deux versions de la technique QIM (QIM-DD et QIM-MS-VQ) au niveau du quantificateur S-MSVQ des paramètres ISF du G.722.2. Par la suite, nous avons appliqué la QIM à division de dictionnaires au niveau de la quantification de l'excitation du filtre de synthèse (dictionnaire algébrique) du AMR-WB G.722.2. D'autres parts, nous avons implémenté une autre technique de dissimulation au niveau du bloc quantification du délai tonal (pitch) du codeur G.722.2. Il s'agit la technique de dissimulation LSB adaptative appliquée sur la valeur du délai tonal. Les procédures de dissimulation et d'extraction de ces techniques de stéganographie implémentées seront décrites en détails dans ce chapitre et les résultats de simulation seront présentés dans le chapitre suivant.

3.2 Schémas de dissimulation dans la quantification vectorielle

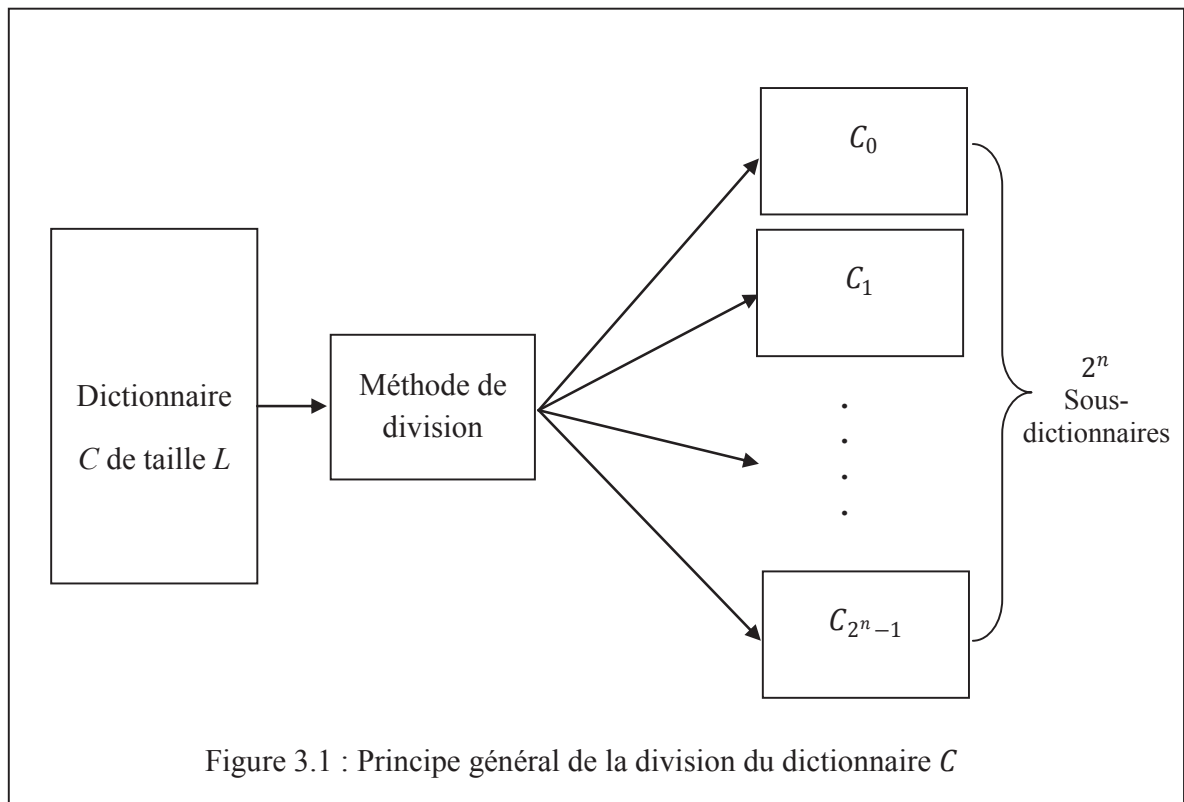
Ces dernières années, plusieurs chercheurs ont proposé divers schémas de dissimulation de données sur des systèmes à QV [6, 22, 48, 49]. Dans ces techniques proposées, le principe de base consiste à dissimuler les données en utilisant l'indice de quantification. Une catégorie importante de ces techniques est la dissimulation par modulation d'indice de quantification QIM. En effet, la QIM est largement utilisée dans la dissimulation des messages secrets dans les codeurs de parole [16, 17, 25, 48, 50], qui utilisent la QV.

En général, il existe deux techniques QIM. La QIM basée sur la division de dictionnaires QV [16, 22] et la QIM basée sur la modification des schémas de quantificateurs vectoriels [48, 50]. Comme nous l'avons déjà mentionné dans l'introduction, ces deux techniques QIM seront

implémentées dans le codeur de parole AMR-WB G.722.2 au niveau de la quantification des paramètres ISF et de l'excitation du filtre de synthèse LP.

3.2.1 Principe de la QIM basée sur la division de dictionnaire QV (QIM-DD)

Le principe de base de cette technique se résume comme suit : soit le dictionnaire de QV C contenant L vecteurs-code. Pour dissimuler un message de n bits, on doit diviser C en 2^n sous-dictionnaires (figure 3.1).



Par exemple, pour dissimuler un message de 1 bit, on divise donc le dictionnaire C en 2 sous-dictionnaires (C_0 , C_1). Par la suite, selon la valeur des données secrètes, l'encodeur peut choisir le meilleur vecteur-code à partir de l'un des deux sous-dictionnaires pour coder le vecteur courant. C'est-à-dire, le vecteur d'entrée est codé en utilisant le plus proche vecteur-code de C_0 ou de C_1 selon que le bit secret soit 0 ou 1 respectivement.

Les schémas proposés peuvent être décomposés en trois phases [51] : prétraitement, dissimulation du message et extraction du message. Pour décrire ces phases, nous donnerons un

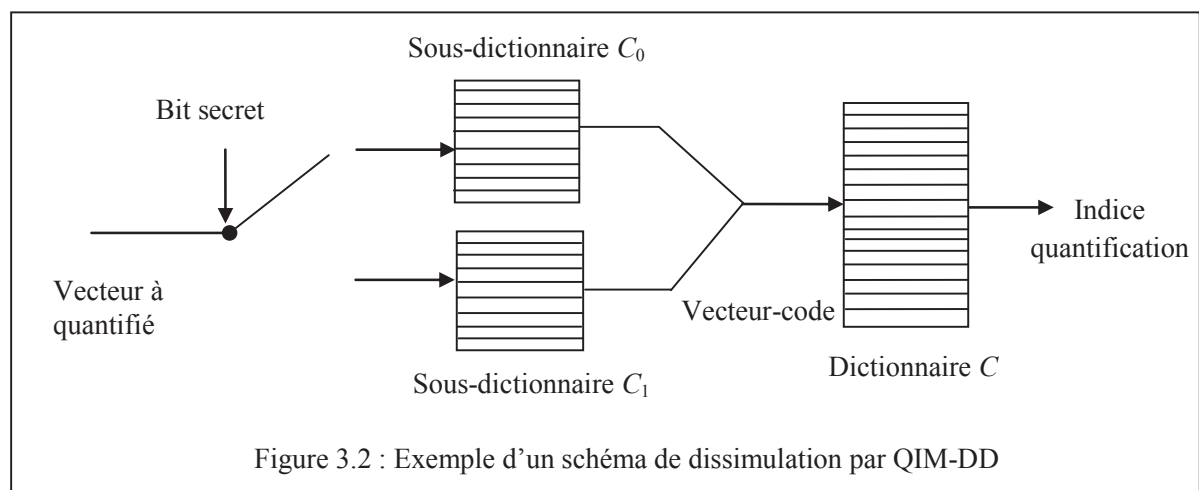
exemple de dissimulation d'un seul bit qui représente un schéma de base très simple. Cependant, il existe des schémas plus pratiques comme celui utilisé dans le paragraphe 3.

3.2.1.1 Phase de prétraitement

La phase de prétraitement est la phase la plus importante dans notre travail car cette phase va influencer directement sur les performances de nos schémas de dissimulation. La phase de prétraitement consiste à diviser un dictionnaire de QV C de taille L , en 2^n sous dictionnaires ($C_0, C_1, \dots, C_i, \dots, C_{2^n-1}$) où n est le nombre de bits à dissimuler dans chaque indice. Chaque sous-dictionnaire C_i représente un message i . La division doit respecter un critère d'évaluation pour assurer une dégradation minimale [22]. Dans notre travail, nous divisons le dictionnaire C en deux sous-dictionnaires (C_0, C_1) afin de dissimuler un seul bit (0 ou 1). Nous montrerons plus loin comment cette division sera réalisée.

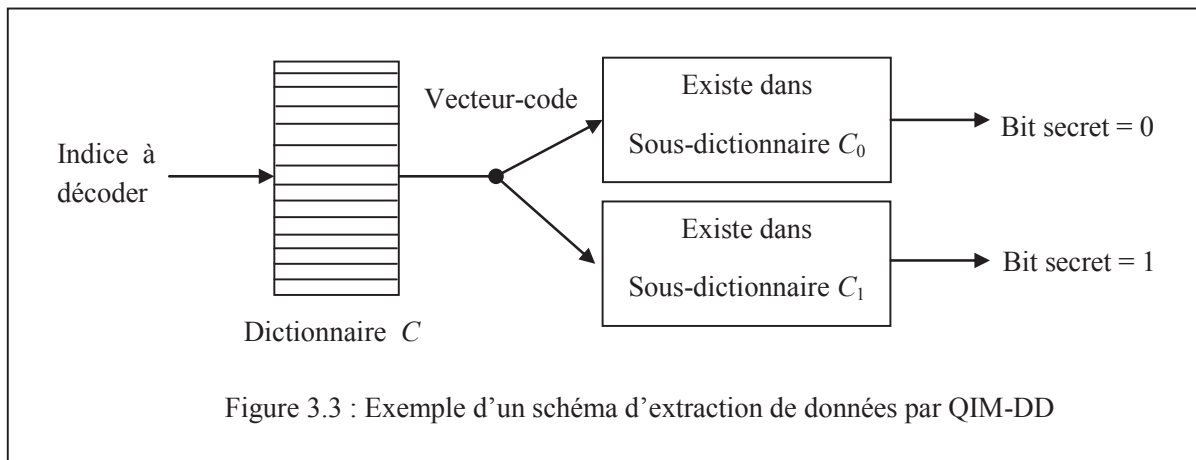
3.2.1.2 Phase de la dissimulation

La dissimulation est effectuée durant la procédure de quantification. La figure 3.2 présente un schéma de dissimulation de messages d'un seul bit par la technique QIM. Selon les données secrètes, l'encodeur peut choisir l'un des deux sous-dictionnaires pour quantifier le vecteur d'entrée. C'est-à-dire que le quantificateur utilisera le sous-dictionnaire C_0 si le bit à dissimuler est 0 ou le sous-dictionnaire C_1 si le bit à dissimuler est 1. Une fois le vecteur-code trouvé, l'encodeur cherche son indice dans le dictionnaire C et l'envoi au récepteur.



3.2.1.3 Phase de l'extraction

L'extraction de message sera effectuée en deux étapes comme nous le montre la figure 3.3. La première étape est le décodage de l'indice reçu par la méthode conventionnelle qui donne le vecteur-code décodé. La deuxième étape est la recherche de ce vecteur-code dans les sous-dictionnaires C_0 et C_1 . Si ce vecteur existe dans le sous-dictionnaire C_0 alors le bit secret est égal à 0, sinon le vecteur-code existe forcément dans C_1 et le bit secret est égal à 1.



Cette technique exploite les propriétés et la structure de chaque schéma de quantificateurs vectoriels (VQs) structurés comme le MSVQ, le SVQ, le S-MVQ,...etc. Pour chaque schéma, cette technique utilise une idée de dissimulation différente tout en respectant toujours la règle de base : la dégradation causée par la dissimulation doit être la plus minimale possible.

Nous utiliserons cette technique de dissimulation dans le quantificateur S-MSVQ des ISFs de la norme G.722.2.

3.3 Dissimulation dans les indices de quantification des paramètres ISF

Les paramètres ISF jouent un rôle très important dans les codeurs de parole en large bande comme le AMR-WB G.722.2. Dans le standard G.722.2, ces paramètres sont codés par un S-MSVQ de 46 bits/trame à des modes débits supérieurs à 6.6 kbps. A un débit de 12.65 Kbits/s, ces 46 bit représentent 18,1% de l'information codée, c.-à-d, 2.3 Kbits/s [38]. Donc, on peut exploiter une partie de ce débit pour dissimuler un message.

Comme nous l'avons vu dans le chapitre précédent, les paramètres ISF du G.722.2 sont codés à l'origine par un quantificateur S-MSVQ qui utilise 7 dictionnaires (deux dictionnaires dic11 et dic12 du premier étage de la quantification et 5 dictionnaires dic21, dic22, dic23, dic24, dic25 du deuxième étage) [38]. La dissimulation dans ce cas peut être effectuée soit par la technique QIM à base de division de dictionnaire ou par QIM basée sur la modification de l'architecture du S-MSVQ.

3.3.1 QIM à base de division du dictionnaire de quantification des ISFs

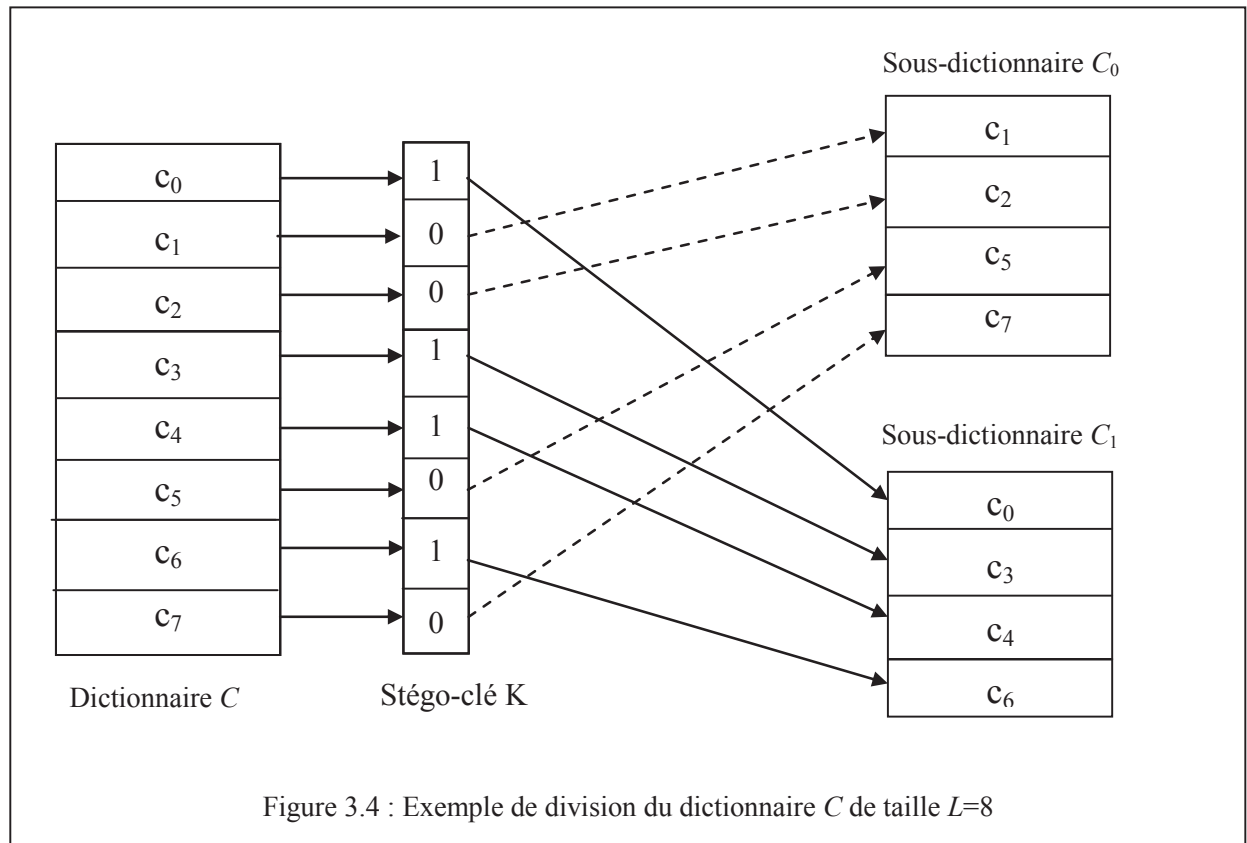
Dans cette partie, nous allons détailler la méthode de dissimulation par QIM basée sur le principe de division du dictionnaire de quantification des paramètres ISF (QIM_DD) que nous dénotons par dictionnaire C . Par analogie, cette méthode pourra être appliquée à tous les dictionnaires ISFs selon la même procédure utilisée pour le dictionnaire C qui représente un des dictionnaires de la quantification des ISFs (dic11, dic12, ..., etc.). Cette méthode a été inspirée des travaux de Wang et al. [31, 51].

