

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
Scientifique

Université A. Mira de Bejaia

Faculté de technologies



Mémoire de fin d'études

En vue de l'obtention du Diplôme MASTER

Département : Génie Electrique

Filière : Télécommunications

Spécialité : Réseaux et télécommunication

THEME

**Amélioration de la Qualité de Service (QoS) de la voix
sur IP (VoIP)**

Réalisé par :

- BAAR Meriem
- HADDADI Sara

Promoteur :

- Mr. BENAMIROUCHE.N
Docteur à l'université de Bejaia

Soutenu le 12/10/2020 devant le jury composé de :

- Mme. HAMZAOUID Docteur à l'université de Bejaia Présidente
- Mr. TOUNSI.M Docteur à l'université de Bejaia Examineur

Année universitaire : 2019/2020

Remerciement

Ce travail est l'aboutissement d'un dur labeur et de beaucoup de sacrifices ; nos remerciements vont d'abord à Dieu le tout puissant et le Miséricordieux, qui nous a doté d'intelligence, et nous a maintenu en santé pour mener à bien ces années d'études.

Nous tenons aussi à remercier vivement notre encadreur Monsieur BENAMIROUCHE Nadir, qui malgré les conditions difficile, il nous a guidé et aidé à travers ses remarques pertinentes, ses conseils avisés et précieux. Et sa supervision éclairée tout au long de la rédaction de ce mémoire.

Nous adressons notre profonde gratitude à Madame HAMZAOUID et Monsieur TOUNSI.M, qui ont acceptés d'examiner ce travail et d'y apporter leur témoin. Et dont nous tenons à avoir de leur part des critiques constructives.

Nous souhaitons exprimer nos remerciements les plus sincères au corps professoral et administratif de l'Université Targa ou Zemmour Bejaïa, pour la richesse et la qualité de leur enseignement et qui déploient de grands efforts pour assurer à leurs étudiants une formation actualisée.

Enfin, nous ne pouvons passer outre notre reconnaissance envers nos parents, nos sœurs et frères et tous ce qui ont contribués à la rédaction de ce modeste travail.

Merci beaucoup

Au nom de dieu clément et miséricordieux

Dédicace



Avant tout, je tiens à remercier le bon dieu, et l'unique qui m'offre le courage et la volonté nécessaire pour affronter les différentes difficultés de la vie,

Je dédie ce modeste travail :

A ceux qui sont les plus chers du monde, Mon père Nadjim, et ma mère Sakina, à qui je n'arriverai jamais à exprimer ma gratitude et ma reconnaissance, pour ses amours ses soutiens et leur patience incommensurable tout au long de mes études.

A mon frère Yahia malgré son absence, et ma chère sœur Kenza. A toute ma famille. A mes amis surtout mes chères amies Nawel et Sílía qui étaient toujours proche de moi et tous mes collègues.

A ma binôme Meriem, qui est plus d'un binôme c'est une vraie amie je la souhaite de la réussite pendant toute sa vie.

A mon cher fiancé Halim, pour sa patience et pour tous ses encouragements durant ma période d'études.

A tous ceux que j'aime

HADDADI Sara

Au nom de dieu clément et miséricordieux

Dédicace



Avec l'expression de ma reconnaissance, je dédie ce modeste travail

A ma chère mère,

A mon cher père,

Qui n'ont jamais cessé de formuler des prières à mon égard, de me soutenir et encourager durant ces années d'études pour que je puisse atteindre mes objectifs.

A mon cher frère, Yasser

A mes très chères sœurs, Amina, Hadjer

Khawla, Bouchra et Marya

Qui me donnent de l'amour et de la vivacité, pour leur soutien moral et leurs conseils précieux durant mon parcours.

A ma chère binôme, Sara

A tous mes amis et amies (El Madani, Louiza, Khedidja, Wissem)

Qui m'ont chaleureusement supporté et encouragé tout au long de mon parcours, et à qui je souhaite plus de succès.

A tous ceux que j'aime

BAAR Meriem

Table des matières

Introduction générale

Introduction générale	1
Organisation du mémoire	2

1 Chapitre 1 La voix sur IP : protocoles, réseaux, avantages et contraintes

1.1 Introduction	4
1.1.1 Historique	4
1.1.2 Définition du la Voix IP	4
1.2 Architecture et scénarios	5
1.2.1 Architecture	5
1.2.2 Scénarios.....	8
a. Ordinateur à ordinateur.....	8
b. Ordinateur vers le téléphone.....	8
c. Téléphone à téléphone	9
1.3 Principe de fonctionnement.....	9
1.4 Protocoles	10
1.4.1 Protocoles de signalisation	11
a. Protocole H.323	11
b. Protocole SIP.....	11
1.4.2 Protocoles de transport	12
a. Protocoles TCP/UDP	12
b. Protocoles RTP/RTCP	12
1.5 Les contraintes.....	13
1.5.1 La gigue	14
1.5.2 Latence.....	14
1.5.3 Perte de paquets	15
1.5.4 Le délai de transit.....	15
1.6 Conclusion.....	15

2 Chapitre 2 Codage de la parole : analyse par synthèse

2.1 Introduction	16
2.2 Définition.....	16
2.3 Codage basé sur la prédiction linéaire : analyse par synthèse.....	18
2.3.1 Prédiction à court terme.....	18

Table des matières

a.	Détermination des paramètres du filtre de prédiction	20
b.	Quantification des coefficients des paires de rai spectrales	20
2.3.2	Pondération perceptuelle des formants	21
2.3.3	Prédiction à long terme	22
2.3.4	Modélisation du signal	23
a.	Estimation du délai tonal en boucle ouverte	23
b.	Mise en forme du bruit harmonique	23
c.	Estimation de la tonie en boucle fermée	23
2.3.5	Le Répertoire adaptatif conventionnel	23
2.3.6	Le répertoire fixe des excitations	24
2.4	Prédiction linéaire à excitation multi impulsienne (MP-LPC)	24
2.5	Prédiction linéaire excité par code algébrique (ACELP)	25
2.6	Codeur de la parole basé sur CELP utilisé en VoIP	26
2.6.1	Le choix du codeur G.723.1	28
2.6.2	Le score MOS	29
2.6.3	Calcul du facteur R de l'E-modèle	29
2.6.4	Avantages et inconvénients	30
2.7	Les méthodes de masquage de pertes de paquets (PLC)	30
2.7.1	Techniques basés sur la réception	31
2.7.1.1	Réplication de forme d'onde	31
2.7.1.2	Utilisation de modèle de parole	32
2.7.2	Techniques basés sur l'émission	32
2.7.2.1	Correction d'erreur en avale	32
2.7.2.2	Entrelacement de trames	33
2.7.2.3	Retransmission	34
2.8	Conclusion	34
3	Chapitre 3 Amélioration de la qualité de service : simulation	
3.1	Introduction	36
3.2	Qualité de service	36
3.3	Types de signaux	37
3.3.1	Le son voisé	37
3.3.2	Le son non voisé	38
3.4	Modèle et outils de mesure	38
3.4.1	Modèle Gilbert Elliot	38
3.4.2	Outils de mesure	39

Table des matières

3.5	Génération du signal d'excitation.....	40
3.5.1	Génération du signal d'excitation avec des zéros.....	40
3.5.2	Génération du signal d'excitation avec la trame passée	41
3.5.3	Génération du signal d'excitation PLC du codec G.723.1	43
3.6	Régénération du signal d'excitation par un filtrage adaptatif.....	45
3.6.1	Algorithme RLS.....	45
3.6.2	Principe de base des méthodes proposées	46
a.	Expérience 1 : RLS-PLC	46
b.	Expérience 2 : RLS-TP.....	50
c.	Expérience 3 : RLS comme FEC.....	52
3.6.3	Résultats généraux et discussion	56
3.7	Conclusion.....	57
 Conclusion générale		
	Conclusion générale	59
 Bibliographie		
		61

