

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieure et de la Recherche Scientifique
Université des Sciences et de la Technologie Houari Boumediene
Faculté d'Electronique et d'Informatique



THESE DE DOCTORAT EN SCIENCES

Présentée pour l'obtention du grade de **Docteur**

En : **Electronique**

Spécialité : **Signal et Communication**

Par : **Monsieur Hocine CHAOUCH**

Thème :

Amélioration de la robustesse des codeurs ACELP large bande

Soutenue publiquement, le 20/11/2019, devant le jury composé de :

Mr	Abderrahmane AMROUCHE	Professeur, à l'USTHB	Président
Mme	Fatiha MERAZKA	Professeur, à l'USTHB	Directrice de Thèse
Mr	Abdelhakim DAHIMENE	Professeur, à l'UMBB	Examineur
Mr	Mourad ABBAS	Directeur de Recherche, au CRSTDLA	Examineur
Mr	Djamel ADOU	Maitre de Conférences/A, à l'USTHB	Examineur
Mr	Abdelmadjid MAALI	Maitre de Conférences/A, à l'EMP	Examineur

Résumé

Les systèmes VoIP offrent plusieurs avantages par rapport aux réseaux téléphoniques classiques (PSTN), comme : l'intégration de la voix et des données en un seul réseau, l'optimisation et l'efficacité d'exploitation de la bande passante, la diminution des tarifs et des coûts des communications, la facilité d'administration, de supervision et de maintenance des réseaux IP et l'augmentation et l'amélioration des services.

Lorsqu'un ou plusieurs paquets sont perdus, et aucun effort n'est fait pour les récupérer, la qualité perceptuelle de la parole reçue peut considérablement se détériorer. Pour qu'ils deviennent une alternative crédible aux réseaux téléphoniques traditionnels, les systèmes VoIP doivent offrir la même fiabilité et qualité de la voix. Une bonne qualité de la voix de bout en bout dans les réseaux à commutation de paquets dépend principalement des facteurs dits facteurs de Qualité de Service (QoS).

Ces facteurs ne sont pas garantis par le réseau Internet qui fournit un service d'acheminement des paquets avec meilleur effort « Best-Effort ». Lorsque des paquets de parole sont envoyés en temps réel à travers des réseaux IP, il n'y a aucune garantie de les recevoir dans une manière appropriée, ce qui est dû à la nature « meilleur effort » (best-effort) des réseaux IP. La perte de paquets dégrade la qualité de la voix et se traduit par des ruptures au niveau de la conversation et une impression de hachure de la parole. Il est, par conséquent, indispensable de mettre en place un mécanisme de dissimulation de perte de paquets. Plusieurs algorithmes de masquage des pertes de paquets sont utilisés aussi bien au niveau de l'émetteur qu'au niveau du récepteur. Les standards conçus pour des applications VoIP sont présentés par ITU G.732.1 à double débit et le ITU G.729 à qualité "toll" et le G.722.2 AMR-WB (Adaptive Multi-Rate WideBand) également appelée voix HD (Haute Définition).

L'objectif de notre travail est d'améliorer la robustesse des codeurs ACELP (Algebraic code-excited linear prediction) large bande AMR-WB G.722.2. Ils reposent sur la prédiction à long terme (LTP) ce qui provoque des erreurs de propagation dans les trames de parole et, par conséquent, rend les codeurs ACELP plus sensibles aux pertes de paquets. Ainsi, la qualité de la parole reconstituée sera dégradée dans des conditions de perte de paquets. Nous avons apporté deux contributions. La première consiste à concevoir deux nouveaux modes 5 et 6 (18,25 et 19,85 kbps, respectivement) en se basant sur le codage à description multiples pour le codec G.722.2. Notre méthode est meilleure que la méthode incorporée dans le codec G.722.2. La deuxième contribution concerne la proposition d'un nouveau modèle de Markov de second ordre, simulé les pertes de paquets du réseau IP.

Notre modèle à quatre états est irréductible et apériodique. Il est donc ergodique et convergent. Il est plus général que le modèle de Gilbert à 2 états communément utilisé pour simuler les pertes de paquets. Les tests subjectifs et objectifs (WB-PESQ, EMBSD, MOS et MUSHRA) confirment l'efficacité de notre approche proposée et montrent une amélioration considérable de la qualité de la voix comparée à l'algorithme intégré dans la norme ITU-T G.722.2 pour différents taux de pertes.

Mots-clés :VoIP ; ITU G.722.2 ; Codage par Description Multiple ; Modèle de Chaîne de Markov ; WB-PESQ.

Abstract

VoIP systems offer several advantages over telephone networks (PSTN), such as integrating voice and data into one network, optimization and efficiency of bandwidth, reduction of prices and costs of communications, ease of administration, supervision and maintenance of IP networks and increase and improvement of services.

When one or more packets are lost and no effort is made to recover them, the perceptual quality of the received speech can be considerably deteriorate. To make it a credible alternative to traditional telephone networks VoIP systems must offer the same voice quality. A Good voice quality end-to-end in the packets switched networks depend mainly on factors so-called quality of service (QoS).

These factors are not guaranteed by the Internet that provides a service routing of packages with best effort "Best Effort". When Speech packets are sent in real time across IP networks, there is no guarantee to receive them in an appropriate manner, which is due to the nature of "best effort" IP networks. Packet loss degrades the quality of the voice and results in breaks in the conversation and hatching speech. It is therefore essential to put in place a mechanism concealment of packet loss.

The goal of our work is to improve the robustness of ACELP coders (Algebraic Code-Excited Linear Prediction) broadband AMR-WB G.722.2. They rest on Long-Term Prediction (LTP) which causes errors propagation in frames of speech and, therefore, makes ACELP coders more sensitive to packet losses. Thus, the quality of the reconstituted speech will be degraded in packet loss conditions. We made two contributions. The first consists of designing two new modes 5 and 6 (18.25 and 19.85 kbps, respectively) based on multi-description coding for G.722.2 codec. Our method is better than the method embedded in the G.722.2 codec. The second contribution concerns the proposition of a new second order Markov model to simulate packet loss of the network IP.

Our four states model is irreducible and aperiodic. It is ergodic and it converge. It is more general than the two-state Gilbert model commonly used to simulate packet loss. Subjective and objective tests (WB-PESQ, EMBSD, MOS and MUSHRA) confirm the effectiveness of our proposed approach and show a considerable improvement in the voice quality compared to the algorithm integrated in ITU-T G.722.2 standard for different loss rates.

Keywords : VoIP ; ITU G.722.2 ; Multiple Description Coding ; Markov Model ; WB-PESQ.

Remerciements

Je voudrais exprimer ma reconnaissance et mes remerciements les plus sincères à toutes les personnes qui ont eu l'amabilité de me prodiguer aide, critiques, suggestions et encouragements dans l'accomplissement du présent travail.

Je tiens à remercier en particulier, la directrice de thèse, le professeur Fatiha ME-RAZKA pour sa patience, ses encouragements et ses suggestions, le professeur Philippe MARTHON, de l'ENSEEIH, Toulouse, France, pour son aide et ses conseils précieux et tous les membres de mon jury de thèse pour s'être intéressés à mon travail, pour l'évaluation et la correction de cette thèse.

Je remercie également tous mes collègues et amis qui ont contribué de près ou de loin à ce projet. Je leurs adresse toute ma gratitude et je leurs souhaite beaucoup de succès dans leur carrière. Un remerciement particulier à ma femme et toute la famille et en particulier ma mère, pour leur soutien inconditionnel. Je vous remercie d'avoir toujours été présent et de tous vos encouragements.

Dédicace

Je dédie ce modeste travail, fruit de mes efforts en exprimant ma profonde gratitude à mes proches, et particulièrement :

Ma mère pour son éducation et ses prières.

A mon cher poussin et fils Wail

A la mémoire de mon père.

Ma femme pour son aide et soutien.

Monsieur, HAOULI Mohammed Larbi.

Ma directrice de thèse, professeur Fatiha MERAZKA qui a fourni des efforts énormes par son soutien, ses conseils et ses encouragements.

Monsieur le professeur Philippe MARTHON pour son aide et ses orientations.

A tous mes amis : Djamel, Belkacem, Youcef et tous les collègues de travail.

Table des matières

Résumé-Abstract	v
Remerciements	v
Liste des figures	xii
Liste des tableaux	xiii
Liste des acronymes	xiii
Introduction générale	1
1 CODAGE DE LA PAROLE	5
1.1 Introduction	7
1.2 Le signal parole	7
1.3 Processus de la phonation	7
1.4 Modèle de production de la parole	9
1.5 La prédiction linéaire	11
1.5.1 Méthode d'autocorrélation	12
1.5.2 Méthode de covariance	14
1.5.3 Considérations pratiques	15
1.5.4 Représentation des paramètres de prédiction	16
1.6 Principe de la quantification	17
1.6.1 Quantification vectorielle	17
1.6.2 Quantification Vectorielle par Split	18
1.7 Techniques de codage de la parole	18
1.7.1 Codage de forme d'onde	19
1.7.2 Codage paramétrique	19
1.7.3 Codage hybride	20
1.8 Critères de performance dans le codage de la parole	20

1.8.1	Qualité du signal	20
1.8.2	Débit binaire	20
1.8.3	Complexité	21
1.8.4	Retard de communication	21
1.9	Mesure de la qualité	21
1.9.1	MUSHRA	22
1.9.2	Evaluation de la qualité vocale perçue	22
1.9.3	MOS	23
1.10	Conclusion	23
2	TRANSMISSION DE LA VOIX SUR LES RESEAUX IP	25
2.1	Introduction	27
2.2	La VoIP	27
2.3	Avantages de la VoIP	27
2.4	Transmission de la VoIP	28
2.4.1	Les phases de transmission de la VoIP	28
2.5	Les composants de la VoIP	28
2.5.1	Les protocoles de transmission de la VoIP	29
2.6	Les paramètres affectant la VoIP	30
2.6.1	Qualité du codage	30
2.6.2	Délai de bout en bout	30
2.6.3	Gigue	31
2.6.4	Perte des paquets	31
2.6.5	Largeur de la bande passante	31
2.7	Solutions pour améliorer la qualité de service	31
2.8	Les CoDecs Audio	32
2.8.1	G.711	32
2.8.2	G.729	33
2.8.3	G.723.1	33
2.8.4	G.726	33
2.8.5	G.722	34
2.8.6	Description du CoDec AMR-WB	34
2.8.7	Description du codeur	36
2.8.8	Description du décodeur	37
2.8.9	Pertes de paquets dans un réseau IP	38
2.9	Conclusion	39

3	TECHNIQUES DE RECOUVREMENTS DE PAQUETS	41
3.1	Introduction	42
3.2	Etat de l'art sur les mécanismes de masquage de pertes des paquets	42
3.3	Techniques de masquage des pertes de paquets	44
3.3.1	Les méthodes basées sur l'émetteur	45
3.3.2	Les méthodes basées sur le récepteur	48
3.4	Conclusion	52
4	LES SOLUTIONS PROPOSEES POUR LE G.722.2	53
4.1	Introduction	54
4.2	Techniques MDC pour améliorer la robustesse de codeurs basés sur ACELP AMR-WB	55
4.2.1	Panorama sur l'AMR-WB G.722.2	55
4.2.2	Modélisation de la perte de paquets par la Chaîne de Markov du second ordre	56
4.2.3	Méthode de dissimulation basée sur MDC	58
4.2.4	Simulations et discussions	62
4.3	Amélioration de la robustesse des codeurs basés sur ACELP AMR-WB G.722.2	72
4.3.1	Quantification	72
4.3.2	Méthode proposée	73
4.3.3	Simulations et discussions	73
4.4	Etude comparative sur les méthodes de dissimulation pour améliorer la robustesse des codeurs basés sur ACELP AMR-WB	78
4.4.1	Méthode basée sur MDC	78
4.4.2	Méthode basée sur le FEC	79
4.4.3	Méthode basée sur l'entrelacement	79
4.4.4	Simulations et discussions	79
4.5	Conclusion	83
	Conclusion générale	85

Table des figures

1.1	Système de production de la voix	8
1.2	Forme d'onde et spectre du son voisé.	9
1.3	Forme d'onde et spectre du son non voisé.	9
1.4	Modèle simplifié de production de la parole [6].	10
1.5	Spectre LP avec LSF superposés.	17
1.6	Quantificateur Vectoriel.	18
2.1	Les étapes de transmission de la voix.	27
2.2	Architecture d'un réseau VoIP [8].	29
2.3	Schéma fonctionnel du codeur AMR-WB [73].	36
2.4	Schéma fonctionnel du décodeur AMR-WB [73].	37
3.1	Classification des techniques de réparations basées à l'émetteur.	45
3.2	FEC indépendant du média.	46
3.3	FEC dépendant du média.	46
3.4	Entrelacement sur plusieurs paquets.	47
3.5	Schéma de base de MDC.	49
3.6	Classification des techniques de réparations basées au récepteur [36].	49
3.7	Système de masquage de pertes de paquets proposé par Gunduzhan et al. [52].	50
4.1	Modèle de Markov de second ordre à quatre états (00 : bon, 10 : rupture, 11 : mauvais, 01 : récupéré).	57
4.2	Schéma de base pour le MDC.	58
4.3	Schéma fonctionnel de la transmission VoIP.	62
4.4	Processus de paquetsisation/dépaquetsisation basé sur deux descriptions de MDC : mode 5	62
4.5	Simulation de trames pour un taux de perte de 11,11%.	63
4.6	Simulation de trames pour un taux de perte de 18,68%.	63
4.7	Simulation de trames pour un taux de perte de 32,32%	64

4.8	Simulation de trames pour un taux de perte de 42,42%	64
4.9	Valeurs EMBSD de l'AMR-WB sans et avec paquet perte pour les modes 5 et 6.	65
4.10	Valeurs WB-PESQ de l'AMR-WB sans et avec paquet perte pour les modes 5 et 6.	66
4.11	Valeurs MOS de l'AMR-WB sans et avec perte de paquets pour les modes 5 et 6.	66
4.12	Valeurs EMBSD de l'AMR-WB sans et avec paquet perte pour les modes 5 et 6.	67
4.13	Valeurs WB-PESQ pour différents taux de perte comparant le G.722.2 d'origine (modes 5 et 6) à notre méthode de dissimulation proposée basée sur MDC (nouveaux modes 5 et 6).	68
4.14	Valeurs MOS pour différents taux de perte comparant le G.722.2 d'origine (modes 5 et 6) à notre méthode de dissimulation proposée basée sur MDC (nouveaux modes 5 et 6).	68
4.15	Valeurs MUSHRA pour différents taux de perte comparant le G.722.2 d'origine (modes 5 et 6) à notre méthode de dissimulation proposée basée sur MDC (nouveaux modes 5 et 6).	69
4.16	Portion audiogramme mode 5 pour le taux de perte 42,42%	70
4.17	Portion audiogramme mode 6 pour le taux de perte 42,42%	71
4.18	Modèle de Gilbert à deux états : "0" indique la reception du paquet et "1" indique la perte du paquet.	74
4.19	Résultats avec l'évaluation EMBSD.	75
4.20	Résultats avec l'évaluation PESQ.	75
4.21	Résultats des tests avec EMBSD pour différents taux de perte.	76
4.22	Résultats des tests avec PESQ pour différents taux de perte.	76
4.23	Audiogramme de la parole avec EMBSD pour différents taux de perte. . .	77
4.24	Audiogramme de la parole avec PESQ pour différents taux de perte. . . .	78
4.25	Evolution de la valeur EMBSD vs taux de perte.	80
4.26	Evolution de la valeur PESQ vs taux de perte.	80
4.27	Résultat EMBSD pour différents taux de perte.	81
4.28	Résultat PESQ pour différents taux de perte.	81
4.29	Résultat de comparaison des tests avec EMBSD pour différent taux de perte.	82
4.30	Résultat de comparaison des tests avec PESQ pour différent taux de perte.	83

Liste des tableaux

1.1	Qualité avec le test MUSHRA (UIT-R BS.1534)	22
1.2	Qualité avec le test MOS	23
2.1	Les délais recommandés par IUT-T pour la voix.	31
2.2	Répartition des bits de l'AMR-WB pour chaque trame de 20 ms.	35
4.1	G.722.2. Affectation de bits de l'algorithme de codage AMR-WB modes 5 et 6 pour une trame de 20 ms.	56
4.2	Affectation de bits de l'AMR-WB G.722.2 modifié.	60
4.3	Quantification de l'ISP pour 6,05, 7,25, 12,2 et 12,6 kbps.	61
4.4	Taux de perte simulés avec le modèle de Markov de second ordre proposé. .	65
4.5	Résultats des tests avec EMBSD, WB-PESQ, MOS et MUSHRA.	67
4.6	AMR-WB G.722.2 modifié.	73
4.7	Taux de perte simulés du modèle de Markov à 2 états.	74
4.8	Résultats EMBSD et PESQ	75
4.9	Taux de perte simulés.	79
4.10	Résultats des tests avec EMBSD et PESQ	80

INTRODUCTION GENERALE

Depuis l'apparition en 1969 du premier réseau longue distance WAN (Wide Area Network) ancêtre de l'Internet et reliant à ses débuts quatre universités disséminées sur le territoire des Etats-Unis (projet ARPANET), la nébuleuse Internet a connu une très forte croissance et une mutation. L'Internet a non seulement changé en volume du fait d'une forte augmentation du nombre d'utilisateurs et de l'apparition de nouvelles applications, mais également en nature. En effet, les réseaux étaient à l'origine conçus pour des applications « primaires » comme par exemple, le transfert de fichiers entre ordinateurs, l'accès à des machines distantes, l'échange de courriers électroniques, ou bien encore à la navigation sur le WEB lors de son apparition au début des années 1990. Toutes ces applications ont en commun qu'elles ne requièrent aucune contrainte temporelle. Cependant, elles ne tolèrent pas la perte de paquets.

Même si ces applications constituent toujours de nos jours une bonne partie du trafic de l'Internet, l'apparition de nouvelles applications ayant de fortes contraintes temporelles pose de nouveaux défis à la communauté Internet. Les applications de nature temps réel, telles que la téléphonie sur IP (Internet Protocol), la visioconférence ou d'une manière plus générale les applications multimédias intégrant des services audio et vidéo, ont des exigences strictes concernant les délais de bout en bout et la variation de ces délais (gigue). Elles sont cependant, capables de tolérer un certain taux de perte de paquets.

Néanmoins, l'Internet actuel ne fournit qu'un service dit « au mieux » (best effort service) qui est loin d'être satisfaisant pour ces nouvelles applications. Le service au mieux est un service primaire qui vise à acheminer les paquets entre deux nœuds du réseau sans aucune garantie de qualité de service. Les phénomènes de congestion survenant dans l'Internet rendent difficile le contrôle du taux de perte de paquets et les variations des délais de bout en bout, ce qui ne peut point satisfaire les applications de type temps-réel.

Afin de transmettre des trafics de nature hétérogène (données, audio ou vidéo), et afin de garantir un certain niveau de qualité de service (fiabilité, allocation suffisante de bande passante, faibles délais de bout en bout, peu ou pas de pertes), il a donc fallu développer des algorithmes de codage audio plus tolérants et introduire des mécanismes de contrôle de la qualité de service dans les réseaux de données. Le basculement de différents types de

données sur un même réseau permet en plus, de simplifier son administration.

Nos travaux se sont concentrés sur le codeur G.722.2, retenu parmi les recommandations de l'IUT (International Union of Télécommunications), qui standardise un codage de la parole à 9 débits différents, basé sur des techniques de Prédiction Linéaire de type ACELP (Algebraic Code Excited Linear Prediction). Le choix de ce dernier a été validé car il présentait déjà des caractéristiques particulièrement adaptées aux applications mobiles HD (Haute Définition), offrant des appels vocaux de haute qualité.

Par rapport au CoDec vocal à bande étroite, l'algorithme de compression du G.722.2 double la bande passante de la voix et produit de meilleurs résultats. De plus, la faible complexité d'implantation du G.722.2 en fait un choix intéressant pour la téléphonie sur Internet et les réseaux 3G++.

En effet, le G.722.2 a de faibles exigences en ce qui concerne le débit. Cependant, une éventuelle perte de trames nuit considérablement à son efficacité. Un algorithme de masquage des pertes, basé sur la méthode d'interpolation, lui est incorporé, mais il n'introduit pas une grande amélioration.

L'objectif recherché a été menée pour améliorer les performances du codeur en implémentant des méthodes de masquage (FEC ; Forward Error Correction, entrelacement et MDC ; Multiple Description Coding) des trames perdues plus performantes que celle adoptée par ce dernier pour alléger les effets des pertes de paquets.

Dans les travaux précédents, l'approche MDC a été appliquée aux codeurs à bande étroite à très faibles, faibles et moyens débits et sur des hauts débits AMR-WB pour la comparaison analytique et expérimentale du MDC et du FEC. Notre approche basée sur MDC, vise à améliorer la détérioration de la qualité de la parole causée par la perte de paquets pour le codeur AMR-WB à deux nouveaux débits de haute qualité conçus qui correspondent respectivement aux mode 5 et 6 (18,25 kbps et 19,85) kbps par rapport à la technique intégrée dans la norme ITU-T G.722.2. Notez que, les modes appropriés sont sélectionnés en fonction du débit de transmission requis. En outre, une nouvelle modélisation de la perte de paquets sous forme de chaîne de Markov de second ordre à quatre états est proposée et utilisée à la place du modèle de Gilbert ; chaîne de Markov de premier ordre à deux états.

Le modèle de Markov de second ordre à quatre états proposé pour modéliser les pertes de parquets et tester les performances de notre schéma MDC proposé. A ce titre, nous avons évalué ses performances par des mesures objectives basées sur WB-PESQ (WideBand Perceptuel Evaluation of Speech Quality, PESQ), EMBSD (Enhanced Modified Bark Spectral Distortion), MOS (Mean Opinion Score). Aussi, des tests subjectifs ont également été effectués par la méthodologie MUSHRA. La simulation des pertes des trames a été effectuée pour des différents taux de pertes : 11,11%, 18,68%, 32,32% et 42,42%.

Les résultats de simulation obtenus confirment que la méthode proposée améliore nettement la qualité du signal parole synthétisé.

3- Organisation du manuscrit

Cette thèse est composée de quatre chapitres, à savoir :

- le premier chapitre présentera des notions générales sur le codage de la parole, le processus de la phonation, les techniques de codage et les mesures objectives et subjectives d'évaluation de la qualité du signal parole.
- le deuxième chapitre sera consacré à la transmission de la voix sur les réseaux IP, l'architecture, les caractéristiques et les standards utilisés.
- le troisième chapitre donne une vue générale sur les techniques de dissimulation de pertes de paquets servant à masquer les pertes de paquets.
- le quatrième chapitre présente les solutions proposées pour le CoDec G.722.2 afin d'améliorer la qualité de performance. Les résultats de simulation obtenus et leurs interprétations sont expliqués.

Enfin, une conclusion générale résume et achève notre travail.

Chapitre 1

CODAGE DE LA PAROLE

Sommaire

1.1	Introduction	7
1.2	Le signal parole	7
1.3	Processus de la phonation	7
1.4	Modèle de production de la parole	9
1.5	La prédiction linéaire	11
1.5.1	Méthode d'autocorrélation	12
1.5.2	Méthode de covariance	14
1.5.3	Considérations pratiques	15
1.5.4	Représentation des paramètres de prédiction	16
1.6	Principe de la quantification	17
1.6.1	Quantification vectorielle	17
1.6.2	Quantification Vectorielle par Split	18
1.7	Techniques de codage de la parole	18
1.7.1	Codage de forme d'onde	19
1.7.2	Codage paramétrique	19
1.7.3	Codage hybride	20
1.8	Critères de performance dans le codage de la parole	20
1.8.1	Qualité du signal	20
1.8.2	Débit binaire	20
1.8.3	Complexité	21
1.8.4	Retard de communication	21
1.9	Mesure de la qualité	21
1.9.1	MUSHRA	22
1.9.2	Evaluation de la qualité vocale perçue	22

1.9.3 MOS	23
1.10 Conclusion	23

1.1 Introduction

Le codage de parole permet la réduction de débit de transmission du signal dans des canaux à largeur de bande limitée. La largeur de bande du canal de transmission doit être minimisée tout en préservant la qualité du signal vocal reconstruit. Dans le cas de la transmission de la voix, la réduction du débit limite le nombre des paquets à transmettre sur le réseau IP [1].

Le codage consiste à prélever d'échantillons du signal de parole analogique à un taux d'échantillonnage et une résolution suffisamment grande pour préserver la qualité désirée. Le signal est alors analysé en utilisant des techniques de traitement numérique du signal pour en extraire une représentation compacte sous forme d'un nombre restreint de bits qui seront transmis à travers le canal de transmission. Au niveau du décodeur, les données reçues sont utilisées pour restituer le signal de parole [2].

Dans ce chapitre, nous présenterons le processus de la phonation, les techniques de codage de la parole, les critères de performance et les mesures de la qualité l'évaluation du signal parole.

1.2 Le signal parole

Le signal de parole est généré par l'appareil phonatoire [3]. C'est un organe d'une grande complexité mécanique. Il se compose de deux parties anatomiquement distinctes. Le poumon et le larynx, partie supérieure de la trachée artère, constituent l'essentiel du générateur sonore. Le larynx est un ensemble de muscles et de cartilages mobiles qui entourent une cavité située à la partie supérieure de la trachée.

1.3 Processus de la phonation

Les principaux organes composant l'appareil phonatoire sont : les poumons, la trachée, artère, le pharynx, les cavités buccales et nasales (FIGURE 1.1).

Le larynx a une fonction qui lui est propre : c'est la production des sons, ou "phonation". Les cordes vocales sont en fait deux lèvres symétriques placées en travers du larynx ; ces lèvres peuvent fermer complètement le larynx et, en s'écartant, déterminer une ouverture triangulaire appelée glotte. Pendant la respiration, l'air y passe librement et aussi pendant la phonation des sons sourds ou non voisés. Les sons voisés résultent au contraire d'une vibration périodique des cordes vocales ; des impulsions périodiques de pression sont ainsi appliquées au conduit vocal qui s'étend du pharynx jusqu'aux lèvres.

La voix résulte du fonctionnement simultané des poumons, du larynx et de la cavité de

la bouche et du nez qui modifie sa forme et ses dimensions suivant le son émis et qui, avec la poitrine, jouent le rôle de caisse de résonance. Toutes les voix se ressembleraient si la voix était seulement laryngée. Or, ce sont les modifications de forme et de dimensions que subissent la bouche, le pharynx pendant l'émission de la voix, qui donnent au contraire à celle-ci un timbre qui est particulier à chacun d'entre nous.

On peut remarquer que, d'après ce que nous avons vu précédemment, nous avons deux générateurs de sons, voisés (larynx), et non voisés (nez, bouche), et d'un filtre (le conduit vocal) capable d'amplifier ou d'amortir certains sons. Un son voisé est un signal quasi périodique et un son non voisé peut être considéré comme un bruit blanc.