3.3.1.1 Division du dictionnaire

Le dictionnaire C contenant L vecteurs-code sera donc divisé en deux sous-dictionnaires C_0 et C_1 suivant une certaine clé (stégo-clé) $K = \{k_1, k_2, \dots, k_L\}$ tel que $k_i \in \{1, 0\}$ avec $1 \leq i \leq L$. Chaque vecteur-code c_i ($1 \leq i \leq L$) du dictionnaire C est affecté au sous-dictionnaire C_{k_i} . C'est-à-dire, si la clé $k_i = 0$ donc $c_i \in C_0$ sinon $c_i \in C_1$ ($k_i = 1$). Cette division est effectuée d'une façon implicite où certains vecteurs-code de C seront affectés à C_0 et les autres à C_1 et la stégo-clé remplace les deux sous-dictionnaires C_0 et C_1 . Nous donnons un exemple de division du dictionnaire C ayant une taille $L = 8$ dans la figure 3.4. Nous présenterons dans le paragraphe 3.3.1.4 la méthode qui permettra de sélectionner les vecteurs-code de C qui seront affectés soit à C_0 ou C_1 . Les relations entre le dictionnaire initial C et les deux sous-dictionnaires C_0 et C_1 sont :

$$\begin{cases} C = C_0 \cup C_1 \\ C_0 \cap C_1 = \emptyset \end{cases} \quad (3.1)$$

Nous utilisons les sous-dictionnaires C_0 et C_1 pour dissimuler un bit "0" et un bit "1" respectivement.



La division du dictionnaire a une complexité très grande pour les dictionnaires ayant une taille L grande. Par exemple pour $L = 64$, le nombre de stégo-clés possibles est calculé par une combinaison C_{64}^{32} égale à $1,83 \times 10^{18}$. Donc, il est pratiquement impossible de tester toutes les stégo-clés. Pour remédier à ce problème, nous avons utilisés trois méthodes de divisions qui sont détaillées dans le paragraphe 3.3.1.4.

3.3.1.2 Procédure de la dissimulation

La procédure de dissimulation dans ce cas est basée sur l'utilisation des deux sous-dictionnaires trouvés dans l'étape de division. Un de ces deux sous dictionnaires remplacera le dictionnaire initial C . Le choix de sous-dictionnaire à utiliser se fait en fonction du message à dissimuler.

Les étapes permettant à un système conventionnel VQ d'avoir la capacité de dissimuler un message binaire secret sont résumées dans l'algorithme 3.1.

Algorithme 3.1: Procédure de la dissimulation par QIM basée sur la division de dictionnaire

Données : - Dictionnaire C contenant L vecteurs-code

- Stégo-clé $K = \{k_1, k_2, \dots, k_L\}$ telle que $k_i \in \{1,0\}$, $1 \leq i \leq L$

Etape 1:

- Soit le vecteur à quantifier x et le bit m à dissimuler

Etape 2 :

- Trouver l'indice i du vecteur-code y le plus proche de x dans le dictionnaire C

avec la condition : $k_i = m$

(Cette condition est l'équivalent d'utiliser le sous-dictionnaire C_0 Si $m = 0$ ou C_1 si $m = 1$)

Etape 3:

- Envoyer au récepteur l'indice i du vecteur-code y .

Algorithme 3.2 : Procédure de l'extraction

Données :

- Dictionnaire VQ initial C contenant L vecteurs-code.

- Stégo-clé: $K = \{k_1, k_2, \dots, k_L\}$ telle que $k_i \in \{0, 1\}$ avec $1 \leq i \leq L$.

- Indices de quantification reçus des vecteurs-code porteurs: $I = \{i_1, i_2, \dots\}$

Etape 1:

- Pour le $j^{\text{ème}}$ indice de quantification i_j , trouver la valeur k_{i_j} dans la stégo-clé K .

- Déterminer le $j^{\text{ème}}$ bit caché du message binaire m : $m_j = k_{i_j}$

Etape 2:

- Répéter l'étape précédente jusqu'à l'extraction de tous les bits cachés (message m).

3.3.1.3 Procédure de l'extraction

La procédure de l'extraction est très simple. Elle nécessite la connaissance de la stégo-clé utilisée dans l'étape de dissimulation et les indices de quantification reçus. Pour extraire le bit dissimulé dans un indice porteur reçus i , il suffit de connaître la valeur de k_i dans la stégo-clé K . Les étapes de l'algorithme 3.2 sont alors appliquées.

3.3.1.4 Méthodes de la division du dictionnaire

Comme nous l'avons indiqué dans la section 3.2.1.1, la méthode utilisée pour diviser les dictionnaires est très importante dans cette technique. L'utilisation d'une mauvaise division peut provoquer une grande dégradation pendant la procédure de dissimulation. Afin d'assurer que la division de dictionnaire a un effet minimale sur la qualité, il a été constaté que la distorsion totale causée par la procédure de dissimulation doit être minimale. Cette distorsion totale, dénotée par D_T , est calculée par [51]:

$$D_T = \sum_{i=1}^L ED(c_i, C_0) + ED(c_i, C_1) \quad (3.2)$$

où,

ED : est la fonction qui donne la distance euclidienne minimale entre un vecteur et les éléments d'un dictionnaire.

c_i : est le $i^{\text{ème}}$ vecteur dans le dictionnaire C .

C_0 et C_1 : sont les deux sous-dictionnaires de C .

L : est nombre des vecteur-code du dictionnaire C .

Le calcul de cette distorsion globale n'est pas un problème simple. Il existe plusieurs techniques de division de dictionnaire qui utilisent les clés secrètes et les algorithmes génétiques.

a. Division par un algorithme génétique

Plusieurs schémas de division génétique de dictionnaire (DGD) ont été proposés dans la littérature [51, 52]. Ils emploient les algorithmes génétiques (AG) pour trouver la meilleure façon de diviser le dictionnaire de sorte que l'imperceptibilité du système stéganographique soit

améliorée. Dans ce qui suit, nous présentons les étapes de base de l'algorithme génétique que nous avons utilisé pour réaliser la division de dictionnaire de notre système stéganographique.

Algorithme 3.3 : Procédure de la division de dictionnaire par un algorithme génétique

Données :

- Nombre d'itération : $nb_itérations$.
- Compteur : $i=0$.

Etape 1:

- Générer aléatoirement S stégo-clés $\{K_1, K_2, \dots, K_S\}$ comme individus initiaux de l'AG.

Etape 2 :

- Dissimuler le message à cacher suivant les S stégo-clés $\{K_1, K_2, \dots, K_S\}$ respectivement en utilisant la procédure de dissimulation (section 3.3.1.2, algorithme 3.1).

Etape 3:

- Évaluer la performance de chaque individu (stégo-clé K_i ; $i = \{1 \dots S\}$) en calculant la distorsion partielle causée par chaque clé dans la distorsion totale D_T (équa. 3.2).

Etape 4:

- Sélectionner et sauvegarder la meilleure stégo-clé (l'individu) parmi les S stégo-clés ayant donné la meilleure distorsion partielle minimale.
- Sélectionner $S/2$ les meilleures stégo-clés comme parents, pour S nouveaux stégo-clés $\{K_1, K_2, \dots, K_S\}$ et régénérer les S stégo-clés.
- incrémenter le compteur i .
- **Si** ($i < nb_itérations$) : réitérer les étapes 2, 3 et 4.
- **Sinon** mettre fin à la procédure.

Dans la première étape, l'objectif de la procédure de DGD est de trouver la meilleure façon de diviser le dictionnaire utilisé. La stégo-clé K est considéré donc comme un individu de l'AG. On génère aléatoirement S stégo-clés comme les individus initiaux de l'AG. Par exemple, si $S = 3$ et $L = 6$, alors les S stégo-clés telles que $K_1 = \{0, 1, 1, 0, 1, 0\}$, $K_2 = \{0, 1, 0, 1, 0, 1\}$ et $K_3 = \{1, 1, 0, 0, 1, 0\}$ peuvent être utilisées.

Cette méthode assure une dégradation globale minimale mais elle n'assure pas une dégradation minimale pour chaque vecteur-code du dictionnaire C . Pour remédier à cet inconvénient, nous avons proposé une autre méthode de division qui assure une dégradation globale minimale tout en assurant aussi une dégradation minimale pour chaque vecteur-code du dictionnaire C . Contrairement à la méthode DGD qui engendre une division équilibrée du dictionnaire, cette méthode divise le dictionnaire C en deux sous-dictionnaires qui n'ont pas le même nombre des vecteurs-code. Nous avons appelé cette méthode par "division non équilibrée de dictionnaire".

b. Division non équilibrée du dictionnaire

Cette technique est basée sur l'idée de trouver une dégradation minimale pour chaque indice du dictionnaire C . Pour chaque indice i ($1 \leq i \leq L$), nous devons trouver l'indice j ($1 \leq j \leq L, i \neq j$) telle que la distance entre la paire des vecteur-codes c_i et c_j du dictionnaire C soit minimale. Les deux indices i et j seront donc attribués à deux groupes différents. Ces deux groupes vont permettre de générer la stégo-clé K qui sera utilisée dans les procédures de dissimulation et de l'extraction. Les étapes de cette méthode sont dans l'algorithme 3.4.

c. Division aléatoire

Dans cette méthode, nous divisons d'une façon aléatoire le dictionnaire C pour étudier soigneusement l'influence de la méthode de division sur la qualité de la quantification et de la parole publique en comparant les résultats de cette division aléatoire avec les résultats des deux autres méthodes (division équilibrée et non équilibrée). Un exemple simple de division aléatoire consiste à mettre les $L/2$ premiers vecteurs-code de C dans le sous-dictionnaire C_0 et le reste dans C_1 .

Algorithme 3.4 : Procédure de la division non équilibrée de dictionnaire**Données :**

- Dictionnaire VQ initial C contenant L vecteurs-code.

Etape 1:

- Trouver toutes les L paires d'indices (i, j) : $1 \leq i, j \leq L$, $i \neq j$) telles que les distances entre les paires de vecteurs-code (c_i, c_j) soient minimales.

Etape 2:

- Pour chaque paire d'indice (i, j) , mettre l'indice i dans un groupe (0) et j dans l'autre groupe (1), de sorte que chaque indice doit appartenir seulement à un seul groupe.
- Etiqueter chaque indice par le numéro de son groupe (0 ou 1).

Etape 3:

- Générer la stégo-clé $K = \{k_1, k_2, \dots, k_L\}$ en utilisant les étiquettes des L indices du dictionnaire C :

$$k_i = \text{étiquette}(i), \text{ pour } i = 1, \dots, L$$

3.3.1.5 Débit de la dissimulation

Le débit (ou capacité) de dissimulation R a une relation directe avec le nombre des indices utilisés et la longueur de la trame d'analyse du codeur de parole. Chaque indice de quantification est calculé à chaque 20 ms (longueur de trame T_l). Donc, si on dissimule un bit dans l'indice de quantification alors on dissimule un bit chaque 20 ms. Le débit R_1 de l'utilisation d'un seul indice est donné par :

$$R_1 = \frac{1}{T_l} = \frac{1}{20 \times 10^{-3}} = 50 \text{ bits/s} \quad (3.3)$$

La relation entre le débit de dissimulation total R_d et le débit R_1 est :

$$R_d = n \times R_1 \quad (3.4)$$

Où n est le nombre d'indices utilisés dans la dissimulation. Donc si on utilise les sept indices de la quantification des paramètres ISFs du G.722.2, le débit sera égal à 350 bps, qui représente le débit de dissimulation maximum par cette méthode.

3.3.2 QIM basée sur la modification des schémas VQs des ISFs

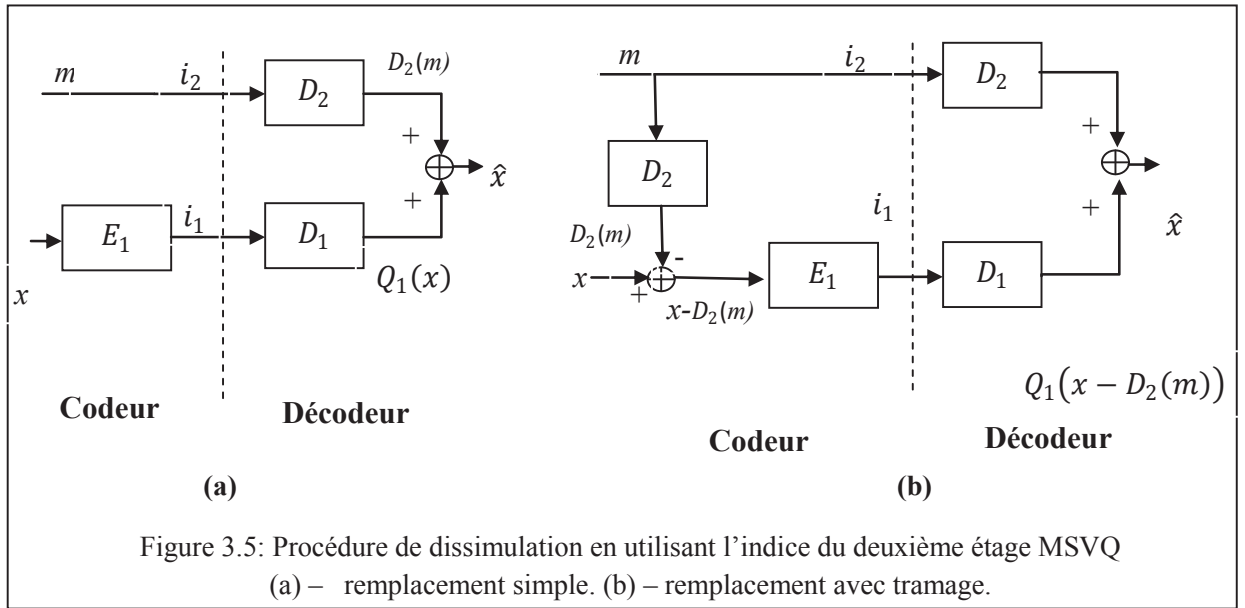
Plusieurs travaux utilisent la modification de l'architecture des quantificateurs pour la dissimulation des bits de messages secrets. Nous nous sommes intéressés particulièrement aux techniques de dissimulation appliquées aux quantificateurs MSVQ [48, 49, 50] puisque les paramètres ISF du G.722.2 sont quantifiés par un quantificateur de type S-MSVQ.

Dans un MSVQ, les signaux dans les derniers étages ont tendance à être moins corrélées. Par conséquent, si l'indice de quantification du dernier étage est remplacé par un message binaire à dissimuler, le dernier étage peut être considéré comme un bruit aléatoire qui génère des données non corrélées avec les étages précédents. Il a été démontré dans [48], qu'en soustrayant ce bruit aléatoire de l'entrée du codeur MSVQ et en l'additionnant plus tard au niveau du décodeur MSVQ, ceci permettra de réduire la dégradation causée par la dissimulation des messages secrets comparée à la méthode traditionnelle de remplacements simples. Cette technique est applicable sur un quantificateur S-MSVQ en exploitant les propriétés du MSVQ.

3.3.2.1 Procédure de la dissimulation

Dans un MSVQ à deux étages (deux paires d'encodeurs et de décodeurs E_1 et D_1 du premier étage Q_1 , E_2 et D_2 du deuxième étage Q_2), le vecteur d'entrée x est quantifié par l'encodeur VQ du premier étage E_1 qui génère le premier indice i_1 . L'approximation quantifiée $D_1(i_1)$ est ensuite soustraite de x . Le vecteur résultant sera appliqué au deuxième encodeur E_2 qui génère le deuxième indice i_2 . Comme nous l'avons vu dans le chapitre 2, le vecteur total quantifié $\hat{x}$ du vecteur d'entrée x est formé en sommant les vecteurs sorties du premier et du deuxième étage ($\hat{x} = D_1(i_1) + D_2(i_2)$).