Liste des figures

Figure 1.1. Architecture générale de la VoIP	6
Figure 1.2. Scénario PC à PC	8
Figure 1.3. Scénario PC à téléphone	9
Figure 1.4. Téléphone à téléphone	9
Figure 1.5. En-tête d'un datagramme RTP	12
Figure 1.6. En-tête d'un datagramme RTCP	13
Figure 2.1. Un système de codage de la voix	17
Figure 2.2. Analyse et Synthèse LP	19
Figure 2.3. Spectres du filtre LP et du filtre de pondération des formants	22
Figure 2.4. Système d'analyse par Synthèse à excitation multi-impulsions (MP-LPC)	25
Figure 2.5. Fonctionnement à l'encodeur du modèle ACELP.....	26
Figure 2.6. Classification de techniques de masquage basées à l'émetteur	32
Figure 2.7. Exemple du principe FEC	33
Figure 2.8. Exemple du principe d'entrelacement de trames	34
Figure 3.1. Signal de parole d'un son voisé	37
Figure 3.2. Signal de parole d'un son non voisé	38
Figure 3.3. Modèle de Gilbert-Elliot	39
Figure 3.4. L'excitation générée par implémentation des zéros	40
Figure 3.5. Excitation générée par implémentation de la trame passée	42
Figure 3.6. Excitation générée par le PLC du G723.1, trame passée non voisée	43
Figure 3.7. Excitation générée par le PLC du G723.1, trame passée voisée	44
Figure 3.8. Schéma bloc de décodeur modifié pour la méthode proposé (RLS-PLC)	46

Liste des figures

Figure 3.9. Signal estimé par RLS-PLC	49
Figure 3.10. Signal estimé par RLS-TP	52
Figure 3.11. Schéma du codeur modifié pour la méthode RLS-PLC comme FEC	53
Figure 3.12. Schéma du décodeur modifié pour la resynchronisation du ACB	53
Figure 3.13. L'évolution du PESQ pour chaque méthode	57

Liste des tableaux

Tableau 1.1. Délai requis pour la VoIP en fonction de la classe d'appartenance.....	14
Tableau 2.1. Le bilan de débit des différents codecs	28
Tableau 2.2. La variation du score MOS et leurs indication	29
Tableau 3.1 Les résultats du PESQ avec implémentation des zéros	41
Tableau 3.2 Les résultats du PESQ avec la trame passée	42
Tableau 3.3 Les résultats du PESQ avec le PLC de G.723.1	44
Tableau 3.4 Les résultats du PESQ en fonction du facteur d'oubli, λ , et de la longueur, L, de RLS-PLC associés aux différents taux de pertes de paquets (PLR)	48
Tableau 3.5 Les résultats du PESQ avec RLS-PLC	49
Tableau 3.6 Les résultats du PESQ en fonction du facteur d'oubli, λ , et de la longueur, L, de RLS-TP associés aux différents taux de pertes de paquets (PLR)	50
Tableau 3.7 Les résultats du PESQ pour la méthode RLS-TP	51
Tableau 3.8 Les résultats du PESQ en fonction du facteur d'oubli, λ et de la longueur, L	54
Tableau 3.9 Les résultats du PESQ avec RLS-PLC comme FEC	54
Tableau 3.10 Les résultats du PESQ en fonction du facteur d'oubli, λ et de la longueur, L	55
Tableau 3.11 Les résultats du PESQ avec la méthode RLS-TP comme FEC	56
Tableau 3.12 Les résultats des méthodes utilisées	56

Liste des abréviations

Les termes les plus couramment utilisés dans ce mémoire et dans le domaine de la voix sur IP généralement sont :

A

ACB	Adaptive CodeBook
ACELP	Algebraic Code Excited Linear Prediction
ADSL	Asymmetric Digital Subscriber Line

B

BE	Best Effort
-----------	-------------

C

CAN	Convertisseur Analogique-Numérique
CELP	Code Excited Linear Prediction
CNA	Convertisseur Numérique- Analogique
CNG	Comfort Noise Generation
Codec	Coder Decoder
cpl	Conditional Probability of Loss
CS-ACELP	Conjugate structure Algebraic Code Excited Linear Prediction
CSRC	Contributing Source Real-time Control protocol

D

DTMF	Dual-Tone Multi-Frequency
DTX	Discontinuous transmission

E

ETSI	European Telecommunication Standards Institute
-------------	--

F

FAI	Fournisseurs d'Accès à Internet
------------	---------------------------------

FCB Fixed CodeBook

FEC Forward Error Correction

G

GAFAM Google, Appel, Facebook, Amazon, Microsoft

I

IETF Internet Engineering Task Force

IoT Internet of Things

IP Internet Protocol

IPBX Internet Protocol Private Branch eXchange

ISP Internet Service Provider

ITSP Internet Telephony Service Provider

ITU International Telecommunication Union

K

Kbps kilo bit per second

L

LAN Local Area Network

LP Linear Prediction

LPC Linear Prediction Coding

LSF Line Spectral Frequencies

LTP Long Term Prediction

M

MMUSIC Multiparty Multimedia Session Control

MP-LPC Multi-Pulse Linear Prediction Coding

MELP Mixed Excitation Linear Prediction

MOS Mean Opinion Score

O

OLA	OverLap and Add
OSI	Open Systems Interconnection

P

PABX	Private Automatic Branch eXchange
PCM	Pulse Code Modulation
PESQ	Perceptual Evaluation of Speech Quality
PLC	Packet Loss Concealment

Q

QOS	Quality of Service
------------	--------------------

R

RLS	Recursive Least Square
RNIS	Réseau Numérique à Intégration de Services
RTC	Réseau Téléphonique Commuté
RTCP	Real-Time Control Protocol
RTP	Real time Transport Protocol

S

SSRC	Synchronization Source Real-time Control protocol
-------------	---

T

TCP	Transport Control Protocol
TDM	Time division multiplexing
TIC	Technologies de l'information et de la communication
ToIP	Telephony over Internet Protocol
TIMIT	Texas Instruments and Massachusetts Institut of Technology

Liste des abréviations

U

UDP	User Datagram Protocol
upl	Unconditional Probability of Loss

V

VAD	Voice Activity Detection
VoIP	Voice over Internet Protocol
VPN	Virtual Private Network

W

Wi-Fi	Wireless Fidelity
WAN	Wide Area Network

Introduction

Générale

Introduction générale

Depuis des siècles, l'être humain n'a cessé d'élaborer ses propres stratégies et moyens de communication afin de pouvoir transmettre un message à son semblable, et cela grâce à des efforts intellectuels et physique très remarquables, à un point qu'il en est arrivé à mettre en place un réseau de téléphonie classique assez satisfaisant pour qu'il soit à ce jour applicable. Néanmoins, ce réseau présente des inconvénients majeurs en matière de coût et d'infrastructure vu la demande croissante et les exigences du marché des télécommunications. Durant les dernières années, le réseau Internet a évolué d'une manière extraordinaire pour devenir un réseau disponible mais aussi indispensable pour l'accès à l'information et la transmission de données. De ce fait, nous vivons une large migration de multiples services vers la grande toile d'Internet. En effet, d'innombrables applications de communications à distance ont été développées, tel que, la transmission de la voix et de la vidéo, qui sont devenues très populaire. Ainsi, chaque utilisateur possédant un ordinateur équipé d'une carte vocal, d'un microphone et d'un logiciel multimédia a la possibilité de se connecter au réseau et effectuer des appels à très bas coût. Par conséquent, plusieurs protocoles complémentaires sont développés et ajoutés pour superviser le transport de données en temps réel, afin de resynchroniser les paquets et éviter les retards de congestion et les pertes. Ces protocoles sont essentiellement, RTP, RTCP, H.323 et SIP parmi d'autres.