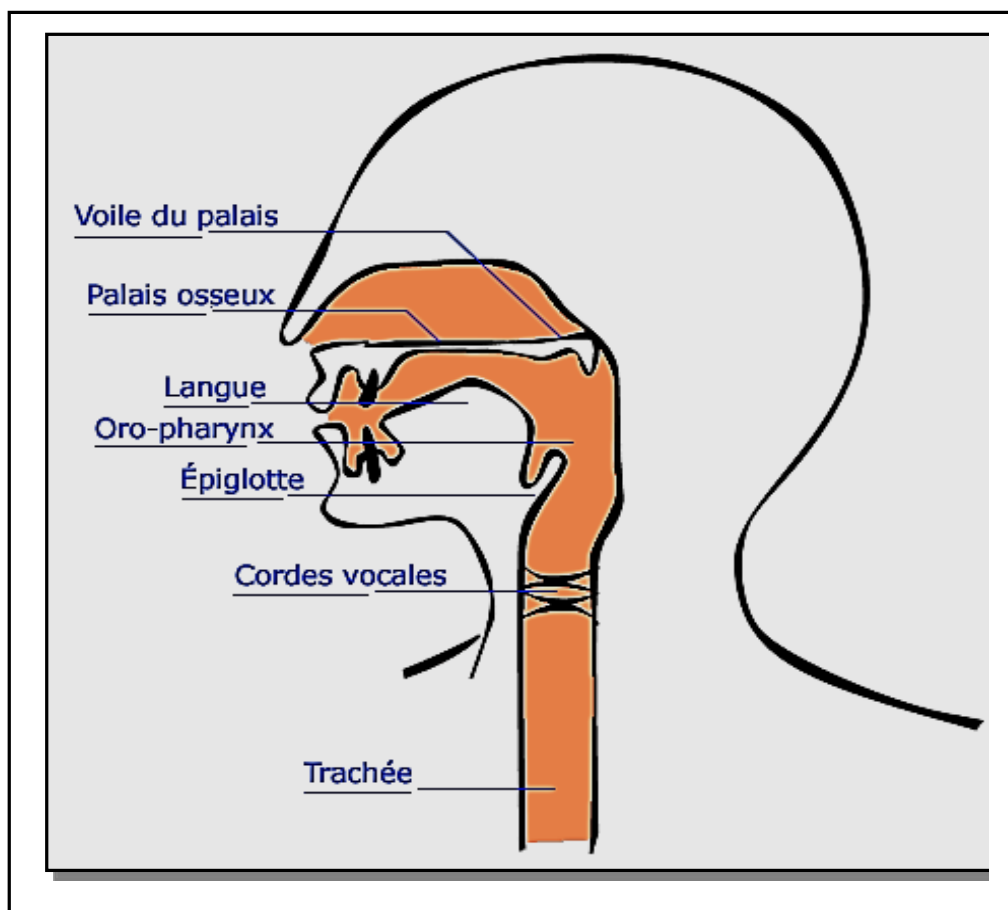


FIGURE 1.1 – Système de production de la voix

L'intensité du son émis est liée à la pression de l'air en amont du larynx ; sa hauteur est fixée par la fréquence de vibration des cordes vocales, appelée fréquence du fondamentale ou pitch. La fréquence du fondamentale peut varier [4] :

- de 80 à 200 Hz pour les hommes ;
- de 150 à 450 Hz pour les femmes ;

— de 200 à 600 Hz pour les enfants.

Un son voisé est un signal quasi périodique dont le spectre est illustré à la FIGURE 1.2. On y observe les raies qui correspondent aux harmoniques du fondamentale F_0 (structure de pitch), l'enveloppe de ces raies présente des maximums appelés formants et qui correspondent aux fréquences propres $F_i (i = 1, 2, 3, \dots)$ du conduit vocal (structure formantique) [5].

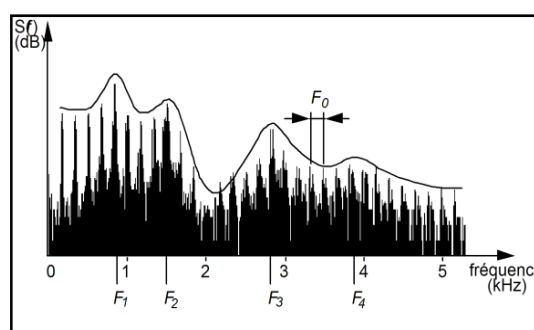


FIGURE 1.2 – Forme d'onde et spectre du son voisé.

Un son non voisé ne présente pas de structure périodique, il peut être considéré comme un bruit blanc filtré par la transmittance de la partie du conduit vocal situé entre la constriction et les lèvres (FIGURE 1.3). son spectre ne présente donc pas de structure de pitch.

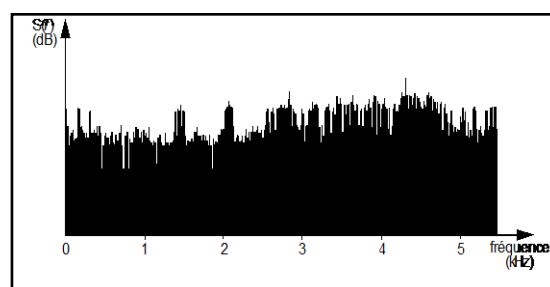


FIGURE 1.3 – Forme d'onde et spectre du son non voisé.

1.4 Modèle de production de la parole

Fant a proposé en 1960 [6] un modèle de production dont nous résumons ici la version numérique. Un signal voisé peut être modélisé par le passage d'un train d'impulsions $u(n)$ à travers un filtre numérique récursif de type tous pôles [7]. On montre que cette modélisation

reste valable dans le cas de sons non voisés, à condition que $u(n)$ soit cette fois un bruit blanc. Le modèle final est illustré à la FIGURE 1.4. Il est souvent appelé modèle auto régressif (AR), parce qu'il correspond dans le domaine temporel à une régression linéaire de la forme :

$$X(n) = \sigma u(n) + \sum_{i=1}^p -a_i X(n-i) \quad (1.1)$$

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

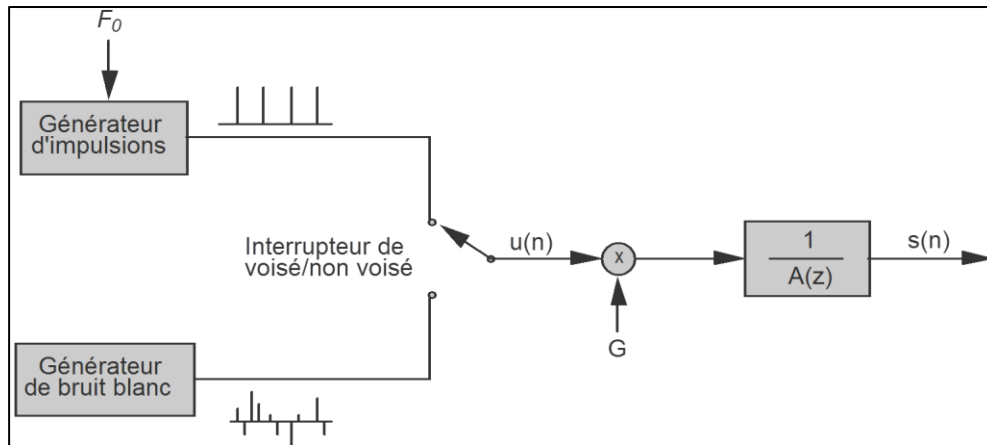


FIGURE 1.4 – Modèle simplifié de production de la parole [6].

Ce modèle comprend :

- un générateur périodique d'impulsions ;
- un générateur de bruit ;
- un interrupteur servant à choisir les sons voisés ou non ;
- un gain proportionnel à la valeur efficace du signal $s[n]$;
- un modèle du conduit vocal.
- un filtre tous pôles $H(z) = 1/A(z)$.

Le problème de l'estimation d'un modèle AR, souvent appelée analyse LP (Linear Prediction) revient à déterminer les coefficients d'un filtre tout pôles dont on connaît le signal de sortie, mais pas l'entrée. Il est par conséquent nécessaire d'adopter un critère, afin de faire un choix parmi l'infinité de solutions possibles. Le critère classiquement utilisé est celui de la minimisation de l'énergie de l'erreur de prédiction.

1.5 La prédiction linéaire

La prédiction linéaire est l'une des méthodes les plus puissantes dans l'analyse du signal de la parole pour l'estimation des paramètres essentiels du signal vocal, son succès est dû au fait qu'elle représente une solution linéaire au problème de l'estimation du modèle de la production de la parole [8], [9].

Le principe fondamental de la prédiction linéaire est qu'un échantillon du signal $S(n)$ peut être modélisé comme la sortie d'un système Auto Régressif à Moyenne Ajustée (ARMA) avec une entrée $u(n)$:

$$s(n) = \sum_{k=1}^p a_k s(n-k) + G \sum_{i=0}^q b_i u(n-i) \quad (1.2)$$

où, a_k et b_i le gain G sont les paramètres du système. L'équation (1.2) prédit la sortie courante en utilisant une combinaison linéaire des sorties antérieures et les entrées courantes et antérieures.

Dans le domaine fréquentiel, la fonction de transfert du modèle de prédiction linéaire de la parole est de la forme :

$$H(z) = \frac{B(z)}{A(z)} = \frac{G(1 + \sum_{i=0}^q b_i z^{-i})}{1 - \sum_{k=0}^p a_k z^{-k}} \quad (1.3)$$

Les racines du dénominateur et numérateur sont, respectivement, les pôles et les zéros du système ou modèle pôle-zéro $H(z)$.

Si $a_k = 0$ pour $1 < k < P$ pour, $H(z)$ devient un modèle tout zéro ou modèle à moyenne ajustée (MA). Si $B_i = 0$ pour $1 < i < q$ $H(z)$ devient un modèle tout pôle ou modèle AR :

$$H(z) = \frac{1}{A(z)} \quad (1.4)$$

Dans l'analyse de la parole, les classes de phonèmes comme les fricatives et les nasales contiennent des vallées spectrales qui correspondent aux zéros dans $H(z)$. Par contre, les voyelles contiennent des résonances qui peuvent être modélisées par le modèle tout-pôle. Pour des raisons de simplicité, ce modèle est préféré pour l'analyse par prédiction linéaire de la parole.

Ainsi, le signal prédit est égal à :

$$s(n) = \sum_{k=1}^p a_k s(n-k) \quad (1.5)$$

et l'erreur de prédiction ou résiduel du signal est la sortie $e(n)$:

$$e(n) = s(n) - \sum_{k=1}^p a_k s(n-k) \quad (1.6)$$

L'ordre p du système est choisi de façon que l'estimation de l'enveloppe spectrale soit adéquate [10]. Une façon de procéder est d'allouer une paire de pôles pour chaque formant présent dans le spectre. On ajoute 2 ou 3 pôles pour approximer les zéros dus aux sons non voisés.

Quand la prédiction linéaire est basée sur les échantillons de parole passés $s(n)$, celle-ci, est dite Prédiction Linéaire Adaptative Progressive (Forward) et dans ce cas les coefficients de prédiction doivent être transmis au décodeur. Si la prédiction linéaire est basée sur les échantillons de parole reconstruits antérieurs, celle-ci, est dite Prédiction Linéaire Adaptative Régressive (Backward). Pour avoir les coefficients du filtre court-terme a_i du processus AR, la méthode classique des moindres carrés peut être utilisée. La variance ou l'énergie, du signal erreur $e(n)$ est minimisée sur une trame de parole. Deux grandes approches sont utilisées pour le codage par prédiction linéaire LP court-terme [11], [12] : la méthode d'auto corrélation et la méthode de covariance [8], [9].

1.5.1 Méthode d'autocorrélation

La méthode d'autocorrélation garantit la stabilité du filtre LP. Les suppositions de cette méthode sont les suivantes :

Le signal est défini pour toutes les valeurs du temps ; il est identiquement nul en dehors d'une séquence de N échantillons, où N est un entier ; ceci équivaut à multiplier le signal de parole par une fenêtre de longueur finie correspondant à N échantillons.

$$\begin{cases} S_f(n) = W(n).S(n), & \text{pour } 0 \leq n \leq N-1 \\ S_f = 0, & \text{ailleurs.} \end{cases} \quad (1.7)$$

La fonction de pondération la plus courante est la fenêtre de Hamming :

$$\begin{cases} W(n) = 0.54 - 0.46\cos(\frac{2\pi n}{N-1}), & \text{pour } 0 \leq n \leq N-1 \\ W(n) = 0, & \text{ailleurs.} \end{cases} \quad (1.8)$$

Chaque échantillon peut être prédit approximativement à partir de p échantillons précédents. Ceci est valable pour toutes les valeurs du temps : $-\infty \leq n \leq +\infty$.

L'erreur quadratique totale entre le signal fenêtre et le modèle (signal prédit) est minimisée sur l'ensemble des échantillons. Après la multiplication du segment parole fenêtre d'analyse, les coefficients d'autocorrélations du segment parole fenêtre sont calculés. La

fonction d'autocorrélation du signal fenêtré est :

$$R(i) = \sum_{n=i}^{N-1} s_f(n)s_f(n-i), \quad 1 \leq i \leq p \quad (1.9)$$

La fonction d'autocorrélation est une fonction paire : $R_i = R_{-i}$. Pour trouver les coefficients du filtre LP, l'énergie du résiduel de prédiction sur l'intervalle fini $0 \leq n \leq N-1$ doit être minimisée :

$$E = \sum_{n=-\infty}^{+\infty} e^2(n) = \sum_{n=-\infty}^{+\infty} (s_f(n) - \sum_{k=1}^p a_k s_f(n-k))^2 \quad (1.10)$$

En annulant les dérivations partielles par rapport aux coefficients du filtre :

$$\frac{\partial E}{\partial a_k} = 0, \quad 1 \leq i \leq p \quad (1.11)$$

On obtient p équations linéaires avec "p" coefficients inconnus :

$$\sum_{k=1}^p a_k \sum_{n=-\infty}^{+\infty} s_f(n-i)s_f(n-k) = \sum_{n=-\infty}^{+\infty} s_f(n-i)s_f(n), \quad 1 \leq i \leq p \quad (1.12)$$

Alors, les équations linéaires peuvent être écrites sous la forme :

$$\sum_{k=1}^p R(|i-k|)a_k = R(i), \quad 1 \leq i \leq p \quad (1.13)$$

Sous la forme matricielle, l'ensemble des équations linéaires est représenté par $R.a = v$ qui peut être ré-écrit en :

$$\begin{bmatrix} R(0) & R(1) & \cdots & R(p-1) \\ R(1) & R(2) & \cdots & R(p-2) \\ R(2) & R(0) & \cdots & \cdots \\ \vdots & \cdots & \cdots & \cdots \\ \vdots & \cdots & \cdots & \cdots \\ R(p-1) & R(p-2) & \cdots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ \vdots \\ R(p) \end{bmatrix} \quad (1.14)$$

La matrice d'autocorrélation $p \times p$ obtenue est une matrice Toeplitz. L'algorithme de Wiener Levinson-Durbin est utilisé pour trouver les coefficients de prédiction minimisant la moyenne quadratique de l'erreur de prédiction.

1.5.2 Méthode de covariance

La méthode d'autocorrélation et de covariance diffèrent dans l'emplacement de la fenêtre d'analyse. Dans la méthode de covariance, le signal erreur est fenêtré au lieu du signal parole de façon que l'énergie à minimiser soit :

$$E = \sum_{n=-\infty}^{+\infty} e_f^2(n) = \sum_{n=-\infty}^{+\infty} e^2(n)w^2(n) \quad (1.15)$$

En annulant les dérivations partielles par rapport aux coefficients du filtre

$$\frac{\partial E}{\partial a_k} = 0 \quad (1.16)$$

pour $1 \leq i \leq p$, on a "p" équations linéaires.

$$\sum_{k=1}^p \phi(i, k) = \phi(i, 0), \quad 1 \leq i \leq p \quad (1.17)$$

où la fonction de covariance $\phi(i, k)$ est définie par :

$$\phi(i, k) = \sum_{n=-\infty}^{+\infty} w^2(n)s(n-1)(s-k) \quad (1.18)$$

Sous la forme matricielle, les p équations deviennent :

$$\phi_a = \Psi$$

$$\begin{bmatrix} \phi(1,1) & \phi(1,2) & \cdots & \phi(1,p) \\ \phi(2,1) & \phi(2,2) & \cdots & \phi(2,p) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(p,1) & \phi(p,2) & \cdots & \phi(p,p) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \Psi(1) \\ \Psi(2) \\ \vdots \\ \Psi(p) \end{bmatrix} \quad (1.19)$$

où :

$$\Psi(i) = \phi(i, 0) \quad (1.20)$$

pour $0 \leq i \leq p$.

La matrice ϕ n'est pas une matrice Toeplitz, elle est symétrique et définie positive. La matrice de covariance peut être décomposée en matrices triangulaires supérieures et

inférieures :

$$\phi = LU \quad (1.21)$$

La décomposition de Cholesky est utilisée pour convertir la matrice de covariance en :

$$\phi = CC^T \quad (1.22)$$

où $C = L$ et $C^T = U$. Le vecteur a est trouvé en résolvant d'abord l'équation :

$$Ly = \Psi \quad (1.23)$$

puis :

$$Ua = y \quad (1.24)$$

1.5.3 Considérations pratiques

Pour mener à bien une analyse LP, il faut pouvoir choisir [13] :

- la fréquence d'échantillonnage f_e .
- la méthode d'analyse et l'algorithme correspondant.
- l'ordre P de l'analyse LP.
- le nombre d'échantillons par tranche N et le décalage entre tranches successives L .

La fréquence d'échantillonnage est de 8 kHz pour les signaux téléphoniques, 10 kHz pour les applications de reconnaissance, et 16 kHz pour les applications de synthèse.

L'ordre de prédiction P , est choisi de façon à ce qu'il permette de bien représenter toute séquence de signal de parole.

Il a été montré que pour donner une représentation satisfaisante des pôles de la fonction de transfert du conduit vocal, la durée de mémorisation du prédicteur linéaire doit être le double du temps mis par l'onde de parole pour se propager de la glotte jusqu'aux lèvres.

Lorsque la fréquence d'échantillonnage f_e (exprimée en échantillon/sec), la période de 1ms correspond à $f_e/1000$ échantillons. A la fréquence d'échantillonnage de 8 kHz, la valeur correspondante de P doit être au moins égale à 8. Elle trouve d'ailleurs une justification expérimentale dans le fait que l'énergie de l'erreur de prédiction diminue rapidement lorsqu'on augmente p à partir de 1, pour tendre vers une asymptote autour de ces valeurs : il devient inutile d'augmenter encore l'ordre, puisqu'on ne prédit rien de plus.

La durée des tranches d'analyse et leur décalage sont souvent fixés à 30 et 10 ms respectivement. Ces valeurs ont été choisies empiriquement ; elles sont liées au caractère quasi-stationnaire du signal de parole.

Enfin, pour compenser les effets de bord, on multiplie en général préalablement chaque tranche d'analyse par une fenêtre de pondération $w(n)$, la plus souvent utilisée est celle de Hamming (1.8).

1.5.4 Représentation des paramètres de prédiction

Les coefficients de prédiction linéaire ne sont pas toujours codés directement, mais sont transformés en un ensemble de paramètres qui ont des propriétés désirables. Plusieurs représentations des coefficients ont été proposées. Les plus populaires actuellement sont les paires de fréquences spectrales LSF (Line Spectral Frequency) [14]. D'autres représentations incluent le coefficient de réflexion, les rapports d'aires logarithmiques LAR (Log Area Ratio), les coefficients spectraux, la réponse impulsionnelle du filtre $LP, \dots, etc.$ Les paramètres LSFs se prêtent mieux à la quantification que les autres représentations du filtre LP [6].

* Paires de raies spectrales (LSP)

Connues aussi sous le nom de fréquences de raies spectrales (LSF), la représentation LSP (Line Spectral Pair) a été introduite par [6], [14].

Soient les polynômes et définis par :

$$\begin{cases} P(z) = A(z) + z^{-(p+1)}A(z^{-1}) \\ Q(z) = A(z) - z^{-(p+1)}A(z^{-1}) \end{cases} \quad (1.25)$$

donc

$$A(z) = \frac{1}{2}(P(z) + Q(z)) \quad (1.26)$$

Les zéros des polynômes $P(z)$ et $Q(z)$ sont appelés les LSP. Ces polynômes ont les propriétés suivantes :

- tous les zéros de $P(z)$ et $Q(z)$ se trouvent sur la cercle unité.
- les zéros $P(z)$ et $Q(z)$ sont entrelacés les uns aux autres, les LSPs sont dans un ordre croissant.

Il a été montré dans [8] que le filtre LP est à phase minimum si et seulement si les LSPs satisfont les deux propriétés citées plus haut, donc la stabilité du filtre de synthèse est facilement vérifiable. De plus, les caractéristiques suivantes ont été relevées de [6], [8], [9] :

- * Comme illustré à la FIGURE 1.5, il y a une relation évidente entre les LSPs et le spectre du filtre LP. Une concentration des LSPs dans une certaine bande de fréquences correspond approximativement à une résonance dans cette bande.

- * Sensibilité spectrale : Un changement d'une LSP cause seulement un changement dans la forme du filtre d'analyse dans une petite gamme de fréquence autour de cette LSP.

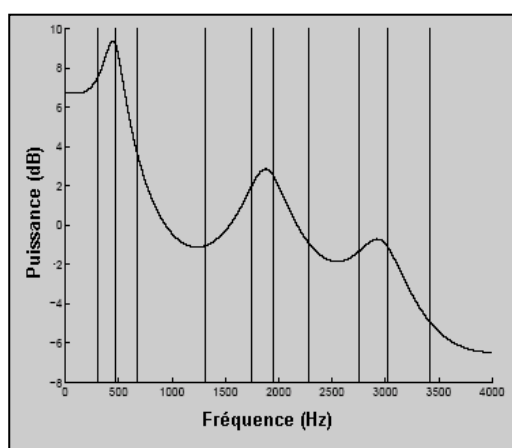


FIGURE 1.5 – Spectre LP avec LSF superposés.

1.6 Principe de la quantification

La quantification est l'opération de numérisation d'une ou plusieurs variables. Pour une seule dimension, le processus est bien connu sous le nom de quantificateur scalaire. Pour plusieurs dimensions, le quantificateur est dit vectoriel.

1.6.1 Quantification vectorielle

Contrairement à la quantification scalaire, le quantificateur vectoriel (QV) s'applique sur des vecteurs. Il fait correspondre à tout vecteur d'entrée x de dimension k décrit comme suit :

$x = (x(0), x(1), \dots, x(k-1))$ un vecteur de même dimension.

$y_i = (y_i(0), y_i(1), \dots, y_i(k-1))$ choisi parmi un ensemble fini B de N vecteurs de reproduction ou mots de code. Dans la littérature, on attribue à cet ensemble B l'appellation de dictionnaire.

Le débit binaire R d'un QV utilisant un dictionnaire B est défini par l'équation suivante :

$$R = \frac{1}{k} \log_2 N \quad (1.27)$$

Le débit représente le nombre de bits par échantillon à coder ($\log_2 N$) bits qui sont nécessaires pour représenter l'indice d'un vecteur de dimension k . Cette formulation du débit permet de faire des études comparatives de quantificateurs opérant sur des vecteurs de dimension différentes. La FIGURE 1.6 illustre le principe de la quantification vectorielle [15].

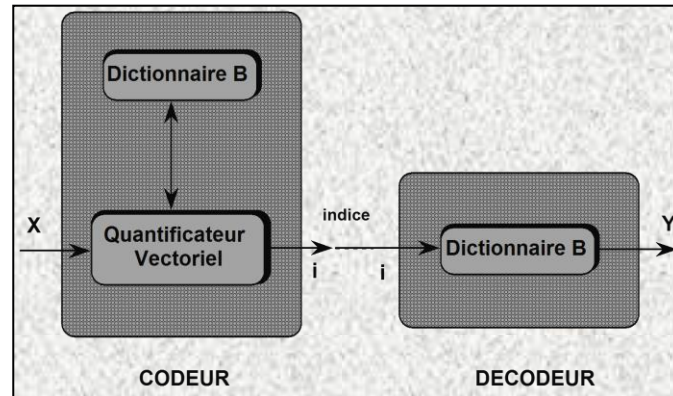


FIGURE 1.6 – Quantificateur Vectoriel.

Dans le codeur, on associe au vecteur d'entrée x un mot de code Y_i du dictionnaire selon le critère du plus proche voisin. Seul l'indice i est transmis au décodeur. On a N indices correspondant aux N mots de code du dictionnaire B . Le même dictionnaire est utilisé au décodage. Le mot de code Y_i , représentant du vecteur x , est retrouvé à partir de l'indice i reçu.

1.6.2 Quantification Vectorielle par Split

Dans la Quantification Vectorielle par Split (SVQ, Split Vector Quantization), le vecteur des paramètres à quantifier x est divisé en plusieurs (k) blocs $[x_0, x_1, \dots, x_{k-1}]$ et chaque bloc est quantifié séparément par un quantificateur vectoriel à recherche exhaustive [8], [15], [16]. On sait que la SVQ réduit la complexité au prix d'une dégradation de performance. D'où il y a un compromis entre la complexité et les performances, qui déterminent le nombre de blocs à prendre dans le split [15].

1.7 Techniques de codage de la parole

Un système de codage de la parole comprend deux parties : le Codeur et le Décodeur (CoDec). Le codeur analyse le signal pour en extraire un nombre réduit de paramètres pertinents qui sont représentés par un nombre restreint de bits pour archivage ou transmission.

Le décodeur utilise ces paramètres pour reconstruire un signal de parole synthétique. Les algorithmes de codage de la parole peuvent être divisés en trois catégories [17], [18] :

- codage de forme d'onde (waveform coding) ;
- codage paramétrique (parametric coding) ;
- codage hybride (hybrid coding).

1.7.1 Codage de forme d'onde

Les codeurs de ce type fonctionnent à hauts débits (supérieurs à 16 kbps). Ils essaient de reconstruire la forme du signal de manière aussi proche que le signal d'origine et basés sur des échantillons du signal d'origine. En théorie, cela signifie que ces codeurs sont indépendants du signal, et peuvent fonctionner avec des signaux non vocaux, de type modem ou fax par exemple. Ces codeurs sont relativement simple à mettre en œuvre, et produisent une qualité acceptable jusqu'à des débits de 16 kbps. En deçà, la qualité du signal reconstruit se dégrade rapidement.

La PCM (Pulse Coded Modulation) est un exemple de cette technique. Dans le cas d'une quantification linéaire, au moins 12 bits par échantillon sont nécessaires pour assurer une bonne qualité, ce qui conduit à un débit de 96 kbps (8000 échantillons de 12 bits). Cependant, la nature de la parole et de l'oreille humaine ne suit pas une échelle linéaire. La plupart des signaux vocaux sont de faible amplitude, et l'oreille humaine n'est pas sensible à l'amplitude absolue d'un son, mais au logarithme de l'amplitude. La représentation binaire (nombre de bits) est alors plus importante pour les signaux de faible amplitude que pour les signaux de forte amplitude. Cela conduit à un débit de 64 kbps.

D'autres techniques de ce type existent comme la DPCM (Differential PCM), ADPCM (Adaptative Differential Pulse Code Modulation) et ADM (Adaptive Delta Modulation).

1.7.2 Codage paramétrique

Les codeurs paramétriques sont destinés à fonctionner pour des débits de quelques dizaines de bits par seconde à 4 Kbps. La performance de ce type de codage ; connu aussi sous le nom de codage de source ou vocodeurs, dépend aux modèles de production de la parole. Ces codeurs sont désignés pour des applications à bas débit et sont destinés à maintenir l'intelligibilité de la parole. La plupart de ces codeurs sont basés sur le codage linéaire prédictif LP.

Le codage LP consiste à synthétiser des échantillons à partir d'un modèle d'un système de production vocal et d'une excitation [19]. Pour la voix humaine, le système de production vocal est l'ensemble poumons-cordes vocales -trachée -gorge -bouche -lèvres. En pratique, on modélise ce système par un ensemble de cylindres de diamètres différents, 10 dans le

cas de LP-10, excités par un signal qui est soit une sinusoïde, soit un bruit blanc. Le choix de la fonction d'excitation (sinusoïde ou bruit blanc) épend des caractéristiques, voisée ou non voisée, du signal [20].