La figure 3.1- (a) montre la méthode de dissimulation classique de message par remplacement appliquée à un MSVQ de deux étages où l'indice du deuxième étage est utilisé pour dissimuler un message secret m . La partie (b) de la figure 3.1 montre le principe de la méthode de dissimulation avec tramage (Dithering). Le tramage est une technique d'amélioration de données numériques basée sur l'application volontaire d'un bruit. Cette technique est souvent utilisée pour améliorer les erreurs de quantification dans un système numérique [53]. Dans notre travail, le message à dissimuler, considéré comme le bruit de tramage, est soustrait du premier étage de l'encodeur MSVQ puis ajouté par la suite au niveau du décodeur. Dans ce cas, l'encodeur MSVQ génère les deux indices i_1 et i_2 , où ce dernier est égal exactement au message dissimulé. Cette méthode est appelée dissimulation par remplacement avec tramage (DRT). Le décodage se fait simplement par la méthode MSVQ standard.



En basant sur le critère d'erreur quadratique moyenne (EQM), on calcule l'erreur du système avec remplacement simple et avec remplacement avec tramage (soustraction du message à cacher comme un bruit):

Pour le système avec remplacement simple, l'erreur en termes d'EQM1 est :

$$EQM_1 = E(|x - \hat{x}|^2) = E(|x - Q_1(x) - D_2(m)|^2) \quad (3.5)$$

Pour le système avec remplacement avec tramage, l'EQM2 est :

$$EQM_2 = E(|x - \hat{x}|^2) = E(|x - Q_1(x - D_2(m)) - D_2(m)|^2) \quad (3.6)$$

Supposant que Q_1 est un quantificateur à haute résolution et que ε_1 et ε_2 représentent les erreurs de quantification du premier et deuxième étages, respectivement :

$$\varepsilon_1 = x - Q_1(x) \quad (3.7)$$

$$\varepsilon_2 = (x - D_2(m)) - Q_1(x - D_2(m)) \quad (3.8)$$

Où ε_1 et $D_2(m)$ sont i.i.d (indépendants, identiquement distribués) [48].

En remplaçant les équations (3.7) et (3.8) dans les équations (3.5) et (3.6) respectivement, on trouve :

$$\begin{aligned} EQM_1 &= E(|x - Q_1(x) - D_2(m)|^2) \\ &= E(|\varepsilon_1 - D_2(m)|^2) \\ &= E(|\varepsilon_1|^2) - 2E(|\varepsilon_1 D_2(m)|) + E(|D_2(m)|^2) \\ &= E(|\varepsilon_1|^2) + E(|D_2(m)|^2) \end{aligned} \quad (3.9)$$

et

$$\begin{aligned} EQM_2 &= E(|(x - D_2(m)) - Q_1(x - D_2(m))|^2) \\ &= E(|\varepsilon_2|^2) = E(|\varepsilon_1|^2) \end{aligned} \quad (3.10)$$

On remarque bien que $EQM_1 > EQM_2$. Ainsi, la distorsion causée par la procédure de dissimulation par remplacement avec tramage (DRT) est plus petite que la distorsion causée par la méthode de dissimulation par remplacement simple classique (DRS).

Notons que dans une situation réelle, cette conclusion n'est pas toujours vraie si x , ε_1 et $D_2(m)$ ne sont pas i.i.d et que Q_1 n'est pas un quantificateur à haute résolution [48].

3.3.2.2 Débit de dissimulation

Dans cette technique, le message est dissimulé dans l'indice du deuxième étage de la quantification. Le débit de dissimulation est donc le même débit du codage de cet indice. Le nombre de bits du code de chaque indice est présenté dans le chapitre précédent (cf. Tableau 2.3).

Rappelons que les paramètres ISF du G.722.2 sont codés par un quantificateur S-MSVQ qui utilise 7 dictionnaires (deux dictionnaires dic11 et dic12 au premier étage de la quantification et 5 dictionnaires dic21, dic22, dic23, dic24, dic25 au deuxième étage). Donc on peut utiliser les dictionnaires du deuxième étage pour dissimuler le message. Le débit de dissimulation est le résultat de division de ce nombre par la durée de la trame (20 ms). Le tableau 3.1 donne les débits de dissimulation (en bits/s) lors de l'utilisation de chaque indice.

Tableau 3.1: Débits de dissimulation dans chaque indice du deuxième étage S-MSVQ

Dictionnaire	Nombre de bits par trame	Débit de dissimulation (bits/s)
dic21	6 bits	300
dic22	7 bits	350
dic23	7 bits	350
dic24	5 bits	250
dic25	5 bits	250

Il faut bien noter que ce tableau présente les débits de la dissimulation dans les indices du deuxième étage d'une façon individuelle. En réalité, le débit de dissimulation total est la somme de ces débits individuels. Donc, si on utilise les indices des dictionnaires dic21 et dic22, le débit de dissimulation est égal à 650 bits/s. Selon les combinaisons possibles, on peut trouver plusieurs débits de dissimulation variant entre 250 bits/s (débit minimum) et 1500 bits/s (débit maximum).

3.4 Dissimulation dans l'indice de l'excitation

Dans cette technique, on utilise l'indice de l'excitation du dictionnaire algébrique du AMR-WB G.722.2 pour dissimuler des messages secrets. En pratique, le dictionnaire algébrique, noté ici C , est généralement très grand. La recherche de l'indice se fait donc d'une façon non-exhaustive; c.-à-d., seulement une partie du dictionnaire qui est explorée durant le processus de codage (voir

chapitre 2, §. 2.8.3). Ceci peut être exploité dans la conception de la division du dictionnaire C [16]. La dissimulation des messages secrets se fait dans la boucle d'analyse par synthèse lors de la recherche au niveau du dictionnaire algébrique. Par conséquent, la dissimulation de N bits par sous-trame est obtenue en divisant le dictionnaire algébrique C en 2^N sous-dictionnaires.

3.4.1 Préparation du message et débit de dissimulation

Pour chaque sous-trame de 5 ms, le vecteur d'excitation algébrique est divisé en quatre pistes de positions des impulsions. Le message à dissimuler doit être présenté sous forme d'un vecteur *message* contenant quatre éléments où chaque élément sera dissimulé dans une piste. Le nombre de bits des éléments du message ne sont pas forcément égaux. Dans nos travaux, nous avons utilisé le AMR-WB G.722.2 en mode 12.65 Kbits/s où le vecteur d'excitation *vec* contient huit impulsions non nulles (deux impulsions dans chaque piste). Les positions admissibles pour la deuxième impulsion sont calculées en fonction de la première impulsion et le message à dissimuler dans chaque piste.

Le débit de dissimulation R est calculé en fonction du nombre de bits à dissimuler dans chaque sous-trame par la formule suivante :

$$R = \frac{N}{T} \quad (3.11)$$

où N est le nombre de bits à dissimuler dans chaque sous-trame qui est égal à la somme des bits à dissimuler dans chaque piste et T est la durée de la sous trame, égal à 5 ms.

Rappelons que pour le mode 12.65 Kbits/s du WB-AMR G722.2, les positions des impulsions par piste sont codées avec quatre bits. Ainsi, le nombre de bits à dissimuler dans chaque piste doit être inférieur ou égal à quatre. Dans une sous-trame, le nombre de bits total à dissimuler N est compris entre 1 et 16; ce qui nous donne un débit de dissimulation entre 200 bits/s et 3200 bits/s.

3.4.2 Division du dictionnaire de l'excitation

Dans notre application, nous limitons les positions d'impulsion admissibles pour la seconde impulsion ($i_4, \dots, i_7$) dans chaque piste. Le nombre des positions d'impulsion admissibles dans une piste k ($k = 0, \dots, 3$) est lié au nombre de bits à dissimuler dans cette piste. Ce nombre

est calculé en divisant le nombre de positions standard (16 positions) par 2^{n_k} , où n_k représente le nombre de bits à dissimuler dans la piste k . Par exemple, si on veut dissimuler $n_k = 1$ bit alors le nombre des positions admissibles est égal à $16/2 = 8$.

On couple les positions admissibles de la seconde impulsion i_{k+4} dans chaque piste à la position i_k de la première impulsion respective en utilisant les règles données par les équations (3.12) – (3.15). Naturellement, cette stratégie requiert que i_k soit connue avant que les positions admissibles i_{k+4} pour la deuxième impulsion puissent être calculées. Un algorithme adapté de recherche dans le dictionnaire algébrique peut assurer cette propriété [16].

Dans ce qui suit, nous donnons les positions possibles dans chaque piste avec dissimulation du vecteur *message*. Le nombre de bits n_k du *message* joue un rôle très important dans cette étape.

Pour dissimuler $n_k = 1$ de *message*, le nombre de positions admissibles est égal à 8 :

$$i_{k+4} = \begin{cases} G^{-1} \left(G \left(\frac{i_k}{4} \right) \oplus \text{message}[k] \right) \times 4 + k \\ G^{-1} \left(G \left(\frac{i_k}{4} \right) \oplus (\text{message}[k] + 2) \right) \times 4 + k \\ G^{-1} \left(G \left(\frac{i_k}{4} \right) \oplus (\text{message}[k] + 4) \right) \times 4 + k \\ G^{-1} \left(G \left(\frac{i_k}{4} \right) \oplus (\text{message}[k] + 6) \right) \times 4 + k \\ G^{-1} \left(G \left(\frac{i_k}{4} \right) \oplus (\text{message}[k] + 8) \right) \times 4 + k \\ G^{-1} \left(G \left(\frac{i_k}{4} \right) \oplus (\text{message}[k] + 10) \right) \times 4 + k \\ G^{-1} \left(G \left(\frac{i_k}{4} \right) \oplus (\text{message}[k] + 12) \right) \times 4 + k \\ G^{-1} \left(G \left(\frac{i_k}{4} \right) \oplus (\text{message}[k] + 14) \right) \times 4 + k \end{cases} \quad k = 0, 1, 2, 3 \quad (3.12)$$

Pour dissimuler $n_k = 2$ de *message*, le nombre de positions admissibles est égal à 4 :

$$i_{k+4} = \begin{cases} G^{-1} \left(G \left(\frac{i_k}{4} \right) \oplus \text{message}[k] \right) \times 4 + k \\ G^{-1} \left(G \left(\frac{i_k}{4} \right) \oplus (\text{message}[k] + 4) \right) \times 4 + k \\ G^{-1} \left(G \left(\frac{i_k}{4} \right) \oplus (\text{message}[k] + 8) \right) \times 4 + k \\ G^{-1} \left(G \left(\frac{i_k}{4} \right) \oplus (\text{message}[k] + 12) \right) \times 4 + k \end{cases} \quad k = 0, 1, 2, 3 \quad (3.13)$$

Pour dissimuler $n_k = 3$ de *message*, le nombre de positions admissibles est égal à 2 :

$$i_{k+4} = \begin{cases} G^{-1} \left(G \left(\frac{i_k}{4} \right) \oplus \text{message}[k] \right) \times 4 + k \\ G^{-1} \left(G \left(\frac{i_k}{4} \right) \oplus (\text{message}[k] + 8) \right) \times 4 + k \end{cases} \quad k = 0, 1, 2, 3 \quad (3.14)$$

Pour dissimuler $n_k = 4$ de *message*, le nombre de positions admissibles est égal à 1 :

$$i_{k+4} = G^{-1} \left(G \left(\frac{i_k}{4} \right) \oplus \text{message}[k] \right) \times 4 + k \quad k = 0, 1, 2, 3 \quad (3.15)$$

L'opération $\oplus$ représente l'opération d'addition modulo 2 sans retenue (ou exclusif XOR). La fonction G est la fonction qui permet de convertir une valeur en code Gray et G^{-1} est la fonction inverse de G . L'utilisation de ces deux fonctions G et G^{-1} dans le calcul des positions admissibles contribue dans l'augmentation de la robustesse de la méthode de dissimulation aux erreurs de transmission [16].

3.4.3 Procédure de dissimulation

La procédure de dissimulation suit la même stratégie de la recherche standard du AMR-WB G722.2. La seule différence se trouve dans les positions admissibles et le nombre de combinaisons testées. Cette procédure est détaillée dans l'algorithme (3.5).

Algorithme 3.5 : Procédure de recherche des impulsions en mode 12.65 Kbits/s avec dissimulation du message

Données : $i = 0, k = 0, nb_itérations = 4, Q_p = 0, nb_impulsions = 8,$

$Ordre = \{0,1,2,3,1,2,3,0\}, nb = \{4,8,8\}, message[], vec.$

Etape 1:

- Même étape 1 de la recherche standard des impulsions (algorithme 2.1, chap.2).

Etape 2:

- Même étape 2 de la recherche standard des impulsions (algorithme 2.1, chap.2).

Etape 3:

-Si ($k < nb_impulsions/2$) :

- Chercher les deux meilleurs impulsions m_k et m_{k+1} par la méthode standard.
- Calculer les positions admissibles de m_{k+4} et m_{k+5} pour dissimuler $message[k]$ et $message[k+1]$.

Sinon

- Chercher les deux meilleurs impulsions m_k et m_{k+1} dans les positions admissibles des piste $ordre[k]$ et $ordre[k+1]$.

FinSi

- Mise à jour de l'énergie E et la corrélation R .
- Incrémenter k par 2.
- Si ($k < nb_impulsions$) : réitérer l'étape 3.

Etape 4:

- Calculer $Q = R^2/E$

- Si ($Q > Q_p$) :

- $Q_p = Q$.
- Sauvegarder les impulsions $\{m_0 \dots m_7\}$ dans le vecteur vec .

FinSi

- Incrémenter i par 1.
- Décalage à droite cyclique des deux sous-vecteurs $\{ordre[0], \dots, ordre[3]\}$ et $\{ordre[4], \dots, ordre[7]\}$.
- Si ($i < nb_itérations$) : réitérer les étapes 2, 3 et 4.
- Sinon** : Créer l'indice du vecteur d'excitation à partir du vecteur vec .

3.4.4 Procédure d'extraction

L'extraction d'un message $message[k]$ existant dans la piste k nécessite la connaissance de i_k , i_{k+4} , n_k avec $k = \{0, 1, 2, 3\}$, les deux positions des impulsions et le nombre de bits de message dissimulé dans la piste k . Selon la valeur du nombre de bits n_k dissimulés par piste, les formules utilisées dans l'extraction du message sont données par les équations (3.16) - (3.19).

Pour $n_k = 1$:

$$message[k] = \left(G\left(\frac{i_k}{4}\right) \oplus G\left(\frac{i_{k+4}}{4}\right) \right) \bmod 2 \quad (3.16)$$

Pour $n_k = 2$:

$$message[k] = \left(G\left(\frac{i_k}{4}\right) \oplus G\left(\frac{i_{k+4}}{4}\right) \right) \bmod 4 \quad (3.17)$$

Pour $n_k = 3$:

$$message[k] = \left(G\left(\frac{i_k}{4}\right) \oplus G\left(\frac{i_{k+4}}{4}\right) \right) \bmod 8 \quad (3.18)$$

Pour $n_k = 4$:

$$message[k] = \left(G\left(\frac{i_k}{4}\right) \oplus G\left(\frac{i_{k+4}}{4}\right) \right) \bmod 16 \quad (3.19)$$

3.4.5 Complexité de la méthode de dissimulation

La complexité de cette méthode est donnée en fonction du nombre de combinaisons possibles nbc pour générer le vecteur d'excitation. Ce nombre est égal à $nbc = 4 \times (4 \times 16 + 8 \times 16 + 8 \times 16) = 256 \times (1 + 4 \times (\frac{1}{2^1 \times 2^0} + \frac{1}{2^1 \times 2^0})) = 1280$ combinaisons dans le codeur standard.

Dans nos applications, ce nombre nbc dépend du débit de dissimulation et du nombre de bits dissimulés n_k dans chaque piste. Ce nombre est donc calculé par la formule suivante :

$$nbc = 4 \times \left(4 \times 16 + \frac{16}{2^{n_0}} \times \frac{16}{2^{n_1}} + \frac{16}{2^{n_2}} \times \frac{16}{2^{n_3}} \right)$$

$$\begin{aligned}
 nbc &= 256 + 4 \left(\frac{16}{2^{n_0}} \times \frac{16}{2^{n_1}} + \frac{16}{2^{n_2}} \times \frac{16}{2^{n_3}} \right) \\
 nbc &= 256 \times \left(1 + 4 \times \left(\frac{1}{2^{n_0} \times 2^{n_1}} + \frac{1}{2^{n_2} \times 2^{n_3}} \right) \right) \quad (3.20)
 \end{aligned}$$

Il est bien évidant que le nombre de combinaisons dépend non seulement du débit de dissimulation mais aussi de la distribution des bits dissimulés dans les pistes. Donc pour le même débit de dissimulation, on peut trouver deux nombres de combinaisons différents. Si on veut dissimuler par exemple 7 bits par sous-trame, on peut utiliser plusieurs distributions. Supposons qu'on prend $n_0 = 1, n_1 = 1, n_2 = 3$ et $n_3 = 2$, alors $nbc = 828$ combinaisons. Si on prend $n_0 = 1, n_1 = 3, n_2 = 1$ et $n_3 = 2$, alors $nbc = 729$ combinaisons. On remarque que si on veut dissimuler 2 bits par la distribution $n_0 = 1, n_1 = 0, n_2 = 1$ et $n_3 = 0$ $nbc = 1280$ combinaisons qui égale exactement au nombre de combinaisons dans le codeur standard.