L'exploitation croissante des réseaux basés sur le protocole IP pour les nouveaux services de communications y compris la téléphonie, sont devenus des domaines importants non seulement pour l'industrie des télécommunications mais aussi pour de nombreuses institutions commerciales ou gouvernementales. De toute évidence, il est très avantageux de réaliser un réseau privé pour assurer le transport conjoint de la voix et des données, avec un partage de bande passante. Cela constitue un point de convergence entre deux technologies distinctes, à savoir, la commutation de circuits utilisée actuellement dans les réseaux classiques de téléphonie et la commutation de paquets pour la transmission de la parole et de l'image. Cette dernière est appelée, Voice over IP, elle consiste généralement à transporter de la voix traditionnelle sur un réseau IP. La migration des entreprises vers ce genre de technologie est dans le but de minimiser le coût des communications tout en exploitant le même réseau pour offrir des services de données, de voix, et d'images, et simplifier les coûts de configuration et d'assistance. En dehors des systèmes de codage et de décodage très complexes écartés, cette tâche peut être menée avec un matériel très économique, et peuvent être disponibles chez tout

Introduction générale

les utilisateurs au sein de la même entreprise. A ce propos, les deux codecs à basés sur des trames, tout comme, le G.723.1 et le G.729 sont très convenables pour la téléphonie IP. D'ailleurs, ils peuvent fournir une qualité téléphonique de la parole avec de faibles débits binaires, d'ordre de 5.3 à 6.3 kbps et 8 kbps, pour chacun respectivement. Contrairement au codec PCM conventionnel à lui seul consomme 64 kbps de la bande passante disponible. Cependant, les exigences sur la capacité du réseau pour une diffusion à grande échelle peuvent être réduites considérablement. Actuellement, les réseaux à commutation de paquets sont basés généralement sur le principe dit, best effort, qui ne garantit pas l'arrivée à temps de tous les paquets envoyés. De plus, même si les paquets seront reçus, il est fortement possible qu'ils soient retardés ou mal ordonnés. Cela introduit de diverses influences désavantageuses sur la qualité objective et subjective de la parole. Surtout, lorsque les nœuds de réseau sont encombrés de paquets ce qui implique leur suppression ou abandon. Dû à la contrainte d'une transmission en temps réel pour la parole, il est généralement impossible que le récepteur demande la retransmission des paquets perdus.

Ajouter aux problèmes du trafic, le codage prédictif adopté dans la norme G.723.1 [1], rend la tâche plus complexe vu sa vulnérabilité aux pertes de paquets. Malheureusement, ces pertes induisent une désynchronisation entre le codeur et le décodeur. D'où des erreurs de décodage seront introduites et se propagent aussi sur les trames bien reçues.

C'est pour cela, il est nécessaire d'implémenter une technique de masquage de pertes des paquets perdus, appelée, PLC, pour mieux estimer le segment du signal de parole des trames perdues afin d'offrir une qualité de service satisfaisante aux utilisateurs.

Donc, nous devons proposer des solutions pour diminuer ces dégradations dans la qualité de service. En effet, dans notre projet, nous avons proposé des méthodes afin d'améliorer l'algorithme de masquage de pertes de paquets, PLC, de la norme G723.1.

Organisation du mémoire

Le mémoire s'articule autour de trois chapitres, à savoir,

Le premier chapitre est consacré à une présentation globale de la transmission de la voix sur les réseaux IP, avec les modules constituants de ce réseau avec ses contraintes et les différents protocoles utilisés dans la gestion de la voix sur IP.

Le deuxième chapitre est axé sur le codage prédictif de la parole de l'analyse par synthèse. En outre, ce chapitre se scinde en deux grandes sections qui sont :

Introduction générale

- Dans la première, on étudie en détail la prédiction linéaire basée sur l'analyse par synthèse, la prédiction MP-LPC et le système de fonctionnement de codeur ACELP, ainsi que, les différents codeurs basés sur CELP, avec une description générale du codec G.723.1.
- La deuxième section, décrit la technique de masquage de pertes, PLC, et ces différents algorithmes, dont certains sont utilisés en émission et d'autres en réception.

Et enfin, le dernier chapitre qui s'ajoutera pour présenter nos réalisations et méthodes PLC proposées sur le codec G.723.1 afin d'améliorer la qualité de service de la technologie VoIP. A ce propos, nous commencerons par une description du model de transmission adopté dans nos tests. Ainsi, nous avons défini le canal de transmission avec pertes utilisé pour simuler un réseau IP et les outils utilisés pour juger la qualité objective, PESQ. De plus, des analyses et discussions sont ajoutées pour les multiples graphes, tableaux, figures et résultats obtenus. Et nous terminerons par une conclusion générale sur l'ensemble du travail accompli.

Chapitre 1

La voix sur IP : Protocoles, réseaux, avantages et contraintes

1.1 Introduction

Ces dix dernières années, la VoIP est devenue presque l'unique moyen de communication dans la téléphonie moderne. Certes la plus utilisée, mais c'est aussi l'outil numérique le plus populaire.

En effet, une grande majorité des entreprises modernes dans le marché numérique et des TICs les plus puissantes au monde, les cinq sœurs du numérique, connues sous l'acronyme GAFAM (Google, Apple, Facebook, Amazon et Microsoft) l'utilisent et l'exploitent davantage. Et ce même, à travers l'offre d'une variété d'applications mobiles, telles que : *Messenger, Skype, Vibre, WhatsApp, Zoom, Team et Duo,...*etc. Et surtout, avec les nouveaux défis du télétravail que nous avons vécu plus récemment liés à la crise sanitaire mondiale suite à l'apparition de la pandémie Covid-19.

De ce fait, elle est devenue une technologie de communication vocale en pleine émergence. Elle fait partie d'un tournant dans le monde de la communication. De plus, la convergence du triple Play (voix, données et vidéo) fait partie des enjeux principaux des acteurs de la télécommunication aujourd'hui. Plus récemment l'Internet des objets (IoT), qui s'est étendue partiellement dans l'espace de travail de chaque organisation ou établissement et entreprise, voyant le trafic total basé sur un transport réseau de paquets IP surpassant le trafic traditionnel de la téléphonie RTC (Réseau à Commutation de Circuits).

Il est devenu clair et remarquable que dans le sillage de cette avancée technologique, les opérateurs, entreprises ou fournisseurs de TICs se livrent une bataille sans merci pour bénéficier davantage des avantages du réseau de transport unique IP pour introduire de nouveaux services voix et vidéo pour tirer profit de ce premier acteur du marché numérique.

1.1.1 Historique

Pendant plusieurs décennies, la transmission analogique de la voix fut la seule technologie maîtrisée. Mais au milieu du vingtième siècle, grâce à l'application de nouvelles techniques d'échantillonnage, de quantification, de compression et de codage qui ont rendu possible la transmission de gros volumes de données pour les applications multimédias (données, voix et vidéo), nécessitant de débits élevés a donné naissance à de nouvelles technologies telles que le RNIS, l'ADSL, la téléphonie mobile et le fameux réseau sans frontières, avec ces technologies innovantes, Internet.

Dans les années 80s, une première évolution majeure a été le passage à la transmission numérique (TDM). Les données émises par le signal électrique sont converties en une suite numérique, ces derniers sont exploitables sur un ordinateur via Internet. Ce fût en 1996 la naissance de la première version voix sur IP appelée H.323 [2]. Normalisée par l'organisation internationale de standardisation dans les télécoms ITU-T sur la base de la signalisation voix RNIS (Q.931) [2].

Cependant, il est remarquable que, les fournisseurs d'accès à Internet (FAI) adoptent toute technologie nouvelle plus performante liée à la VoIP, qui se sert d'internet au lieu des réseaux téléphoniques traditionnels pour passer des appels téléphoniques gratuitement. Depuis son invention, la téléphonie mobile ne cesse de marquer de grands pas dans le domaine des télécommunications, et d'envahir le monde des TICs. Aujourd'hui, il y a plus de 1.2 milliards de lignes de téléphone fixes, et plus de 140 milliards de minutes de communication téléphoniques internationales chaque année [3].

1.1.2 Définition de la voix IP

VoIP est un abrégé de l'anglais Voice over Internet Protocol. La transmission de la voix sur les réseaux fondés sur le protocole Internet (IP), est une technique qui permet de communiquer par la voix (ou via des flux multimédia : audio ou vidéo) sur des réseaux compatibles IP, qu'ils s'agissent de réseaux privés ou publics, filaires (câble/ADSL/fibre optique) ou non (satellite, réseaux mobile, Wi-Fi). [4]

La VoIP (Voice over IP), comme son nom l'indique, est le fait de transmettre de la voix sur un réseau IP qui transporte les données sous forme de paquets. La voix soumise à des traitements spécifiques afin qu'elle puisse être envoyée sur un réseau IP, elle est digitalisée, compressée puis envoyée au récepteur par paquets de 10 à 30 ms [4] de données. Les données reçues par le destinataire sont décompressées et converties en voix audible.