1.7.3 Codage hybride

Les codeurs hybrides sont destinés à fonctionner pour des débits moyens de 4 kbps à 16 kbps, ils combinent les caractéristiques des deux techniques précédentes. Le principal représentant de cette classe est le codage Prédicatif Linéaire à Excitation par séquences Codées CELP (Code Excited Linear Prediction) [8].

Le codage CELP est le plus efficace et le plus utilisé actuellement [21]. C'est est une extension du codage LP. L'idée de base est que la série des impulsions utilisées comme excitation du filtre prédicatif est un ensemble discret et fini. On peut donc essayer d'opérer une quantification vectorielle de cet ensemble, de constituer un dictionnaire. Il suffit alors de transmettre le numéro dans le dictionnaire de la forme d'excitation la plus proche [22], [23].

1.8 Critères de performance dans le codage de la parole

Le problème essentiel dans la compression du signal est de minimiser le débit binaire dans la représentation numérique du signal toute en maintenant des niveaux adéquats de qualité du signal, de complexité d'implantation et de retard de communication [24].

1.8.1 Qualité du signal

La qualité du signal perçu est souvent évaluée sur une échelle de 5 point qui est connue comme étant l'échelle MOS dans les tests de la qualité de la parole : une moyenne à travers un grand nombre d'entrée parole, locuteurs d'écoute évaluant la qualité du signal. Les cinq points de la qualité sont associés à un ensemble d'adjectifs de description : mauvais, médiocre, inacceptable, bon, excellent. On attribue ainsi un seul niveau à chaque signal parole à évaluer durant la procédure d'évaluation subjective.

1.8.2 Débit binaire

On mesure le débit binaire d'une représentation digitale en bits par échantillon, ou bit par seconde (kbps) selon le contexte. Le débit en bits par seconde n'est que le produit de la fréquence d'échantillonnage et le nombre de bits par échantillon. La fréquence d'échantillonnage doit être au moins deux fois plus grande que la largeur de bande du signal

correspondant. Dans le cas de la téléphonie, pour une bande de 3.2 KHz (200-3400 Hz), la fréquence d'échantillonnage de 8 KHz est utilisée.

1.8.3 Complexité

La complexité d'un algorithme de codage est l'effort de calcul exigé pour implanter les processus de l'encodage et du décodage dans les cartes de traitement du signal (hardware), mesuré en termes de la capacité arithmétique (évalué en MIPS) et l'espace mémoire utilisé. D'autres mesures de complexité peuvent être signalées telles que la taille physique de l'encodeur ou du décodeur ou codec, le prix et la consommation de puissance, ce dernier étant un important critère dans un système portable.

1.8.4 Retard de communication

La complexité dans un algorithme de codage est souvent accompagnée d'une augmentation de la durée de traitement dans l'encodeur et le décodeur. Bien que l'évolution des capacités des processeurs de traitement du signal soit un facteur en faveur d'utilisation d'algorithme plus sophistiqué, le besoin de limiter le retard de communication ne doit pas être d'une importance moindre. Selon l'environnement de communication, le retard total permis à un sens peut être aussi bas qu'une milliseconde (comme en réseaux téléphoniques sans annulateur d'écho).

Le retard de codage à un seul sens est défini comme étant le temps écoulé entre l'instant où l'échantillon du signal de parole arrive à l'entrée de l'encodeur et l'instant où le même échantillon apparaît à la sortie du décodeur, moins tout retard introduit par les autres équipements de communication (comme les MODEM) entre la paire encodeur-décodeur et le retard de propagation du signal qui dépend de la distance. En d'autres termes, c'est comme si l'encodeur et le décodeur sont directement connectés par fils sans aucun équipement entre eux. Cette définition fait que le retard de codage dépend seulement de l'algorithme de codage.

Avec cette définition, le retard de codage des codeurs CELP peut être grossièrement déterminé en fonction de la taille de la trame du signal de parole utilisée.

1.9 Mesure de la qualité

La qualité par rapport au débit est l'une des propriétés principales de tout codeur de parole. En fonction des modifications effectuées au signal, de la disponibilité d'un signal de référence et de la nature du résultat voulu, la qualité sonore peut être évaluée de manière objective ou subjective.

Les méthodes les plus utilisées pour tester de la qualité de la parole Les méthodes les plus utilisées sont les suivantes :

1.9.1 MUSHRA

Le test MUSHRA "Multi Stimulus test with Hidden Reference and Anchor" UIT-R BS.1534 [25] est utilisé pour évaluer la qualité de la parole avec une grande précision, en utilisant des signaux de référence (aussi appelés ancrage). Ce test comporte deux étapes ; la première consiste à identifier le signal d'ancrage (la référence cachée) proposé parmi les signaux dégradés et la deuxième consiste à demander aux sujets de donner leurs jugements de qualité vocale sur une échelle continue représentée dans la Table 1.1, par rapport au signal d'ancrage. La première étape est évidente pour des signaux comportant de fortes dégradations, mais elle est aussi très utile dans le cas de faibles dégradations afin de vérifier les capacités des sujets à bien détecter les défauts.

TABLE 1.1 – Qualité avec le test MUSHRA (UIT-R BS.1534)

80-100	Excellente
60-79	Bonne
40-59	Passable
20-39	Médiocre
0-19	Mauvaise

1.9.2 Evaluation de la qualité vocale perçue

La technique basée sur l'évaluation de la qualité vocale perçue (Perceptuel Evaluation of Speech Quality, PESQ) est devenue une recommandation de l'UIT-T (P.862) pour les tests objectif de la qualité de la parole.

C'est un outil qui permet d'évaluer la qualité de la parole communiquée sur réseau téléphonie large-bande ou internet, WB-PESQ est une méthode de test de la qualité vocale applicable aux tests de qualité vocale de bout en bout.

Il prend en compte divers facteurs de perception (tels que la distorsion du codec, les erreurs, la perte de paquets, le retard variable, gigue et filtration) pour tester objectivement la qualité de la parole et fournit une méthode de test quantifiable de la qualité de la parole.

Il a été normalisé par l'Union Internationale des Télécommunications sous le nom de P.862.2 Evaluation de la qualité vocale perçue : méthode objective d'évaluation de la qualité vocale de bout en bout des codecs vocaux et des réseaux téléphoniques à bande étroite.

1.9.3 MOS

Le MOS [26] est une technique de test de la qualité de la parole dans un système de communication. Elle est basée des tests effectués par un certain nombre d'auditeurs sur un système de télécommunications dans différentes conditions, afin de noter sur une échelle la qualité vocale de ce système. Ainsi, une note d'évaluation moyenne peut ainsi être attribuée au réseau. Cette méthode est simple, mais coûteuse, a été utilisée pendant de nombreuses années et clairement définie par la spécification ITU P.800.

TABLE 1.2 – Qualité avec le test MOS

MOS	Qualité
1	Mauvais
2	Médiocre
3	Passable
4	Bon
5	Excellent

1.10 Conclusion

Le codage de la parole consiste à réduire le volume d'informations à transmettre en gardant une qualité de la parole acceptable. La connaissance de la façon de la production de la parole chez l'être humain permet de pouvoir utiliser les propriétés de ce signal pour la réduction du débit de l'information.

Ainsi, la prédiction linéaire essaye d'exploiter la redondance dans le signal et d'extraire des coefficients (paramètres LP) qui caractérisent le comportement du signal. La simplicité de concept et la résolution linéaire dans la prédiction linéaire et ses performances dans le codage de la parole sont sans doute celles qui la rendent la méthode la plus communément admise et la plus largement utilisée dans le codage du signal parole.

Chapitre 2

TRANSMISSION DE LA VOIX SUR LES RESEAUX IP

Sommaire

2.1	Introduction	27
2.2	La VoIP	27
2.3	Avantages de la VoIP	27
2.4	Transmission de la VoIP	28
2.4.1	Les phases de transmission de la VoIP	28
2.5	Les composants de la VoIP	28
2.5.1	Les protocoles de transmission de la VoIP	29
2.6	Les paramètres affectant la VoIP	30
2.6.1	Qualité du codage	30
2.6.2	Délai de bout en bout	30
2.6.3	Gigue	31
2.6.4	Perte des paquets	31
2.6.5	Largeur de la bande passante	31
2.7	Solutions pour améliorer la qualité de service	31
2.8	Les CoDecs Audio	32
2.8.1	G.711	32
2.8.2	G.729	33
2.8.3	G.723.1	33
2.8.4	G.726	33
2.8.5	G.722	34
2.8.6	Description du CoDec AMR-WB	34
2.8.7	Description du codeur	36

2.8.8	Description du décodeur	37
2.8.9	Pertes de paquets dans un réseau IP	38
2.9	Conclusion	39

2.1 Introduction

La voix sur IP (VoIP ou Voice over Internet Protocol) constitue actuellement un domaine de recherche très actif dans le but de minimiser les coûts de communications par l'utilisation d'un seul réseau pour différents services (données, voix, vidéo). L'objectif de ce chapitre est de présenter les différents aspects de cette technologie, ses avantages, ses inconvénients ainsi que ses protocoles de signalisation et de transport. Nous présenterons également les différents codeurs qui travaillent en milieu IP.

2.2 La VoIP

La transmission de la VoIP consiste à découper le signal numérique obtenu par numérisation de la voix en paquets qui seront transmis sur un réseau IP c'est-à-dire sur un réseau de données par opposition à une transmission sur le réseau téléphonique classique ou réseau téléphonique commuté (RTC).

Le protocole IP présente la caractéristique de transporter des données sous forme de paquets. La voix est donc numérisée, compressée, envoyée sous forme de paquets de données sur le réseau conformément au protocole IP. Les données reçues sont décompressées et converties en voix audible (FIGURE 2.1).

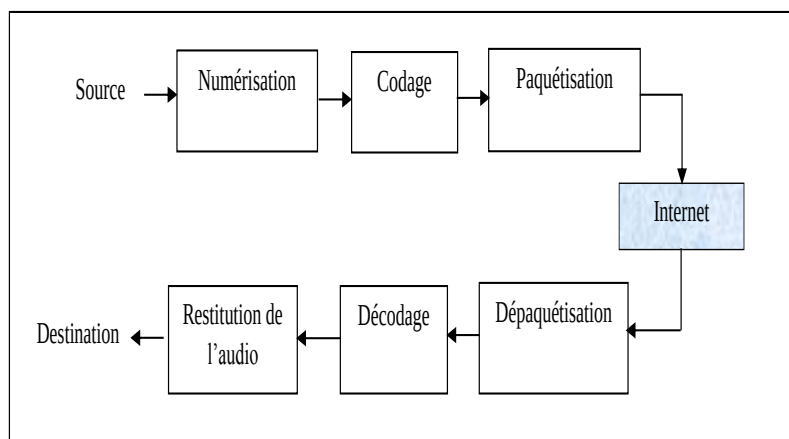


FIGURE 2.1 – Les étapes de transmission de la voix.

2.3 Avantages de la VoIP

La VoIP offre de nombreuses nouvelles possibilités aux opérateurs et utilisateurs qui bénéficient d'un réseau basé sur IP. Les avantages les plus marqués sont les suivants :

- réduction des coûts ;
- flexibilité ;
- interopérabilité multi fournisseurs.

2.4 Transmission de la VoIP

Dans ce paragraphe nous expliquons les différentes phases de la VoIP, ainsi que les différents protocoles utilisés.

2.4.1 Les phases de transmission de la VoIP

Les opérations de transmission de la VoIP et sur le réseau téléphonique commuté sont deux entièrement différentes [8]. Sur le RTC, la voix est transmise sous forme d'un flux constant, alors que sur un réseau IP cette opération passe par les étapes suivantes [14] :

- * numérisation de la voix : C'est la conversion de la voix de la forme analogique vers la forme numérique.
- * annulation du bruit : Cette étape permet de séparer le signal de la voix de l'environnement du bruit.
- * Compression de la voix : La compression des paquets permet de réduire la largeur de la bande passante nécessaire pour la transmission de la voix. Les deux dernières opérations (annulation du bruit, compression de la voix) sont assurées par l'utilisation des CoDecs (Codeurs/Décodeurs), tels que : G.722, G.722.2, G.723, G.728, G.729, G.711,..., etc.
- * Transmission : Après l'annulation du bruit et la compression, la voix est transmise sous forme de paquets IP en utilisant des protocoles de signalisation tels que H.323, SIP.
- * Restitution de l'audio : Cette étape permet la transformation de la voix de la forme numérique vers la forme analogique.

La transmission de la VoIP nécessite des protocoles de transport des paquets telque le protocole RTP/RTCP et des protocoles de signalisation pour l'établissement et la libération des sessions.

2.5 Les composants de la VoIP

Pour transférer de la voix sous forme de données sur le même réseau transportant les e-mails et les pages Web, un nouvel ensemble de composant est rajouté à ceux déjà existants. Parmi ces composants, on peut citer :

- les CoDecs ;
- les TCP/IP (Transmission Control Protocol)/Internet Protocol) ;
- les protocoles VoIP ;
- les serveurs de téléphonie IP et les PBXs (Private Branch eXchange) ;
- les routeurs et les gateways VoIP ;
- les gatekeepers ;
- les téléphones IP et les softphones.

La FIGURE 2.2 montre une architecture générale d'un réseau VoIP.

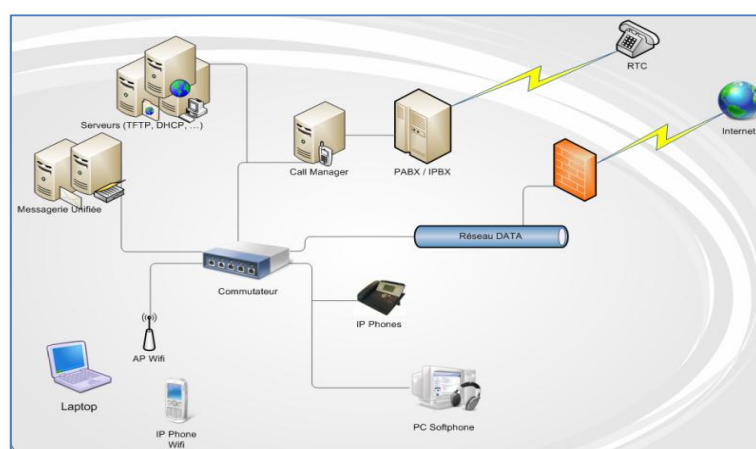


FIGURE 2.2 – Architecture d'un réseau VoIP [8].

2.5.1 Les protocoles de transmission de la VoIP

- * Le protocole RTP/RTCP.
- * Les TCP/IP(Transmission Control Protocol)/Internet Protocol).

L'Internet Engineering Task Force (IETF) a adopté le protocole RTP (Real time Transport Protocol) avec son compagnon RTCP comme protocole de transport et de contrôle adapté aux applications ayant des propriétés temps réel. Le protocole RTP est employé au-dessus du protocole de datagramme comme UDP. Il fonctionne de bout en bout. Il ne permet pas la réservation des ressources et n'assure pas la retransmission automatique des paquets. Le protocole RTP offre à une application les possibilités pour [18] :

- reconstituer la base de temps des flux de données audio, vidéo et temps réel engénéral ;
- détecter rapidement les pertes de paquets et informer la source dans des délais compatibles avec le service ;

- identifier le contenu des données pour en permettre la transmission sécurisée ;
- * Le protocole RTCP.

Le protocole RTCP est basé sur la transmission périodique des messages de contrôle à l'ensemble des participants de la même session. Il ne transporte que les informations de contrôle. Un paquet RTCP est constitué d'un entête fixe similaire à l'entête de RTP suivie par d'autres éléments qui dépendent du paquet RTCP transporté.

2.6 Les paramètres affectant la VoIP

2.6.1 Qualité du codage

L'opération principale des CoDecs est la transformation d'un signal vocale en signal numérique. La première fonction du codage est l'échantillonnage du signal analogique à une certaine fréquence d'échantillonnage avec une certaine précision. La précision est caractérisée par le nombre de bits utilisés pour coder l'amplitude de chaque échantillon analogique [18]. Le choix de la fréquence et du nombre de bits utilisés représente un compromis débit/qualité du signal codé.

Le théorème d'échantillonnage établit qu'un signal analogique peut être restitué à partir des échantillons si la fréquence d'échantillonnage est égale au moins deux fois la largeur de la bande passante du signal original. L'oreille humaine est capable de percevoir une gamme de fréquences de 20 Hz à 20 KHz environ, correspondant à une bande passante de 20 KHz [18]. Généralement plus le taux de compression est élevé par rapport à la référence 64 kbps utilisée dans RTC moins la qualité de la voix est bonne [6].

2.6.2 Délai de bout en bout

Le temps nécessaire pour qu'un paquet traverse le réseau d'un point d'entrée à un point de sortie [6]. Il dépend des trois paramètres suivants :

- type du media de transmission (temps de propagation) par exemple une liaison satellite est beaucoup plus lent qu'une liaison fibre optique ;
- temps de traitement dans les équipements traversés dans le réseau ;
- taille des paquets mesuré depuis l'émission du premier bit par la source jusqu'à la réception du dernier bit du même paquet au niveau du récepteur.

Ce facteur est très important en particulier pour les applications interactives comme la téléphonie IP qui exige un délai de bout en bout inférieur ou égal à 150 ms. Le Tableau 2.1 résume les différentes valeurs recommandées par IUT-T.

TABLE 2.1 – Les délais recommandés par IUT-T pour la voix.

Délais (ms)	Tolérance
Moins de 150	Bonne interactivité
Du 150 à 400	Les usagers peuvent noter quelques pertes dans l'interactivité
Plus de 400	Perte de l'interactivité

2.6.3 Gigue

La gigue est définie par le temps inter arrivé entre les paquets. Elle dépend de :

- type et volume du trafic sur le réseau ;
- type et nombre des équipements sur le réseau.

Généralement, dans les applications de la voix, la gigue doit être inférieure ou égal à 20 ms.

2.6.4 Perte des paquets

La gestion des pertes des paquets est devenue très difficile lorsqu'il s'agit des données multimédias interactives car il est impossible par exemple de réémettre un paquet de voix perdu.

2.6.5 Largeur de la bande passante

Dans les réseaux, la bande passante est partagée entre plusieurs nœuds donc la largeur de la bande disponible est diminuée avec l'augmentation des nœuds s'accédant à la bande.

2.7 Solutions pour améliorer la qualité de service

Le problème de la QoS dans les réseaux constitue un vaste sujet de recherche et a fait l'objet de nombreuses techniques proposées dans la littérature. En dehors des techniques tendant à organiser l'architecture de communication de façon à réduire les délais d'acheminement des messages tout en garantissant une bonne disponibilité du support de transmission (en travaillant notamment sur le déploiement du réseau en utilisant des algorithmes de placement appropriés ou la prise en compte de la QoS au niveau de la couche liaison), les autres peuvent être classées dans les catégories suivantes : le contrôle de congestion, le contrôle de débit, la réservation de ressources, la différenciation de services et l'intégration de la QoS dans les décisions de routage des paquets.

2.8 Les CoDecs Audio

Le transport de la VoIP nécessite au préalable toutes ou une partie des étapes suivantes [26], [27] :

- numérisation : Dans le cas où les signaux téléphoniques à transmettre sont sous forme analogique, ces derniers doivent d'abord être convertis sous forme numérique suivant le format PCM ou MIC (Modulation d'Impulsion Codée, PCM : Pulse Coded Modulation ; en anglais) à 64 Kbps. Si l'interface téléphonique est numérique (accès RNIS, par exemple), cette fonction est omise ;
- codage et compression : Le signal numérique PCM à 64 Kbps est compressé selon l'un des formats de compression / décompression, données sur le Tableau 2.1, puis inséré dans des paquets IP. La fonction du CoDec est le plus souvent réalisée par un DSP. Selon la bande passante à disposition, le signal voix peut également être transporté dans son format original à 64 Kbps ;
- décodage et Décompression : Côté réception, les informations reçues sont décompressées. Il est nécessaire pour cela d'utiliser le même CoDec que pour la compression, puis reconverties dans le format approprié pour le destinataire (PCM 64 Kbps, analogique).

L'objectif de mettre un CoDec est d'obtenir une bonne qualité de voix avec un débit et un délai de compression le plus faibles possibles. Les principaux CoDecs standards de l'IUT sont :

2.8.1 G.711

Le G.711 est une norme de compression audio de l'IUT-T qui définit les codages PCMU et PCMA, s'appuyant sur les lois de quantification A (Europe, Afrique) ou μ (Amérique du Nord, Japon).

- échantillonnage : 8000 Hz pour une bande passante du téléphone entre 300 et 3400 Hz.
- bande passante sur le réseau : 64 ou 56 kbps ;
- type de codage : MIC

Son principe repose sur une grille de quantification non linéaire, permettant de diminuer le rapport signal sur bruit de l'erreur de quantification pour les sons de faible amplitude. Un codage sur 8 bits en G.711 correspond à une quantification sur 13 ou 14 bits en MIC.

La norme G.711 a été révisée en 2000 [28]. Elle est la base de transport de la voix sur le réseau téléphonique commuté (RTC, PSTN en anglais) ou sur le RNIS (ISDN en anglais) et est également utilisée pour le transport de la voix avec peu de compression dans les réseaux IP. Elle est en revanche assez peu utilisée pour faire de la téléphonie directement sur des réseaux étendus comme Internet de par l'utilisation importante de bande passante

[29], [28].

2.8.2 G.729

La recommandation UIT-T G.729 définit un codage de la parole à 8 kbps par prédiction linéaire avec excitation par séquences codées à structure algébrique conjuguée [30]. Le CoDec G.729 est moins consommateur en bande passante que G.711. Il est utilisé pour obtenir une téléphonie de qualité. Le CoDec G.729 est :

- supporté par la plupart des PABX IP ;
- utilisé pour le codage de la partie audio d'une visioconférence ;
- rencontré aussi pour transporter de la VoIP sur les WAN ;
- type de codage : MIC ;
- utilisé préférentiellement par les opérateurs de téléphonie.

2.8.3 G.723.1

Le G.723.1 est normalisé pour deux taux de compression, à savoir ; 5.3 kbit/s et 6.4 kbit/s. Ces deux débits correspondent à deux types de codage différents. La version à 6.4 kbit/s est d'une qualité légèrement supérieure à la version 5.3 kbit/s, la note MOS du 6.4 kbit/s atteint 3.9 contre 3.65 pour le 5.3 kbit/s. Ces deux CoDecs ont une latence algorithmique de 30 ms [31]. Le G.723.1 est surtout utilisé pour la visioconférence et la téléphonie sur IP, que ce soit sur signalisation H.323, H.320 ou SIP.

La compression audio G.723.1 est surpassée aujourd'hui, en termes de qualité acoustique et de taux de compression. D'autres CoDecs ont bénéficié d'avancées en mathématiques et en moyens informatiques qui permettent d'améliorer le compromis qualité/débit. Une des raisons de la pérennité du G.723 reste la contrainte de compatibilité et d'interopérabilité avec les équipements existants.

Un autre aspect militant en sa faveur est sa relative simplicité algorithmique qui lui permet d'entrer dans des terminaux téléphoniques (IP-phone) à très bas coût où la puissance de calcul est limitée.

2.8.4 G.726

C'est une norme de compression audio de l'IUT-T, de type Modulation par Impulsions et Codage Différentiel Adaptatif (MICDA) à 40, 32, 24, 16 kbps [32].

2.8.5 G.722

La norme de codage mondiale G.722 normalisée par l'IUT-T en 1987 permet d'obtenir en VoIP une qualité de voix HD (dite téléphonie large-bande). Cette qualité est obtenue par doublement de la bande de fréquence codée (50-7000 Hz) par rapport à la qualité téléphonique usuelle dite bande étroite (300-3400 Hz) produite par le format de codage G.711 utilisé en téléphonie "classique" sur les réseaux RTC. L'utilisateur bénéficie donc d'une sensation de présence de son interlocuteur, d'un confort d'écoute et d'une intelligibilité fortement améliorés [33].

Cette qualité est obtenue au même débit de 64 kbps que le G.711 "bande étroite" grâce à la technologie MICDA de codage en sous bande : les parties basses fréquences [0,4 kHz] et hautes fréquences [4,8 kHz] du signal sont séparées puis codées et décodées séparément selon un algorithme de codage MIC différentiel adaptatif.

Cette technologie permet en outre au G.722 de fonctionner de manière très flexible aux débits inférieurs de 56 kbps et 48 kbps. Les opérations de codage et de décodage se font échantillon par échantillon et n'induisent donc quasiment aucun retard de codage susceptible de nuire à l'interactivité des conversations et présentent l'avantage d'être peu complexes (environ 10 MIPS) et donc peu coûteuses à implémenter. Afin d'optimiser l'usage de G.722 en VoIP, des systèmes de protection de la qualité audio contre les pertes de paquets ont venus récemment enrichir le standard. En outre, aucun coût de licence n'est associé à ce CoDec. Le format de codage UIT-T G.722 a donc été choisi début 2007 par l'organisme de normalisation européen ETSI comme format obligatoire pour les terminaux DECT de nouvelle génération et haute qualité audio. Certains opérateurs ont ainsi pu lancer avec succès des services de téléphonie IP en voix utilisant le format de codage G.722.

La qualité d'un CoDec est mesurée de façon subjective en laboratoire par une population test de personnes. Ces dernières écoutent tout un ensemble de conversations compressées selon les différents CoDecs à tester et les évaluent qualitativement sur une échelle de 5. 1 pour Insuffisante, 2 pour Pauvre, 3 pour Correcte, 4 pour une Bonne qualité et 5 pour une Excellente qualité de la parole. Le score de l'opinion moyenne de la qualité d'écoute ou le MOS est ensuite calculé pour chaque CoDec.

2.8.6 Description du CoDec AMR-WB

Le CoDec AMR-WB [34] est standardisé dans la recommandation de l'IUT-T G.722.2 [35] et sous la norme 3 GPP TS 26 190 V4 [36]. Le codeur effectue un codage de la parole large bande (50-7000 Hz) échantillonné à 16 kHz à des débits variables, de 6,6 à 23,85 kbps

(neuf modes), sur des trames de signal de 20 ms. L'adaptation du débit et le traitement sur la bande élargie permettent une qualité de voix supérieure. En adaptant son débit en fonction du codage de la source et du canal, le codeur est robuste face aux erreurs de transmission. Il utilise aussi un détecteur d'activité vocale (VAD). L'allocation des bits pour chacun des modes du codec est présentée dans le Tableau 2.2.

TABLE 2.2 – Répartition des bits de l'AMR-WB pour chaque trame de 20 ms.

Mode	Débit (Kbps)	VAD	ISF	Filtrage LPC	Pitch	Code algébrique	Gains	Total par trame (bits)
0	6,6	1	36	0	23	48	24	132
1	8,85	1	46	0	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	447
Du 150 à 400	Les usagers peuvent noter quelques pertes dans l'interactivité	/	/	/	/	/	/	/
		/	/	/	/	/	/	/
		/	/	/	/	/	/	/
Plus de 400	Perte de l'interactivité	/	/	/	/	/	/	/

Le CoDec est interopérable avec des systèmes sans fils GSM (avec l'emploi d'une fonction de transmission discontinue DTX) et de troisième génération (avec un contrôle du débit de la source SCR). Deux bandes de fréquences (50-6400 Hz et 6400-7000 Hz) sont codées séparément pour réduire la complexité et concentrer les bits dans la plage de fréquences la plus significative. L'AMR-WB est basé sur la technologie ACELP [37]. Le codeur fournit des paramètres issus d'un codage par synthèse au décodeur. L'allocation des bits pour chacun des modes du CoDec est présentée dans le Tableau 2.2.

Les FIGURES 2.3 et 2.4 représentent respectivement les schémas blocs du codeur et du décodeur AMR-WB [27].