La distribution des bits de message sur les pistes contrôle ainsi la manière de la dissimulation de message, le nombre de combinaisons nbc , la complexité de la méthode et la qualité de la couverture.

3.5 Dissimulation dans le délai tonal

Cette méthode de dissimulation est basée sur la quantification du délai tonal (pitch) [54]. La sensibilité du système auditif humain en ce qui concerne la différence du délai tonal est connue pour être rudement proportionnelle au rapport de la différence de fréquence. En d'autres termes, un auditeur humain a une difficulté pour détecter une petite différence de délai tonal relativement à une grande valeur de délai tonal. Par conséquent, La dissimulation d'un message dans les bits moins significatifs (LSB) ne dégrade pas fortement la qualité de la parole publique. Le nombre de bits à dissimuler par cette méthode est proportionnel à la valeur du délai tonal. Donc, cette méthode utilise la technique LSB adaptative appliquée sur la valeur du délai tonal.

Dans notre travail, rappelons-le, nous utilisons le mode 12.65 Kbits/s. Le délai tonal est calculé deux fois dans les sous-frames 1 et 3. Il est quantifié sur neuf bits, donc la valeur quantifiée varie dans l'intervalle $[0, \dots, 512]$. La méthode consiste à comparer la valeur quantifiée

Algorithme 3.6 : Procédure de la dissimulation LSB adaptative dans le délai tonal quantifié**Données** :

- Nombre de bits à dissimuler : n .
- Seuil : $seuil$

Etape 1:

- Pour une trame donnée, trouver les valeurs du délai tonal dans les sous-frames 1 et 3.

Etape 2 :

- Si la valeur du délai tonal $> seuil$
 - Changer les n LSB du délai tonal par les n bits du message à dissimuler.
- FinSi

Etape 3:

- Envoyer la valeur du délai tonal porteur.

Algorithme 3.7 : Procédure de l'extraction à partir du délai tonal quantifié**Données** :

- Nombre de bits à dissimuler : n .
- Seuil : $seuil$

Etape 1:

- Trouver les valeurs du délai tonal dans les sous-frames 1 et 3

Etape 2 :

- Si la valeur du délai tonal $> seuil$
 - Le message est le n LSB du délai tonal
- FinSi

du délai tonal à un certain seuil. Si elle est supérieure à ce seuil, on dissimule les n bits du message. Nous remarquons que la dissimulation dans ce cas dépend de la couverture. Donc, le débit de dissimulation n'est pas constant entre deux couvertures différentes, même si nous utilisons les mêmes conditions sur le seuil et le nombre de bits à dissimuler. Les procédures de la dissimulation et de l'extraction sont données dans les algorithmes 3.6 et 3.7 respectivement.

3.6 Génération du message secret

Dans nos applications, nous utilisons le codeur de parole MELP V.1.2 de 2.4 Kbits/s (voir chapitre 2) 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). Généralement, ce message passe par un registre "buffer" dans lequel toutes les adaptations nécessaires seront effectuées. A la réception, le message reçu passe par une procédure d'extraction qui permet de retrouver le message généré par le MELP.

3.7 Conclusion

Les différentes techniques de stéganographie présentées dans ce chapitre seront réadaptées pour les implémenter sur le codeur de parole AMR-WB G.722.2. Chacune de ces techniques de dissimulation sera évaluée selon ses propriétés en fonction des critères d'évaluation d'un système stéganographique (indétectabilité, imperceptibilité ...).

La quantification des paramètres ISF donne la possibilité d'utiliser deux techniques de la QIM. La première technique utilise la division de dictionnaire et son débit de dissimulation est entre 50 et 350 bits/s. La deuxième est basée sur la modification de schéma de quantificateur et son débit de dissimulation varie entre 250 et 1500 bits/s. L'indice du vecteur d'excitation du dictionnaire fixe est un candidat très important pour une procédure de dissimulation par stéganographie. En effet, il occupe 56.9 % de débit total dans le mode 12.65 kbps et le dictionnaire algébrique contient un nombre très grand d'éléments. Le débit de dissimulation de cette technique est entre 200 et 3200 bits/s. La dissimulation dans l'indice du délai tonal est une autre possibilité. Le débit de dissimulation de cette technique n'est pas constant, car il dépend de la valeur du délai tonal.

Dans le chapitre suivant, nous allons implémenter ces techniques de dissimulation dans le codeur de parole AMR-WB G.722.2. Ces techniques seront évaluées en utilisant les critères d'évaluation adaptés à la technique et au codage de la parole en large bande.

Chapitre 4

Implémentation de systèmes stéganographiques dans un codeur de parole en large bande

4.1 Introduction

Dans ce chapitre, nous présentons les différentes implémentations des techniques de stéganographie mises au point en utilisant les deux codeurs de parole AMR-WB G.722.2 et le MELP V 1.2, fonctionnant respectivement à 12.65 kbps et 2.4 kbps. L'objectif principal est de dissimuler la parole secrète codée par le MELP dans la parole publique codée par le AMR-WB. Les performances de nos systèmes stéganographiques implémentés seront évaluées selon deux critères : la capacité de dissimulation représentée par le débit de dissimulation de la parole secrète et l'imperceptibilité (transparence) représentée par la qualité perceptuelle de la parole publique synthétisée. Pour l'imperceptibilité, nous avons choisi les méthodes d'évaluation objectives, connues sous le nom de mesures perceptuelles. Elles sont basées sur

la modélisation de l'audition en prenant en considération certaines caractéristiques de la perception humaine.

Dans une première partie, nous évaluons les techniques stéganographiques appliquées aux quantificateurs des paramètres ISF (ISFs). Il s'agit de la technique QIM basée sur la division des dictionnaires que nous avons appelée "QIM-DD" et la technique QIM basée sur la modification des schémas de quantificateurs vectoriels (appelée "QIM-MS-VQ"). Au début, nous avons cherché à optimiser nos systèmes stéganographiques pour les mêmes débits de dissimulation en se basant sur les meilleurs choix de dictionnaires utilisés dans la dissimulation. Pour évaluer les performances du quantificateur des paramètres ISF, nous avons utilisé la distorsion spectrale (SD) moyenne comme mesure objective. L'évaluation de la qualité de la parole publique, avec les choix optimaux trouvés dans l'étape précédente, est réalisée en utilisant la mesure objective perceptuelle WB-PESQ (Wide Band Perceptual Speech Quality Measure).

Dans la deuxième partie, nous évaluons les performances de la technique stéganographique QIM que nous avons appliquée au niveau des indices d'excitation du dictionnaire algébrique du AMR-WB G.722.2. Nous évaluons aussi les performances de la technique de dissimulation LSB adaptative appliquée sur la valeur du délai tonal.

4.2 Mesures d'évaluation des performances

La qualité de la parole publique décodée (parole synthétique) est l'un des attributs les plus importants d'un système stéganographiques de parole. En effet, elle représente l'**imperceptibilité** de la technique stéganographique utilisée. Il y'a plusieurs dimensions dans la perception de la qualité, certains d'entre eux sont décrits comme suit :

- Intelligibilité : Que le message sous-jacent du signal de parole peut être bien entendu.
- Naturalité : Cela est souvent lié à la présence ou l'absence de différents types de distorsion ou d'un aspect artificiel dans la parole synthétisée, tels que le bruit, les échos, le silence et la voix métallique.
- Reconnaissance du locuteur : Est-ce que la parole de synthèse permet une simple identification du locuteur d'origine?

Afin que la qualité de la parole soit significative, il est important qu'elle soit mesurée dans différentes conditions selon le domaine d'application du codeur, telles que:

- Dépendance avec le locuteur : Beaucoup de codeurs à faible débit binaire donnent une qualité qui dépend du locuteur. Une évaluation de la qualité approfondie devrait impliquer la parole des hommes adultes, femmes adultes et les enfants.
- Dépendance avec la langue : Plusieurs langues ont des phonèmes différents qui peuvent ou ne peuvent pas être modélisés de manière appropriée par le codeur. Concevoir un codeur en utilisant une langue donnée peut découvrir une faiblesse vers un langage différent.
- Dépendance avec le niveau du signal parole : La qualité d'un codeur n'est pas uniforme en ce qui concerne la variation de la puissance d'entrée. En particulier, le codeur doit être testé pour le niveau extrêmement bas, ainsi que pour le très haut niveau.
- Dépendance avec le bruit de fond : l'étude des performances du codeur de parole avec le bruit de fond est importante dans les applications portables. La source des bruits peuvent être les voitures, la rue, restaurant, bibliothèque...etc.
- Erreurs du Canal : Tous les canaux de communication contiennent une sorte d'erreurs. La qualité du codeur peut être mesurée dans différents taux d'erreur binaire.
- Reproduction des signaux non-parole : La fidélité dans la reproduction de signaux non-parole, comme la musique ou tonalités de signalisation pourrait être importante dans certaines applications.

4.2.1 Mesures subjectives

Les mesures subjectives sont les plus fiables pour évaluer la qualité perçue de la parole. Parmi ces mesures, on note le test subjectif MOS.

Tableau 4.1 : Echelle de variation du score MOS.

Score	Qualité de la parole
5	Excellente
4	Bonne
3	Acceptable
2	Mauvaise
1	Catastrophique

Ce test fait appel à des participants humains qui donnent leurs opinions et notent la qualité du signal synthétique sur une échelle de qualité subjective (tableau 4.1). Les notes obtenues sont par la suite moyennées pour obtenir le score moyen d'opinion MOS (Mean Opinion Score).

Dans le test DMOS (dégradation du score moyen d'opinion), les auditeurs sont présentés avec le signal d'origine comme référence, avant qu'ils écoutent les signaux de synthèse. Ils sont priés de comparer les deux et de donner une note en fonction de la dégradation perçue. Les choix sont notés dans le tableau suivant.

Tableau 4.2 : Echelle de variation du score DMOS.

Score	Niveau de dégradation
5	Imperceptible
4	Juste perceptible, mais pas gênant
3	Perceptible et légèrement gênant
2	Gênant, mais tolérable
1	Très gênant et inacceptable

La fiabilité des résultats des tests subjectifs dépend du nombre d'auditeurs. Dans la plupart des tests, le nombre minimum est de 16 avec des auditeurs choisis de la population générale de manière à mieux simuler les conditions dans lesquelles le codeur pourra être utilisé.

4.2.2 Mesures objectives

Les méthodes de mesure objectives se présentent comme une alternative aux méthodes subjectives. Elles peuvent être classées selon le domaine dans lequel elles opèrent. En effet, il est possible de mesurer la ressemblance des signaux dans le domaine temporel ou spectral. Dans ce qui suit, nous allons présenter seulement les principales mesures objectives y compris celles utilisées dans nos différentes évaluations.

4.2.2.1 Mesures objectives : domaine spectral

Dans le domaine spectral, la performance de la quantification des coefficients LPC est habituellement évaluée par la distorsion spectrale SD_i [48] entre deux vecteurs de la $i^{\text{ème}}$ trame (original et quantifié). Elle est donnée par la formule suivante :

$$SD_i = \sqrt{\frac{1}{n_1 - n_0} \sum_{n=n_0}^{n_1-1} \left[10 \log_{10} \left(\frac{|A_q(e^{j2\pi n/N})|^2}{|A(e^{j2\pi n/N})|^2} \right) \right]} [dB] \quad (4.1)$$

où A et A_q représente la densité spectrale de puissance des coefficients LPC originaux et quantifiés respectivement. Dans notre travail, on s'intéresse à l'encodage des coefficients LPC du codeur de la parole en large bande AMR-WB G.722.2 [38]. Ce dernier utilise des signaux parole sous échantillonnés à 12,8 kHz et fenêtrés chaque 20 ms. D'où, une Transformé de Fourier Rapide FFT (Fast Fourier Transform) de $N = 512$ points sera utilisée pour calculer A et A_q avec $n_0 = 2$ et $n_1 = 256$ qui correspondent à 50 Hz et 6,4 kHz respectivement. C'est l'intervalle dans lequel on effectue la mesure de distorsion, car le signal parole encodé par le G.722.2 est filtré entre ces deux fréquences.

Pour N_f trame, la distorsion spectrale moyenne est donnée par :

$$SD = \frac{1}{N_f} \sum_{i=1}^{N_f} SD_i [dB] \quad (4.2)$$

Les conditions suivantes sont nécessaires pour réaliser une quantification de qualité transparente des LPCs, dans le codage de la parole en bande étroite [36]. Selon [55], des tests d'écoute ont montré que ces conditions de transparence, s'appliquent également dans le codage en large bande

1. La distorsion spectrale moyenne SD devrait être environ 1 dB.
2. Le nombre de trame ayant une SD dans la plage 2-4 dB est moins de 2%.
3. Aucune trame n'a une distorsion spectrale SD_i plus grande que 4 dB.

Les trames qui ne vérifient pas les deux dernières conditions sont connues sous le nom de "Outliers". Les deux conditions des trames Outliers sont ajoutées pour assurer une distorsion plus homogène sur toutes les fenêtres puisqu'ils sont audibles par l'oreille humaine.

4.2.2.2 Mesures objectives : domaine temporel

Dans le domaine temporel, le rapport signal sur bruit (RSB) représente la normalisation de la puissance du signal original sur la puissance du bruit; c-à-d, différence entre signal original et signal évalué, mais ça ne correspond pas aux mesures subjectives. Dans le codage de la parole, on utilise plutôt le rapport signal sur bruit segmental moyen RSBseg. Il est défini

comme la moyenne des RSB de tous les segments du signal de parole, où un segment correspond à un ensemble de N échantillons de 20 à 30 ms. Sa formalisation pour une fenêtre est donnée comme suit [56] :

$$RSB_{seg} = 10 \log_{10} \left[\frac{\sum_{n=1}^N x(n)^2}{\sum_{n=1}^N (x(n) - \hat{x}(n))^2} \right] [dB] \quad (4.3)$$

Le RSB et le RSBseg n'ont de sens que pour les codeurs qui gardent la forme d'onde du signal (PCM, ADPCM,...). A la fois, le RSB et le RSBseg sont extrêmement sensibles aux désalignements et aux déphasages qui ne sont pas toujours perceptives par l'oreille humaine. D'autre part, la plupart des codeurs à faible débit binaire ne préservent pas la forme d'onde d'origine et donc le RSB et le RSBseg n'ont pas de sens pour l'évaluation de ces codeurs.

4.2.3 Mesures perceptuelles objectives

Il est clair que les tests subjectifs sont coûteux à mettre en œuvre et prennent beaucoup de temps. Par conséquent, les efforts des recherches actuels sont dirigés vers des tests perceptuels basés sur des mesures objectives.

Afin d'évaluer la qualité globale de la parole synthétique, une autre méthode d'évaluation objective est utilisée. Il s'agit de l'algorithme PESQ (Perceptual Speech Quality Measure) et sa version adaptée en large bande WB-PESQ (Wide Band Perceptual Speech Quality Measure). Cet algorithme, standardisé par l'Union Internationale des Télécommunications ITU-T, est connu sous la recommandation P.862.2 [57]. Il est utilisé indifféremment du test MOS mais il donne une note similaire à ce dernier. Le PESQ et WB-PESQ permettent d'évaluer la qualité d'écoute dans de nombreuses conditions de dégradation (perte de paquets, distorsion due au codage et bruit,...etc).

Cet algorithme doit avoir le signal original comme référence, car l'évaluation WB-PESQ compare un signal initial $X(t)$ à un signal dégradé $Y(t)$ produit par le passage de $X(t)$ à travers un système de communication. Il permet de prévoir les notes de qualité de perception qu'attribueraient les participants à un essai d'écoute subjectif au signal $Y(t)$. Du fait que cet algorithme est largement utilisé dans l'évaluation des systèmes stéganographique de la parole [27, 28, 49, 58], nous avons choisi de l'utiliser comme mesure d'évaluation des performances de nos implémentations.

4.3 Système QIM-DD appliqué au quantificateur des ISFs

Cette technique de dissimulation peut être appliquée séparément sur les dictionnaires du quantificateur des ISFs du AMR-WB G.722.2. Dans la phase de division, chaque dictionnaire ISF est évalué individuellement pour trouver la meilleure division possible. Il s'agit d'évaluer les trois techniques de division que nous avons détaillées dans le chapitre précédent : la division équilibrée par DGD (division génétique de dictionnaire), la division aléatoire et la division non équilibrée proposée. D'autre part, pour concevoir notre système stéganographique QIM-DD optimal qui assure une dégradation minimale, nous avons fait une évaluation des performances de ce système pour différents choix des dictionnaires ISF utilisés.