La tendance récente est de remplacer les réseaux de commutation de circuits, tels que le RTC public qui utilise une connexion physique pendant toute la durée d'une communication, par une transmission de la voix par des paquets sur les réseaux internet, aucune connexion physique permanente n'est établie.

1.2 Architecture et scénarios

1.2.1 Architecture :

La VoIP étant une nouvelle technologie de communication, elle n'a pas encore de standard unique. En effet, chaque constructeur apporte ses normes et ses fonctionnalités à ses solutions.

Il existe donc plusieurs approches pour offrir des services de téléphonie et de visiophonie sur des réseaux IP. Certaines placent l'intelligence dans le réseau alors que d'autres préfèrent une approche Peer to Peer avec l'intelligence répartie à la périphérie (terminal de téléphonie IP, passerelle avec le réseau RTC... etc.) [5].

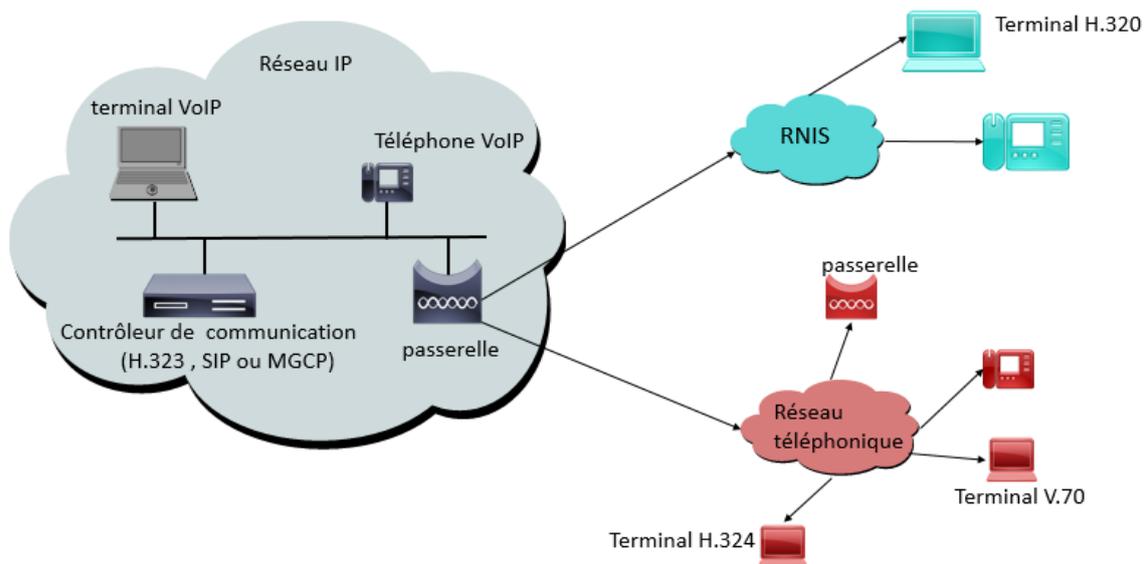


Fig. 1.1. Architecture générale de la VoIP

De façon générale, la topologie d'un réseau de téléphonie IP comprend toujours des terminaux, un serveur de communication et une passerelle, appelé aussi Gateway, vers les autres réseaux. Chaque norme a ensuite ses avantages et inconvénients pour garantir une meilleure, plus ou moins bonne qualité de service (QoS). Tout en retrouvant les éléments communs suivants :

- **Le routeur :** assure la commutation ou encore acheminement de paquets d'un réseau vers un autre réseau. Certains routeurs permettent de simuler un Gatekeeper grâce à l'ajout de cartes spécialisées supportant les protocoles VoIP [6].
- **Le PABX :** (Private Automatic Branch eXchange) c'est le commutateur du réseau téléphonique classique. Il permet de faire le lien entre la passerelle ou le routeur et le réseau RTC. Une mise à jour du PABX est aussi nécessaire. Si tout le réseau devient IP, il n'y a plus besoin de ce matériel, on parle alors de IPBX.
- **Les terminaux :** sont généralement de type logiciel (software phone) ou matériel (hardphone) [6]. Le softphone est un système de téléphonie via VoIP installé dans

les ordinateurs des utilisateurs, spécialisé pour les appels téléphoniques en temps réel, sur Internet et les réseaux d'entreprises. Pour qu'ils soient opérationnels, les softphones nécessitent une connexion Internet haute débit. Les hardphones sont des téléphones IP qui utilisent la VoIP pour permettre des appels téléphoniques sur un réseau IP.

▪ **Gateway et Gatekeeper** : Les passerelles ou Gateway en téléphonie IP sont des ordinateurs qui fournissent une interface où se fait la convergence entre les réseaux téléphoniques commutés (RTC) et les réseaux basés sur la commutation de paquets, assurent les fonctions de codage, décodage et la mise en paquet de la voix et disposent d'interface d'interconnexion analogique et numérique. C'est une partie essentielle de l'architecture du réseau de la téléphonie IP. Le Gatekeeper est l'élément qui fournit de l'intelligence à la passerelle, et le compagnon logiciel de la Gateway. Le Gatekeeper (garde-barrière) répond aux aspects suivants de la téléphonie IP :

- Le routage des appels : en effet, le Gatekeeper est responsable de la fonction de routage. Non seulement, il doit tester si l'appel est permis et faire la résolution d'adresse, mais il doit aussi rediriger l'appel vers le bon client ou la bonne passerelle.
- Administration de la bande passante : le Gatekeeper alloue une certaine quantité de bande passante pour un appel et sélectionne les codecs à utiliser.
- Tolérance aux fautes et sécurité : le Gatekeeper est aussi responsable de la sécurité dans un réseau de téléphonie IP. Il doit gérer les redondances des passerelles afin de faire aboutir tout appel. Il connaît à tout moment l'état de chaque passerelle et route les appels vers les passerelles accessibles et qui ont des ports libres.
- Gestion des différentes Gateway : dans un réseau de téléphonie IP, il peut y avoir beaucoup de Gateway. Le Gatekeeper, de part ses fonctionnalités de routage et de sécurité, doit gérer ces Gateway pour faire en sorte que tout appel atteigne sa destination avec la meilleure qualité de service (QoS) possible.

1.2.2 Scénarios

Une communication dans un système de téléphonie VoIP est établie selon trois types :

a. *Ordinateur à Ordinateur*

Dans ce scénario le but sera de transformer chaque hôte en un poste téléphonique, et consiste à équiper chacun d'une carte son full-duplex, pour garantir une conversation simultanée ; d'un microphone, d'un haut-parleur et d'un logiciel de la VoIP compatible. Cette configuration est fréquemment couplée à des fonctionnalités de visioconférence à partir d'une webcam connectée à l'ordinateur. Il existe deux modes de connexion possible pour ce cas sont ainsi

- **connexion directe** : en composant l'adresse IP du correspondant, les deux usagers doivent ainsi fixer un rendez-vous préalable, à moins qu'ils soient connectés en permanence.
- **connexion serveur** : en sélectionnant le correspondant sur une liste d'utilisateurs en ligne. Si quelqu'un se connecte au réseau, ses coordonnées (email, adresse IP, ...etc.) sont automatiquement inscrites dans l'annuaire en ligne.