Le codeur effectue une analyse par synthèse à chaque trame vocale de 20 ms pour en extraire les paramètres de prédiction CELP : les paramètres de codage LPC qui correspondent aux coefficients du filtre de prédiction linéaire, ainsi que les paramètres de l'excitation, qui incluent les indices et les gains des dictionnaires adaptatifs et fixes à la fréquence d'échantillonnage de 12,8 KHz (le signal d'entrée ayant subi un sous échantillonnage de 16 KHz à 12,8 KHz) [38], [39]. Les paramètres de prédiction linéaire sont calculés une seule fois par trame, tandis que les paramètres de l'excitation sont calculés en quatre sous-frames de 5

ms chacune. En plus de ces paramètres, des indices de gain dans les hautes fréquences sont calculés dans le mode 23,85 kbps. Les paramètres encodés sont transmis sur un canal avec ou sans erreurs. A la réception, le décodeur restitue un signal vocal synthétisé par filtrage du signal d'excitation reconstruit au moyen du filtre de synthèse LP [40]. Par la suite, le signal vocal reconstruit subit un post-traitement et un sur-échantillonnage. Enfin, le signal dans la bande des hautes fréquences est produit dans la bande de 6 à 7 KHz.

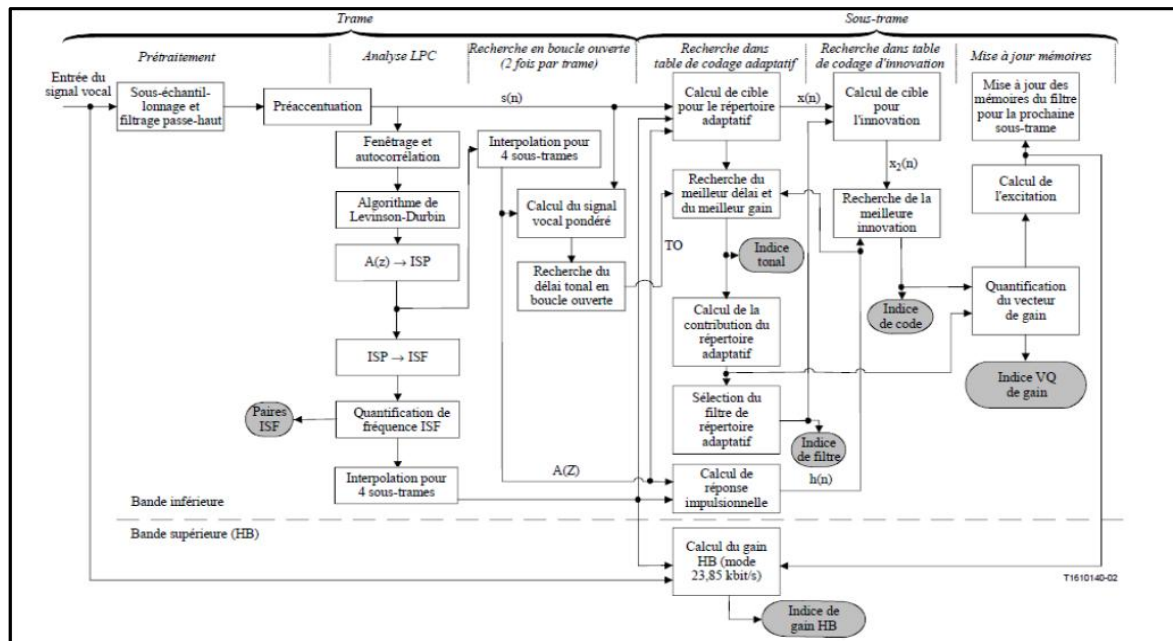


FIGURE 2.3 – Schéma fonctionnel du codeur AMR-WB [73].

2.8.7 Description du codeur

Le codeur effectue l'analyse des paramètres de codage LPC et du dictionnaire fixe à une fréquence d'échantillonnage de 12,8 kHz [19]. Pour ce faire, le signal d'entrée subit une décimation de 16 kHz à 12,8 kHz. Le délai introduit par le filtrage est compensé par l'insertion de zéros à la fin du vecteur d'entrée. Le signal décimé subit tout d'abord, un filtrage passe-haut, contre des composantes parasites basses fréquences en coupant à 50 Hz, puis une préaccentuation à l'aide d'un filtre passe-haut [41]. L'analyse de prédiction linéaire court-terme est effectuée à chaque trame vocale de 20 ms au moyen d'une autocorrélation sur le signal fenêtré. Les autocorrélations des signaux fenêtrés sont converties en coefficients de prédiction linéaire d'ordre 16 au moyen de l'algorithme de Levinson-Durbin. Ces coefficients LP sont ensuite transformés en paires d'immitances spectrales ISP (Immitance Spectrum

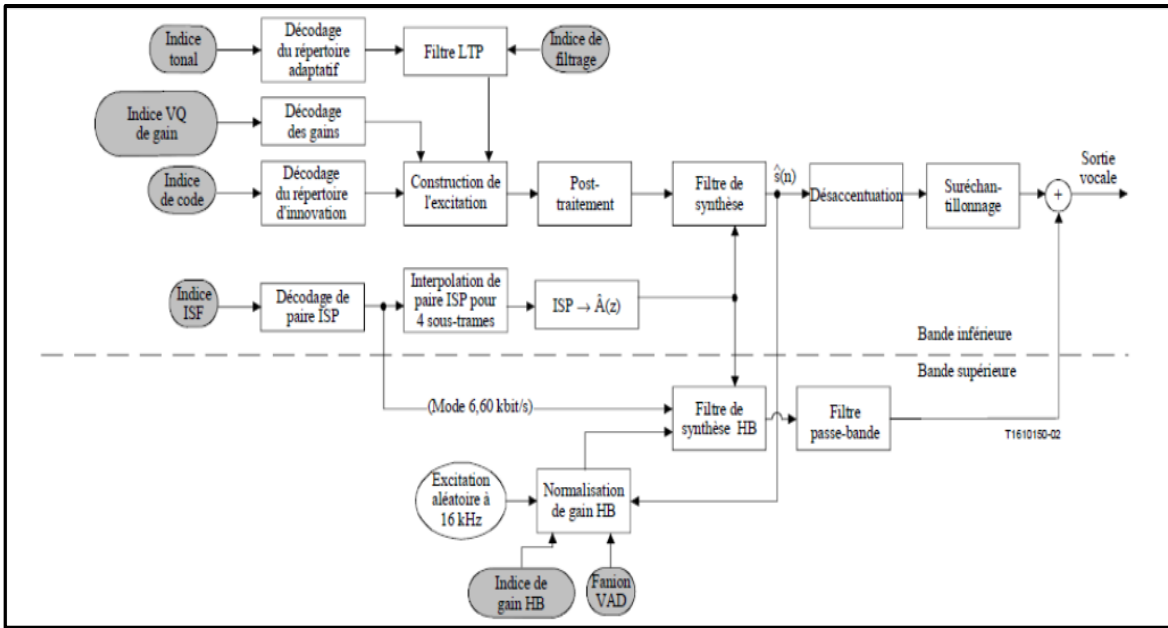


FIGURE 2.4 – Schéma fonctionnel du décodeur AMR-WB [73].

Pairs [18]), qui deviennent les coefficients d'ISF (Immitance Spectrum Frequency) dans le domaine des fréquences. Les ISF sont quantifiés et interpolés pour quatre sous-trames. Les ISFs subissent une quantification vectorielle séparée sur deux étages SVQ et MSVQ. Les ISF sont quantifiés sur 46 bits dans les modes 8,85 à 23,85 kbps (avec 8 bits pour les deux sous vecteurs du premier étage, puis, 6 bits, 7 bits, 7 bits, 5 bits et 5 bits pour les six sous vecteurs du second étage) et sur 36 bits dans le mode 6,6 kbps (8 bits pour les deux sous vecteurs du premier étage puis, 7 bits, 7 bits et 6 bits pour les trois sous-vecteurs du second étage). Les filtres interpolés quantifiés et non quantifiés sont reconvertis en coefficient LP pour permettre la construction de filtres de synthèse (pour la recherche dans les dictionnaires adaptatifs et fixes) et de pondération (pour la recherche du pitch en boucle ouverte) dans chaque sous-trame de 5 ms.

2.8.8 Description du décodeur

Le décodage des paramètres LP est effectué. Les indices reçus de la quantification des ISP permettent de reconstruire le vecteur de paires ISP. L'interpolation permet l'obtention de quatre vecteurs ISP interpolés pour chacune des quatre sous-trames. Pour chacune d'elles, le vecteur ISP interpolé est converti en coefficients LP pour le filtre de synthèse. Par la suite, le même processus est répété dans chaque sous-trame. Les étapes successives

sont : le décodage du vecteur adaptatif, le décodage du vecteur innovateur, le décodage des gains adaptatifs et innovateurs, la reconstruction du signal vocal, le traitement anti-dispersion, le renforcement du bruit, le renforcement du pitch et le post-traitement.

- Le décodage du vecteur adaptatif : l'indice du pitch reçu permet de trouver les parties entières et fractionnaires du pitch. Le vecteur adaptatif $v(n)$ est trouvé par interpolation de l'excitation antérieure (au pitch actuel) avec le filtre de synthèse.
- Le décodage du vecteur innovateur : l'indice reçu permet de trouver les positions et les amplitudes des impulsions dans le vecteur d'excitation $c(n)$.
- Le décodage des gains : l'indice qui est reçu au décodeur permet de trouver le facteur de correction du gain du dictionnaire fixe. Le gain prédit est retrouvé au moyen de l'énergie prédite pour chaque sous-trame, sa multiplication avec le facteur de correction permet de trouver le gain quantifié du dictionnaire fixe.
- La reconstruction du signal : le signal reconstruit est de la forme : $u(n) = gpv(n) + gcc(n)$, avec $v(n)$ le vecteur adaptatif, $c(n)$ le vecteur fixe ainsi que leurs gains respectifs.
- Le traitement anti-dispersion : cette procédure est appliquée au vecteur fixe pour réduire les artefacts issus de la dispersion du vecteur de code fixe avec seulement quelques échantillons non nuls par sous-trame. Le traitement effectue une convolution circulaire du vecteur fixe avec une réponse impulsionnelle parmi trois réponses mémorisées. Le choix de l'une des trois est fonction de la valeur du gain innovateur quantifié et selon le cas, elle apportera une modification forte, moyenne ou nulle.

2.8.9 Pertes de paquets dans un réseau IP

Les codeurs ACELP sont souvent utilisés dans les systèmes VoIP car ils offrent une bonne qualité vocale en l'absence de perte de paquets. Cependant, leur dépendance à la prédiction à long terme provoque la propagation des erreurs à travers les trames vocales, et donc rendent les codeurs ACELP plus sensibles aux pertes de paquets. Les pertes de paquet se traduisent par des ruptures au niveau de la conversation avec des dégradations du signal de la parole [28]. De plus, ces détériorations affectent non seulement les trames perdues mais même les trames qui sont convenablement reçues. Bien que, la QoS des réseaux IP est considérablement améliorée au cours de ces dernières années, mais malheureusement la VoIP ne fournit pas encore une qualité de transmission de voix équivalente à celle offerte par le réseau public commuté traditionnel [30]. En fait, le trafic vocal est très sensible aux problèmes de réseaux, tels que la congestion (gigue), le retard (tampon) et la perte de paquets [31]. Par conséquent, les paquets manquants doivent être masqués au niveau du décodeur à l'aide d'un algorithme de masquage de perte de paquets. Pour cela, des techniques de masquage des pertes paquets ont été élaborées et développées afin d'assurer

une qualité de service d'une transmission de parole sur IP la meilleur possible. Ainsi, nous pouvons distinguer deux grandes catégories de méthodes de masquage de paquets qui s'appliquent, soit à l'émission ou à la réception. Et d'ailleurs, le chapitre suivant traitera en détail les différentes techniques de masquage.

2.9 Conclusion

Les réseaux IP se montrent robustes et flexibles aux pertes se produisant sur des données de type fax, images, bande de son. En effet, les données perdues peuvent être réémises à partir de la source et réintroduites à la destination sans que l'intégralité ou que la qualité des fichiers reçus soit altérée. Cependant, dans des applications temps réel, telle que la VoIP, la procédure de réémission risque d'être impraticable tant les exigences temporelles sont strictes.

Le codage du signal, et plus particulièrement le traitement de la parole, est nécessaire pour effectuer des communications sur des réseaux quelqu'ils soient. Dans le cas de transmissions sur un réseau sans fil, des codeurs bas, moyens et hauts débits sont un choix judicieux. En général, un encodeur sépare le signal vocal en trames afin d'être traitées, puis transmises au décodeur.

Dans ce chapitre, nous avons décrit le principe de la VoIP, les codeurs couramment utilisés. Le CoDec AMR-WB est plus détaillé que les autres, puisqu'il a été choisi pour encoder la parole à transmettre sur un réseau IP. Ces réseaux ont des prédispositions aux erreurs. Ils introduisent des délais et des pertes de paquets en raison de l'engorgement et de la mobilité des nœuds. Cependant, les paquets arrivant trop tard sont considérés comme perdus. A cet effet, il existe des méthodes de masquages permettant atténuer les effets des pertes. Dans le prochain chapitre, nous aborderons certaines techniques de recouvrement de paquets.

Chapitre 3

TECHNIQUES DE RECOUVREMENTS DE PAQUETS

Sommaire

3.1	Introduction	42
3.2	Etat de l'art sur les mécanismes de masquage de pertes des paquets	42
3.3	Techniques de masquage des pertes de paquets	44
3.3.1	Les méthodes basées sur l'émetteur	45
3.3.2	Les méthodes basées sur le récepteur	48
3.4	Conclusion	52

3.1 Introduction

Dans un milieu IP, il arrive qu'une ou plusieurs trames soient perdues en raison de la congestion du réseau, la corruption de données, ou arrivées suffisamment en retard.

Le masquage de pertes de paquets est un domaine de recherche avec beaucoup d'implications industrielles.

Les causes de pertes de paquets sont multiples telles que le désordre à l'arrivée des paquets, la congestion dans les nœuds du réseau, les erreurs de transmission et le retard à l'arrivée des paquets. Pour pallier l'effet de ces pertes sur la qualité du signal parole, des techniques de masquage des pertes de paquets ou PLC (Packet Loss Concealment) sont introduits au niveau du décodeur afin de reconstruire les trames perdues [42].

Les techniques de masquage de pertes des paquets peuvent être classées en deux catégories principales, à savoir : les techniques basées sur le récepteur et les techniques basées sur l'émetteur. Les techniques basées sur le récepteur sont les techniques qui ne nécessitent pas d'informations supplémentaires et n'impliquent pas l'intervention de l'expéditeur. Par contre, les techniques basées sur l'expéditeur sont toutes les techniques qui nécessitent l'intervention de l'expéditeur [36].

3.2 Etat de l'art sur les mécanismes de masquage de pertes des paquets

Lorsqu'un signal de parole est transmis par paquets, il arrive que des paquets soient perdus ou qu'ils arrivent avec un grand délai pour être traités. Il en résulte un manque d'informations pour faire la reconstruction, ce qui entraîne une dégradation du signal. Cette dégradation peut être locale, c'est-à-dire n'affecte qu'une seule trame, ce qui est le cas pour un codeur non-prédictif.

Par contre, dans le cas où le codeur est prédictif, la dégradation affecte la trame perdue et se propage aux trames suivantes. Pour augmenter la qualité du signal au décodeur, il est possible de combler les trames manquantes de diverses façons.

Les différents types de masquage se classent en deux grandes catégories ; le masquage basé au codeur et celui basé au décodeur. Plusieurs techniques sont exposées en revue par Wah et al. [43] et Perkins [36].

Le masquage basé au codeur peut être du type actif ou passif. Le type actif est simplement une retransmission permettant souvent de recevoir le paquet perdu, mais nécessite un long temps de traitement. Une méthode efficace pour limiter les dégâts inévitables aux pertes des paquets est la technique d'entrelacement ou interleaving qui consiste à mélanger les trames et à former des paquets hétérogènes. Une séquence de 16 trames réparties en 4

paquets peut être envoyée selon un ordre chronologique ([1-2-3-4],[5-6-7-8],[9-10-11-12],[13-14-15-16]), mais aussi selon un ordre ordonné d'une trame sur 4 ([1-5-9-13],[2-6-10-14],[3-7-11-15],[4-8-12-16]).

Aussi, si le troisième paquet est perdu, les trames perdues ne seront pas successives et il sera plus facile de les masquer au niveau du décodeur. L'inconvénient de cette technique est d'imposer un certain délai entre l'émetteur et le récepteur, ce qui engendre l'augmentation de la latence.

Une autre technique de masquage basée au codeur, selon Perkins [36], est celle de la Correction d'Erreur en Direct ou FEC. Cette méthode sert à réparer le signal lorsque des trames sont disparues. Le FEC se divise en deux méthodes en vertu de la dépendance ou l'indépendance entre le masquage et les caractéristiques du signal. La méthode indépendante du signal utilise n paquets pour créer un paquet supplémentaire qui sera lui aussi envoyé au décodeur. Parmi les différentes techniques rencontrées, l'opération ou-exclusif (XOR) est fréquemment employée. L'opérateur est appliqué à plusieurs paquets pour générer un paquet de parité. Lorsqu'un seul des paquets est perdu, il est possible de reconstruire le paquet perdu grâce au paquet de parité. La méthode dépendante du signal quant à elle, consiste à envoyer certaines caractéristiques du signal avec le paquet suivant.

Certainement, ces masquages permettent de récupérer le signal, ils entraînent des délais additionnels et une augmentation de la bande passante nécessaire selon la quantité d'informations supplémentaires transmises. D'autres techniques peuvent être appliquées au codeur pour diminuer l'impact d'une trame corrompue. Kabal [44] propose d'ajouter de la redondance dans l'information transmise. La redondance est basée sur les paramètres de l'excitation de la trame suivante selon l'importance de la trame.

Cette importance est caractérisée par une différence importante de l'énergie entre les paramètres d'excitation des deux trames (courante et subséquente). La redondance est incluse dans les environs de 11% des trames envoyées. Le fait de ne pas envoyer de redondance à toutes les trames permet de limiter l'augmentation de débit introduit par cette technique.

Andersen [45] propose un prédicteur long terme indépendant d'une trame à l'autre. Ainsi, suite à une trame perdue, il n'y a pas de désynchronisation entre le codeur et le décodeur, ce qui limite la propagation d'erreurs.

Eksler [46] remplace le prédicteur long terme par un dictionnaire d'impulsions glottales non-prédictif pour les trames qui suivent les transitoires qui sont plus sensibles à la perte parce que la construction de la partie périodique du signal ne peut être basée sur la dernière bonne trame reçue qui est non voisée. Le remplacement du prédicteur long terme par un dictionnaire d'impulsions glottales permet de limiter plus rapidement la propagation d'erreurs lorsque la trame perdue est une transitoire.

Au niveau du décodeur, le masquage se fait par insertion. Perkins [36] et Bhute [47] résument plusieurs approches possibles. Il est possible d'insérer des zéros, du bruit ou répéter la trame précédente.

Ces techniques sont faciles à implémenter, mais ne sont pas très efficaces. Il est aussi possible de remplacer la trame perdue par une combinaison de la trame précédente et de la trame suivante. La première partie de la trame perdue est remplacée par la dernière moitié de la trame précédente et la deuxième moitié par la première partie de la trame suivante. Lorsqu'une trame perdue est située entre deux trames voisées, la reconstruction peut se faire en fonction du pitch de la trame précédente et celui de la trame suivante. De cette manière, la reconstruction produit un pitch cohérent. Lorsque la trame perdue est entre deux trames non voisées, il est possible de répéter la trame précédente.

Lorsqu'une trame est perdue, la méthode de masquage utilisée estime la trame manquante. Avec un codeur prédictif, une estimation du filtre de prédiction à long terme et du filtre de synthèse est faite par le masquage. Cette estimation peut faire en sorte qu'il y a propagation de l'erreur sur plusieurs trames puisque l'état des filtres du codeur et du décodeur n'est plus identique. Pour réduire la propagation de l'erreur lorsqu'une trame est en retard, Gournay et al.[48] propose d'utiliser les trames qui arrivent en retard pour mettre à jour l'état des filtres de prédiction. Le masquage de perte des paquets à l'aide des modèles de Markov cachés (Hidden Markov Model, HMM) est utilisé en reconnaissance de la parole est largement répandue [49]. A cet effet, ces modèles peuvent être considérés comme une modélisation du signal vocal.

Aussi, la technique de masquage basée sur MDC est largement utilisée. Dans [25], Orozco et al. ont fait une comparaison entre le codeur CELP et le codeur sinusoidal. Dans leur travaux, les auteurs ont appliqué au codeur sinusoidal présenté, un schéma de paquets basé sur MDC. Cette méthode de masquage a été appliquée aux codeurs de très faibles débits.

3.3 Techniques de masquage des pertes de paquets

Les pertes de paquets survenant dans les réseaux de transmission se traduisent par des pertes d'une ou de plusieurs de trames dans le signal de parole perçu comme des coupures. Pour pallier à l'effet de ces pertes sur la qualité vocale, des mécanismes de masquage des pertes de paquets sont introduits au niveau du décodeur afin de reconstruire les trames perdues. Ces techniques sont classées en deux principales catégories à savoir : les techniques basées sur l'émetteur et le récepteur. Les techniques basées sur l'émetteur nécessitent l'intervention de l'expéditeur. Par contre, les techniques basées sur le récepteur ne nécessitent

pas d'informations secondaires et n'impliquent pas l'intervention de l'expéditeur [35], [36].

3.3.1 Les méthodes basées sur l'émetteur

Ces méthodes reposent sur l'émetteur et s'appuient sur l'ajout de paquets de redondance (méthodes passives) ou de la retransmission (méthodes actives). La classification des différentes méthodes basées sur l'émetteur est montrée dans la FIGURE 3.1 [50].

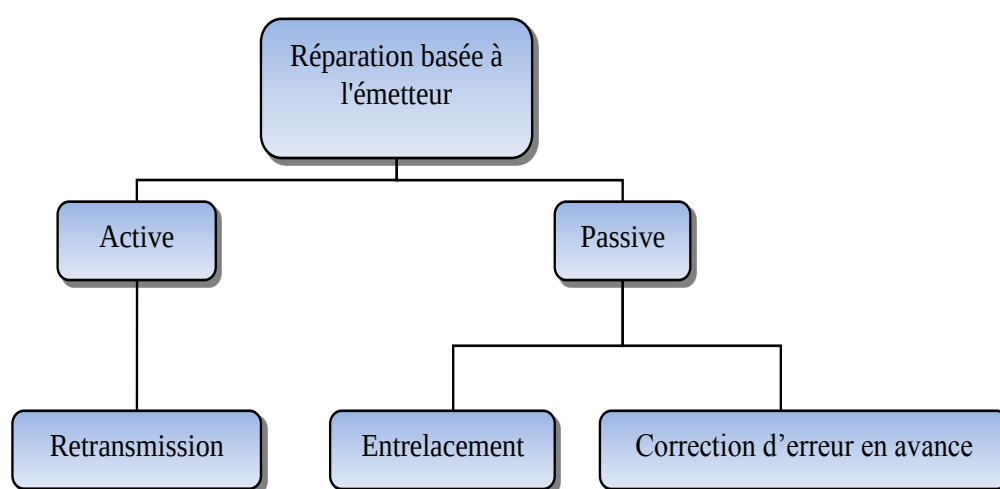


FIGURE 3.1 – Classification des techniques de réparations basées à l'émetteur.

A Les méthodes passives

Les méthodes passives transmettent de l'information de redondance en plus de l'information originale de telle sorte qu'au moins une partie de l'information perdue peut être reconstruite à partir de l'information de redondance [51].

— Correction d'erreur direct

Le FEC repose sur l'ajout d'une redondance dans le flux, de façon à être en mesure de restituer au moins une partie des informations manquantes en raison de perte des paquets dans le réseau. Il existe deux classes de FEC, à savoir :

(a) FEC indépendant du média

Cette technique utilise des méthodes algébriques pour produire des paquets supplémentaires qui contiennent suffisamment d'informations afin de restituer les paquets manquants (des exemples de ce type de FEC sont : parity

coding, Reed Solomon codes, ou complex XOR schemes) [51], comme l'indique la FIGURE 3.2.

(b) **FEC dépendant du média**

Cette technique exploite la connaissance de la structure du flux et les caractéristiques de codage pour fournir une protection efficace. Les données supplémentaires introduites par le FEC ne sont généralement pas très significatives et leurs avantages sont très remarquables. Plus de détails sur les techniques du FEC peuvent être trouvés dans [51]. La technique FEC dépendant du média est montrée dans la FIGURE 3.3.

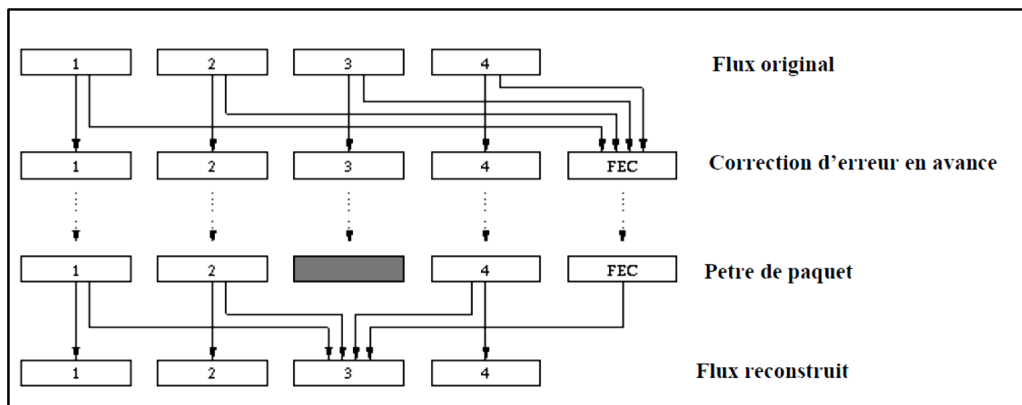


FIGURE 3.2 – FEC indépendant du média.

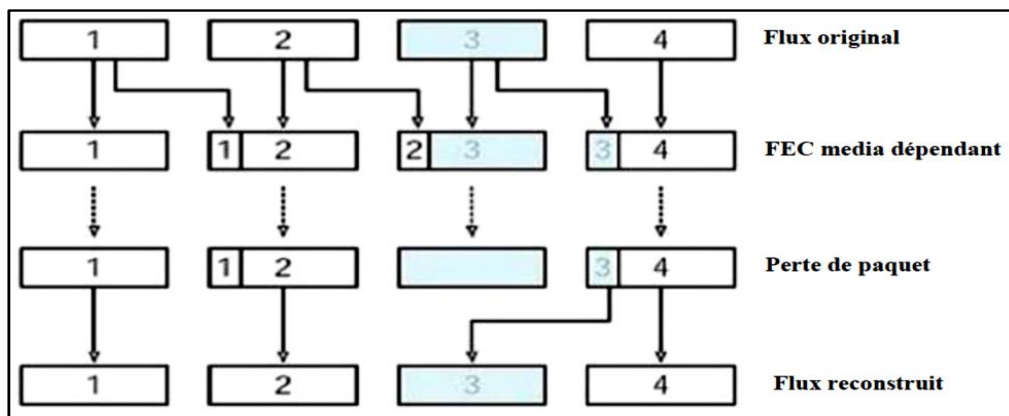


FIGURE 3.3 – FEC dépendant du média.