Nous présentons, ci-dessous, les performances en termes de SD et "Outliers" des systèmes QIM-DD, conçus selon les meilleurs choix pour le même débit de dissimulation. La base de données parole utilisée dans les simulations se compose de 60 minutes de parole prise de la base de données internationale TIMIT [59] dont la fréquence d'échantillonnage est de 16 kHz. Pour construire la base de données des vecteurs ISFs, on a utilisé la même fonction d'analyse par prédiction linéaire (LP) du codeur G.722.2. Cette fonction effectue une analyse LP d'ordre 16 par la méthode d'auto-corrélation sur chaque trame d'analyse de 20 ms. Les paramètres ISF sont obtenus en utilisant la fonction de conversion du codeur G.722.2. D'où, une base de 180000 vecteurs ISFs a été construite. Notons que cette base est obtenue pour différents locuteurs.

4.3.1 Performances des systèmes stéganographiques QIM-DD

Les performances des systèmes QIM-DD, conçus selon les trois méthodes de division, sont présentées sur les tableaux 4.3, 4.4 et 4.5. Les vecteurs ISFs de dimension $k = 16$ sont quantifiés par le même quantificateur du G.722.2 qui contient sept dictionnaires. Pour un débit de dissimulation fixe, ces dictionnaires sont notés implicitement dans les tableaux de la gauche à la droite (dic11, dic12, dic21, ..., dic25). La quantification se fait à des débits de dissimulation variant de 1 à 7 bits/trame où l'on présente le choix optimum des dictionnaires utilisés dans la dissimulation de message selon la distance spectrale SD . La notation **2** (0-0-0-1-1-0-0) par exemple signifie que pour un débit de dissimulation de 2 bits/trame, les dictionnaires dic22 et dic23 du 2^{ème} étage du S-MSVQ ont été sélectionnés comme meilleurs choix pour dissimuler les 2 bits (1, 1). Notons, enfin, que les bits du message à dissimuler ont été générés d'une façon complètement aléatoire.

Tableau 4.3: Performance du système QIM-DD conçu par la méthode de division non équilibrée.

Débit de dissimulation (Bits/trame)	SD Moy. [dB]	SD "Outliers" (en %)	
		2-4 dB	> 4 dB
1 (0-0-0-1-0-0-0)	0.95	1.13	0.0005
2 (0-0-0-1-1-0-0)	0.99	1.45	0.002
3 (0-0-0-1-1-1-0)	1.03	1.80	0.005
4 (0-0-1-1-1-1-0)	1.09	2.48	0.008
5 (0-0-1-1-1-1-1)	1.15	3.19	0.009
6 (0-1-1-1-1-1-1)	1.23	5.57	0.017
7 (1-1-1-1-1-1-1)	1.30	7.61	0.035

Tableau 4.4: Performance du système QIM-DD conçu par la méthode de division équilibrée

Débit de dissimulation (Bits/trame)	SD Moy. [dB]	SD "Outliers" (en %)	
		2-4 dB	> 4 dB
1 (0-0-0-1-0-0-0)	0.95	1.11	0.001
2 (0-0-0-1-1-0-0)	1.00	1.42	0.001
3 (0-0-0-1-1-1-0)	1.05	1.83	0.007
4 (0-0-1-1-1-1-0)	1.11	2.55	0.007
5 (0-0-1-1-1-1-1)	1.17	3.35	0.007
6 (0-1-1-1-1-1-1)	1.25	6.14	0.027
7 (1-1-1-1-1-1-1)	1.33	8.85	0.041

Tableau 4.5: Performance du système QIM-DD conçu par la méthode de division aléatoire.

Débit de dissimulation (Bits/trame)	SD Moy. [dB]	SD "Outliers" (en %)	
		2-4 dB	> 4 dB
1 (0-0-0-1-0-0-0)	0.96	1.32	0.001
2 (0-0-0-1-1-0-0)	1.03	1.93	0.005
3 (0-0-0-1-1-1-0)	1.09	2.66	0.010
4 (0-0-1-1-1-1-0)	1.17	4.43	0.005
5 (0-0-1-1-1-1-1)	1.24	5.84	0.015
6 (0-1-1-1-1-1-1)	1.37	11.03	0.120
7 (1-1-1-1-1-1-1)	1.51	17.46	0.465

Ces résultats de simulation montrent que notre méthode de division non équilibrée de dictionnaire assure des performances légèrement meilleures que les deux autres méthodes de division conventionnelles. La quantification transparence est assurée jusqu'à un débit de dissimulation de 4 bits/trame; c.-à-d., un débit de 200 bits/s (bps). L'utilisation d'une division aléatoire provoque une plus grande dégradation comparée aux deux autres méthodes. Par exemple, pour un débit de dissimulation de 7 bits/trame, on remarque que la division non équilibrée donne une SD égale à 1.33 dB, la division équilibrée donne une SD égale à 1.35 dB et la division aléatoire donne une SD égale à 1.51 dB. D'où, des gains SD de 0.02 dB et de 0.18 dB ont été obtenus par la méthode de dissimulation avec division non équilibrée comparé aux deux autres méthodes respectivement. Il en est de même en ce qui concerne le pourcentage des trames Outliers, où l'utilisation de la division non équilibrée réduit aussi le nombre des trames ayant une SD supérieure à 2 dB.

Dans les trois méthodes de division, la capacité de dissimulation varie entre 1 bit/trame et 7 bits/trame, c.à.d. entre 50 bits/s et 350 bits/s

4.3.2 Performance du G.722.2 avec implémentation du QIM-DD

Dans nos simulations, nous avons utilisé le G.722.2 en mode débit 12.65 kbps; c'est-à-dire, 46 bits/trame sont utilisés pour coder les paramètres ISF. La base de données utilisée dans les évaluations suivantes est composée de 10 séquences de parole de 32 secondes extraites de la base de données TIMIT. Les performances du codeur G.722.2, avec implémentation du système QIM-DD au niveau du quantificateur S-MSVQ des ISFs, sont présentées sur le tableau 4.6. Pour chaque méthode de division utilisée, les résultats sont donnés en termes de SD moyenne pour évaluer le codage des ISFs avec dissimulation du message et en notes WB-PESQ pour évaluer la qualité globale de la parole synthétisée par le G.722.2. Le choix optimum de dictionnaires utilisés dans la procédure de dissimulation est le même choix trouvé précédemment. Notons que le message binaire secret dissimulé a été généré par le codeur de parole MELP de 2.4 kbps. On a utilisé une base de données arabe phonétiquement équilibrée [60] de 143s prononcé par 3 locuteurs masculins et 3 locuteurs féminins où chaque locuteur a prononcé 10 phrases. Notons que nous avons effectué un sous-échantillonnage à 8 kHz puisque au départ le corpus de parole a été enregistré avec une fréquence de 10 kHz. Il faut préciser que les performances données pour un débit de dissimulation 0 bits/trame sont exactement les performances du codeur standard original G.722.2 sans modification.

Tableau 4.6 : Performance du codeur global G.722.2 avec implémentation du QIM-DD.

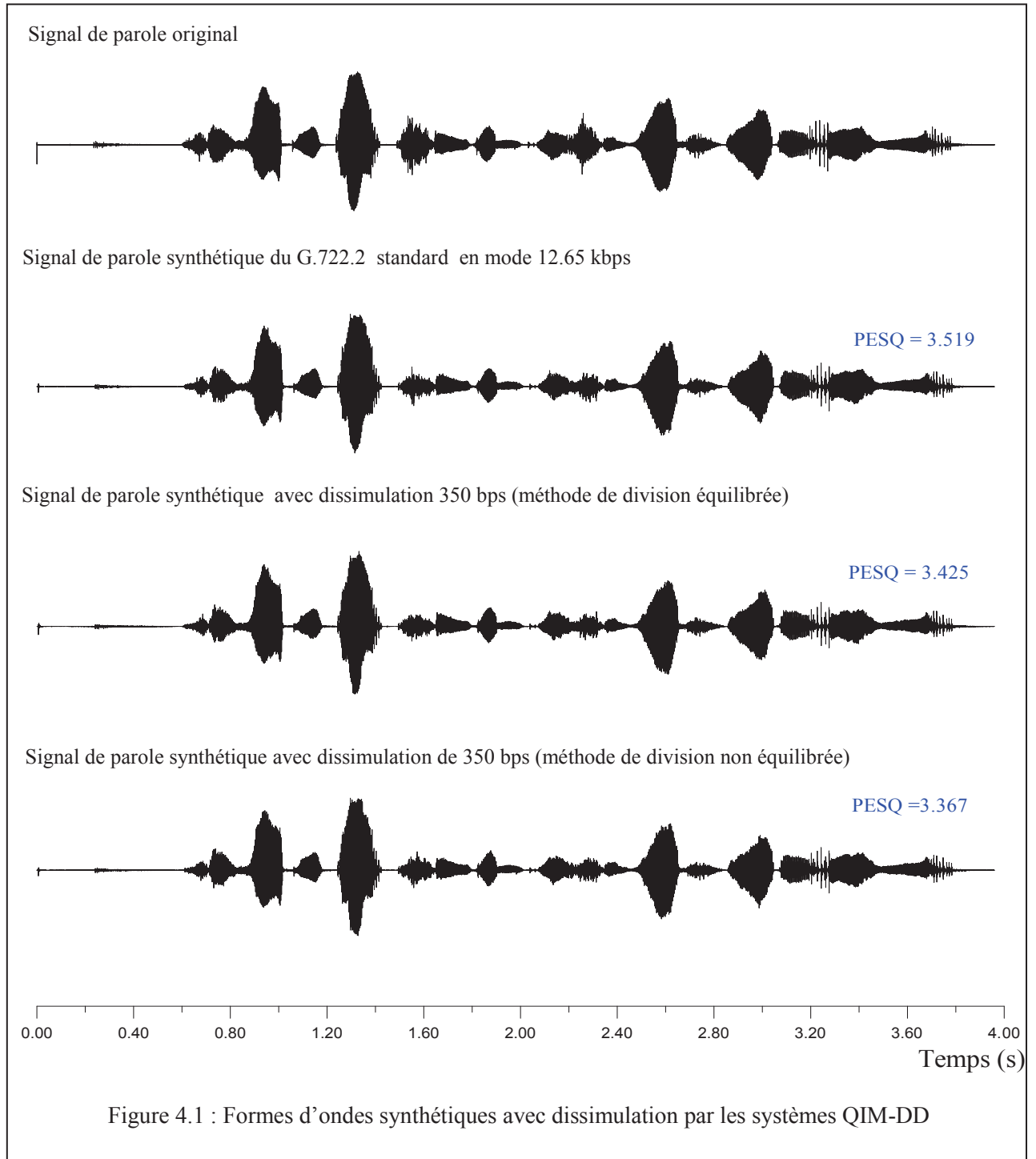
Débit de dissimulation (Bits/trame)	QIM-DD par division équilibrée		QIM-DD par division non-équilibrée	
	<i>SD</i> Moy.	WB-PESQ	<i>SD</i> Moy.	WB-PESQ
0	0.91	3.790	0.91	3.790
1	0.95	3.798	0.95	3.705
2	1.00	3.823	0.99	3.744
3	1.05	3.687	1.03	3.814
4	1.11	3.719	1.09	3.680
5	1.17	3.756	1.15	3.747
6	1.26	3.766	1.23	3.720
7	1.34	3.676	1.32	3.720

Pour tous les débits de dissimulation, ces résultats montrent que la technique QIM-DD avec la méthode de division non équilibrée apporte une légère amélioration aux performances WB-PESQ du G.722.2 comparée à la QIM-DD par division équilibrée. D'autre part, la qualité globale de la parole publique avec dissimulation est presque identique, en termes de WB-PESQ, à la qualité de la parole publique synthétisée par le codeur original G.722.2 sans dissimulation. Ce qui signifie que nos techniques stéganographiques implémentées sont pratiquement imperceptibles.

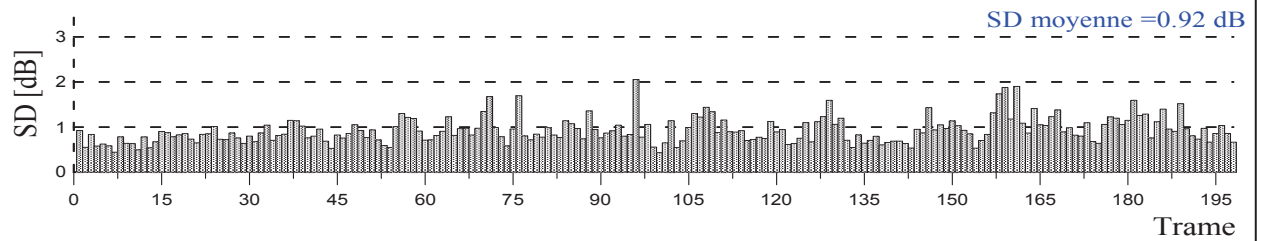
Nous donnons dans les figures 4.1 et 4.2, un simple exemple de résultat de simulation obtenu lors du codage de la phrase "She had your dark suit in greasy wash water all year " de durée de 3.96 s. Cette phrase qui représente la parole publique a été extraite de la base TIMIT. Les performances du codeur G.722.2 (original et avec dissimulation du message) sont données en termes de WB-PESQ sur les figures, où l'on représente aussi les différents formes d'onde synthétiques ainsi que l'évolution des *SD* correspondants.

Les résultats montrent que les deux méthodes permettent d'assurer une faible dégradation pour un débit de dissimulation de 350 bps. La division non équilibrée donne un résultat légèrement supérieur à celui donné par la méthode de la division équilibrée. Pour un débit de dissimulation de 350 bps, la note WB-PESQ est diminuée de 0.09 et 0.15 pour la méthode de division équilibrée et non équilibrée respectivement. Cette faible dégradation n'influence pas la qualité globale de la parole qui reste toujours acceptable ($3 < \text{WB-PESQ} <$

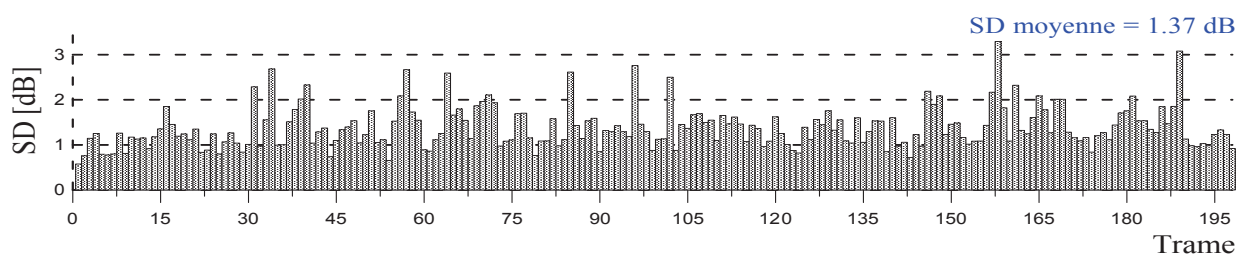
4). Pratiquement, la qualité globale de la parole est identique à la qualité obtenue par le codeur standard G.722.2. Donc la QIM-DD a une très bonne imperceptibilité avec une capacité de dissimulation maximale de 350 bps.