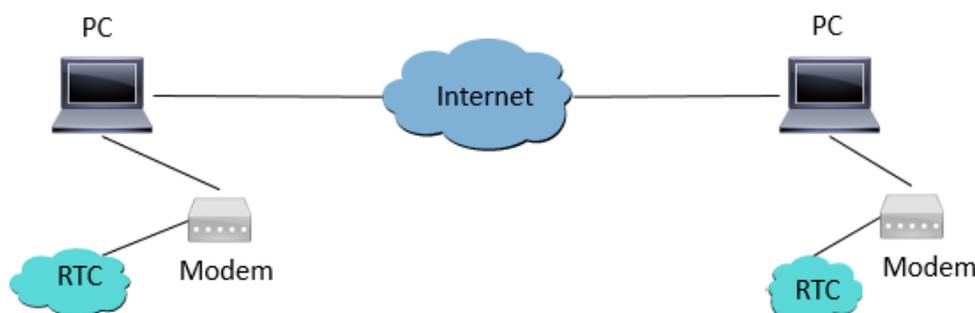


Fig. 1.2. Scénario PC à PC

b. *Ordinateur vers le téléphone*

Dans ce scénario, l'un des utilisateurs est un ordinateur qui se connecte à Internet via un fournisseur d'accès à Internet (FAI), et il doit disposer d'un logiciel approprié. Tandis que l'autre utilisateur est un abonné normal d'un réseau téléphonique fixe ou mobile. Lorsque l'ordinateur souhaite appeler un correspondant sur un poste téléphonique, il doit d'abord se connecter à internet de manière classique grâce au réseau de voix FAI, (ISP-Internet Service Provider en

anglais). Une fois connecté, il utilise le service FAI, qui opère une "passerelle" permettant d'accéder au plus près du central téléphonique de l'abonné demandé. [7]

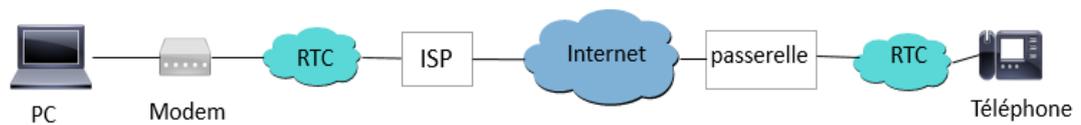


Fig. 1.3. Scénario PC à téléphone

c. Téléphone à téléphone

Dans ce cas, l'appelant et l'appelé sont tous les deux des abonnés du réseau téléphonique commuté public (RTCP) et utilisent de manière classique (analogique) leur appareil téléphonique pour la communication vocal. Il s'agit de la technologie ToIP. L'objectif donc est de fusionner sur un même réseau, un réseau téléphonique commuté (RTC) et d'un réseau informatique, on mises en place soit des passerelles ou des boitiers. Des téléphones classiques doivent en effet être alimenté par courant au contraire de téléphone IP. Il est capable de numériser la voix pour la transmettre sur des réseaux IP et peut, à l'inverse, rassembler les paquets entrants pour interpréter la voix reçue. La téléphonie sur IP circule sur des réseaux privés (LAN ou VPN) ou publics.



Fig. 1.4. Téléphone à téléphone

1.3 Principe de fonctionnement

Le principe de la VoIP est la transmission de données parlées par un réseau IP. La conversation téléphonique VoIP ne circule plus par les fils de cuivre du réseau téléphonique standard analogique, mais elle passe par un réseau de données numérique en empruntant les

câbles Ethernet au départ et à l'arrivée. Les données transmises sont sous forme de paquets (un lot de données numérique), en clair, une série de 0 et 1 sous forme d'impulsions électriques ou optiques. Dans la communication par VoIP, la voix est traitée de la même manière que les autres données numériques transmises via le réseau Internet. Elle est d'abord captée par le microphone sous forme d'un signal analogique, puis envoyée via le réseau Internet. Ce signal analogique doit être converti en un signal numérique à l'aide d'un CAN, puis il est ensuite comprimé par un codec et transmis. Pour être utilisable, il doit être transformé de nouveau en un signal analogique, à l'aide d'un CNA.

Au départ, la modulation sonore produite par la voix est numérisée en temps réel ensuite compressé, mais elle ne circule pas en flux continu. L'écart est très bref pour que cela ne puisse pas avoir d'impacts dans la conversation, la recommandation G.114 de l'ITU préconise un temps de transmission inférieur à 300 ms pour avoir une bonne qualité de conversation [5]. Ensuite, les données numériques sont regroupées et envoyées régulièrement dans le réseau IP en wagons d'informations numériques successifs, ce sont les paquets. A l'arrivée, c'est l'inverse qui se passe. Il est à noter que les paquets renferment en plus des données vocales l'adresse réseau des deux interlocuteurs. Le format numérique est plus facile à contrôler, il peut être compressé, routé et converti en un nouveau format meilleur. Le signal numérique est plus tolérant au bruit que l'analogique.

En plus, le son capté par un terminal VoIP est numérisé. Les données sont transmises par le réseau de deux manières selon le type de terminal du destinataire :

- Si le destinataire dispose d'un terminal VoIP la transmission des données se fait par Internet.
- Si le destinataire possède un terminal traditionnel, une passerelle entre l'Internet et le réseau téléphonique commuté classique permet au destinataire d'appel d'entendre et de répondre à son interlocuteur.

Ainsi, le destinataire peut enfin écouter le message émis par l'émetteur sur son terminal.

1.4 Protocoles

Avant de pouvoir communiquer directement, les membres de la discussion doivent établir de nombreux protocoles pour la démarrer, chacun pour une fonction précise. Les principaux protocoles utilisés sont :

1.4.1 Protocoles de signalisation

La signalisation désigne la transmission d'un ensemble de signaux et d'informations de contrôle échangés entre les intervenants d'une communication. Ces intervenants peuvent être des entités en bout de liaison (terminaux) ou des entités intermédiaires de contrôle et de gestion des communications [8]. Les deux protocoles les plus utilisés actuellement dans la signalisation VoIP présents sur le réseau, sont le H.323 et le SIP.

a- Protocol H.323

Le standard H.323 fournit, depuis son approbation en 1996, un cadre pour les communications audio, vidéo et de données sur les réseaux IP. Il a été initié par l'ITU-T pour le traitement de la signalisation des données multimédias avec de fortes contraintes temporelles dans les réseaux qui ne garantissent pas une qualité de service (QoS), tels qu'IP ou IPX sur Ethernet, Fast-Ethernet et Token-Ring.

Le H.323 n'est en réalité que la référence du protocole, il fait partie de la série H.32x. Son nom complet est Packet-based Multimedia Communications Systems, ou (Systèmes de communication multimédia fonctionnant en mode paquet) [8]. Le protocole H.323 est actuellement le plus utilisé par les constructeurs, il normalise le type du codage, la procédure de la signalisation et les protocoles en temps réel sur IP.

b- Protocol SIP

SIP (Session Initiation Protocol) est un standard proposé par le groupe de travail de l'IETF MMUSIC (Multiparty MULTimedia Session Control) qui a été conçu pour établir, maintenir, modifier et terminer des sessions multimédias. Le protocole n'assure pas le transport des données utiles échangées durant la session comme la voix ou la vidéo, mais assure simplement la signalisation. SIP étant indépendant de la transmission des données, tout type de protocole peut être utilisé pour cet échange.

Le protocole SIP se situe au niveau de la couche applicative (Niveau 7) du modèle de référence OSI et fonctionne selon une architecture client-serveur. Le client émettant des requêtes et le serveur exécutant en réponse les actions sollicitées par le client [8]. Il remplace progressivement le protocole H.323.

1.4.2 Protocoles de transport

a- Protocoles TCP/UDP

Les protocoles TCP (Transport Control Protocol) et UDP (User Datagram Protocol) supportent la transmission de donnée sur une large gamme de réseaux, depuis les liaisons filaires câblées jusqu'aux réseaux commutés. Le TCP assure la transmission de données fiable et orienté connexion avec une détection et correction d'erreur de bout en bout, tandis que l'UDP offre un service de transmission de datagrammes sans connexion et sans correction.

Ces protocoles s'interfaçent juste au-dessus du protocole IP et fournissent un service de transfert de segments de données ou de voix, encapsulés dans un paquet internet.

b- Protocoles RTP/RTCP

▪ Protocole RTP

Le RTP (Real time Transport Protocol), standardisé en 1996 et développé par l'IETF afin de faciliter le transport temps réel de bout en bout, des flots de données audio et vidéo sur les réseaux IP. Il permet ainsi, de détecter les pertes de paquet et d'en informer la source ou encore d'identifier le contenu des données pour leur associer un transport sécurisé, et assure l'horodatage des paquets et possibilité de resynchronisation des flux par récepteur. RTP est un protocole qui se situe au niveau de l'application et qui utilise les protocoles sous-jacents de transport TCP ou UDP [6]. Mais l'utilisation de RTP se fait généralement au-dessus d'UDP ce qui permet d'atteindre plus facilement le temps réel. Le format des paquets RTP se présentent comme suit :

V=2	P	X	CC	M	PT	Sequence number
Horodatage (time stamp)						
Synchronization Source Identifier (SSRC)						
Contributing Source Identifier (CSRC)						

Fig. 1.5. En-tête d'un datagramme RTP

▪ Le protocole RTCP

Le protocole RTCP (Real-time Transport Control Protocol) c'est un protocole de contrôle et de supervision du réseau. Son objectif est d'offrir aux participants d'une session une vision sur l'état du réseau et de s'y adapter de façon dynamique [6].