L'entrelacement ou interleaving permet de disperser les effets de la perte de paquets sur le flux, afin de les rendre moins apparentes. L'idée est d'arranger les échantillons de telle manière que les échantillons successifs se retrouvent dans différents paquets. Aussi, si un paquet est perdu, au lieu d'avoir un grand vide dans le flux, on obtient plusieurs interruptions plus courtes. Si les interruptions sont assez courtes, cependant, elles seront difficiles à repérer [52]. Par exemple, prenant un paquet composé de 4 trames d'une durée 20 ms, la perte d'un paquet engendra la perte de toutes ces trames (perte de 20 ms), par contre l'entrelacement des paquets répartira cette perte sur l'ensemble des 4 paquets (perte de 5 ms), ainsi, cette perte ne sera pas détectée par le récepteur. L'augmentation de latence constitue un sérieux inconvénient à l'utilisation de l'entrelacement dans des applications interactives. Alors que le maintien d'une bande passant stable avant et après son implémentation représente son avantage majeur. La FIGURE 3.4 illustre l'entrelacement d'un groupe de 4 paquets où chaque paquet est scindé en 4 sous trames [51], [37], [53].

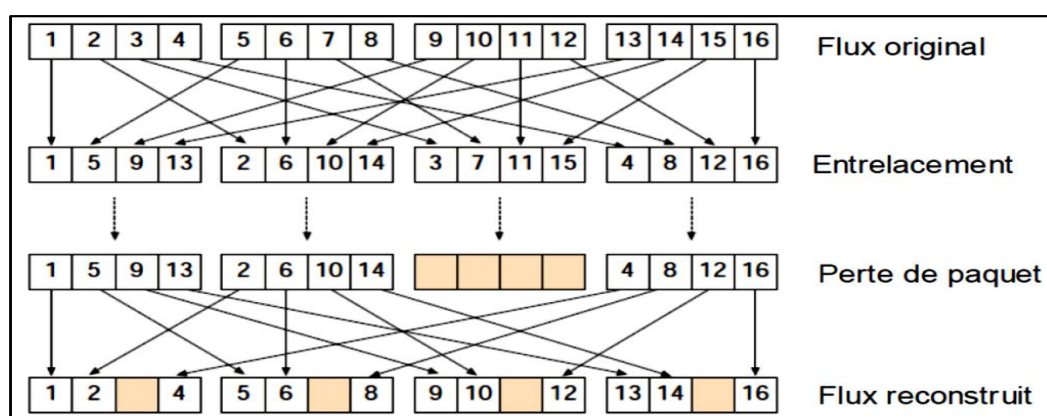


FIGURE 3.4 – Entrelacement sur plusieurs paquets.

B Les méthodes actives

Les méthodes actives s'appuient sur le principe de la retransmission des paquets perdus. Parmi ces méthodes, on peut citer :

(a) Requête automatique de répétition

La requête automatique de répétition ou ARQ (Automatic Repeat reQuest) permet de contrôler les erreurs pour la transmission de données. Elle utilise des requêtes de retransmission et des timeouts pour parvenir à une transmission efficace du flux.

Une requête de retransmission est un message envoyé par le récepteur vers l'émetteur afin de lui demander la retransmission du paquet perdu. Un timeout est un instant précis situé après l'instant d'émission et dont l'écart avec ce dernier est égal à une

durée spécifique ; si le récepteur ne reçoit pas le paquet avant le timeout, il envoie une requête de retransmission du paquet perdu à l'émetteur. Lors de la réception de cette requête, l'émetteur retransmet le paquet cité dans la requête.

Cette technique est difficilement applicable pour les applications interactives dont lesquelles les délais de transmissions de bout en bout sont réduits (le délai de transmission de bout en bout doit être inférieur à 250 ms) [35], [32].

Néanmoins, elle peut être considérée comme une bonne solution dans certaines situations en fournissant des avantages tels que l'économie de la bande passante et l'adaptation aux variations des réseaux.

(b) Requête de répétition automatique hybride

La requête de répétition automatique hybride (Hybrid-Automatic Repeat request, HARQ) est une combinaison des mécanismes du FEC et l'ARQ décrits ci-dessus. Après la transmission d'un premier bloc FEC, avec des paquets utiles et de redondance, l'HARQ autorise l'émetteur à transmettre des paquets de redondance supplémentaires pour le cas où la récupération de l'ensemble des paquets utiles n'a pas été possible du côté récepteur [37].

(c) Codage par descriptions multiples

Le Codage par descriptions multiples ou MDC est une technique basée sur l'émetteur et elle a pour objet de combattre contre les pertes de paquets liées aux pertes par congestion ou aux pertes suite aux erreurs de transmission. La technique MDC repose sur la division de l'information originale (parole d'entrée) en plusieurs descriptions corrélées qui seront transmises indépendamment sur deux ou plusieurs canaux indépendants. Le récepteur doit recevoir au moins une qui lui garantit une reconstruction de qualité minimale [32]. La FIGURE 3.5 illustre le schéma de base de MDC.

3.3.2 Les méthodes basées sur le récepteur

L'ensemble de ces techniques repose sur le principe de remplacer les paquets perdus par des paquets similaires aux originaux. Ceci reste possible du fait de la similarité à court terme du flux (stationnarité court terme du signal de la parole). La FIGURE 3.6 illustre les différentes techniques de masquage basées sur le récepteur [32], [54].

La technique basée sur le modèle de parole est parmi les techniques que nous avons choisies et la plus développée dans le domaine de masquage de pertes de paquets, c'est

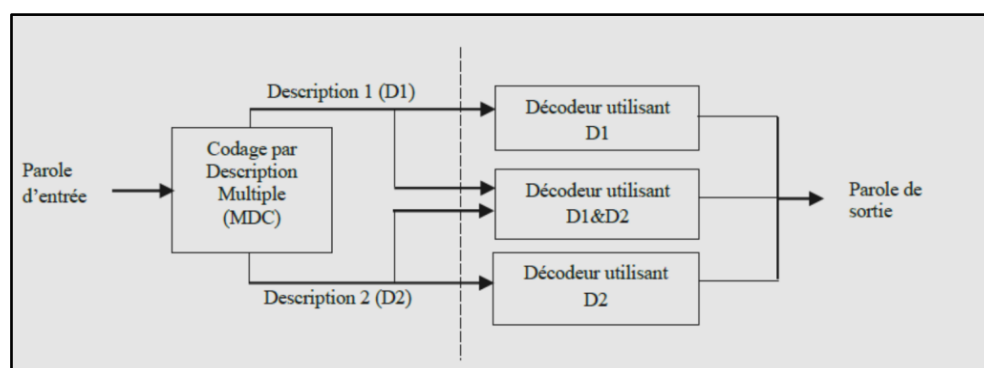


FIGURE 3.5 – Schéma de base de MDC.

l'un des modèles utilisés dans le codage CELP, tel qu'il est détaillé dans le chapitre II. Un signal d'excitation est généré pour simuler la source (située au niveau du larynx) puis filtré par le système acoustique (conduit vocal) pour produire de la voix [4]. C'est ainsi, que les systèmes de masquage de pertes de paquets ont été conçus en se basant sur le principe de fonctionnement de ce modèle de parole.

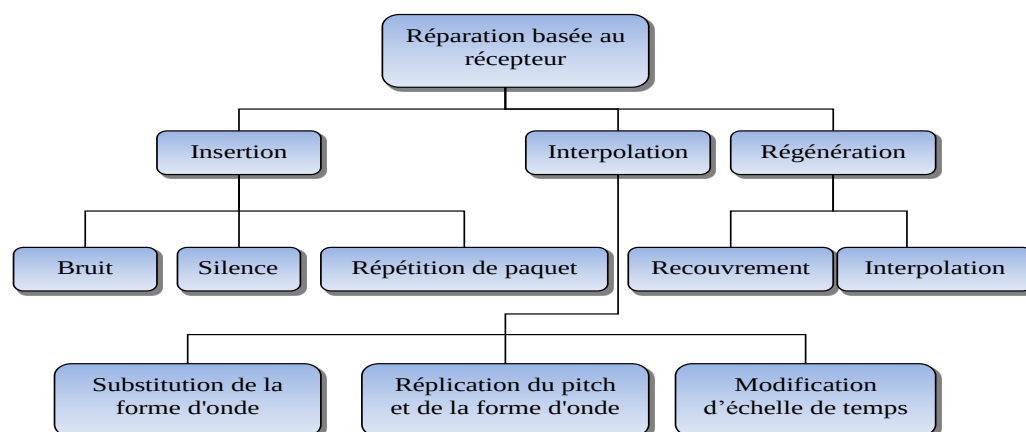


FIGURE 3.6 – Classification des techniques de réparations basées au récepteur [36].

— Système de Gunduzhan et al.

Gunduzhan et al. [48] proposent un système PLC qui se base sur la prédiction linéaire et un modèle source-filtre. Dans un premier temps, les filtres des formants qui possède des coefficients estimés à base de la dernière trame reçue $Tr(n)$ à l'aide d'un modèle AR. Le signal résiduel $e(x)$ en est extrait. Sa période fondamentale T_0 de ce signal résiduel est alors calculée à l'aide d'une méthode par autocorrélation. Cette période sert à générer la trame résiduelle manquante $e(n)$ que nous allons noter par $bx(n)$, elle

est ensuite filtrée avec le filtre formantique. La trame synthétisée $\hat{b}_x(n)$ est alors fondue avec les trames adjacentes par une méthode de superposition/admission [6]. C'est le modèle imbriqué dans le G.711. Le schéma global du système proposé par Gunduzhan est illustré dans la FIGURE 3.7.

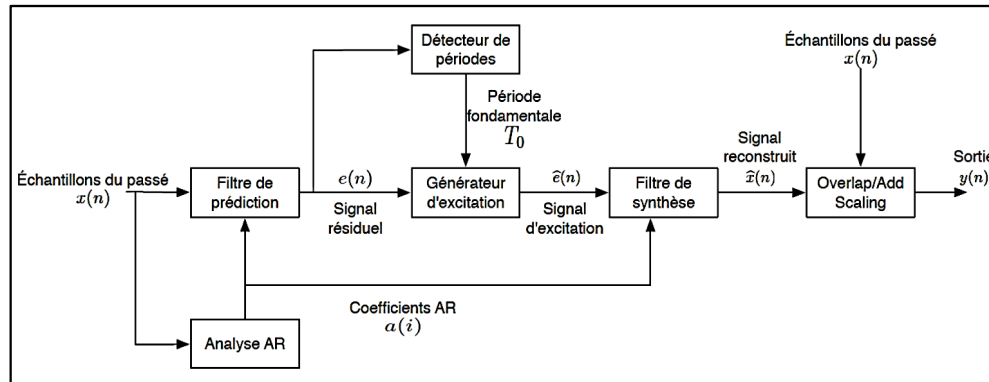


FIGURE 3.7 – Système de masquage de pertes de paquets proposé par Gunduzhan et al. [52].

- technique basée des modèles harmoniques plus bruit Ce modèle de signal de parole est de type "harmoniques plus bruit" est représenté comme suit :

A_k est l'amplitude de la k ème harmonique, avec N est le nombre d'harmoniques au temps t , f_0 est la fréquence fondamentale (pitch), $b(t)$ correspond au bruit du modèle. Les amplitudes, fréquences et phases du signal sont estimées sur la dernière trame reçue. Ces paramètres servent ensuite à la génération de la trame manquante. Le bruit additif $b(t)$ est obtenu en filtrant un bruit blanc à l'aide d'un filtre auto régressif.

Lindblom et al. suggèrent dans [36], [50] une extension de l'approche de Gunduzhan et al. [48] exprimée préalablement, tout en modélisant le signal résiduel $e(n)$ au moyen d'un modèle harmonique plus bruit. Aussi, $e(n)$ est représenté par une somme de sinusoides à laquelle, on ajoute une composante aléatoire et sert d'excitation au filtre de synthèse [6].

Méthodes basées sur l'insertion

Cette technique de réparation génère un remplacement du paquet perdu en insérant une simple donnée de remplacement. Il est à mentionner que cette technique ne prend pas en compte les caractéristiques du signal, ce qui la rend simple à implémenter. La donnée de remplacement peut être de natures différentes, à savoir un silence, un bruit ou bien une version répétée de la dernière bonne trame reçue. Ces techniques sont faciles à implémenter,

mais à l'exception de la technique répétitive, elles possèdent de faibles performances [37], [55], [53].

1. L'omission

Les paquets perdus sont remplacés par des trames de silence. Cette méthode possédant un coût calculatoire nul, ce qui produit un résultat de qualité médiocre : dès que la taille du vide dépasse les 20 ms, l'auditeur perçoit une dégradation [37].

2. L'épissage

Les paquets perdus peuvent être dissimulés par l'aboutement des paquets qui précèdent le paquet perdu avec les paquets qui le suivent. Cette technique est bonne seulement pour des pertes inférieures à 16 ms, autrement l'utilisation de cette méthode peut également interférer avec le tampon adaptatif de reproduction requise dans un système de paquets audio, car il fait une réduction de la quantité de données disponibles pour le tampon. Les tampons sont utilisés au niveau du récepteur pour diminuer l'impact de la gigue sur la qualité du son, et la mauvaise performance de ce tampon peut nuire la qualité de l'ensemble du système sonore [37].

3. La répétition

Les paquets précédents remplacent les éléments de signal manquant par les paquets précédents. Il s'agit donc de répéter le ou les derniers paquets reçus. Ce type de dissimulation rend imperceptible des vides de petite taille [37].

Méthodes basées sur l'interpolation

Cette technique utilise l'identification de paramètre et l'interpolation pour remplacer les paquets perdus. Elle est plus difficile à implémenter et requière plus de ressource de calcul que la méthode d'insertion, cependant, parce que cette technique prend en compte les changements des caractéristiques du signal, ses performances sont meilleures [37], [56].

1. Substitution de la forme d'onde

Cette technique met à profit le signal d'avant et, optionnellement, d'après perte pour trouver un signal convenable pour combler les pertes [9].

2. Réplication du pitch et de la forme d'onde

Cette méthode est une amélioration de la méthode précédente et semble donner de meilleurs résultats, elle utilise, en plus, un algorithme de détection du pitch des deux côtés d'une perte de paquet [14].

3. Modification d'échelle de temps (Time Scale Modification)

Cette technique permet au signal audio, des deux côtés d'une perte, d'être étiré sur toute la longueur de la perte. Malgré une demande, en ressource de calcul, importante, cette méthode semble travailler mieux que les deux méthodes précédentes [37].

Méthodes basées sur la régénération

Les techniques de régénération profitent de la connaissance à priori de l'algorithme de compression des signaux audio pour récupérer les paramètres du codec. Par conséquent, le signal audio dans un paquet perdu peut être synthétisé. Ces techniques sont plus performantes en raison de la grande quantité d'informations utilisées dans la réparation. Nous citons [37] :

1. Interpolation de l'état transmis

Cette technique est basée sur le codage par transformée ou la prédiction linéaire telles que les codecs de la norme G.722.2 et le G.723.1. Elle utilise le décodeur pour interpoler entre les états.

2. Réparation basée sur un modèle

Cette méthode extrapole ou interpole les paramètres d'un modèle de parole du signal reçu sur le signal manquant. Elle utilise un modèle de génération de la parole couvrant la perte.

3.4 Conclusion

Dans ce chapitre, nous avons décrit les différentes techniques de masquage de pertes des paquets afin de pallier aux effets de ces pertes survenant dans les réseaux, plusieurs techniques ont été mises en œuvre comme le résume les FIGURES 3.1 et 3.6. Outre la retransmission par ARQ des paquets perdus, le FEC et MDC ont été proposés pour masquer la perte des paquets dans le réseau IP, permettant ainsi d'améliorer les performances des applications de type temps-réel. D'autres techniques permettant de réduire les pertes de paquets tout en évitant l'émission des paquets de redondance sont les techniques de l'entrelacement et l'insertion des paquets. Dans le chapitre suivant, nous allons décrire les différentes méthodes que nous avons conçues et implémentées dans le codeur large bande de la norme G.722.2.

Chapitre 4

LES SOLUTIONS PROPOSEES POUR LE G.722.2

Sommaire

4.1	Introduction	54
4.2	Techniques MDC pour améliorer la robustesse de codeurs basés sur ACELP AMR-WB	55
4.2.1	Panorama sur l'AMR-WB G.722.2	55
4.2.2	Modélisation de la perte de paquets par la Chaîne de Markov du second ordre	56
4.2.3	Méthode de dissimulation basée sur MDC	58
4.2.4	Simulations et discussions	62
4.3	Amélioration de la robustesse des codeurs basés sur ACELP AMR-WB G.722.2	72
4.3.1	Quantification	72
4.3.2	Méthode proposée	73
4.3.3	Simulations et discussions	73
4.4	Etude comparative sur les méthodes de dissimulation pour améliorer la robustesse des codeurs basés sur ACELP AMR-WB	78
4.4.1	Méthode basée sur MDC	78
4.4.2	Méthode basée sur le FEC	79
4.4.3	Méthode basée sur l'entrelacement	79
4.4.4	Simulations et discussions	79
4.5	Conclusion	83

4.1 Introduction

Les codeurs ACELP, telle que la norme G.722.2 de UIT-T, sont souvent utilisés dans les systèmes VoIP car, ils offrent une qualité de parole respectable dans les environnements autres que l'apparence des pertes de paquets. Néanmoins, leur dépendance à la Prédiction à Long Terme (LTP) provoque la propagation des erreurs dans les trames de parole et rend ainsi les codeurs ACELP plus sensibles aux pertes de paquets. Plusieurs solutions ont été apportées pour remédier aux problèmes de perte de paquets. Dans ce chapitre, nous allons présenter nos contributions dans ce domaine, notamment les solutions proposées pour le G.722.2.

Ainsi, nous présentons notre méthode de dissimulation basée sur MDC pour améliorer la qualité de la parole provoquée par la perte de paquets pour les codeurs basés sur ACELP. Dans cette méthode, nous avons appliqué au codeur G.722.2 une technique de dissimulation de perte de paquet, qui utilise des schémas de paquets basés sur le MDC. Ce dernier est utilisé avec deux nouveaux modes conçus, à savoir, les modes 5 et 6 (18,25 et 19,85 kbps, respectivement).

Aussi, nous présentons notre nouveau modèle de Chaîne de Markov du second ordre à quatre états afin de simuler les pertes de réseau pour différents taux de perte [57].

Egalement, nous montrons une comparaison des méthodes de dissimulation des pertes de paquets basées sur le FEC, l'entrelacement et le MDC, et ce, afin d'améliorer la qualité de la parole détériorée, provoquée par les pertes de paquets des codeurs ACELP du réseau [58].

Enfin, nous présentons notre contribution que nous avons appliquée au codeur de parole G.722.2 pour évaluer une méthode proposée basée sur MDC. L'approche proposée consiste en un schéma de paquets basé sur MDC au niveau des paramètres. Chaque paquet contenant des informations redondantes sur les futures trames voisines. Cette information redondante n'a pas la qualité de l'original, car elle est quantifiée plus grossièrement, mais elle aide à reconstruire la parole lorsqu'un paquet est perdu. Dans l'approche proposée, nous avons utilisé un modèle de Markov à deux états avec différents taux de perte pour simuler le taux de perte de paquets, en appliquant les trois méthodes proposées précédemment [59].

4.2 Techniques MDC pour améliorer la robustesse de codeurs basés sur ACELP AMR-WB

4.2.1 Panorama sur l'AMR-WB G.722.2

L'algorithme G.722.2 AMR-WB est utilisé comme CoDec audio à large bande sur la VoIP avec une bande audio de 50-7000 Hz au lieu de 200-3400 Hz utilisée dans la téléphonie classique. Lorsque la bande passante est augmentée, l'intelligibilité et le caractère naturel de la parole sont améliorés. Le CoDec AMR-WB G.722.2 est similaire au CoDec 3GPP AMR-WB. Les propriétés 3GPP relatives sont les normes TS 26.190 du CoDec de parole et TS 26.194 pour le VAD [60], [61].

Le G.722.2 décrit le mappage précis des blocs d'entrée de 320 échantillons de parole dans un code d'impulsion uniforme d'un format de modulation de 16 bits (PCM) en blocs codés de 132, 177, 253, 285, 317, 365, 397, 461 et 477 bits et à partir de blocs codés de 132, 177, 253, 285, 317, 365, 397, 461 et 477 bits pour produire des blocs de parole reconstitués de 320 échantillons [62].

Le schéma de codage pour les modes de codage à plusieurs débits est également appelé codeur à prédiction linéaire à excitation de code algébrique, appelé ci-après ACELP. Le codeur ACELP multi-bande large bande est appelé AMR-WB [61]. Le G.722.2 utilise également un VAD [60].

Le taux d'échantillonnage est de 16 000 échantillons relatif à un débit binaire codé de 6,60, 8,85, 12,65, 14,25, 15,85, 18,25, 19,85, 23,05 et 23,85 kbps qui correspondent, dans la pratique, respectivement aux modes 0, 1, 2, 3, 4, 5, 6, 7 et 8.

Dans notre travail, nous nous intéressons aux modes 5 et 6 du codeur, qui correspondent respectivement à des débits binaires de 18,25 kbps et de 19,85 kbps. Les paramètres de ces modes sont donnés dans le Tableau 4.1.

TABLE 4.1 – G.722.2. Affectation de bits de l’algorithme de codage AMR-WB modes 5 et 6 pour une trame de 20 ms.

18.25 Kbps	VAD-flag					1
	ISP					46
	LTP-filtering	1	1	1	1	4
	Pitch delay	9	6	9	6	30
	Algebraic code	64	64	64	64	256
	Gain	7	7	7	7	28
	Total					365
19.85 Kbps	VAD-flag					1
	ISP					46
	LTP-filtering	1	1	1	1	4
	Pitch delay	9	6	9	6	30
	Algebraic code	64	64	64	64	288
	Gain	7	7	7	7	28
	Total					397

4.2.2 Modélisation de la perte de paquets par la Chaîne de Markov du second ordre

Dans cette section, nous présentons un nouveau modèle de perte de paquets. Notre schéma modélise les paquets perdus par une chaîne de Markov du second ordre [57].

Soit X_t le troisième résultat du paquet tandis que $X = 1$ et $X_t = 0$ correspondent respectivement, à une perte de paquets et une transmission sans erreur. Notre modèle peut être considéré comme une chaîne de Markov du second ordre où chaque état de ce modèle est représenté par le couple (X_t, X_{t+1}) . Nous distinguons quatre transitions de probabilité :

$$p_{00} = P(X_{t+1} = 1 | X_t = 0, X_{t-1} = 0) \quad (4.1)$$

$$p_{01} = P(X_{t+1} = 1 | X_t = 0, X_{t-1} = 1) \quad (4.2)$$

$$p_{10} = P(X_{t+1} = 1 | X_t = 1, X_{t-1} = 0) \quad (4.3)$$

$$p_{11} = P(X_{t+1} = 1 | X_t = 1, X_{t-1} = 1) \quad (4.4)$$

Le graphique de transition correspondant est présenté à la FIGURE 4.1. Notre modèle de chaîne de Markov à quatre états est irréductible et apériodique. Il est donc ergodique

et convergent. La distribution de probabilité stationnaire est donnée par :

$$\pi_{00} = P((X_t, X_{t-1}) = (0, 0)) = \frac{1 - p_{01}}{p_{00}} \pi_{01} \quad (4.5)$$

$$\pi_{10} = \pi_{01} \quad (4.6)$$

$$\pi_{11} = \frac{p_{10}}{1 - p_{11}} \pi_{01} \quad (4.7)$$

$$\pi_{01} = \frac{1}{a} \quad (4.8)$$

$$\text{avec : } a = 2 + \frac{1 - p_{01}}{p_{00}} + \frac{p_{10}}{1 - p_{11}}$$

Notons que notre modèle est plus général que le modèle simple de Gilbert, qui est une chaîne de Markov du premier ordre à deux états [51], [63]. Cependant, lorsque $p_{00} = p_{01}$ et $p_{10} = p_{11}$, notre modèle est réduit au simple modèle de Gilbert à deux états. Notez que l'avantage dans notre modèle est le fait que nous pouvons ajouter un état "de rupture" (1, 0) et un état "de récupération" (0, 1) contrairement aux modèles de Gilbert et Gilbert-Elliot [64], [49].

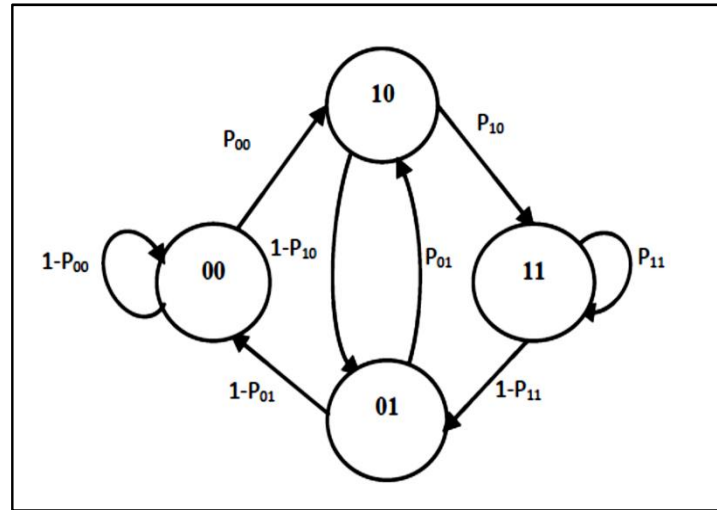


FIGURE 4.1 – Modèle de Markov de second ordre à quatre états (00 : bon, 10 : rupture, 11 : mauvais, 01 : récupéré).

4.2.3 Méthode de dissimulation basée sur MDC

La structure de base du MDC est illustrée dans la FIGURE 4.2. La parole, après le codage, est divisée en deux ou plusieurs descriptions qui sont transmises indépendamment. Chaque description est décodée séparément pour diminuer la reconstruction de la qualité du signal d'entrée. Cependant, si deux ou plusieurs descripteurs sont disponibles, ils peuvent être conjointement décodés pour une reconstruction de meilleure qualité de la parole de sortie [65], [66].

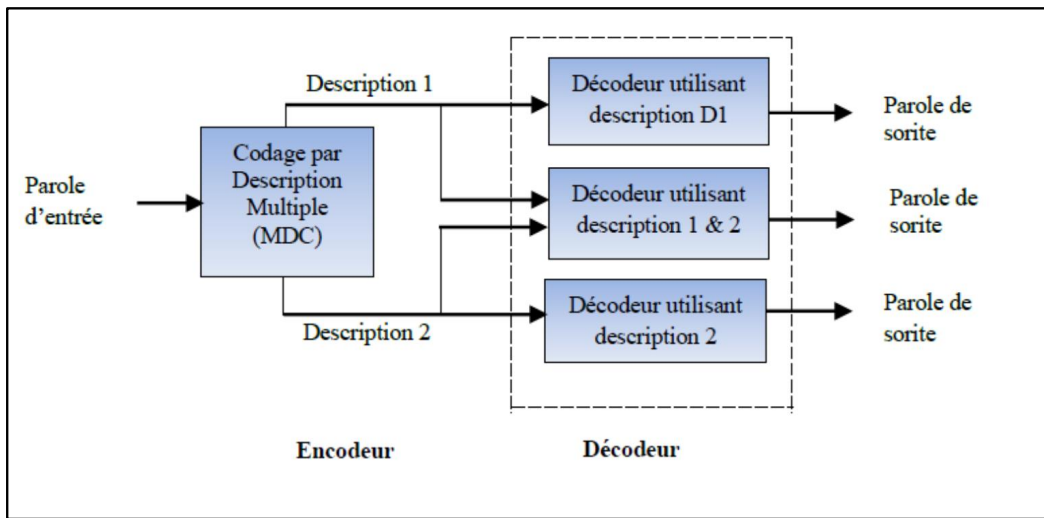


FIGURE 4.2 – Schéma de base pour le MDC.