SD du signal de parole synthétique du G.722.2 standard en mode 12.65 kbps



SD du signal de parole synthétique avec dissimulation 350 bps (méthode de division équilibrée)



SD du signal de parole synthétique avec dissimulation 350 bps (méthode de division non équilibrée)

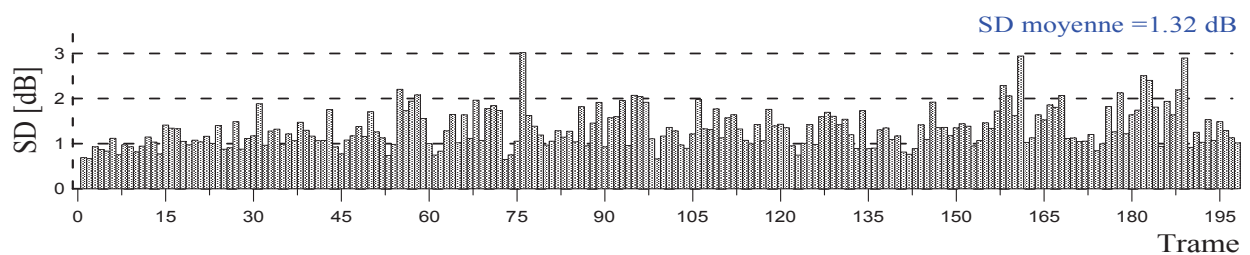


Figure 4.2 : Comparaison de l'évolution de la *SD* entre la division équilibrée et non équilibrée

4.4 Système QIM-MS-VQ appliqué au quantificateur des ISFs

Dans ce qui suit, nous évaluons les performances d'un autre système stéganographique que nous avons conçu à base de la technique QIM par modification de l'architecture des schémas VQs du quantificateur des ISFs du G.722.2. Pour mettre en œuvre ce système, que nous avons appelé QIM-MS-VQ, nous nous sommes intéressés à deux techniques de modification (c.f. chapitre 3). Il s'agit de la procédure de dissimulation par remplacement avec tramage (DRT) et la procédure classique de dissimulation par remplacement simple (DRS). Rappelons qu'une partie du débit du deuxième étage du S-MSVQ des ISFs sera allouée pour dissimuler les messages secrets binaires. D'une façon similaire au système précédent, il est toujours

important d'évaluer les performances de ce quantificateur pour différents choix de dictionnaires utilisés afin de retenir le choix optimal pour chaque débit de dissimulation. Dans les simulations suivantes, nous utilisons la même base de données des ISFs qui contient 180000 vecteurs. Les vecteurs ISFs sont quantifiés par un quantificateur S-MSVQ de G.722.2 modifié selon le principe de la technique QIM-MS-VQ.

4.4.1 Performances des systèmes stéganographiques QIM-MS-VQ

Les performances des systèmes stéganographiques QIM-MS-VQ, conçus selon les deux techniques de modification (DRS et DRT) sont présentées sur les tableaux 4.7 et 4.8. Pour un débit de dissimulation fixe, les dictionnaires notés dans les tableaux sont seulement les dictionnaires du deuxième étage (dic21,..., dic25) du S-MSVQ. Précisons que le "1" signifie que le dictionnaire correspondant est utilisé dans la procédure de dissimulation. Le débit de ce dictionnaire est alors réservé pour le message à dissimuler.

Tableau 4.7: Performance du système QIM-MS-VQ conçu par la méthode DRS.

Débit de dissimulation (Bits/trame)	SD Moy. [dB]	SD "Outliers" (en %)	
		2-4 dB	> 4 dB
5 (0-0-0-0-1)	1.65	21.63	0.08
6 (1-0-0-0-0)	2.34	60.79	3.20
7 (0-1-0-0-0)	1.92	39.39	1.11
10 (0-0-0-1-1)	2.20	58.38	0.78
11 (1-0-0-0-1)	2.72	75.34	6.40
12 (0-0-1-0-1)	2.39	65.97	2.53
14 (0-1-1-0-0)	2.61	71.19	5.66
16 (1-0-0-1-1)	3.09	80.70	12.65
17 (0-1-0-1-1)	2.81	81.93	6.35
18 (1-0-1-0-1)	3.27	76.25	18.82
20 (1-1-1-0-0)	3.41	71.99	24.02
23 (1-1-0-1-1)	3.57	69.77	28.64
24 (0-1-1-1-1)	3.29	79.05	17.93
25 (1-1-1-0-1)	3.67	65.67	33.00
30 (1-1-1-1-1)	3.98	53.12	46.57

Tableau 4.8: Performance du système QIM-MS-VQ conçu par la méthode DRT.

Débit de dissimulation (Bits/trame)	SD Moy. [dB]	SD "Outliers" (en %)	
		2-4 dB	> 4 dB
5 (0-0-0-0-1)	1.48	14.82	0.06
6 (1-0-0-0-0)	2.03	46.36	1.64
7 (0-0-1-0-0)	1.79	31.98	0.64
10 (0-0-0-1-1)	1.92	39.05	0.40
11 (1-0-0-0-1)	2.35	62.14	2.70
12 (0-0-1-0-1)	2.13	51.38	1.37
14 (0-1-1-0-0)	2.40	64.21	3.31
16 (1-0-0-1-1)	2.67	73.41	5.93
17 (0-0-1-1-1)	2.48	69.77	3.20
18 (1-0-1-0-1)	2.84	75.79	9.05
20 (1-1-1-0-0)	3.10	76.91	14.50
23 (1-0-1-1-1)	3.12	77.40	14.95
24 (0-1-1-1-1)	2.95	80.70	9.85
25 (1-1-1-0-1)	3.29	76.03	19.30
30 (1-1-1-1-1)	3.54	69.37	28.53

La quantification se fait à des débits de dissimulation variant de 5 à 30 bits/trame où l'on présente le choix optimum des dictionnaires utilisés dans la dissimulation de message selon la distance spectrale. Les messages à dissimuler sont générés d'une façon aléatoire.

Ces résultats montrent que la dégradation des performances *SD* due à la dissimulation n'est pas proportionnelle au débit de dissimulation. Par exemple, pour un débit de dissimulation de 10 bits/trame, la dégradation causée par la dissimulation dans les indices des dictionnaires dic24 et dic25 est inférieure à celle causée par la dissimulation dans les indices du dictionnaire dic21 du cas de 6 bits/trame. En effet, la dégradation est plutôt liée à l'importance du dictionnaire utilisé. Sachant que la perception humaine de la distorsion causée par les formants dans les bandes de fréquence élevées est négligeable par rapport à celle dans les basses fréquences, donc les dictionnaires qui représentent les hautes fréquences sont moins importants que ceux des basses fréquences. On remarque aussi que les deux techniques ont le même comportement en ce qui concerne l'importance des dictionnaires.

Pour tous les débits de dissimulation, l'évaluation comparative des résultats montre que le système stéganographique QIM-MS-VQ conçu par la méthode DRT assure de meilleures performances que le QIM-MS-VQ conçu par la technique DRS. D'autre part, on constate que pour les deux systèmes stéganographiques, la quantification de qualité transparente n'est pas totalement assurée même pour un débit de dissimulation minimal de 5 bits/trame en utilisant le dictionnaire le moins important (dic25).

4.4.2 Performance du G.722.2 avec implémentation du QIM-MS-VQ des ISFs

Pour les simulations suivantes, nous utilisons toujours le G.722.2 en mode 12.65 kbps où 46 bits/trame sont utilisés pour coder les paramètres ISF. La base de données utilisée est composée de 10 séquences de parole de 32 secondes extraites de la base de données TIMIT.

Les performances du codeur G.722.2, avec implémentation du système QIM-MS-VQ au niveau du quantificateur S-MSVQ des ISFs, sont présentées sur le tableau 4.9. Pour chaque méthode de dissimulation utilisée (DRS ou DRT), les résultats sont donnés en termes de *SD* moyenne pour l'évaluation du codage des ISFs avec dissimulation et en notes WB-PESQ pour l'évaluation de la qualité globale de la parole synthétisée par le G.722.2. Notons que pour chaque débit de dissimulation, les choix optimaux des dictionnaires du 2^{ème} étage utilisés dans la procédure de dissimulation sont les mêmes que ceux trouvés précédemment. Enfin, le message binaire secret dissimulé est le flux binaire généré (bit-stream) par le codeur de parole MELP de 2.4 kbps en utilisant la même base de données arabe précédente [60]. Rappelons que les performances données pour un débit de dissimulation 0 bits/trame sont exactement les performances du codeur standard original G.722.2 sans modification.

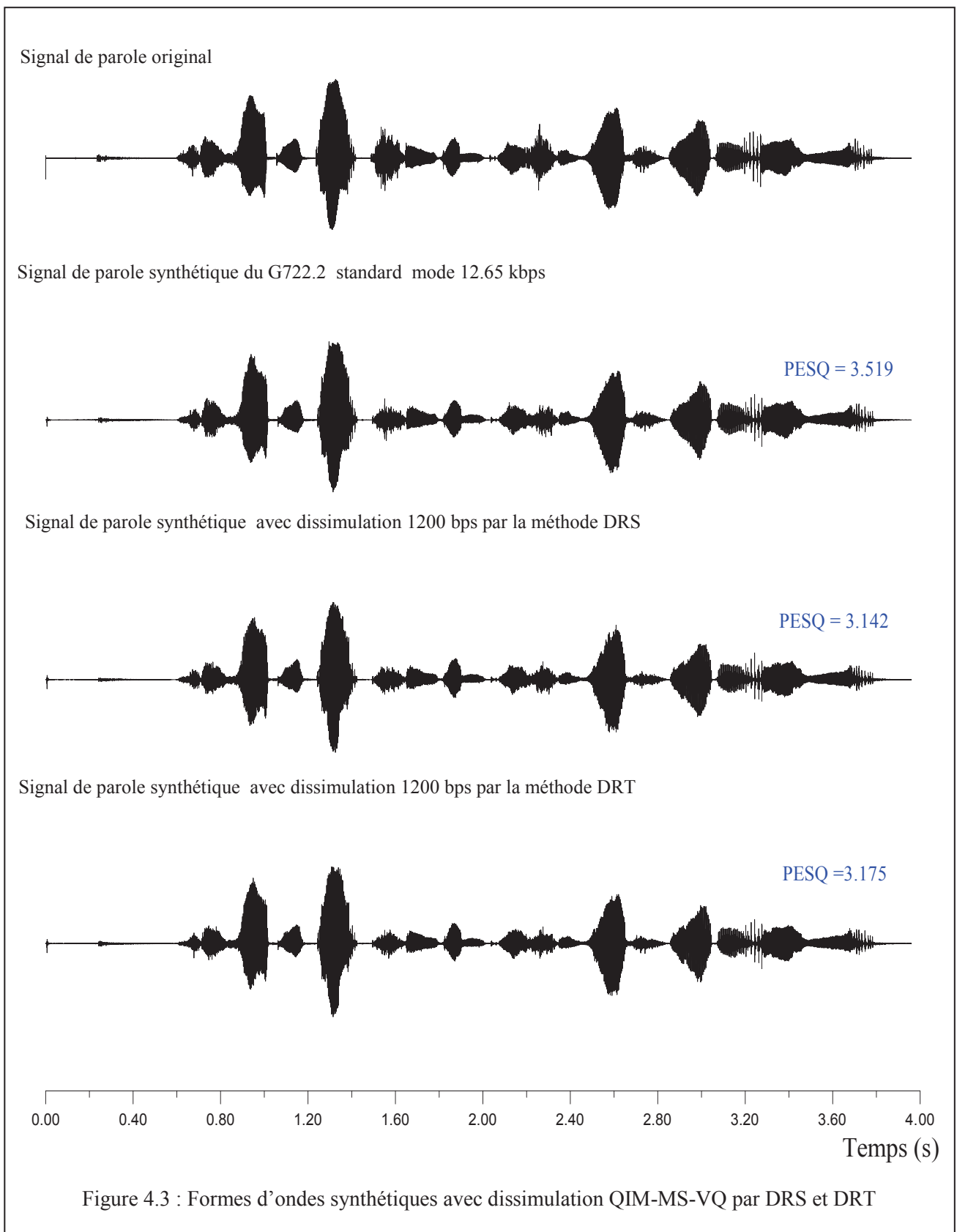
Ces résultats montrent que l'imperceptibilité n'est pas liée seulement à la capacité (débit) de dissimulation mais aussi au choix des dictionnaires utilisés. Par exemple, pour les débits de dissimulation de 23 et 24 bits/trame (1150 et 1200 bps), on remarque que l'imperceptibilité de la dernière est meilleure. Donc dans ce cas, l'utilisation de 24 bits /trame offre plus de capacité et une meilleure imperceptibilité.

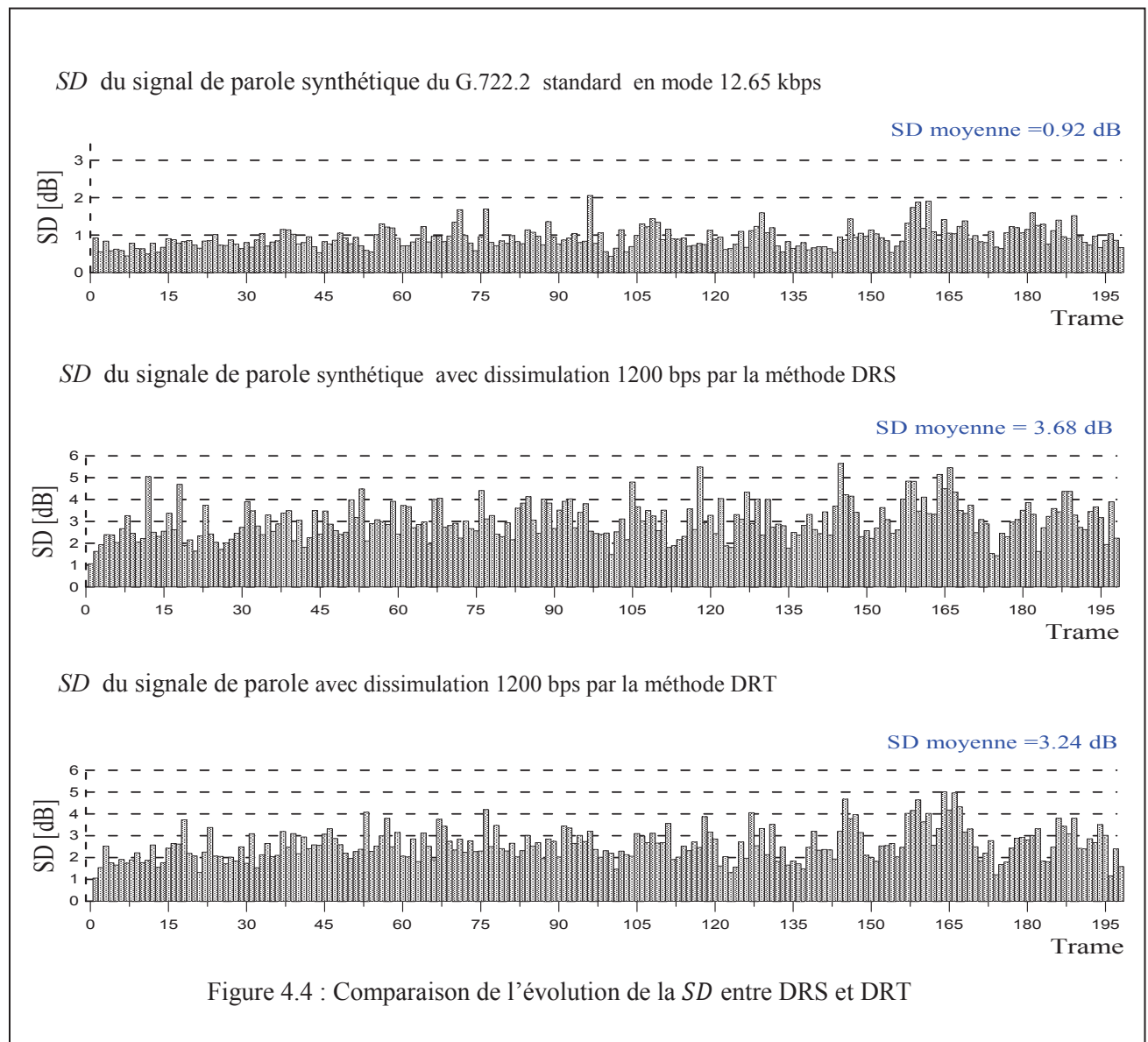
On remarque aussi que l'imperceptibilité de la technique QIM-MS-VQ conçu par DRT est meilleure que celle conçu par DRS pour tous les débits de dissimulation.

Tableau 4.9 : Performance du codeur G.722.2 avec implémentation du système QIM-MS-VQ des ISFs.

Débit de dissimulation (Bits/trame)	QIM-MS-VQ par DRS		QIM-MS-VQ par DRT	
	<i>SD</i> Moy.	WB-PESQ	<i>SD</i> Moy.	WB-PESQ
0	0.91	3.790	0.91	3.790
5	1.58	3.775	1.44	3.738
6	2.06	3.265	1.88	3.321
7	2.08	3.684	1.91	3.728
10	2.09	3.651	1.86	3.651
11	2.51	3.233	2.22	3.337
12	2.52	3.665	2.22	3.700
14	2.89	3.556	2.57	3.568
16	2.86	3.210	2.50	3.279
17	2.89	3.579	2.52	3.574
18	3.10	3.224	2.75	3.312
20	3.54	3.054	3.15	3.110
23	3.51	3.093	3.07	3.161
24	3.52	3.480	3.07	3.553
25	3.77	3.072	3.34	3.105
30	4.02	3.033	3.52	3.139

Nous donnons ci-dessous, un exemple de résultat de simulation obtenu lors du codage de la même phrase utilisée dans la partie précédente (She had your dark suit in greasy wash water all year). Les performances du codeur G.722.2 (original et avec dissimulation du message) sont données en termes de WB-PESQ sur les figures suivantes, où l'on représente aussi les différents formes d'onde synthétiques ainsi que l'évolution des *SD* correspondants pour plusieurs débit de dissimulation.