Ce protocole est généralement utilisé avec le protocole RTP, C'est le protocole UDP (par exemple) qui permet le multiplexage des paquets de données RTP et des paquets de contrôle RTCP. Afin de perfectionner la qualité de distribution, le RTCP fournit un rapport sur cette dernière (gigue mesurée, taux moyen de perte paquet, délai de bout en bout ...etc.).

Ce rapport est envoyé périodiquement de telle sorte que les intervenants disposent d'une mise à jour fréquente de l'état du réseau. Parmi les principales fonctions qu'offre le protocole RTCP :

- Une synchronisation supplémentaire entre les médias.
- L'identification des participants à une session.
- Le contrôle de la session : les participants indiquent leurs départs d'une conférence téléphonique et ils donnent une indication sur leurs comportements.

La figure ci-dessous présente le format des paquets RTCP

	RC	Type	Length
SSRC of packet sender			
SSRC of first source			
Fraction lost		Cumulative number of packet lost	
Interarrival jitter			
Last SR			
Delay since last SR			

Fig. 1.6. En-tête d'un datagramme RTCP

En revanche, RTP et RTCP ne permettent pas d'obtenir des transmissions en temps réel avec une mauvaise qualité pour la VoIP. En effet, RTP ne procure pas de réservation de ressources sur le réseau, de fiabilisation des échanges ou de garantie dans le délai de livraison.

1.5 Les contraintes de la VoIP

Malgré les avantages précieux de la VoIP, mais il faut l'en être conscient de ses limites. La transmission de la voix sur réseau IP se fait par transmission de paquets, ce protocole peut être

sujet à des congestions, auquel cas des paquets seront éliminés aléatoirement et des trames de signal de paroles seront perdus. Les principales causes de ces pertes associées à la VoIP sont :

1.5.1 La gigue

La gigue est la variance statistique du délai de transmission. En d'autres termes, elle mesure la variation temporelle entre le moment où deux paquets auraient dû arriver et le moment de leur arrivée effective. Cette irrégularité d'arrivée des paquets est due à de multiples raisons dont : l'encapsulation des paquets IP dans les protocoles supportés, la charge du réseau à un instant donné, la variation des chemins empruntés dans le réseau, ...etc.

Pour restituer un flux synchrone à l'arrivée, on installe des buffers de compensation de gigue. Cela introduit cependant un retard mais qui est jugé moins gênant que la perte de trame [9]. Afin de garder une qualité acceptable, la taille de ces buffers doit être soigneusement définie, et si possible adaptée de manière dynamique aux conditions du réseau.

1.5.2 La latence

La maîtrise du délai de transmission est un élément essentiel pour bénéficier d'un véritable mode conversationnel et minimiser la perception d'écho. En effet, la latence c'est le retard pour qu'un paquet de données soit transmis de l'émetteur au destinataire et renvoyé à l'émetteur. Plus elle sera élevée, plus elle affectera les performances du réseau.

Les chiffres du tableau 1.1 (tirés de la recommandation UIT-T G.114 [10]) sont donnés à titre indicatif pour préciser les classes de qualité et d'interactivité en fonction du retard de transmission dans une conversation téléphonique. Ces chiffres concernent le délai total de traitement et le temps de transmission de l'information sur le réseau.

Classe n°	Délai	Commentaires
1	0 à 150 ms	Acceptable pour la plupart des conversations
2	150 à 300 ms	Acceptable pour des communications faiblement interactives
3	300 à 700 ms	Devient pratiquement une communication semi duplex
4	Au-delà de 700ms	Inutilisable même pour une conversation semi duplex

Tableau. 1.1. Délai requis pour la VoIP en fonction de la classe d'appartenance

1.5.3 Perte de paquets

L'information qui est transmise à travers les réseaux IP circule sous forme de pièces d'information c'est à-dire paquets. Ces derniers contiennent de l'information nécessaire pour qu'au moment de sa réception, le destinataire soit capable d'ordonner et d'utiliser cette information.

Les pertes de paquets veulent dire qu'une ou plusieurs pièces d'information n'ont pas arrivé à leur destination de façon correcte, se traduisent par des ruptures au niveau de la conversation avec des dégradations du signal de parole. La perte d'un paquet se produit en effet généralement lorsqu'il y a une congestion sur un lien de transmission, qui provoque un débordement des mémoires tampons d'un routeur.

1.5.4 Le délai de transit

Le délai de transit (end-to-end Delay dans la dénomination anglo-saxonne) est un des paramètres critiques influençant fortement la QoS de voix sur IP. C'est le temps que va mettre en moyenne un paquet IP contenant un échantillon de voix pour traverser l'infrastructure entre deux interlocuteurs. Ce temps de transit comporte quatre composantes :

- Le délai d'échantillonnage
- Le délai de propagation
- Le délai de transport
- Le délai des buffers de gigue

1.6 Conclusion

Dans ce chapitre, nous avons présenté un aperçu global sur la transmission de la voix sur le réseau IP, les éléments constituant ce réseau et les différents protocoles utilisés dans la téléphonie IP. Nous avons aussi abordé les limites et contraintes de la VoIP. De plus, nous avons pu démontrer que la VoIP est l'alternative la plus rentable pour effectuer des conversations téléphoniques à courte comme à longue distance sans distinction. D'autant plus, qu'elle présente une technologie en permanente évolution.

La VoIP est la technologie la plus pertinente dans le domaine des télécommunications. Sa fiabilité en termes de réduction des coûts et l'augmentation du facteur de joignabilité des clients malgré leurs mobilités est un atout incontestable. Toutefois, cette technologie pose encore de nombreuses questions quant à l'accroissement en exponentielle du nombre d'utilisateurs et l'exigence d'un meilleur compromis entre le prix et qualité de service (QoS).

Chapitre 2

Codage de la parole : Analyse par synthèse

2.1 Introduction

L'accroissement de la demande au niveau des réseaux de communications numériques a provoqué d'importants travaux de recherches [6]. De ce fait, plusieurs méthodes de traitement de signal sont apparues dans le but de réduire le nombre de bits nécessaires à la représentation du signal de parole et à la transmission des informations tout en maintenant un niveau de qualité suffisant et une complexité de calcul raisonnable. Le développement de calculateurs de plus en plus performants a offert la possibilité de traiter numériquement le signal en remplacement du traitement analogique et a rendu possible la mise en œuvre d'algorithmes de plus en plus sophistiqués pour la compression du signal de parole [11].

La parole est depuis toujours, le moyen privilégié de communication de l'homme. Avec la révolution des machines et notamment des ordinateurs, chercheurs et industriels se sont efforcés d'étendre l'usage de la parole à la communication homme-machine. Le codage de la parole permet la réduction de débit de transmission du signal et des communications dans des canaux à largeur de bande limitée. La largeur de bande d'une transmission devra être minimisée tout en préservant la qualité du signal vocal reconstruit et en répondant aux autres exigences liées à l'application [12].

Le principe de la technique VoIP consiste à encapsuler un signal audio numérisé dans le protocole IP pour le transporter sur un réseau. Pour une communication en VoIP, le signal vocal doit être compressé et codé à l'aide d'un codec audio par exemple G.723.1 [1], G722.1 [13], G722.2 [14], G729 [15]. Ensuite, l'information à transmettre est découpé en paquets par une procédure de paquets avant l'envoi sur le réseau IP. Les paquets d'informations, qui circulent sur internet, empruntent des chemins différents et arrivent fréquemment dans le désordre. Les paquets sont alors stockés dans des mémoires tampons, ou buffer, pour être réordonnés et permettre la décompression et le décodage de l'information, pour sa transformation en signal vocal synthétisé à l'aide d'un décodeur précis.