Dans ce travail, nous avons modifié le code source du CoDec standard d'origine G.722.2 afin d'obtenir de nouveaux débits qui représentent la description de la technique MDC. Pour se faire, nous avons utilisé la base de données DARPA TIMIT Acoustique-Phonétique Discours Continu Corpus (TIMIT), données d'entraînement et de test [67]. La lecture du discours constitue le corpus de la base de données TIMIT qui était destinée à délivrer des données de parole, afin d'effectuer une connaissance phonétique acoustique et de développer, d'améliorer et d'évaluer un mécanisme de reconnaissance automatique de la parole. La base de données est divisée en deux parties, une pour la formation, et l'autre pour le test. Etant donné que notre CoDec comporte 10 dictionnaires de codes de ISF, la réalisation a nécessité l'utilisation d'un test composé de 29150 trames, chacune avec une longueur de trame de 20 ms. L'algorithme de Linde, Buzo et Gray (LBG) a été adopté afin d'avoir les résultats souhaités en taille (en nombre de bits) du dictionnaire des codes [68], [69].

Il est à noter que le type de données de sortie de l'algorithme LBG est un réel qui sera converti en un entier et ensuite utilisé par notre CoDec.

Il est utile de rappeler que le G.722.2 utilise 6 paramètres pour représenter la parole. Pour les autres paramètres comme le codage algébrique, Pitch, Delay et Gain, le module d'un mode a été remplacé par un module d'un autre.

Pour atteindre cet objectif, nous avons généré quatre nouveaux descripteurs, pour quantifier chaque paramètre de notre modèle modifié. Le codeur de parole G.722.2 nécessite l'attribution de bits de donnée comme indiqué dans le Tableau 4.2. Notez que notre objectif est de générer deux nouveaux modes qui donneront le même mode que le CoDec original.

La quantification du vecteur résiduel ISF " r ", dans notre codeur modifié, est basée sur l'utilisation de la quantification vectorielle multi-étages (S-MSVQ), comme indiqué dans le Tableau 4.3. Le vecteur est divisé en deux sous-vecteurs " $r_1(n)$ " et " $r_2(n)$ " de dimensions 9 et 7, respectivement. La quantification de ces sous-vecteurs est réalisée en deux étapes :

- la quantification des débits binaires de 12,6 kbps, 12,2 kbps et 7,25 kbps, pour les deux sous-vecteurs " $r_1(n)$ " et " $r_2(n)$ " dans la première étape, est basée sur l'utilisation de "8" bits pour chacun. Dans la deuxième étape, le vecteur d'erreur de quantification est divisé en "3" et "2" sous-vecteurs, respectivement, c'est-à-dire le nombre des bits est ((8, 8, 6, 7, 6, 3, 3) (8, 8, 6, 7, 7, 5, 4) et (8, 8, 6, 7, 7, 5, 5), respectivement), ce qui donne "41" bits de l'ISF (rep "4" bits et "46" bits).
- la quantification du débit binaire à 6,05 kbps pour les sous-vecteurs " $r_1(n)$ " et " $r_2(n)$ " est basée sur l'utilisation de "8" bits et "6" bits, respectivement.

Dans la seconde étape, nous divisons les vecteurs d'erreur de quantification en sous-vecteurs "2" et "1", c'est-à-dire que le nombre de bits est (8, 8, 6, 7, 6, 3, 3) qui donne "25" bits d'ISF.

Ainsi, nous avons généré deux nouveaux modes (5 et 6) et pour la synchronisation de l'expéditeur et le récepteur, nous avons utilisé deux codeurs synchrones et ajouté un module packetizer comme le montre la FIGURE 4.3. Nous pouvons sélectionner les modes appropriés en fonction des besoins des taux de transmission. Rappelons que la technique MDC avec plusieurs descriptions peut être utilisée [70]. Dans notre travail, deux descriptions ont été utilisées.

Au niveau de l'émetteur, la mise en paquet est effectuée par le codeur G.722.2 du signal de parole d'origine. Dans ce cas, nous avons appliqué deux débits binaires, qui représentent deux descripteurs transmis dans le même paquet. Dans le nouveau mode "5" conçu (18.25 kbps), par exemple, le premier utilise le CoDec G.722.2 pour encoder la trame actuelle à 12,2 kbps, tandis que, le second utilise un autre G.722.2 pour coder la trame suivante à 6,05 kbps [71]. Les deux paquets seront alors disposés comme illustré dans la FIGURE 4.4. La dépaquetisation est effectuée du côté du récepteur.

Lorsque la technique MDC est appliquée avec deux descripteurs, la trame perdue est

TABLE 4.2 – Affectation de bits de l'AMR-WB G.722.2 modifié.

New bit rates for Mode 5 design (18.25 Kbps)	12.2 Kbps	VAD-flag					1
		ISP					45
		LTP-filtering	1	1	1	1	4
		Pitch delay	9	6	9	6	30
		Algebraic code	64	64	64	64	144
		Gain	7	7	7	7	28
		Total					252
	6.05 Kbps	VAD-flag					1
		ISP					46
		Pitch delay	8	5	8	5	26
		Algebraic code	12	12	12	12	48
		Gain	6	6	6	6	24
		Total					145
New bit rates for Mode 6 design (19.85 Kbps)	12.6 Kbps	VAD-flag					
		ISP					
		LTP-filtering	1	1	1	1	4
		Pitch delay	9	6	9	6	30
		Algebraic code	36	36	36	36	144
		Gain	7	7	7	7	24
		Total					244
	7.25 Kbps	VAD-flag					
		ISP					
		Pitch delay	8	5	5	5	23
		Algebraic code	12	12	12	12	48
		Gain	6	6	6	6	24
		Total					121

substituée par celle reçue au décodeur. La FIGURE 4.4 montre la mise en paquets / dépackétisation qui représente un exemple de notre proposition basée sur l'approche MDC, utilisant deux descripteurs (12,2 kbps et 6,05 kbps), afin de construire un nouveau mode

TABLE 4.3 – Quantification de l'ISP pour 6,05, 7,25, 12,2 et 12,6 kbps.

12.2 Kbps	1) Unquantized 16-element-long ISP vector				
	2) Stage 1 (r_1) 8 bits			2) Stage 1 (r_2) 8 bits	
	3) Stage 2	3) Stage 2	3) Stage 2	3) Stage 2	3) Stage 2
	$(r^{(2)}1, 0-2)$	$(r^{(2)}1, 3-5)$	$(r^{(2)}1, 6-8)$	$(r^{(2)}2, 0-2)$	$(r^{(2)}2, 3-6)$
	6 bits	7 bits	6 bits	6 bits	3 bits
6.05 Kbps	1) Unquantized 16-element-long ISP vector				
	2) Stage 1 (r_1) 8 bits			2) Stage 1 (r_2) 6 bits	
	3) Stage 2	3) Stage 2	3) Stage 2	3) Stage 2	3) Stage 2
	$(r^{(2)}1, 0-4)$ 5 bits	$(r^{(2)}1, 5-8)$ 3 bits	$(r^{(2)}1, 3-5)$ 3 bits	$(r^{(2)}1, 0-6)$ 3 bits	$(r^{(2)}1, 3-6)$ 3 bits
12.6 Kbps	1) Unquantized 16-element-long ISP vector				
	2) Stage 1 (r_1) 8 bits			2) Stage 1 (r_2) 8 bits	
	3) Stage 2	3) Stage 2	3) Stage 2	3) Stage 2	3) Stage 2
	$(r^{(2)}1, 0-2)$	$(r^{(2)}1, 3-5)$	$(r^{(2)}1, 6-8)$	$(r^{(2)}2, 0-2)$	$(r^{(2)}2, 3-6)$
	6 bits	7 bits	7 bits	5 bits	4 bits
7.25 Kbps	1) Unquantized 16-element-long ISP vector				
	2) Stage 1 (r_1) 8 bits			2) Stage 1 (r_2) 8 bits	
	3) Stage 2	3) Stage 2	3) Stage 2	3) Stage 2	3) Stage 2
	$(r^{(2)}1, 0-2)$	$(r^{(2)}1, 3-5)$	$(r^{(2)}1, 6-8)$	$(r^{(2)}2, 0-2)$	$(r^{(2)}2, 3-6)$
	6 bits	7 bits	7 bits	5 bits	5 bits

qui est le mode "5". En fait, quand un ou plusieurs paquets successifs sont perdus, le deuxième descripteur permet la reconstruction de la parole synthétisée (reconstruit). La dépackétisation vise à remplacer les trames perdues par le deuxième descripteur.

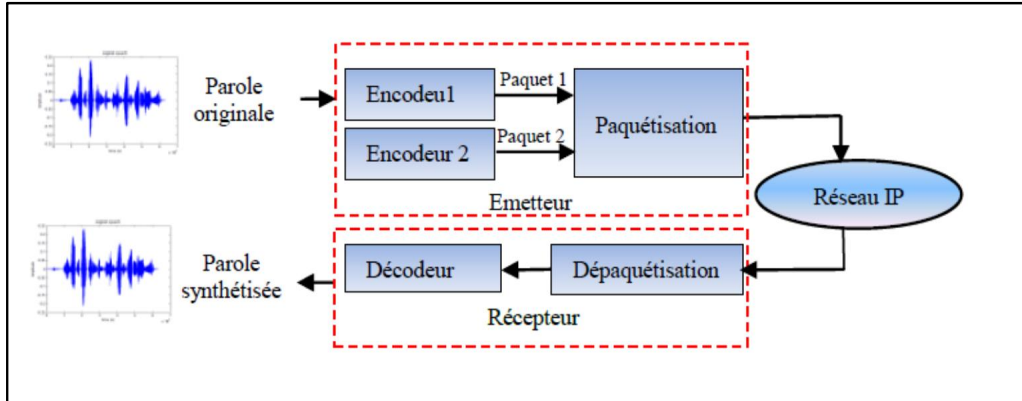


FIGURE 4.3 – Schéma fonctionnel de la transmission VoIP.

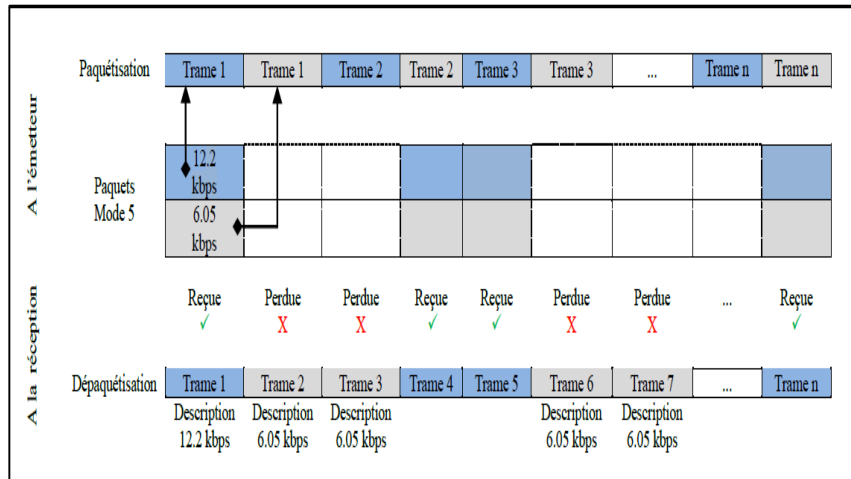


FIGURE 4.4 – Processus de paquetisation/dépaquetisation basé sur deux descriptions de MDC : mode 5

4.2.4 Simulations et discussions

Dans nos expérimentations, un fichier avec 198 trames est utilisé pour tester notre modèle de Markov de second ordre proposé, et ce, pour une variété de taux de perte comme indiqué dans le Tableau 4.4, à savoir, 11,11%, 18,68%, 32,32% et 42,42%, où le nombre de trames perdues est représenté dans les FIGURES 4.5, 4.6, 4.7 et 4.8 avec des segments verticaux blancs, alors que les trames reçues sont en segments verticaux noirs.

Dans les FIGURES 4.5, 4.6, 4.7 et 4.8 l'axe des abscisses représente le numéro de la trame et son ordre de transmission. Si le numéro de la trame k est perdu, on place

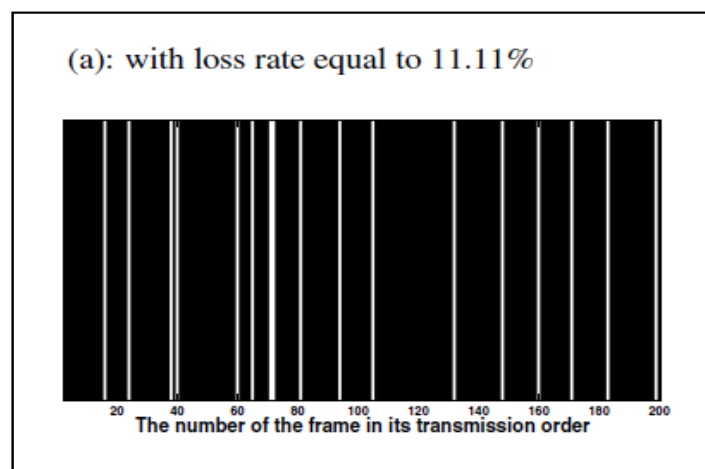


FIGURE 4.5 – Simulation de trames pour un taux de perte de 11,11%.

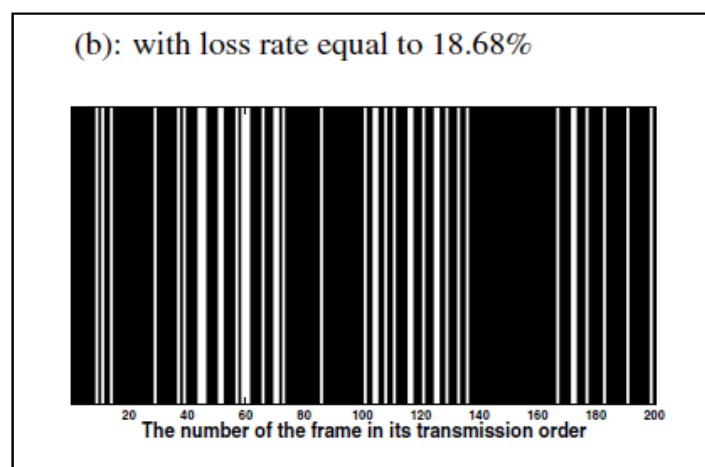


FIGURE 4.6 – Simulation de trames pour un taux de perte de 18,68%.

sur l'abscisse $x = k$, un segment vertical blanc, sinon, si le numéro de la trame k est bien reçu, nous plaçons (à l'abscisse $x = k$) un segment vertical noir. Plus de segments verticaux blancs sont trouvés, plus les trames sont perdus, cependant, le taux de perte est important. Par exemple, dans la FIGURE 4.6 correspondant au taux de perte égal à 18,68%, la trame numéro 60 est perdue, tandis que la trame numéro 180 est bien reçue.

Notre objectif est de quantifier la qualité de la voix perçue en utilisant des algorithmes d'estimation de qualité objective et subjective. Pour la qualité objective, le WB-PESQ [72] et EMBSD [73] ont été utilisés alors que pour la qualité subjective, nous avons utilisé le MOS [74] et confirmé nos résultats avec les tests de MUSHRA [75].

Il est utile de rappeler que :

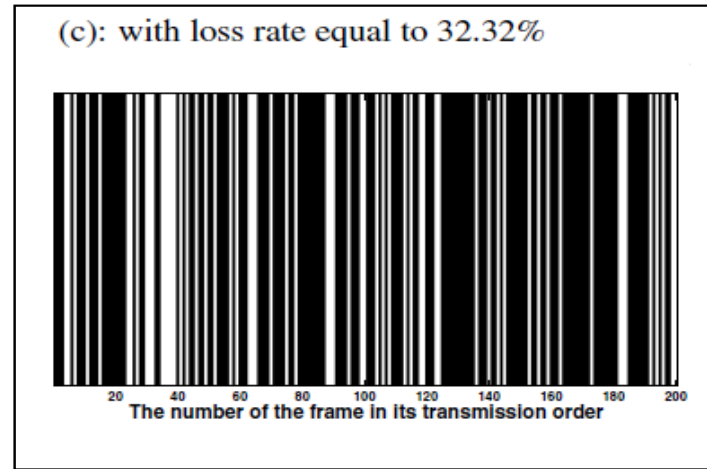


FIGURE 4.7 – Simulation de trames pour un taux de perte de 32,32% .

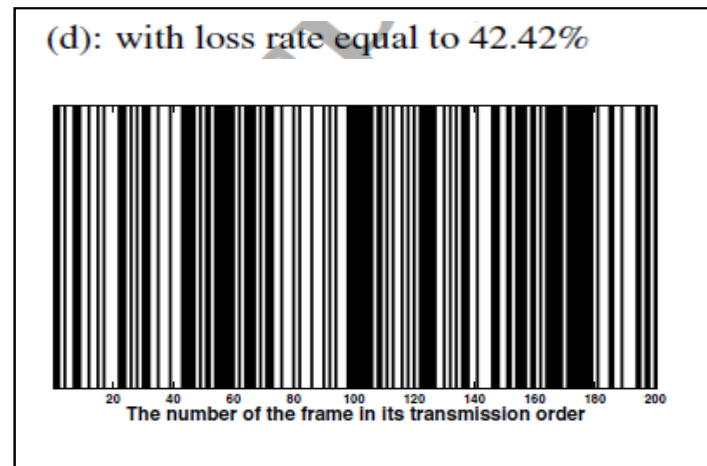


FIGURE 4.8 – Simulation de trames pour un taux de perte de 42,42%.

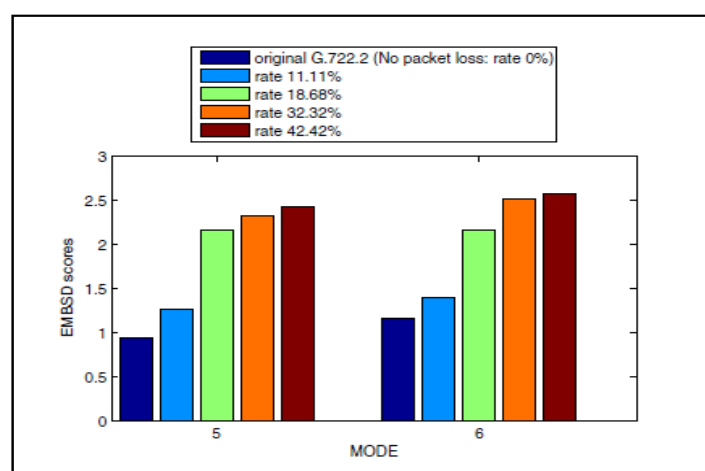
- Les valeurs EMBSD donnent 0 pour une comparaison avec le même fichier ;
- Les valeurs PESQ qui sont compris entre 0,5 et plus à 4,5 donnent une valeur de 4,5 pour une comparaison avec le même fichier ;
- Les échelles MOS sont comprises entre 1 et 5. Cette échelle est définie dans [76] comme étant 5 = Excellent, 4 = Bon, 3 = passable, 2 = Médiocre, 1 = Mauvais et donne un score de 4,54 pour une comparaison avec le même fichier ;
- Les valeurs MUSHRA sont comprises dans une échelle de 0 à 100 pour une comparaison avec le même fichier.

TABLE 4.4 – Taux de perte simulés avec le modèle de Markov de second ordre proposé.

Rate (%)	Lost frames	p_{00}	p_{01}	p_{10}	p_{11}
00	00	00	00	00	00
11.11	22	0.10	0.15	0.10	0.15
18.68	37	0.20	0.30	0.20	0.30
32.32	64	0.30	0.30	0.30	0.40
42.42	84	0.30	0.50	0.30	0.50

Tout d'abord, nous évaluons les performances de l'AMR-WB avec et sans perte de paquets pour les modes 5 et 6. Nous avons utilisé pour la perte de paquets, les taux de perte obtenus avec notre modèle de Markov de second ordre, proposé dans le Tableau 4.4.

Les mesures EMBSD, WB-PESQ et MOS sont illustrées dans les FIGURES 4.9, 4.10, 4.11, respectivement. Nous pouvons confirmer à partir de ces résultats, que la qualité de la parole est généralement bonne, pour les modes 5 et 6, dans le CoDec original sans perte et elle mauvaise avec perte de paquets.

**FIGURE 4.9 – Valeurs EMBSD de l'AMR-WB sans et avec paquet perte pour les modes 5 et 6.**

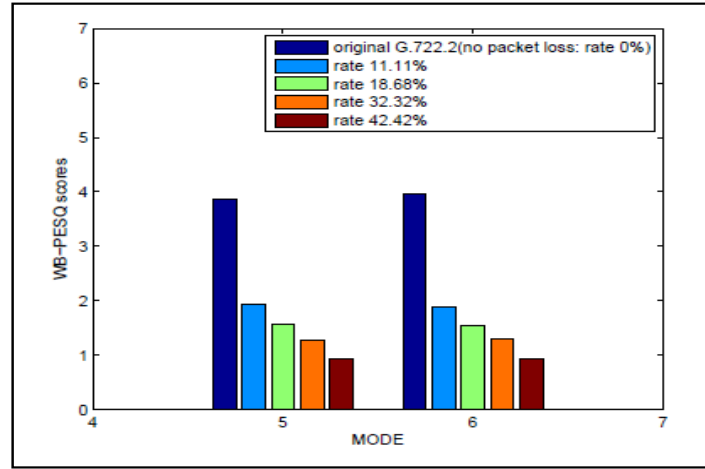


FIGURE 4.10 – Valeurs WB-PESQ de l’AMR-WB sans et avec paquet perte pour les modes 5 et 6.

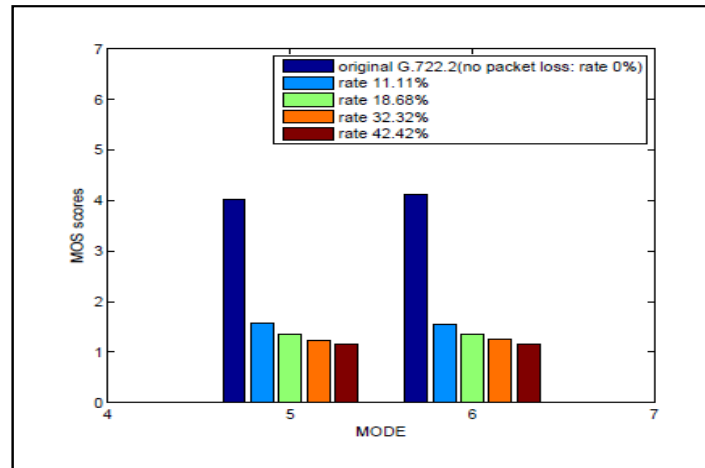


FIGURE 4.11 – Valeurs MOS de l’AMR-WB sans et avec perte de paquets pour les modes 5 et 6.

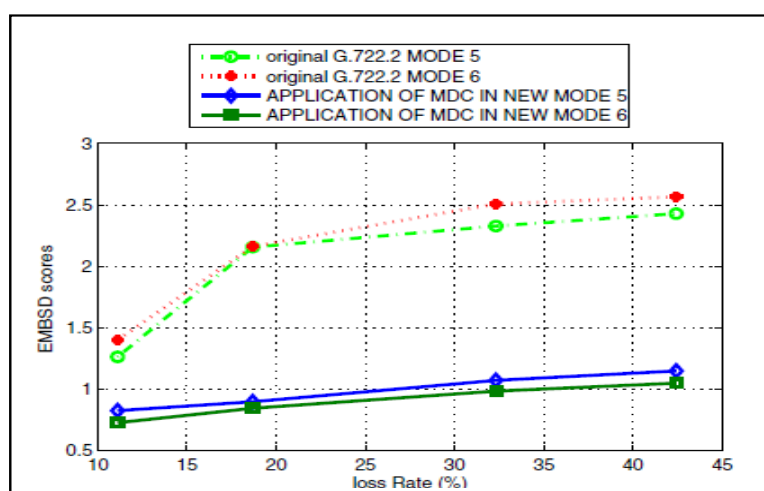
Une deuxième évaluation est réalisée dans le cadre de la proposition de notre méthode de dissimulation basée sur MDC. Premièrement, nous donnons l’évaluation de la performance des quatre nouveaux débits qui figurent dans le Tableau 4.5. Ce tableau montre que la qualité des nouveaux descripteurs est satisfaisante, variant entre juste et bon.

TABLE 4.5 – Résultats des tests avec EMBSD, WB-PESQ, MOS et MUSHRA.

	New bit rates for Mode 5 design (18.25 Kbps)		New bit rates for Mode 6 design (19.85 Kbps)	
	6.05 Kbps	12.2 Kbps	7.25 Kbps	12.6 Kbps
EMBSD	1.315	0.904	1.327	0.869
WB-PESQ	3.297	3.711	3.354	3.779
MOS	3.264	3.830	3.347	3.913
MUSHRA	60	84	61	89

Deuxièmement, l'utilisation du MDC donne les performances illustrées par les FIGURES 4.12, 4.13, 4.14 et 4.15. Les résultats sont presque similaires à ceux du signal original, qui indiquent que la présentation de la méthode a une bonne performance.

Dans la FIGURE 4.12, avec EMBSD pour différents taux de pertes, on constate que la qualité de la parole avec MDC pour les deux modes reste toujours plus élevée que le G.722.2 d'origine sans MDC.

**FIGURE 4.12 – Valeurs EMBSD de l'AMR-WB sans et avec paquet perte pour les modes 5 et 6.**

Dans la FIGURE 4.13, la mesure WB-PESQ confirme l'efficacité de notre méthode de dissimulation basée sur la MDC. Cela montre que, pour le même mode, plus le taux de perte augmente plus la qualité diminue et se détériore dans le codeur original. D'autre part, l'utilisation du MDC améliore la qualité par rapport au CoDec original. Ainsi, l'intelligibilité du signal est préservée selon les échelles MOS décrites à la FIGURE 4.14, avec

des scores variant entre juste et bon dans le CoDec original sans perte de paquet. Par conséquent, la méthode de dissimulation proposée basée sur l'MDC est meilleure que la méthode intégrée dans le codeur G.722.2 standard. C'est une nouvelle approche, qui fournit une excellente qualité de la parole et une grande précision sur les réseaux avec perte.

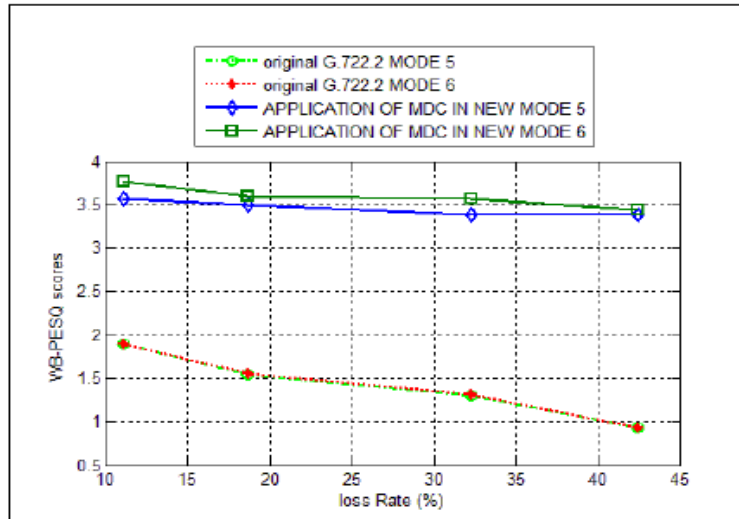


FIGURE 4.13 – Valeurs WB-PESQ pour différents taux de perte comparant le G.722.2 d'origine (modes 5 et 6) à notre méthode de dissimulation proposée basée sur MDC (nouveaux modes 5 et 6).