4.5 Comparaisons entre les méthodes QIM appliquées aux quantificateurs des paramètres ISF

L'un des avantages du système QIM-MS-VQ sur le système QIM-DD est sa grande capacité de dissimulation qui lui permet de dissimuler plus de messages binaires. Cependant ce gain en capacité de dissimulation se paye par une plus grande dégradation qui cause l'imperceptibilité du système.

L'utilisation des paramètres ISF ou une autre conversion des paramètres LPC pour dissimuler un message est possible pour un grand nombre de codeurs de parole récents car ces derniers sont généralement basés sur la prédiction linéaire. Cependant, les performances des

systèmes stéganographiques à implémenter doivent être évaluées avant d'opter pour une telle ou telle méthode. Dans notre cas, nous avons un choix entre un système stéganographique (QIM-MS-VQ) qui offre une grande capacité de dissimulation mais une mauvaise imperceptibilité et un système (QIM-DD) qui assure une bonne imperceptibilité avec une faible capacité de dissimulation.

4.6 Dissimulation dans l'indice de l'excitation du G.722.2

Dans cette méthode, on utilise l'indice du vecteur d'excitation (dictionnaire algébrique) pour dissimuler le message secret. On modifie donc la boucle d'analyse par synthèse du codeur de parole standard AMR-WB G.722.2. Rappelons que le vecteur d'excitation est calculé chaque sous-trame de 5 ms en utilisant une méthode d'entrelacement (c.f. chapitre 2). Chaque sous-trame est représentée par 4 pistes où un nombre maximum de 4 bits peut être dissimulé par piste. On a donc une possibilité de dissimuler jusqu'à 16 bits par sous trame comme on l'a bien expliqué dans le chapitre précédent. Le problème majeur dans cette méthode est de trouver la distribution optimale de ces bits sur les 4 pistes afin d'assurer une dégradation minimale.

Pour un même débit de dissimulation, nous présentons ci-dessous les performances en termes de WB-PESQ des meilleures distributions des bits/sous-trame que nous avons trouvés. La base de données parole utilisée dans les simulations suivantes se compose de 100 phrases de 347 secondes, prises de la base de données TIMIT.

Pour des débits de dissimulation supérieurs à 2 bits par sous-trame, les WB-PESQ moyens obtenus sont légèrement inférieurs au WB-PESQ moyen obtenu par le codeur standard G.722.2 sans dissimulation. Cependant, ces performances restent très acceptables justifiant ainsi l'imperceptibilité de la technique de dissimulation implémentée.

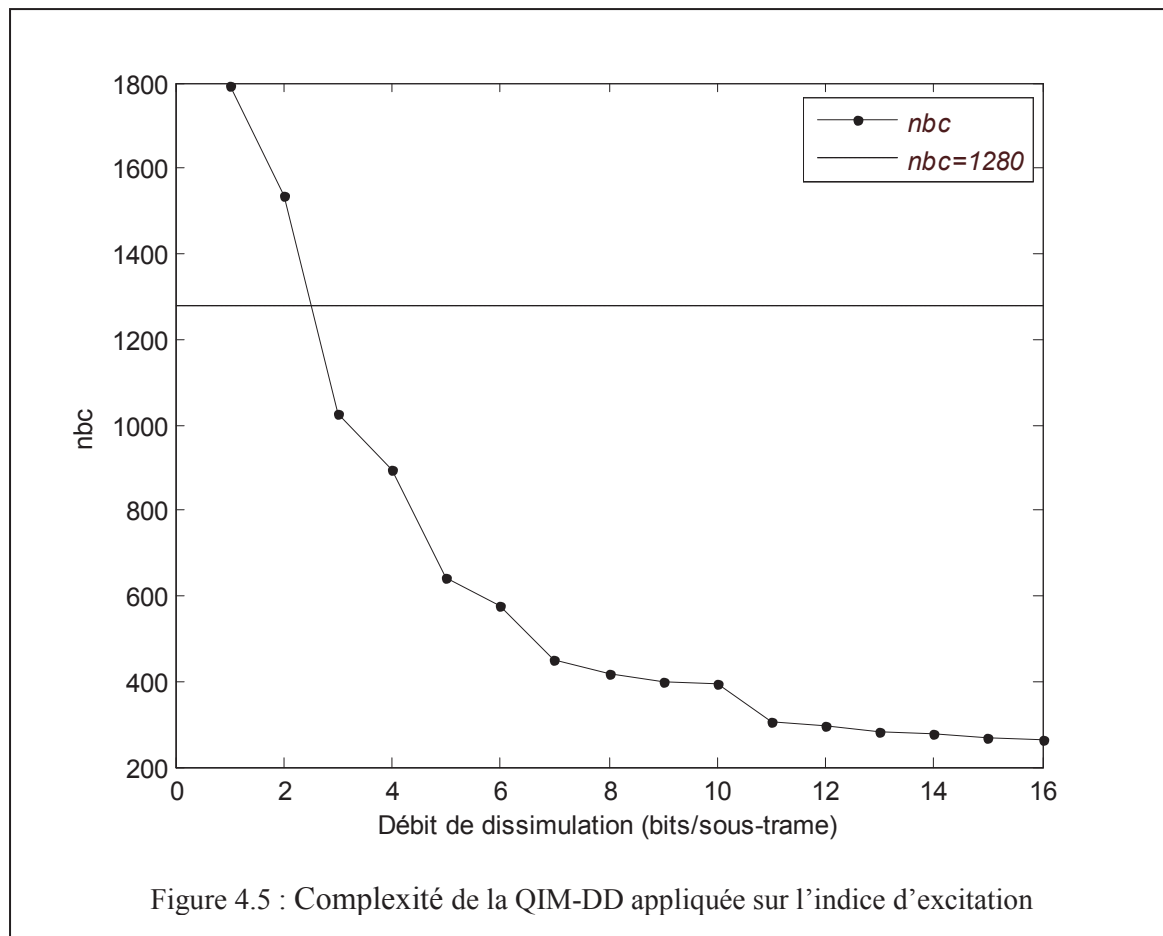
En plus de la dissimulation des messages secrets, cette technique a permis aussi de diminuer la complexité de codage en diminuant le nombre de combinaisons possibles nbc pour générer le vecteur d'excitation. Pour un débit de 16 bits par sous-trame qui est équivalent à 3200 bps le nombre nbc est égal à 256. En le comparant avec le nombre nbc du codeur standard (1280) on conclut que la complexité dans cet étage est diminuée de 20% comparé à celle du codeur standard.

Tableau 4.10: Performance du G.722.2 avec dissimulation dans l'indice de l'excitation

Débit de dissimulation (Bits/sous-trame)	Performance du G.722.2 : WB-PESQ moyen
1 (0+0+1+0)	3.832
2 (0+0+1+1)	3.809
3 (1+1+0+1)	3.797
4 (0+1+1+2)	3.777
5 (3+0+1+1)	3.756
6 (0+2+1+3)	3.740
7 (1+3+3+0)	3.709
8 (2+1+4+1)	3.672
9 (2+1+2+4)	3.659
10 (3+4+2+1)	3.624
11 (2+4+3+2)	3.594
12 (2+3+3+4)	3.554
13 (3+4+4+2)	3.502
14 (2+4+4+4)	3.455
15 (4+3+4+4)	3.414
16 (4+4+4+4)	3.353

Nous présentons, dans la figure 4.5, l'évolution du nombre nbc en fonction des débits de dissimulation. Le nombre nbc est calculé par la distribution optimale des bits obtenu précédemment.

On remarque que le nombre nbc est supérieur au nombre nbc du standard seulement pour les deux débits de dissimulation de 1 et 2 bits/sous-trame. Donc la complexité est plus petite que la complexité du codeur standard G.722.2 pour chaque débit supérieur à 2 bit/sous-trame (400 bps).



4.7 Dissimulation dans le délai tonal

Cette méthode de dissimulation est appliquée au niveau de la quantification du délai tonal (pitch) du G.722.2. Il s'agit de la méthode LSB adaptative qui modifie la valeur du délai tonal en fonction du message à dissimuler (c.f. chapitre3). Rappelons que pour chaque trame le délai tonal est calculé deux fois dans les sous-frames 1 et 3. Il est quantifié sur neuf bits, donc la valeur quantifiée varie dans l'intervalle $[0, \dots, 512]$ dans le mode 12.65 kbps. Par la suite, on utilisera plusieurs combinaisons du nombre de bits à dissimuler et des seuils pour étudier l'influence de ces paramètres sur la qualité de la parole publique.

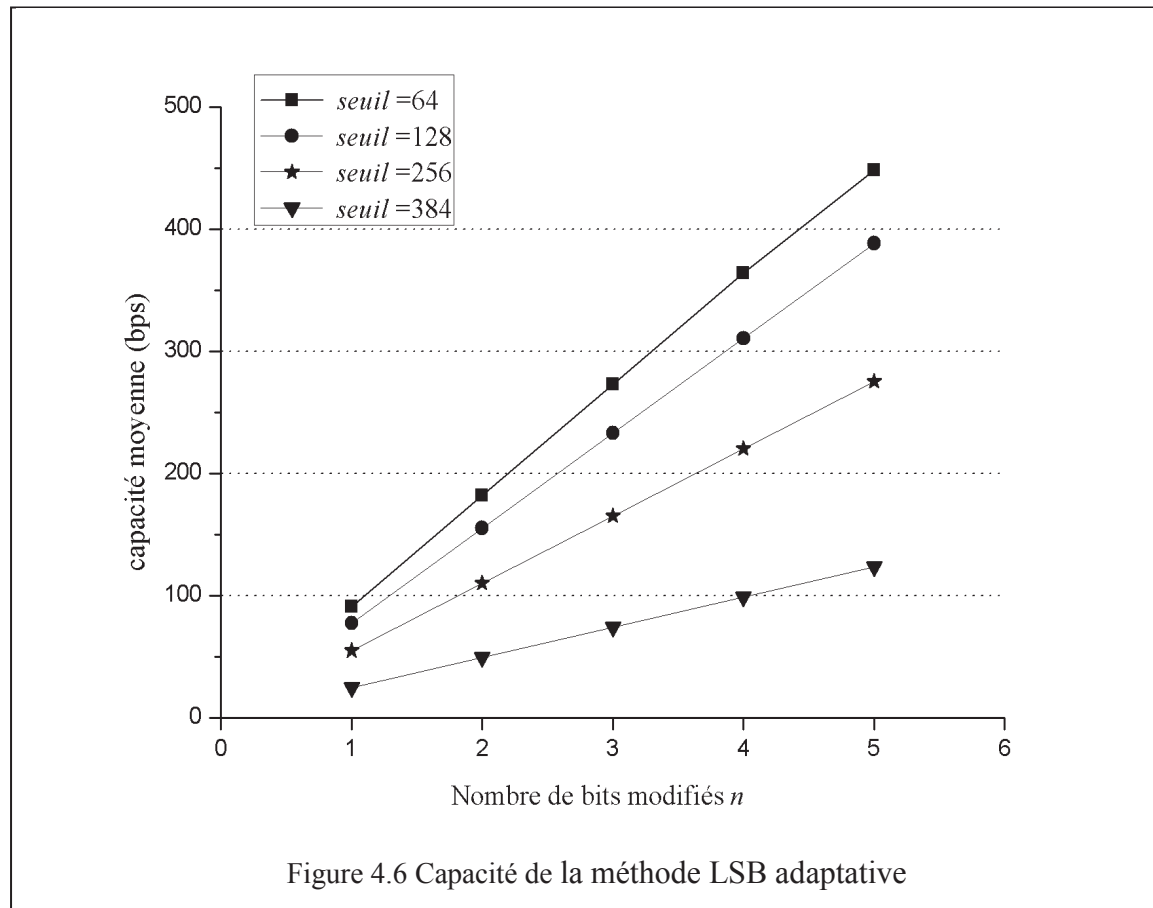
Les performances du codeur G.722.2, avec implémentation de la technique de dissimulation dans le délai tonal sont données sur le tableau 4.11. Nous utilisons la même base de données de parole de 100 phrases. Les performances sont données en termes de WB-PESQ en fonction du nombre de bits dissimulés n et des seuils prédéfinis. Le message secret est généré d'une façon aléatoire.

Tableau 4.11: Performance du G.722.2 avec dissimulation dans le délai tonal

(Seuil, n)	Performances G.722.2 : WB-PESQ moyen
(64, 1)	3.768
(64, 2)	3.595
(64, 3)	2.916
(64, 4)	2.199
(64, 5)	1.706
(128, 1)	3.776
(128, 2)	3.633
(128, 3)	3.077
(128, 4)	2.443
(128, 5)	1.917
(256, 1)	3.789
(256, 2)	3.680
(256, 3)	3.278
(256, 4)	2.828
(256, 5)	2.467
(384, 1)	3.779
(384, 2)	3.713
(384, 3)	3.555
(384, 4)	3.381
(384, 5)	3.181

D'après ces résultats, on remarque que l'imperceptibilité est liée au nombre de bits modifié dans chaque indice présenté dans le tableau par n et au seuil de comparaison. Les meilleurs résultats sont obtenus lorsqu'on utilise un grand seuil et un nombre n petit. Par exemple, l'utilisation d'une paire (seuil, n) égale à (384,1) donne une note WB-PESQ moyen égale à 3.779 qui représente le meilleur résultat possible dans notre cas.

D'autre part, les deux paramètres *seuil* et *n* ont une grande influence sur la capacité moyenne de dissimulation (débit). Dans ce qui suit, nous donnons les résultats de capacité obtenus pour les différentes paires (*seuil*, *n*).



La figure 4.6 montre que l'utilisation d'un petit seuil offre un débit de dissimulation plus grand pour chaque nombre *n*. Par exemple, l'utilisation de seuils égaux à 64 et 384 avec un nombre *n* = 4 donne des débits de dissimulation de 365 bps et de 98 bps respectivement. On remarque que l'utilisation d'une paire (384, 4) donne une capacité moyenne de dissimulation très proche à celle donnée par une paire (64, 1).

D'après l'étude de la capacité moyenne et l'imperceptibilité, on conclut que ces performances dépendent aux paramètres (*seuil*, *n*) sélectionnés. Néanmoins, le plus important dans notre système stéganographique est la relation entre la capacité moyenne et l'imperceptibilité. On remarque que l'utilisation d'un seuil = 64 donne les meilleures performances (capacité moyenne / imperceptibilité). Par exemple si on prend les deux paires (64, 1) et (384, 4), on remarque qu'ils donnent des capacités moyennes de dissimulation très proches (91 bps et 99 bps) avec des WB-PESQ moyens égaux à 3.772 et 3.347

respectivement. Donc l'imperceptibilité de (64, 1) est supérieure avec la même capacité pratiquement.

4.8 Conclusion

Dans ce chapitre, nous avons présenté une évaluation des performances des systèmes stéganographiques que nous avons conçus et implémentés dans un codeur de parole en large bande de type AMR-WB G.722.2. Ce dernier a été utilisé pour générer les signaux de parole publique qui ont servis de supports (hôtes) de dissimulation. Les signaux de parole secrète codés ont été générés par un codeur de parole de bas débit de type MELP fonctionnant à 2.4 kbps.

Nous avons montré que les propriétés des systèmes stéganographiques conçus (la capacité, l'imperceptibilité, ...) dépendent de la technique de dissimulation et de l'étage de quantification du paramètre modifié du codeur publique. Dans la première phase, nous avons mis au point deux systèmes stéganographiques à base de la technique QIM (QIM-DD et QIM-MS-VQ). Ces systèmes ont été appliqués à l'étage de quantification des paramètres ISF du codeur publique G.722.2. L'évaluation des performances a montrée que le système QIM-DD présente une bonne imperceptibilité mais sa capacité de dissimulation est plutôt faible. Par contre, le QIM-MS-VQ peut assurer une grande capacité de dissimulation mais son imperceptibilité est légèrement mauvaise. Par la suite, nous avons conçu un autre système stéganographique à base de la QIM qui a été appliqué à l'étage de quantification de l'excitation du filtre de synthèse (dictionnaire algébrique). Les résultats ont montré que ce système offre une très grande capacité avec une bonne imperceptibilité. Une autre possibilité de dissimulation que nous avons investie est l'utilisation de technique LSB adaptative pour la dissimulation dans l'indice tonal (le pitch) du G.722.2. La capacité de cette technique n'est pas constante et l'imperceptibilité est liée au seuil et au nombre de bits dissimulé dans chaque indice.