Dans ce chapitre, nous allons voir un aperçu sur les différents critères relatifs au codage, les codecs dédiés à la VoIP et ainsi que, les techniques de masquages de perte de paquets (PLC).

2.2 Définition

Le codage de parole est une application de compression de données d'audio numériques des signaux contenant la parole. Le codage de parole diffère des autres formes de codage audio car la parole est un signal plus simple que la plus part des autres signaux audio, avec une bande

Chapitre 2 – codage de la parole : analyse par synthèse

passante plus au moins étroite. Le critère le plus important est la préservation de l'intelligibilité de la parole, avec une quantité limitée de données transmises. C'est pour cela vient le terme, codec [16].

Le codec est une combinaison des mots, codage, et décodage, il permet donc de convertir un signal vocal analogique en un signal vocal numérique. En plus, c'est un logiciel qui numérise et compresse la voix de l'émetteur, ainsi les données numériques sont encapsulées dans des paquets IP et acheminées vers le destinataire. A l'arrivés au destinataire, ce dernier, grâce au même codec décompresse et restitue le signal de la voix synthétisée. La figure 2.1, illustre le principe général d'un système de codage de la parole.

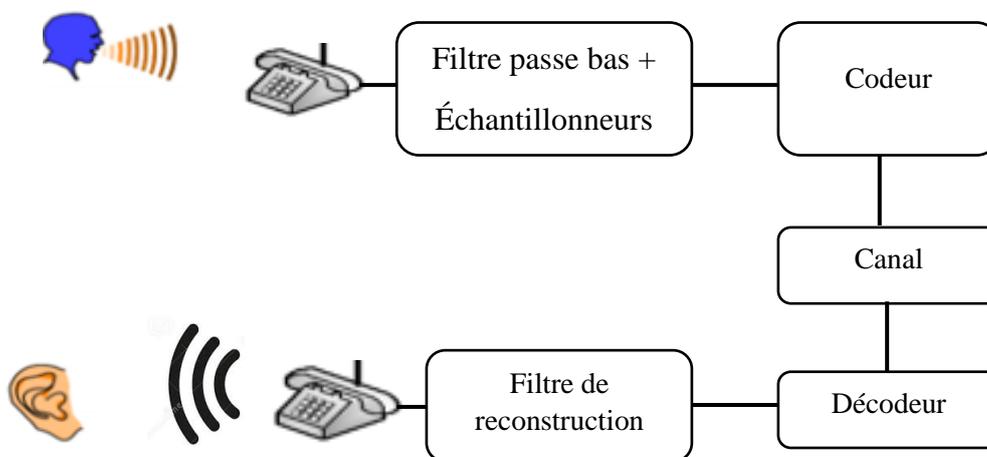


Fig. 2.1. Un système de codage de la voix.

Au fait, les codecs de parole peuvent être répartis en trois classes principales : les codeurs à forme d'onde, les codeurs paramétriques, appelés vocodeurs et les codeurs hybrides. Le choix d'une méthode va dépendre surtout de l'application visée et des contraintes sur le débit. Dans les systèmes de téléphonie filaire classiques, la parole est numérisée à 64 kbps. De nombreux algorithmes ont été proposés pour diminuer ce débit tout en essayant de conserver une qualité subjective suffisante. Cependant, nous pouvons distinguer trois plages de débits correspondant aux classes de codecs qui existent, tels que :

- Les hauts débits, supérieurs à 16 kbps, correspondant à des algorithmes de codage de la forme d'onde non spécifiques à la parole.
- Les bas et très bas débits, de quelques dizaines de bits par seconde à 4 kbps, correspondant aux vocodeurs (Voice coder) spécifiques au codage de la parole.
- Les débits moyens, de 4 kbps à 16 kbps, correspondant à des techniques de codage hybrides utilisant des méthodes de codage de la forme d'onde et prenant en compte

certaines propriétés de la parole ou de la perception auditive. Le principal représentant de cette classe est le codage CELP [5].

2.3 Codage basé sur la prédiction linéaire : analyse par synthèse

Le codage de la parole est une discipline qui a subi une forte évolution depuis 1982, par l'avènement d'un nouveau concept de codage du signal d'excitation. Ce nouveau concept appelé modélisation par analyse-synthèse a vu l'introduction successive du codeur à excitation multi-impulsionnelle (MP-LPC) puis du codeur à excitation par code (CELP), le codeur à excitation mixte (MELP) et ensuite le codeur à excitation par séquence algébrique codée (ACELP, CS-ACELP)

La parole peut être considérée comme étant un signal pseudo-stationnaire, c'est-à-dire, stationnaire sur de courtes durées, des trames allant en général de 10 ms jusqu' à 30 ms. Sur cette période, il est possible de caractériser le spectre du signal par deux attributs :

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

Sur chaque trame, le codeur extrait des paramètres qui représentent l'enveloppe spectrale et des paramètres qui caractérisent l'excitation, tel que, l'énergie, le voisement et la fréquence fondamentale ou le pitch. De même, d'autres paramètres peuvent être aussi calculés pour représenter plus finement l'excitation. Les paramètres les plus souvent utilisés pour l'enveloppe spectrale sont les paires de raies spectrales (LSF) qui sont déduites des coefficients de prédiction linéaire (LPC) et qui possèdent de bonnes propriétés pour la quantification et l'interpolation.

La prédiction linéaire (LP) fait partie intégrante de la plupart des algorithmes de codage de la parole. Le codeur à prédiction linéaire (LPC) est fondé sur le modèle de la production de la parole [17]. Le signal d'entrée est analysé et un signal d'excitation est déterminé. La fonction du codage prédictif consiste à définir les coefficients du filtre de prédiction, tandis que, le signal d'excitation est modélisé avec la technique d'analyse par synthèse en utilisant un filtre (filtre de synthèse) le plus souvent, il s'agit d'un filtre auto-régressif pur. De telle façon que l'erreur entre le signal d'entrée et celui mis en forme par le filtre de synthèse, qui reproduisant les résonances (formants) du conduit vocal soit suffisamment minimisée.

2.3.1 Prédiction à court terme

La prédiction à court terme cherche à éliminer la redondance entre les échantillons voisins. Le filtre utilisé est appelé filtre d'analyse LP. Il supprime la structure formantique du signal

parole et laisse l'erreur de prédiction de sortie à basse énergie qui est connue sous le nom résiduel ou excitation. De ce fait, l'excitation est le signal d'erreur résiduelle entre le signal de parole d'entrée $s(n)$ et son estimation $\check{s}(n)$. L'inverse du filtre d'analyse est le filtre de synthèse, il modélise le conduit vocal et sa fonction de transfert décrit l'enveloppe spectrale du signal parole [18].

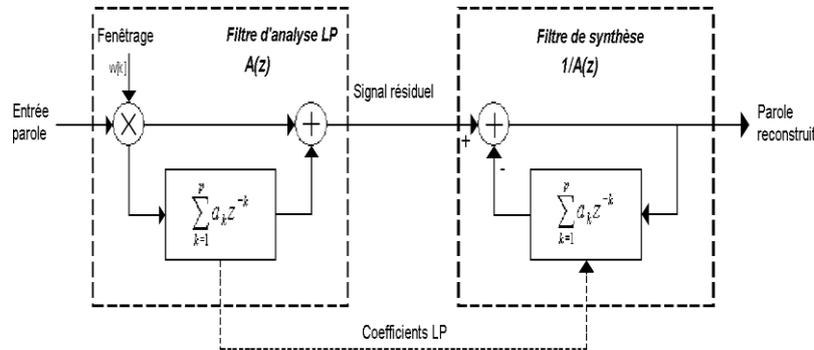


Fig. 2.2. Analyse et Synthèse LP

Le modèle auto-régressif est un modèle pôle-zéro peut-être décomposée en deux parties, tel qu'il est illustré dans la figure (2.2). Ainsi, une partie analyse qui filtre un signal d'entrée avec la fonction de transfert $A(z)$. Ce filtre tout-zéro est défini comme le filtre d'analyse LP, et une partie synthèse qui effectue un filtrage de fonction de transfert $M(z) = 1/A(z)$. Ce filtre tout-pôle, connu sous le nom de filtre de synthèse LP.