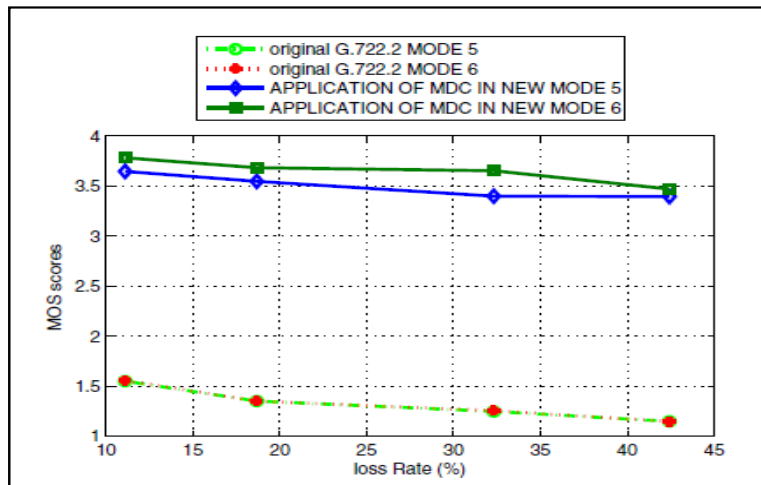


FIGURE 4.14 – Valeurs MOS pour différents taux de perte comparant le G.722.2 d'origine (modes 5 et 6) à notre méthode de dissimulation proposée basée sur MDC (nouveaux modes 5 et 6).

Des tests subjectifs ont également été réalisés pour évaluer les performances de notre

approche proposée. La méthode de test subjective utilisée dans les expériences est la méthodologie MUSHRA.

Ce test a l'avantage d'utiliser moins de participants que le test subjectif MOS, afin d'obtenir des résultats statistiquement significatifs (20 auditeurs sont utilisés) [77]. Les auditeurs doivent comparer l'algorithme standard PLC intégré dans le G.722.2 et notre proposition basée sur MDC, en les écoutant et en les comparant avec le signal non traité (référence) et le signal dégradé avec un taux de perte de 50% (ancre).

Les intervalles de confiance ont été fixés à 95%. La configuration de test est scindée sur 14 phrases évaluées par 22 auditeurs.

Dans nos expériences, les auditeurs ont donné des scores en fonction de la qualité du signal décodé par l'algorithme original et amélioré par l'MDC.

Les phrases de test ont été présentées aux auditeurs à un ordre aléatoire et répété pour quatre taux pertes différentes savoir 11,11%, 18,68%, 32,32% et 42,42%.

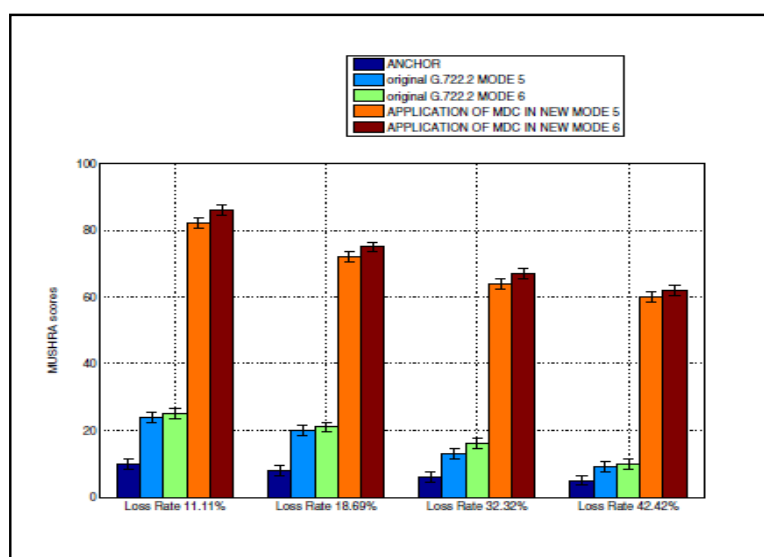


FIGURE 4.15 – Valeurs MUSHRA pour différents taux de perte comparant le G.722.2 d'origine (modes 5 et 6) à notre méthode de dissimulation proposée basée sur MDC (nouveaux modes 5 et 6).

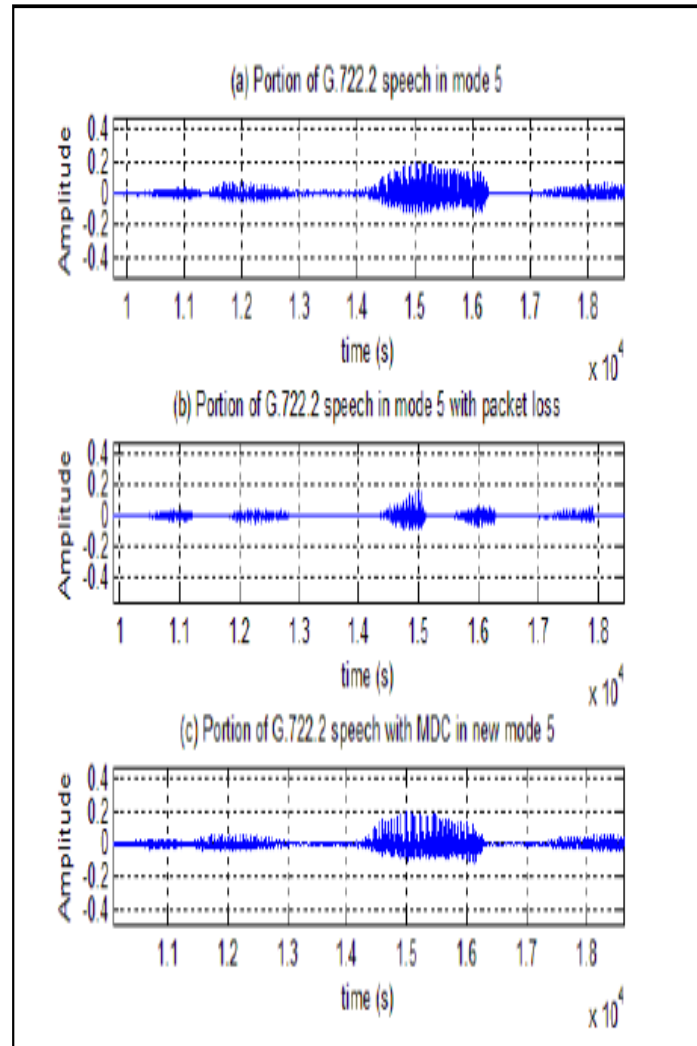


FIGURE 4.16 – Portion audiogramme mode 5 pour le taux de perte 42,42%

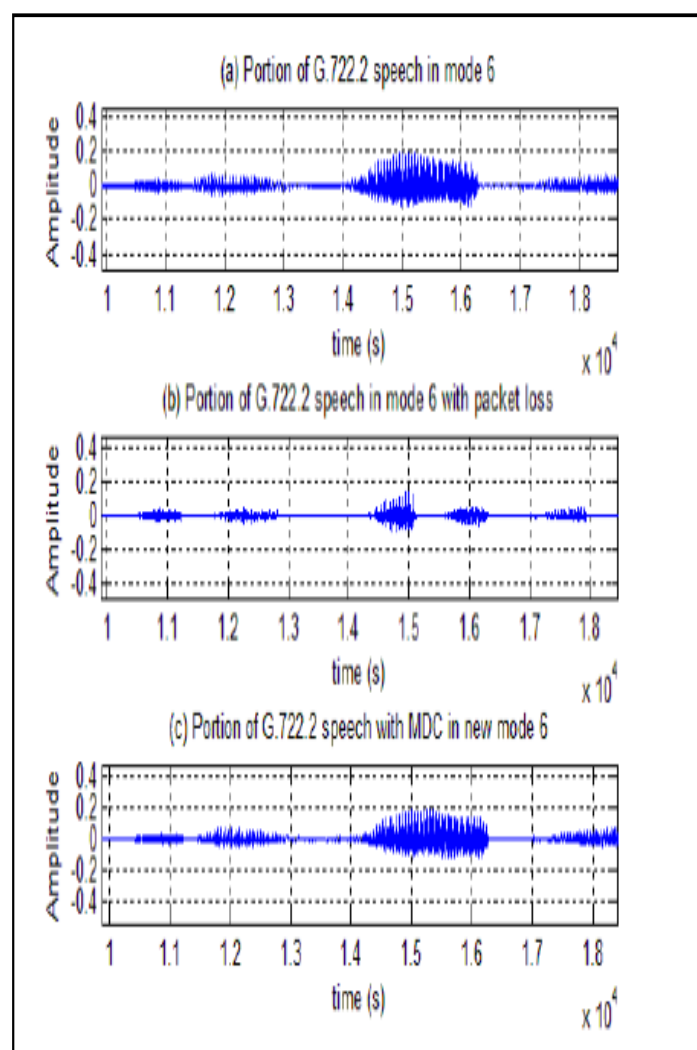


FIGURE 4.17 – Portion audiogramme mode 6 pour le taux de perte 42,42%

4.3 Amélioration de la robustesse des codeurs basés sur ACELP AMR-WB G.722.2

La méthode ACELP, telle que le codeur G.722.2, est souvent utilisée dans la VoIP, car elle offre des solutions de bonne qualité de la parole et la non-apparition des pertes de paquets. Néanmoins, leur dépendance dans LTP provoque la propagation des erreurs dans les trames de parole, et donc, rend les codeurs ACELP plus sensibles aux pertes de paquets. Ainsi, nous avons proposé une méthode basée sur MDC.

L'approche proposée consiste en un schéma basé sur la technique MDC, permettant la mise en paquets où chaque paquet contient des informations redondantes sur les trames voisines. Cette information redondante n'a pas la qualité de l'original qui est plus grossièrement quantifié mais elle permet de reconstruire la parole lorsqu'un paquet est perdu.

Nous avons utilisé le modèle de Markov pour simuler la perte de paquets, avec une comparaison des paquets perdus dans le G.722.2 original et les paquets perdus en appliquant la méthode MDC. L'évaluations PESQ et EMBSD sous diverses conditions de pertes de paquets, confirment que l'algorithme proposé est supérieur à l'algorithme de dissimulation incorporé dans le G.722.2. Les mesures de performance démontrent que la dissimulation basée sur le MDC sont meilleures par rapport au délai supplémentaire [58].

4.3.1 Quantification

L'AMR-WB G722.2 utilise six paramètres pour représenter la parole. Notre codeur vocal AMR-WB G.722.2 modifié nécessite l'affectation de bits comme indiqué dans le Tableau 4.7 pour quantifier chaque paramètre.

Sachant que dans le codeur d'origine, le vecteur ISF résiduel est quantifié à l'aide d'une combinaison de la SVQ et la MSVQ en fonction du nombre de bits (8, 8, 6, 7, 7, 5, 5). Dans notre codeur AMR modifié, le vecteur résiduel "r" est quantifié à l'aide de la quantification vectorielle fractionnée sur plusieurs étages SMSVQ. Le vecteur est divisé en 2 sous-vecteurs $r_1(n)$ et $r_2(n)$ de dimensions 9 et 7, respectivement. Les deux sous-vecteurs sont quantifiés en deux étapes :

- pour 12,2 Kbps : dans la première étape, on quantifie $r_1(n)$ avec 8 bits et $r_2(n)$ avec 8 bits. Dans la deuxième étape, les vecteurs d'erreur de quantification sont scindés en 3 et 2 sous-vecteurs, respectivement ;
- pour 6.05 Kbps : dans la première étape, on quantifie $r_1(n)$ avec 8 bits et $r_2(n)$ avec 6 bits. Dans la deuxième étape les vecteurs d'erreur de quantification sont scindés en 2 et 1 sous-vecteurs, respectivement.

TABLE 4.6 – AMR-WB G.722.2 modifié.

12.2 Kbps	VAD-flag					1
	ISP					41
	LTP-filtering	1	1	1	1	4
	Pitch delay	9	6	9	6	30
	Algebraic code	36	36	36	36	144
	Gain	6	6	6	6	24
	Total					244
6.05 Kbps	VAD-flag					1
	ISP					25
	LTP-filtering	8	5	8	5	23
	Pitch delay	12	12	12	12	24
	Algebraic code	64	64	64	64	288
	Gain	6	6	6	6	24
	Total					121

4.3.2 Méthode proposée

Nous pouvons utiliser MDC avec plusieurs descriptions. Dans notre travail, nous avons utilisé deux descriptions de sorte que si l'un des deux descripteurs est perdu, le décodeur peut estimer la perte en utilisant le descripteur reçu. Notre schéma proposé basé sur MDC est donné par la FIGURE 4.4 qui représente le processus de récupération des paquets via la méthode MDC proposée.

4.3.3 Simulations et discussions

Nous avons utilisé le modèle de Gilbert à deux états pour modéliser les pertes de paquets [51]. Ce modèle est une chaîne de Markov à temps discret, d'ordre 1. La discrétisation temporelle se fait par rapport aux trames successives du signal. L'ordre du modèle correspond à l'entropie des différents états. L'entropie est le taux de variation d'un système de la mesure de sa complexité. Dans notre application elle correspond aux différents liens existants entre les états. Dans le cas où l'entropie est de 0, la prédiction de l'état consécutif suivant est parfaitement fiable, tandis que, plus l'entropie est élevée, plus la prédiction est aléatoire.

Le modèle de Gilbert comprend deux états indiqués par un état "0" dans lequel le paquet est transmis, et l'état "1" dans lequel le paquet est perdu comme le montre la FIGURE 4.18 pour différents taux de perte donnés, dans le Tableau 4.7. P est la probabilité de passer de l'état 0 à 1, et Q est la probabilité de passer de l'état 1 à 0. Ce modèle est alors caractérisé par la matrice de transition T définie comme suit :

$$T = \begin{pmatrix} 1-P & P \\ Q & 1-Q \end{pmatrix}$$

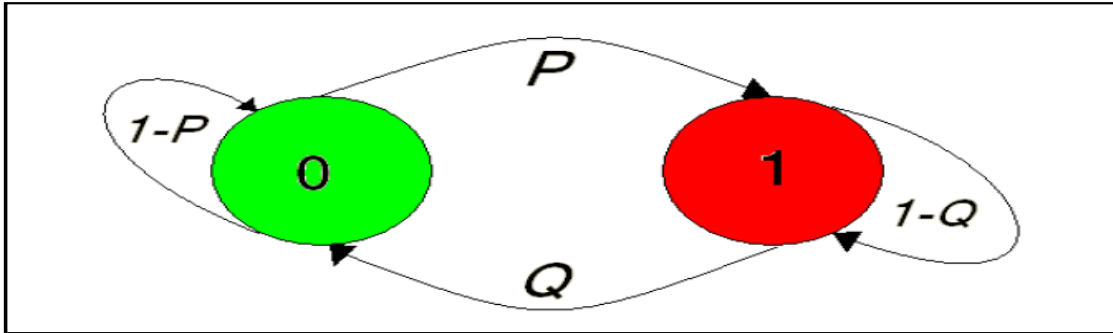


FIGURE 4.18 – Modèle de Gilbert à deux états : "0" indique la réception du paquet et "1" indique la perte du paquet.

TABLE 4.7 – Taux de perte simulés du modèle de Markov à 2 états.

Taux (%)	p_{00}	p_{01}
00	00	00
11	0,10	0,15
19	0,20	0,30
33	0,30	0,35
43	0,30	0,50

Premièrement, nous avons évalué la performance du codeur AMR-WB en utilisant l'EMBSD et le PESQ comme indiqué dans les FIGURES 4.19 et 4.20.

L'objectif est de quantifier la qualité de la perception de ces codeurs avant qu'ils ne soient implémentés à l'aide de MDC, et ce, en mesurant les algorithmes d'estimation de la qualité objective telles que, l'évaluation EMBSD [73] et PESQ [36].

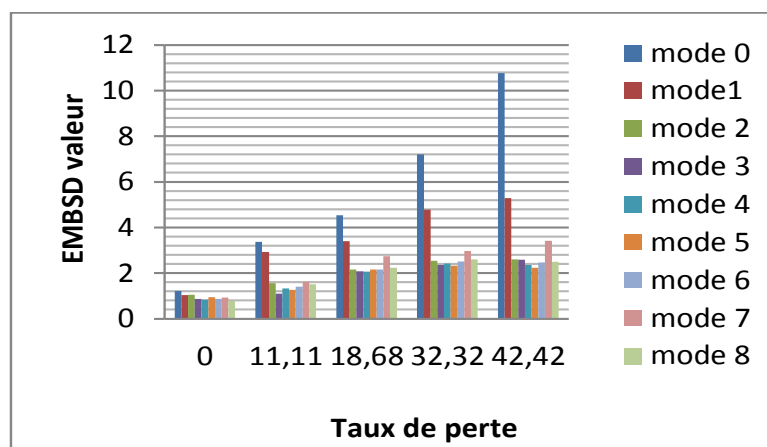


FIGURE 4.19 – Résultats avec l'évaluation EMBSD.

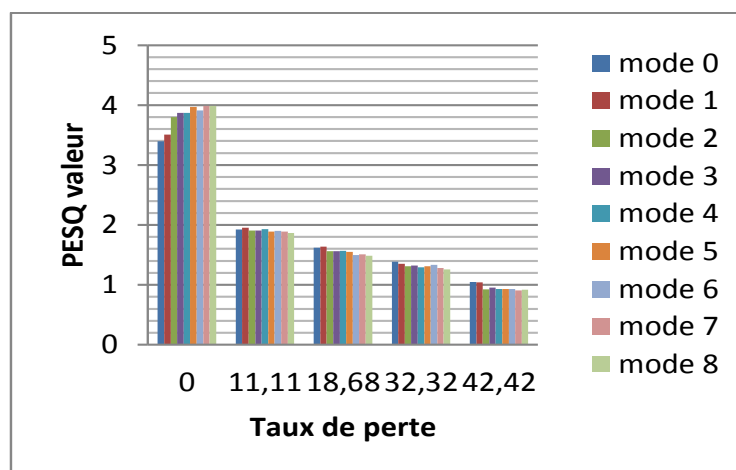


FIGURE 4.20 – Résultats avec l'évaluation PESQ.

Une deuxième évaluation a été réalisée à l'aide de MDC. Nous donnons l'évaluation de la performance des deux nouveaux codeurs appliqués séparément dans le Tableau 4.8.

TABLE 4.8 – Résultats EMBSD et PESQ

Codeur	6.05 Kbps	12.2 Kbps
EMBSD	3.297	3.711
PESQ	0.10	0.15

Le G.722.2 prend en charge l'adaptation dynamique aux conditions du réseau, en utilisant des débits binaires plus faibles en cas de congestion du réseau ou de dégradation,

tout en préservant la qualité audio. Après l'utilisation du MDC qui a donné la performance montrée dans la FIGURE 4.21 et la FIGURE 4.22. Les résultats en termes de l'EMBSD et de PESQ sont presque semblables au signal original, ce qui indique que la méthode présentée donne de bonnes performances.

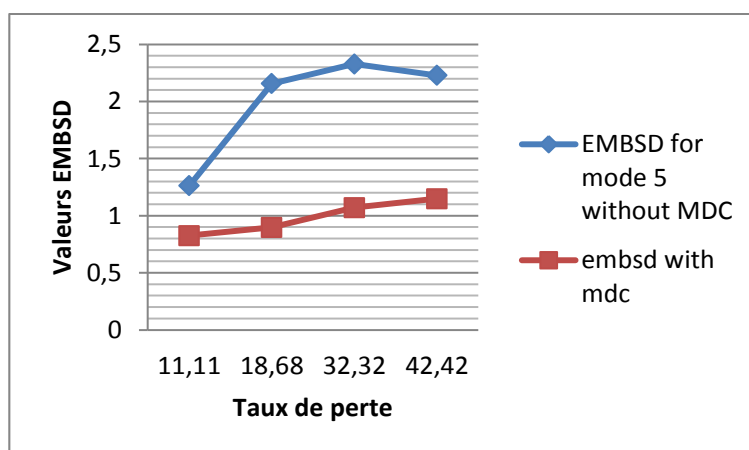


FIGURE 4.21 – Résultats des tests avec EMBSD pour différents taux de perte.

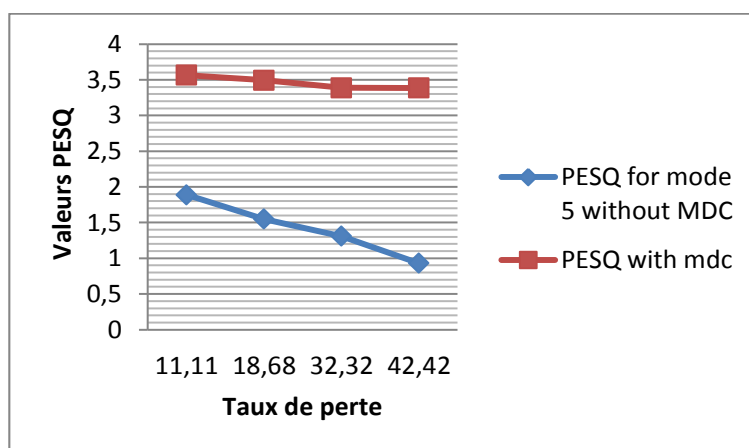


FIGURE 4.22 – Résultats des tests avec PESQ pour différents taux de perte.

Compte tenu du test avec EMBSD pour différents taux de perte, on peut constater dans la FIGURE 4.21, l'évolution de la valeur obtenue par EMBSD avec et sans MDC. Cette FIGURE montre que la qualité EMBSD avec le mode MDC reste toujours supérieure que EMBSD pour le mode 5 sans MDC.

Dans notre simulation, nous avons utilisé comme technique d'évaluation une méthode appelée PESQ. Les résultats sont illustrés dans La FIGURE 4.22 montrent le taux de perte des trames basé sur PESQ évaluation pour les deux cas, à savoir : PESQ pour le mode 5 sans et avec MDC.

La méthode MDC avec augmentation des taux de perte, se présente comme la meilleure méthode comparée à la technique PESQ sans MDC. L'intelligibilité du signal est préservée en fonction de l'échelle PESQ ($\text{PESQ} > 3$) pour le MDC avec test PESQ pour des taux de différence.

Enfin, l'augmentation de la valeur PESQ et la diminution de EMBSD avec MDC améliore l'intelligibilité et de la parole de manière significative.

Dans ce cas, des tests audio ont été réalisé par le logiciel Gold Wave, comme l'illustre les FIGURES 4.23 et 4.24, ce qui confirme l'efficacité et la robustesse de notre approche.

En conclusion, nous avons présenté la méthode MDC qui assure une bonne qualité de la parole pour toute perte de paquets. Le principal but est de réaliser un débit global de 18,25 kbps des deux débits binaires 6,05 kbps et 12,2 kbps.

Les tests EMBSD et PESQ dans diverses conditions de perte de paquets prouvent que l'algorithme proposé est supérieur à l'algorithme de dissimulation incorporé dans G.722.2.

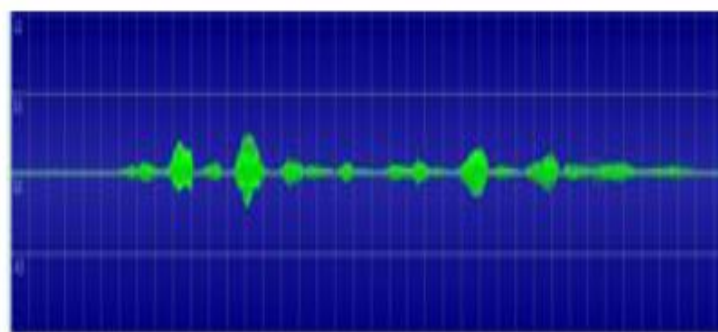


FIGURE 4.23 – Audiogramme de la parole avec EMBSD pour différents taux de perte.

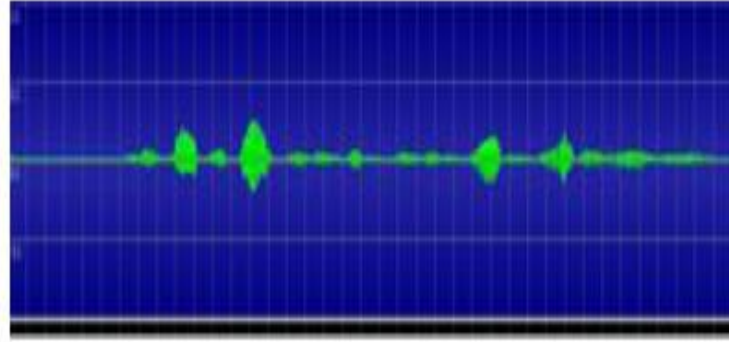


FIGURE 4.24 – Audiogramme de la parole avec PESQ pour différents taux de perte.

4.4 Etude comparative sur les méthodes de dissimulation pour améliorer la robustesse des codeurs basés sur ACELP AMR-WB

Dans cette partie, nous avons comparé des méthodes de dissimulation de perte paquets basées sur le FEC, l'entrelacement et le MDC pour améliorer la qualité de la parole dégradée causée par des pertes de paquets pour les codeurs ACELP, sur le réseau.

Nous avons appliqué notre approche au codeur de parole G.722.2 pour évaluer sa performance par rapport au PESQ et EMBSD sous diverses conditions de pertes de paquets. Les résultats obtenus confirment que l'algorithme proposé, à savoir le MDC est supérieur à l'algorithme de dissimulation intégré dans le G.722.2. Les mesures de performance démontrent que la méthode de dissimulation FEC est mieux que les méthodes l'entrelacement et MDC [59].

4.4.1 Méthode basée sur MDC

Dans notre approche proposée, nous avons utilisé deux descripteurs pour la MDC.

La FIGURE 4.2 de la section 4.2.3 représente notre approche proposée basée sur MDC avec deux descriptions. Cependant, si l'une des descriptions est perdue, le décodeur peut estimer la description perdue en utilisant le signal reçu.

4.4.2 Méthode basée sur le FEC

Le média indépendant FEC est souvent implémenté à l'aide de blocs ou codes algébriques, comme les codes de parité et les codes de Reed-Solomon. Ces codes fonctionnent en prenant un mot de code de k paquets de données pour générer $n \longleftrightarrow k$ paquets de contrôle supplémentaires pour la transmission de n paquets sur le réseau. Les paquets supplémentaires sont destinés pour aider à la récupération des paquets perdus chez le destinataire.

4.4.3 Méthode basée sur l'entrelacement

La méthode d'entrelacement ou interleaving permet de disperser les effets de la perte de paquets sur le flux, afin de les rendre moins apparent. L'idée est d'arranger les échantillons de telle sorte que les échantillons consécutifs se trouvent dans différents packages. Ainsi, si un paquet est perdu, au lieu d'avoir un grand vide dans le flux, nous obtenons plusieurs interruptions plus petites. Si les interruptions sont assez petites, elles ne seront pas faciles à détecter [52].

L'augmentation de la latence est un inconvénient de l'utilisation de l'entrelacement dans les applications interactives. L'avantage de cette méthode est le maintien d'une bande passante stable.

4.4.4 Simulations et discussions

Nous avons utilisé le modèle de Markov pour simuler les pertes de paquets sur le réseau pour différents taux de perte comme indiqué dans le Tableau 4.9.

TABLE 4.9 – Taux de perte simulés.

Taux (%)	P	q
00	00	00
11	0,10	0,15
33	0,30	0,35
43	0,30	0,50

Premièrement, nous avons évalué la performance du G.722.2. les FIGURES 4.25 et 4.26 montre l'évolution des valeurs EMBSD [73] et PESQ [36] par rapport aux taux de perte.