Conclusion générale

Dans ce travail, nous avons abordé la problématique de la sécurité de transmission de la parole en étudiant en détails la technique de stéganographie qui est devenue l'un des maillons importants dans les systèmes de transmission sécurisés. En effet, les problèmes de sécurité prennent de plus en plus d'importance dans le monde des télécommunications modernes (Mobile 3GP, VoIP, ... etc).

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 d'abord étudié et testé deux types de codeurs de parole normalisés : le codeur AMR-WB G.722.2 et le codeur MELP de 2.4 kbits/s. Ces deux codeurs ont été utilisés par la suite pour concevoir notre système de codage de la parole sécurisé où le G.722.2 servait pour le codage et la transmission de la parole publique (signal hôte) et le MELP pour coder le signal de parole secrète à dissimuler. Un stégo signal de parole codée contenait donc un signal de parole publique codée par le G.722.2 et un autre signal de parole secrète dissimulé dans la parole publique qui est codé par le MELP. Dans le cadre de ce projet, nous avons mis en œuvre quatre systèmes stéganographiques qui ont été implémentés séparément dans plusieurs étages du codeur de parole publique AMR-WB G.722.2.

Le premier système a été conçu à base de la technique de dissimulation par modulation d'indice de quantification (QIM) avec division de dictionnaires. Ce système, que nous avons appelé QIM-DD, a été appliqué à l'étage de quantification des paramètres ISF du G.722.2 pour dissimuler les messages secrets au niveau de l'indice de quantification. Pour réaliser la division des dictionnaires, nous avons proposé une méthode de division non équilibrée. Cette méthode a été comparée aux deux méthodes de division conventionnelles en l'occurrence la division équilibrée et la division aléatoire. Les résultats de simulation ont montré que la méthode proposée de division non équilibrée de dictionnaire assure de meilleures performances que les deux autres méthodes de division conventionnelles. Le deuxième système stéganographique a été conçu à base d'une autre version de la technique QIM. Il s'agit de la QIM basée sur la modification des schémas de quantificateurs vectoriels. Ce système, que nous avons appelé

QIM-MS-VQ, a été aussi appliqué à l'étage de quantification des paramètres ISF du G.722.2 où une modification du quantificateur S-MSVQ original a été effectuée. Deux techniques de modification ont été appliquées : la dissimulation par remplacement avec tramage (DRT) et la dissimulation par remplacement simple (DRS). Pour tous les débits de dissimulation, l'évaluation comparative des résultats a montrée que le système stéganographique QIM-MS-VQ conçu par la méthode DRT assure de meilleures performances que le QIM-MS-VQ conçu pas la technique DRS. Le troisième système consistait à appliquer la QIM à division de dictionnaires au niveau de la quantification de l'excitation du filtre de synthèse (dictionnaire algébrique) pour dissimuler le message secret. En plus de la dissimulation des messages secrets, les résultats ont montré aussi que cette technique permet de réduire la complexité de codage en diminuant le nombre de combinaisons possibles pour générer le vecteur d'excitation. Le quatrième système stéganographique, qui a été conçu à base de la technique de dissimulation LSB adaptative, a été appliqué au niveau de la quantification du délai tonal (pitch) du G.722.2. Il s'agissait de modifier la valeur du délai tonal en fonction du message à dissimuler.

L'évaluation globale des performances du codeur de parole publique AMR-WB G. 722.2, après implémentation séparée de chaque système stéganographique, nous a permit de conclure sur la sensibilité de plusieurs quantificateurs de paramètres du codeur de parole aux modifications apportées par la dissimulation de la parole secrète. En effet, nous avons constaté que le système QIM-MS-VQ, appliqué au quantificateur des paramètres ISF, offre une grande capacité de dissimulation mais son imperceptibilité est légèrement mauvaise. Par contre, le système QIM-DD assure une bonne imperceptibilité avec une faible capacité de dissimulation lors de son application aux mêmes paramètres ISF. D'autre part, le système stéganographique QIM, appliqué à l'étage de quantification de l'excitation du filtre de synthèse (dictionnaire algébrique), offre une très grande capacité de dissimulation avec une bonne imperceptibilité.

Deux perspectives sont envisageables pour enrichir davantage ce travail. Une première extension concerne l'étude de la robustesse de chaque système contre le bruit du canal de transmission. La deuxième est d'explorer des combinaisons des différentes techniques utilisées dans ce travail. Enfin, nous souhaitons que ce travail servira de plate forme pour de futurs travaux dans le domaine de la communication sécurisée de la parole.

Bibliographie

- [1] C.E Shannon, "Mathematical Theory of Communication", Bell System technical Journal, VOL. 27,N°.3-4, pp. 379-423, 623-656, July, October, 1948.
- [2] Friedrich L. Bauer, "Decrypted Secrets Methods and Maxims of Cryptology ", Fourth, Revised and Extended Edition, Springer-Verlag Berlin Heidelberg,2007.
- [3] Alfred J. Menezes, Paul C. van Oorschotet Scott A. Vanstone," Handbook of Applied Cryptography", CRC Press, 1996.
- [4] Bruce Schneier." Applied Cryptography", Second Edition, John Wiley & Sons, 1996.
- [5] Whitfield Diffie, Martin Hellman,"New Directions in Cryptography",IEEE Transactions on Information Theory, VOL. IT-22 , N. 6 ,1976 .
- [6] Yong Feng Huang, Shanyu Tang,"Steganography in Inactive Frames of VoIP Streams Encoded by Source Codec" IEEE Transactions on Information Forensics and Security, VOL. 6, No.2, JUNE 2011.
- [7] Yun Q. Shi et Sabu Emmanuel "Transactions on Data Hiding and Multimedia Security VI", Springer-Verlag Berlin Heidelberg 2011.
- [8] CHEN Ning, ZHU Jie ,"Multipurpose speech watermarking based on multistage vector quantization of linear prediction coefficients", The Journal of China Universities of Posts and Telecommunications, VOL. 14. Issue 4, pp. 64-69 2007.
- [9] Jack Wiles, Russ Rogers , "Techno Security's Guide to Managing Risks for IT Managers Auditors and Investigators", Elsevier ,Inc ,2007 .
- [10] Benjamin MATHON "Développement de méthodes de tatouage sûres pour le traçage de contenus multimédia", Thèse de Doctorat, L'UNIVERSITÉ DE GRENOBLE et L'UNIVERSITÉ CATHOLIQUE DE LOUVAIN.2011.
- [11] Ingemar J. Cox, Matthew L. Miller, Jeffrey A. Bloom, Jessica Fridrich and TonKalker. "Digital Watermarking and Steganography", deuxième Edition, Elsevier Inc, 2008
- [12] Ali Mohammad Al-Haj "Advanced Techniques in Multimedia Watermarking: Image, Video and Audio Applications", IGI Global, 2010.

- [13] G. J. Simmons." The Prisoners' Problem and the Subliminal Channel. Advances in Cryptology", Proceeding of CRYPTO 83, Plenum Press,1983.
- [14] Rainer Böhme. "Advanced Statistical Steganalysis", Springer, 2010
- [15] Xing He "Signal Processing, Perceptual Coding and Watermarking of Digital Audio: Advanced Technologies and Models", SRS Labs Inc, 2012.
- [16] Bernd Geiser, Peter Vary. "High Rate Data Hiding in ACELP Speech Codecs", the International Conference on Acoustics, Speech, and Signal Processing ICASSP'2008, pp. 4005-4008. Las Vegas, Nevada, U.S.A. March 30-April 4, 2008.
- [17] Lan Ma et al "An Information-Hiding Model for Secure Communication Third International Conference on Intelligent Computing, ICIC'2007, pp. 1305–1314, Qingdao, China, August 21-24, 2007.
- [18] A. M. Kondo "Digital Speech: Coding for Low Bit Rate Communication Systems", Second Edition, John Wiley & Sons, 2004.
- [19] A. Singh. Eraser: "An Exploit-Specific Monitor to Prevent Malicious Communication Channels".The annual conference of the Special Interest Group on Data Communication SIGCOMM'2004, Oregon, USA, September 2004.
- [20] T. Sohn, J. Seo, and J. Moon. "A study on the covert channel detection of TCP/IP header using support vector machine". Information and Communications Security,2003
- [21] J. Fridrich, M. Goljan, and D. Soukal."A New Steganographic Method for Palette-Based Images". 43rd Conference on Coding, Communication, and Control In CCCC'2005, September 2005.
- [22] WU Zhi-jun, GAO Wei, YANG We, "LPC parameters substitution for speech information hiding", The Journal of China Universities of Posts and Telecommunications, VOL. 16, Issue 6,pp 103-112, 2009.
- [23]: F. Djebbar, B. Ayad, h. Hamamand K. abed-Meraim, "A view on latest audio steganography techniques," in Proc. International Conference on Innovations in Information Technology (IIT2011), United Arab Emirates University (UAEU) 2011.

- [24] Hamid R. Nemati "Techniques and Applications for Advanced Information Privacy and Security". IGI Global. 2009.
- [25] Peter Vary et Rainer Martin "Digital Speech Transmission Enhancement, Coding and Error Concealment", John Wiley & Sons, 2006.
- [26] Naofumi AOKI. "A Technique of Lossless Steganography for G.711 Telephony Speech". The Fourth International Conference on Intelligent Information Hiding and Multimedia Signal Processing. IIH-MSP'2008, pp 608 - 611 August 15 - 17, 2008 Harbin, China, 2008.
- [27] A. Shahbazi, E. Soltan mohammadi, A. H. Rezaie, A. Sayadiyan, and S. Mosayyebpour, "Content Dependent Data Hiding on GSM Full Rate Encoded Speech", International Conference on Signal Acquisition and Processing ICSAP'2010, pp. 68-72, Bangalore, India , 9-10 Feb. 2010.
- [28] Ali Shahbazi, Amir Hossein, Rezaie Ramin "MELPe Coded Speech Hiding on Enhanced Full Rate Compressed Domain", Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation AMS'2010, pp. 267-270, 26 - 28 May 2010, Kota Kinabalu, Borneo.
- [29] Akira Nishimura "Data Hiding in Pitch Delay Data of the Adaptive Multi-Rate Narrow-band Speech Codec", Fifth International Conference on Intelligent Information Hiding and Multimedia Signal, IIH-MSP'2009, pp. 483-486 , September 12-14, 2009, Kyoto, Japan , 2009
- [30] W. Bender et al, "Techniques for data hiding", IBM Systems Journal, VOL. 35, pp. 313-336, 1966.
- [31] Wang, F.H., Jain, L.C., Pan, J.S.: "A novel VQ-based watermarking scheme with genetic codebook partition". Journal of Network and Computation Applications (JNCA), VOL. 30, issue 1, pp. 4-23. 2007.
- [32] Rabiner, L.R. and Schafer, R.W., "Digital Processing of Speech Signals", Prentice Hall, pp. 419-420, 1978.
- [33] A. Gersho and R. M. Gray, "Vector Quantization and Signal Compression", Klumer Academic Publishers, 1992.
- [34] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion". In IRE National convention Record, pp. 142-163, 1959.

- [35] WAI C. CHU, "Speech Coding Algorithms Foundation and Evolution of Standardized Coders", Mobile Media Laboratory, DoCoMo, USA Labs, San Jose, California, 2003.
- [36] K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC Parameters at 24bits/frame", IEEE Trans. Speech Audio Processing, VOL. 1, pp.3-14, Jan. 1993.
- [37] S. So and K. K. Paliwal, "Efficient vector quantization of line spectral frequencies using the switched split vector quantiser," in *Proc. Int. Conf. Spoken language Processing*, Jeju, Korea, 2004.
- [38] ITU Recommendation G.722.2, "Wideband coding of speech at around 16 kb/s using Adaptive Multi-rate Wideband (AMR-WB)", 2002.
- [39] Rainer Martin, Ulrich Heute et Christiane Antweiler "Advances in Digital Speech Transmission", John Wiley & Sons, 2008.
- [40] Y. Bistriz and S. Pellerin, "Immitance Spectral Pairs ISP for speech encoding", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. II-9–II-12, 1993.
- [41] Schroeder, M. R.; Atal, B. S."Code-Excited Linear Prediction(CELP): High-quality Speech at Very Low Bit Rates", *Proceedings of the IEEE International Conference on Acoustics,Speech, and Signal Processing ICASSP'1985*, pp. 937–940, Tampa, FL, USA, 1985.
- [42] J-P. ADOUL et al., "Fast CELP coding based on algebraic codes", *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP'1987*, pp. 1957-1960, 1987.
- [43] R. SALAMI et al., "Design and description of CS-ACELP: a toll quality 8 kb/s speech coder", *IEEE Trans. on Speech and Audio Proc.*, VOL. 6, No. 2, pp. 116-130, March 1998
- [44] A. McCree, K. Truong, E. B. George, T. P. Barnwell, et V. Viswanathan, "A 2.4 kbps MELP Coder Candidate for the New U.S. Federal Standard". *Proceedings of IEEE ICASSP'1996*, pp. 200-203, 1996.
- [45] S. Ahmadi et A. S. Spanias, "Low-bit-rate speech coding based on an improved sinusoidal model". *Speech Communication* , 2001.
- [46] Praphul Chandra, Lide "Wi-Fi Telephony Challenges and Solutions for Voice over WLANs", Elsevier Inc, 2007.

- [47] B. Bessette, R. Salami, R. Lefebvre, M. Jelínek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Järvinen, "The adaptive multirate wideband speech codec (AMR-WB)," *IEEE Transactions on Speech and Audio Processing*, VOL. 10, N°. 8, pp. 620-636, 2002.
- [48] Pao-Chi Chang, Hsin-Min Yu, "Dither-like Data Hiding in Multistage Vector Quantization of MELP and G.729 Speech Coding", *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, VOL. 2, pp 1199 – 1203, CALIFORNIA, USA, 2002.
- [49] A.U. Yargıçoğlu, H.G. İlk, "Hidden data transmission in mixed excitation linear prediction coded speech using quantisation index modulation", *IET Information Security*, VOL. 4, Issue. 3, pp. 158–166, 2010.
- [50] CHEN Ning, ZHU Jie, "Multipurpose speech watermarking based on multistage vector quantization of linear prediction coefficients", *The Journal of China Universities of Posts and Telecommunications*, VOL. 14, Issue. 4, pp. 64-69 2007.
- [51] F. H. Wang, J.S. Pan, and L.C. Jain: "Innovations in Digital Watermarking Techniques". Springer-Verlag Berlin Heidelberg, 2009
- [52] Cheng-Hsing Yang et al "Grouping strategies for promoting image quality of watermarking on the basis of vector quantization", *Journal of Visual Communication and Image Representation*. pp. 49–55, 2009.
- [53] Xinshan Zhu, Yanming Chen "Performance Analysis of Dither Modulation against Composite Attacks", *Journal of Information Hiding and Multimedia Signal Processing*, VOL. 3, Number 2, April 2012.
- [54] [Nishimura, 2009] Akira Nishimura "Data Hiding in Pitch Delay Data of the Adaptive Multi-Rate Narrow-band Speech Codec", *Fifth International Conference on Intelligent Information Hiding and Multimedia Signal, IHH-MSP'2009*, pp. 483 – 486, September 12 - 14, 2009 Kyoto, Japan, 2009
- [55] G. Biundo, S. Grassi, M. Ansorge, F. Pellandini and P.A. Farine, "Design techniques for spectral quantization in wideband speech coding", in *Proc. of 3rd COST 276 Workshops on Information and Knowledge Management for Integrated Media Communication*, Budapest, pp. 114-119, Oct. 2002.

- [56] S. wang, A. Sekey, A. Gersho, "An objective measure for predicting subjective quality of speech coders", IEEE Journal Sel. Areas in Communications, VOL. 10, pp. 819-829, Jun 1992.
- [57] ITU-T Recommendation P.862.2, "Wideband Extension to Recommendation P.862 for the Assessment of Wideband Telephone Networks and Speech Codecs", Geneva, 2005.
- [58] Peter Vary, Bernd Geiser "Steganographic Wideband Telephony Using Narrowband Speech Codecs", the 41st Asilomar Conference on Signals, Systems, and Computers in Pacific Grove, CA, USA, 2007.
- [59] J. S. Garofolo and al., "DARPA TIMIT Acoustic-phonetic Continuous Speech Database", NIST, Gaithersburg, Oct 1988.
- [60] M. Boudraa, B. Boudraa & B. Guerin, " Mise en place de phrases arabes phonétiquement équilibrées, XIX éme JEP, Bruxelles, Mai 1992.