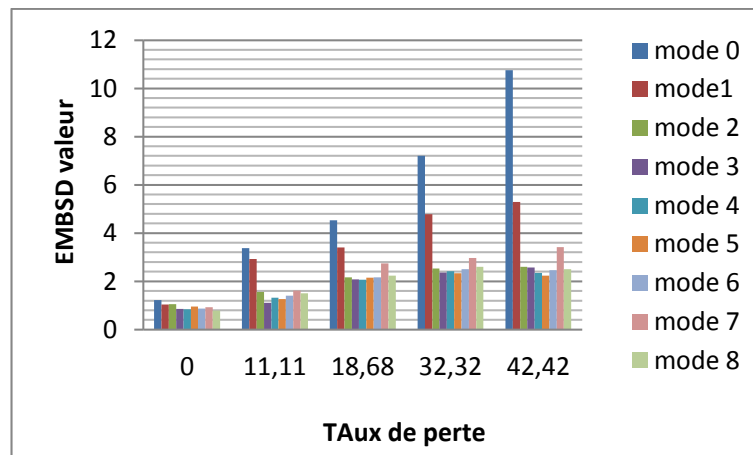


FIGURE 4.25 – Evolution de la valeur EMBSD vs taux de perte.

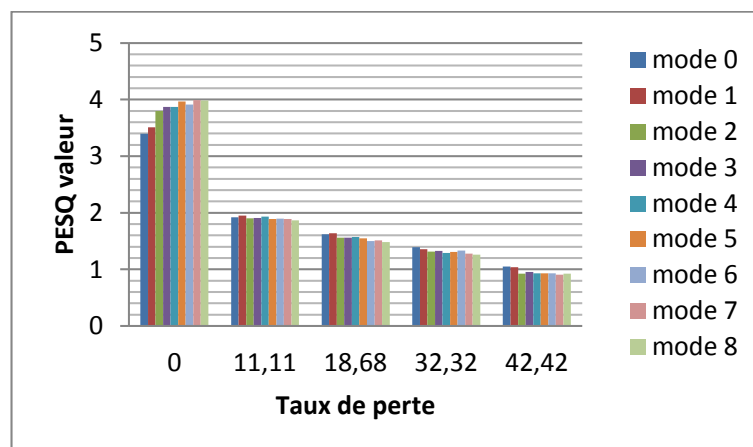


FIGURE 4.26 – Evolution de la valeur PESQ vs taux de perte.

Une deuxième évaluation a été réalisée à l'aide du MDC en montrant dans le Tableau 4.10 l'évaluation de la performance des deux nouveaux débits appliqués séparément.

TABLE 4.10 – Résultats des tests avec EMBSD et PESQ

Codeur	6.05 Kbps	12.2 Kbps
EMBSD	1.315	0.904
PESQ	3.297	3.711

Le G.722.2 prend en charge l'adaptation dynamique des conditions du réseau, en utilisant des débits plus faibles au cours de la congestion du réseau ou la dégradation, tout

en préservant la qualité audio. Les résultats de performance sont représentés par les FIGURES 4.27 et 4.28. Les résultats pour les PESQ et EMBSD sont presque similaires au signal original, ce qui indique que la méthode présentée donne des résultats significatifs avec EMBSD pour différents taux de perte.

On peut constater à partir de la FIGURE 4.27, l'évolution de la valeur EMBSD obtenue, avec et sans MDC. Cette figure, montre la qualité d'EMBSD avec le mode MDC qui reste toujours plus élevée pour le mode 5 EMBSD sans MDC.

Dans nos simulations, les résultats sont illustrés dans la FIGURE 4.28 qui montrent l'évaluation du taux de perte des trames basé sur PESQ pour les deux cas, à savoir : PESQ pour le mode 5 avec et sans MDC.

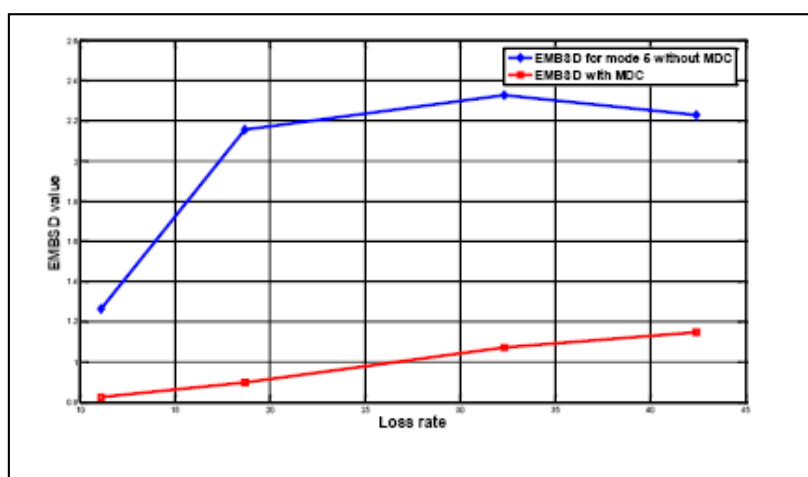


FIGURE 4.27 – Résultat EMBSD pour différents taux de perte.

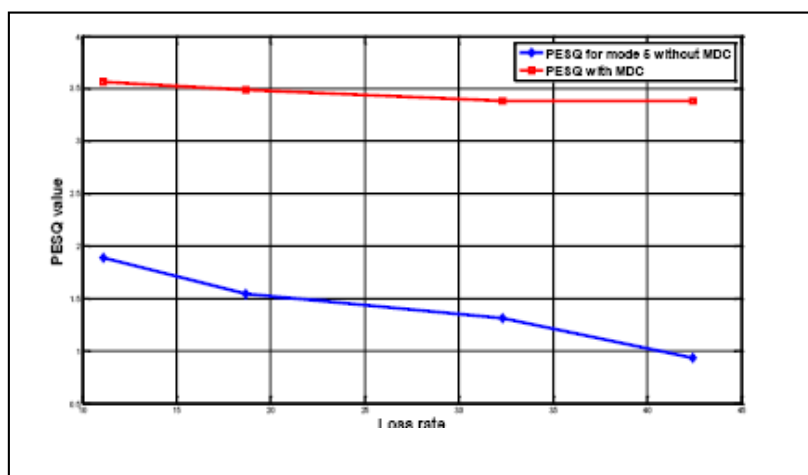


FIGURE 4.28 – Résultat PESQ pour différents taux de perte.

La méthode PESQ basée sur MDC est une nouvelle approche qui fournit une excellente qualité vocale et une haute précision. Enfin, l'augmentation de la valeur PESQ et la diminution de EMBSD avec MDC améliore l'intelligibilité et la nature de la parole de manière significative.

Une troisième évaluation comparant les performances des méthodes MDC, FEC et entrelacement en mode 5 a été réalisée. Nous avons utilisé la méthode PESQ comme mesure objective de la qualité de la voix.

Nous avons effectué un test EMBSD et les résultats sont illustrés dans la FIGURE 4.29. Plus le taux de perte de paquets augmente, l'EMBSD augmente aussi. Il est montré que ces méthodes proposées sont toujours meilleures.

La FIGURE 4.30 montre les résultats de comparaison des tests avec PESQ pour différent taux de perte pour les méthodes de masquage proposées, toujours en mode 5. Lorsque le taux perte de paquets augmente, les valeurs PESQ diminuent.

En conclusion, les valeurs des méthodes proposées sont supérieures à la méthode intégrée dans la norme G.722.2.

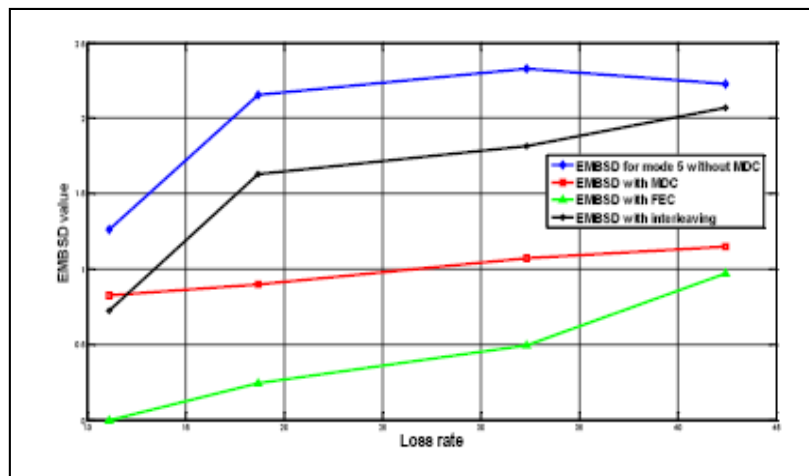


FIGURE 4.29 – Résultat de comparaison des tests avec EMBSD pour différent taux de perte.

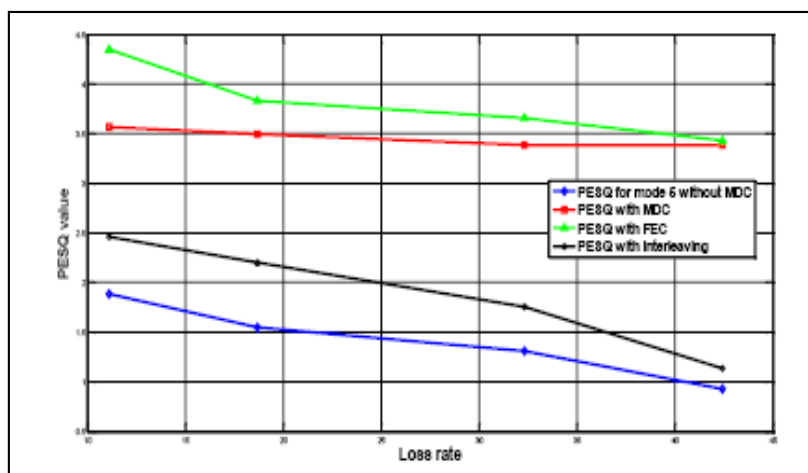


FIGURE 4.30 – Résultat de comparaison des tests avec PESQ pour différent taux de perte.

4.5 Conclusion

Dans ce chapitre, nous avons présenté une technique MDC qui assure correctement une bonne qualité de parole pour un taux de perte de paquets de données. Notre nouveau modèle de chaîne de Markov du second ordre à quatre états (00 : bon, 10 : rupture, 11 : mauvais, 01 : récupéré) a été utilisé pour simuler les pertes de paquets dans un réseau.

L'objectif principal était de réaliser des débits globaux de 18,25 kbps et de 19,85 kbps correspondant aux modes 5 et 6, respectivement. Sur la base des mesures WB-PESQ, MOS, EMBSD et MUSHRA dans diverses conditions, nous pouvons clairement constater que notre méthode proposée améliore considérablement la qualité de la parole par rapport à l'algorithme de masquage intégré dans le codeur G.722.2.

Aussi, des expériences ont été effectuées sur le G.722.2 en modes 5 et 6, le schéma proposé est clairement applicable à d'autres modes et pourrait également être étendu à d'autres codeurs basés sur la technologie ACELP.

Par la suite, sachant que le codeur G.722.2 prend en charge l'adaptation dynamique aux conditions du réseau, en utilisant des débits binaires plus faibles en cas d'encombrement ou de dégradation du réseau tout en préservant la qualité audio, nous avons présenté une technique basée sur MDC pour réaliser un débit global de 18,25 kbps à partir des débits binaires 6,05 kbps et 12,2 kbps.

Enfin, nous avons comparé les performances des trois méthodes telles que ; MDC, FEC et entrelacement avec le G.722.2. En considérant les tests de mesure PESQ et EMBSD dans diverses situations d'effacement de trames, nous avons constaté que la méthode MDC a considérablement amélioré la qualité de la parole par rapport à l'algorithme intégré dans la

norme G.722.2. Les trois méthodes présentent des résultats significatifs lorsque le taux de perte est élevé. En outre, les résultats des tests confirment que le FEC basé sur la méthode de dissimulation est meilleur que les méthodes d'entrelacement et MDC.

CONCLUSION GENERALE

Cette thèse avait comme objectif la proposition de solutions de récupération et de masquage des paquets perdus dans un réseau IP pour le codeur G.722.2. Pour cela, nous nous sommes tout d'abord intéressés au fonctionnement et à l'architecture des réseaux de VoIP. Aussi, nous avons expliqué quelques protocoles et standards utilisés (*RTP, RTCP, H.323, SIP, ..., etc.*) par une application temps réel sur un réseau IP.

Ensuite, nous avons introduit la notion des CoDecs Audio pour le codage/décodage de la parole (G.711, G.723.1, G.729, G.722.2). Nous avons présenté par la suite les problèmes rencontrés par les applications en temps réel (gigue, délais, perte des paquets, ..., etc.).

Nous avons détaillé les techniques de récupération et de masquage des pertes basées sur l'émetteur/récepteur et celles basées sur le récepteur seulement.

Dans le réseau IP, le trafic excessif de paquets, le retard et les pertes de paquets peuvent détériorer la qualité de la communication.

Nous avons proposé un nouveau modèle de Markov à quatre états pour simuler les pertes de paquets dans un milieu IP.

Bien que la téléphonie IP soit très économique, le publique hésite à l'utiliser en raison des éventuelles dégradations dues aux retards et pertes de paquets lors de la transmission. Ces deux facteurs affectent considérablement la qualité de la parole restituée.

Dans les fonctions du CoDec décrites par la norme G.722.2, il existe une procédure de masquage des pertes de paquets (insertion de silence).

Dans le but d'améliorer ses performances, nous avons proposé et implémenté plusieurs techniques de récupération et de masquage des pertes comme solutions pour le standard G.722.2.

Les résultats de simulation obtenus, nous permettent de conclure ce qui suit :

- la méthode d'entrelacement : Cette technique permet de récupérer la trame perdue à partir de la trame suivante. Son avantage réside dans l'économie de la bande passante (taux de redondance nul). Mais elle ne permet pas de récupérer la première trame lorsque deux trames consécutivement sont perdues. Donc, elle n'est pas efficace lorsque le taux de perte est supérieur à 6%.

- la méthode FEC : Cette technique permet de reconstruire une trame perdue dans un bloc de n trames. Mais lorsqu'il y a plus d'une perte, aucune correction n'est possible. Ce qui rend cette méthode pénalisée face à la perte des trames adjacentes.
- la méthode MDC : Cette technique consiste à augmenter le nombre de descripteurs pour s'assurer de la bonne trame. C'est une technique efficace pour récupérer les trames effacées quelque soit le taux de dégradation.

Dans les travaux futurs, nous proposons, l'utilisation des modèles de Markov cachés (Hidden Markov Model, HMM) pour le maquage des pertes pour le G.722.2.

Bibliographie

- [1] A. S. Spanias, “Speech coding : A tutorial review,” *Proceedings of the IEEE* **82**, 1541–1582 (1994).
- [2] A. Gersho and R. Gray, “Vector quantization and signal compression, kluwer academic,” Norwell, MA (1992).
- [3] J.-M. Valin, “Speex : A free codec for free speech,” *arXiv preprint arXiv :1602.08668* (2016).
- [4] J. Gibson, “Speech compression,” *Information* **7**, 32 (2016).
- [5] D. O’shaughnessy, *Speech communication : human and machine* (Universities press, 1987).
- [6] G. Fant, *acoustic theory of speech production* (Mouton, 1960).
- [7] G. Fant, *Acoustic theory of speech production : with calculations based on X-ray studies of Russian articulations*, 2 (Walter de Gruyter, 1970).
- [8] T. Islam, “Interpolation of linear prediction coefficients for speech coding,” Ph.D. thesis, McGill University Libraries (2000).
- [9] J. Wang and J. D. Gibson, “Parameter interpolation to enhance the frame erasure robustness of celp coders in packet networks,” in “2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221),” , vol. 2 (IEEE, 2001), vol. 2, pp. 745–748.
- [10] N. Batri, “Robust spectral parameter coding in speech processing,” Ph.D. thesis, Citeseer (1998).
- [11] M. Hasegawa-Johnson and A. Alwan, “Speech coding : Fundamentals and applications,” *Wiley encyclopedia of telecommunications* (2003).
- [12] J. D. Gibson, “Speech coding methods, standards, and applications,” *IEEE Circuits and Systems Magazine* **5**, 30–49 (2005).
- [13] R. Goldberg and L. Riek, *A practical handbook of speech coders* (CRC press, 2000).
- [14] F. Itakura, “Line spectrum representation of linear predictor coefficients of speech signals,” *The Journal of the Acoustical Society of America* **57**, S35–S35 (1975).

-
- [15] R. Gray, "Vector quantization," (IEEE ASSP Magazine, 1984), vol. 1, pp. 4–29.
 - [16] S. W. Kang, C. Y. Son, W. I. Lee, Y. N. Sung, M. K. Shim, and S. H. Hong, "High-speed search method for lsp quantizer using split vq and fixed codebook of g. 729 speech encoder," (2008). US Patent 7,389,227.
 - [17] W. C. Chu, "Speech coding algorithms," Foundation and evolution of standardized coders (2003).
 - [18] A. P. Bernard, "Source-channel coding of speech," Master's thesis, Citeseer (1998).
 - [19] B. Atal and J. Remde, "A new model of lpc excitation for producing natural-sounding speech at low bit rates," in "ICASSP'82. IEEE International Conference on Acoustics, Speech, and Signal Processing," , vol. 7 (IEEE, 1982), vol. 7, pp. 614–617.
 - [20] P. Kroon, E. Deprettere, and R. Sluyter, "Regular-pulse excitation—a novel approach to effective and efficient multipulse coding of speech," IEEE transactions on acoustics, speech, and signal processing **34**, 1054–1063 (1986).
 - [21] M. Schroeder and B. Atal, "Code-excited linear prediction (celp) : High-quality speech at very low bit rates," in "ICASSP'85. IEEE International Conference on Acoustics, Speech, and Signal Processing," , vol. 10 (IEEE, 1985), vol. 10, pp. 937–940.
 - [22] W.B.Kleijn, "Continuous representation in linear predictive coding"," in "Proc.Int.conf.Acoust. Speech and signal," (1991), pp. 201–204.
 - [23] J. Chiappini, "Performances de la voip sur réseaux wireless." in "Haute Ecole spécialisée Suisse occidentale," (2000).
 - [24] R. M. Gray, P. C. Cosman, and E. A. Riskin, "Image compression and tree-structured vector quantization," in "Image and Text Compression," (Springer, 1992), pp. 3–34.
 - [25] I. Rec, "Multiple description coding for voice over ip using sinusoidal speech coding," IEEE International Conference on Acoustics, Speech and Signal Processing (2006).
 - [26] I. Rec, "P. 800.1, mean opinion score (mos) terminology," International Telecommunication Union, Geneva (2006).
 - [27] D. C. e. a. Jackson, "Communication vocale sur protocole internet localisée," Brevet US N Â° 7.286.521 (23 octobre 2007).
 - [28] I. Rec, "Secteur de la normalisation des télécommunications de l'uit. recommandation uit-t g.711," International Telecommunication Union, Geneva (2001).
 - [29] B. Goode, "Voice over internet protocol (voip)," Proceedings of the IEEE **90**, 1495–1517 (2002).
 - [30] I. Rec, "Secteur de la normalisation des télécommunications de l'uit. recommandation uit-t g.729." International Telecommunication Union, Geneva (2001).

- [31] I. Rec, “Secteur de la normalisation des télécommunications de l’uit. recommandation uit-t g.723.1,” International Telecommunication Union, Geneva (2001).
- [32] I. Rec, “Secteur de la normalisation des télécommunications de l’uit. recommandation uit-t g.726,” International Telecommunication Union, Geneva (2001).
- [33] I. Rec, “Secteur de la normalisation des télécommunications de l’uit. recommandation uit-t g.722,” International Telecommunication Union, Geneva (2001).
- [34] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Jarvinen, “The adaptive multirate wideband speech codec (amr-wb),” *IEEE transactions on speech and audio processing* **10**, 620–636 (2002).
- [35] I.-T. S. G.722.2, “Wideband coding of speech at around 16 kbps using adaptive multi-ratewideband (amr-wb),” International Telecommunication Union, Geneva (2003).
- [36] C. Perkins, O. Hodson, and V. Hardman, “A survey of packet loss recovery techniques for streaming audio,” *IEEE network* **12**, 40–48 (1998).
- [37] I. Rec, “Coding of speech at 8kbit/s using conjugate structure algebraic code excited linear prediction(cs-acelp),” International Telecommunication Union, CH–Geneva (1996).
- [38] K. K. Paliwal and B. S. Atal, “Efficient vector quantization of lpc parameters at 24 bits/frame,” *IEEE transactions on speech and audio processing* **1**, 3–14 (1993).
- [39] O. H. C. Perkins and V. Hardman, “A survey of packet loss recovery techniques for streaming audio,” *IEEE network* **12**, 40–40 (1998).
- [40] D. López-Oller, N. Benamirouche, A. M. Gomez, and J. L. Pérez-Córdoba, “Speech excitation signal recovering based on a novel error mitigation scheme under erasure channel conditions,” *Speech Communication* **97**, 73–80 (2018).
- [41] W. P. LeBlanc, B. Bhattacharya, S. A. Mahmoud, and V. Cuperman, “Efficient search and design procedures for robust multi-stage vq of lpc parameters for 4 kb/s speech coding,” *IEEE Transactions on Speech and Audio Processing* **1**, 373–385 (1993).
- [42] B. Xu, “Packet loss concealment for speech transmissions in real-time wireless applications.” Master’s thesis, Delft, University of Technology (2017).
- [43] X. S. Wah, Benjamin W. and D. Lin., “A survey of error-concealment schemes for real-time audio and video transmissions over the internet,” (ISMSE, 2000).
- [44] L. Tosun and P. Kabal, “Dynamically adding redundancy for improved error concealment in packet voice coding,” in “2005 13th European Signal Processing Conference,” (IEEE, 2005), pp. 1–4.
- [45] e. a. Andersen, Soren Vang, “ilbc-a linear predictive coder with robustness to packet losses.” (IEEE Workshop Proceedings, 2002).

-
- [46] C. Lemyre, “Détection et modification des transitoires d’un signal de parole dans le but de rendre un codec plus robuste aux pertes de paquets,” Ph.D. thesis, Université de Sherbrooke. (2012).
 - [47] V. P. Bhute and U. N. Shrawankar, “Speech packet concealment techniques based on time-scale modification for voip,” in “2008 International Conference on Computer Science and Information Technology,” (IEEE, 2008), pp. 825–828.
 - [48] E. Gunduzhan and K. Momtahan, “Linear prediction based packet loss concealment algorithm for pcm coded speech.” (2001), vol. 9, pp. 778–785.
 - [49] L. R. Bahl, P. F. Brown, P. V. De Souza, and R. L. Mercer, “Maximum mutual information estimation of hidden markov model parameters for speech recognition,” in “proc. icassp,” , vol. 86 (1986), vol. 86, pp. 49–52.
 - [50] H. Sanneck, “Packet loss recovery and control for voice transmission over the internet,” Ph.D. thesis, Université échnique de Berlin (2008).
 - [51] L. Estrada, D. Torres, and H. Toral, “Characterization and modeling of packet loss of a voip communication,” World Academy of Science, Engineering and Technology **66**, 926–930 (2010).
 - [52] L. Koenig, “Masquage de pertes de paquets en voix sur ip,” Ph.D. thesis, Institut National Polytechnique de Toulouse (2011).
 - [53] F. M. T. Gueham, “Enhanced insertion packet loss concealment method for voice over ip network services,” in “International Conference on Telecommunications and Signal Processing (TSP),” , vol. 7 (Barcelona, 2017), vol. 7, pp. 377–382.
 - [54] T. Backstrom, *Packet Loss and Concealment* (2017), pp. 161–184.
 - [55] T. e. a. Romain, “Codage large bande de la parole par encapsulation du codeur itu g.729 (cs-acelp),” (1998).
 - [56] F. Merazka, “Loss concealement by interpolation for speech over ip network services,” in “Constantinides International Workshop on Signal Processing,” , vol. 10 (2013), vol. 10, pp. 560 50 1–4.
 - [57] H. Chaouch, F. Merazka, and P. Marthon, “Multiple description coding technique to improve the robustness of acelp based coders amr-wb,” *Speech Communication* **108**, 33–40 (2019).
 - [58] H. Chaouch and F. Merazka, “Improved robustness of acelp based coders amr-wb g.722.2,” in “The 2nd International Conference on Pattern Analysis and Intelligent Systems PAIS ’16 Khenchela, Algeria,” (IEEE, November 16-17th, 2016).
 - [59] H. Chaouch and F. Merazka, “Concealment methods to improve the robustness of acelp based coders amr-wb : A comparative study,” in “The Third International Conference

- On Electrical Engineering and Control Applications ICEECA'2017 Constantine, Algeria," (IEEE, November 21-22 th 2017).
- [60] M. S. C. S. P. Functions and A. W. S. Codec, "3gpp ts 26.194, amr wideband speech codec ; voice activity detector vad," Cover-54 (2000).
- [61] M. S. C. S. P. Functions and A. W. S. Codec, "Transcoding functions (3gpp ts 26.190 version 5.1. 0 release 5), dec. 2001," Cover-54 .
- [62] ITUT, "G.722.2, 2003. wideband coding of speech at around 16 kbps using adaptive multirate wideband (amrwb)," International Telecommunication Union, Geneva, Switzerland (2003).
- [63] J.-C. Bolot, "End-to-end packet delay and loss behavior in the internet," in "ACM SIGCOMM Computer Communication Review," , vol. 23 (ACM, 1993), vol. 23, pp. 289–298.
- [64] M. Ellis, D. P. Pezaros, T. Kypraios, and C. Perkins, "A two-level markov model for packet loss in udp/ip-based real-time video applications targeting residential users," *Computer Networks* **70**, 384–399 (2014).
- [65] Y. Lang, S. Zhao, and J. Kuang, "A multiple description speech coder based on amrwb," in "Fourth International Conference on Information Technology and Applications, ICITA 2007," (2007), pp. 697–700.
- [66] R. Choupani, S. Wong, and M. R. Tolun, "Unbalanced multiple description wavelet coding for scalable video transmission," *Journal of Electronic Imaging* **21**, 043006 (2012).
- [67] V. Zue, S. Seneff, and J. Glass, "Speech database development at mit : Timit and beyond," *Speech communication* **9**, 351–356 (1990).
- [68] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on communications* **28**, 84–95 (1980).
- [69] F. Merazka, "Enhanced differential split vector quantization of line spectrum pairs for celp-type coders in packet networks," in "Proceedings of the World Congress on Engineering and Computer Science," , vol. 2 (2009), vol. 2.
- [70] Y. Wang, A. R. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proceedings of the IEEE* **93**, 57–70 (2004).
- [71] I. Rec, "Itu-t rec. g.722.2 wideband coding of speech at around 16 kbit/s using adaptive multi-rate wideband (amr-wb)," International Telecommunication Union, Geneva (2003).

- [72] I. Rec, "P. 862.2 : Wideband extension to recommendation p. 862 for the assessment of wideband telephone networks and speech codecs," International Telecommunication Union, CH–Geneva (2005).
- [73] W. Yang, *Enhanced Modified Bark Spectral Distortion : An Objective Speech Quality Measure Based on Audible Distortion and Cognitive Model* (Temple University, 1999).
- [74] I. Rec, "P. 800.1, mean opinion score (mos) terminology," International Telecommunication Union, Geneva (2006).
- [75] I. Recommendation *et al.*, "Method for the subjective assessment of intermediate quality level of coding systems," ITU-R BS pp. 1534–1 (2003).
- [76] I. Rec, "P. 800 : Methods for subjective determination of transmission quality," International Telecommunication Union, Geneva p. 22 (1996).
- [77] M. Y. Kim and W. B. Kleijn, "Comparison of transmitter-based packet-loss recovery techniques for voice transmission," in "Eighth International Conference on Spoken Language Processing," (2004).