Dans ce modèle [19], le signal de parole à chaque instant n est linéairement prédit en combinant les p échantillons passés, tel que,

$$\begin{aligned} \check{s}(n) &= \sum_{k=1}^{\infty} a_k \cdot s(n-k) \\ \check{s}(n) &= \sum_{k=1}^p a_k \cdot s(n-k) \end{aligned} \quad (2.1)$$

Où a_k , $1 \leq k \leq p$, les coefficients de prédiction linéaire (LPC) et p l'ordre de prédiction qui prend la valeur 10 pour les codecs à bande étroite, à bande de fréquence de 4 KHz et l'ordre 16 pour les codecs à large bande à bande de fréquence de 8KHz, [5]. Notons que l'amélioration de la prédiction devient minime lorsque l'ordre p est supérieur à 10. Des modèles LPC les plus courants utilisent un nombre de coefficients de l'ordre de 8 à 16 [20].

Le signal résiduel e_{rs} est l'erreur de prédiction donné par la différence entre le signal original $s(n)$ et le signal de parole prédit $\check{s}(n)$, tel que,

$$e_{rs}(n) = s(n) - \check{s}(n) . \quad (2.2)$$

Qui peut s'écrire aussi sous la forme,

$$e_{rs}(n) = s(n) - \sum_{k=1}^p a_k \cdot s(n-k) \quad (2.3)$$

La transformée en Z de l'équation (2.3) donne,

$$\begin{aligned} E_{rs}(z) &= S(z) - (\sum_{k=1}^p a_k \cdot z^{-k} \cdot \sum_{k=1}^p s(n) \cdot z^{-k}) \\ E_{rs}(z) &= S(z) [1 - \sum_{k=1}^p a_k \cdot z^{-k}] \\ E_{rs}(z) &= S(z) \cdot A(z) \end{aligned} \quad (2.4)$$

D'où, les signaux, $S(z)$ et $E_{rs}(z)$ sont le signal de parole et le signal résiduel donnés par la transformée en Z, respectivement. Sachant que, $A(z)$, est le filtre d'analyse, donné par,

$$A(z) = 1 - \sum_{k=1}^p a_k \cdot z^{-k} \quad (2.5)$$

Son inverse conduit au modèle de synthèse LP,

$$S(z) = E_{rs}(z) \cdot A(z)^{-1} \quad (2.6)$$

a- Détermination des paramètres du filtre de prédiction

Pour avoir les coefficients du filtre à court terme, a_k du processus auto-régressif, la méthode classique des moindres carrés peut être utilisée. Deux grandes approches sont utilisées pour le codage par prédiction linéaire LPC court terme qui sont, la méthode d'autocorrélation et la méthode de covariance [21], [22]. Pour résoudre ce système d'équation et trouver les coefficients du filtre LP, L'énergie du signal erreur $e_{rs}(n)$ est minimisée sur une trame de parole :

$$E = \sum (e_{rs}(n))^2 \quad (2.7)$$

Pour N échantillons (2.7) devient :

$$E = \sum_{n=-\infty}^{+\infty} (e_{rs}(n))^2 = \sum_{n=-\infty}^{+\infty} (s(n) - \sum_{k=1}^p a_k \cdot s(n-k))^2 \quad (2.8)$$

L'algorithme de Wiener Levinson-Durbin est utilisé pour trouver les coefficients de prédiction, en minimisant la moyenne quadratique de l'erreur de prédiction.

b- Quantification des coefficients des paires de raies spectrales

Il est à noter que les coefficients de prédiction linéaire ne sont pas toujours codés directement, mais sont transformés en un ensemble de paramètres qui ont des propriétés désirables. Pour

cela, Itakura dans [23], a proposé une représentation des paramètres LPC plus efficace et plus intéressante, qui sont les raies spectrales de fréquences ou LSF. Ces paramètres possèdent quelques propriétés supplémentaires qui les rendent plus attractifs que les coefficients LPC pour leur qualité de quantification ou interpolation.

Les coefficients, a_k , du filtre LPC, sont donc convertis en coefficients LSFs avec une présentation comme suit :

$$A(z) = \frac{1}{2} (P(z) + Q(z)) \quad (2.9)$$

Tel que, $P(z)$, et $Q(z)$, sont des polynômes prédictif d'ordre, $p+1$, sont respectivement symétriques et antisymétrique définis par,

$$P(z) = A(z) + z^{-(p+1)} A(z^{-1}) \quad (2.10)$$

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

Les coefficients, LSF, sont extraits à partir des racines complexes des polynômes donnés par les formules, (2.10) et (2.11), qui possèdent deux propriétés très intéressantes :

- pour un filtre de synthèse stable, $1/A(z)$, toutes les racines de $P(z)$ et $Q(z)$ sont sur le cercle unité, ainsi, l'ordre des paramètres LSFs est respecté.
- les racines des deux polynômes s'alternent deux à deux sur le cercle d'unité, d'où les LSFs seront toujours dans un ordre croissant.

2.3.2 Pondération perceptuelle de formants

Le signal de parole, appelé aussi, target signal, sera filtré par un filtre passe haut, $H(z)$, passe à travers un filtre de pondération perceptuelle de formants. Ce dernier est un filtre de pôles-zéros avec des coefficients changeant pour chaque sous trame dont les coefficients sont obtenus à partir des coefficients LPC non quantifiés. Le rôle de ce filtre consiste à atténuer les régions des pics, formants, du spectre LP et d'amplifier les régions entre les pics, vallées. L'objectif principal est de masquer le bruit dans ces zones de fréquences vallées [4]. Toutefois, le filtre de pondération des formants peut être défini selon l'équation (2.12),

$$W(z) = \frac{A(z)}{A(\frac{z}{\gamma})} = \frac{1 - \sum_{k=1}^p a_k \cdot z^{-k}}{1 - \sum_{k=1}^p a_k \cdot \gamma^k \cdot z^{-k}} \quad (2.12)$$

Sachant que, $A(z)$, est le filtre d'analyse, a_k , sont les coefficients de prédiction et γ , coefficient de pondération compris entre 0 et 1. Ce facteur vaut, 0.8, pour une valeur typique de fréquence d'échantillonnage, $F_e = 8$ khz, des détails sont donné dans [24] et [25].

- Pour $\gamma=0$, on aura, $w(z) = A(z)$, donc l'erreur pondérée égale à la différence entre le signal résiduel et la suite d'impulsions placées à de multiples positions, c'est à dire $e_w(n) = e_{rs}(n) - u(n)$.
- Pour $\gamma=1$, donc, $w(z) = 1$, aucun effet et l'erreur pondérée égale à la différence entre signal original et signal synthétisé $e_w(n) = s(n) - \check{s}(n)$

De même, la figure (2.3), illustre un exemple d'un filtre de pondération perceptives des formants.

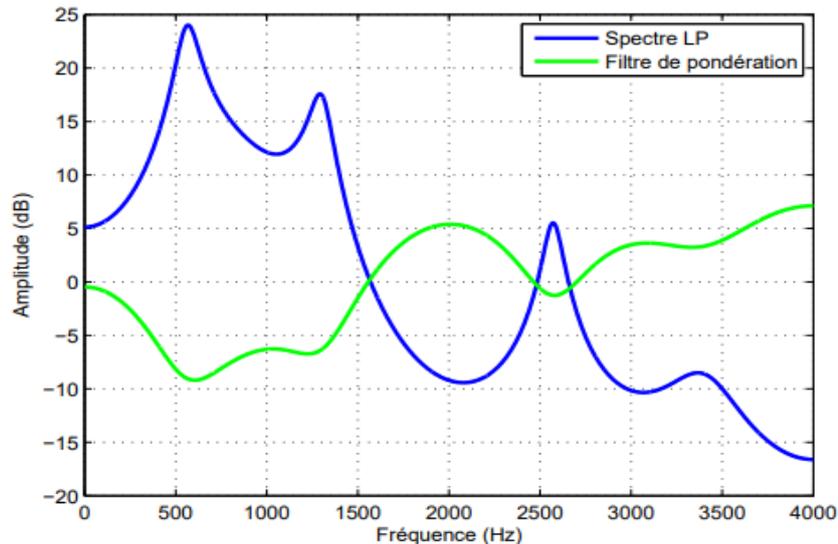


Fig. 2.3. Spectres du filtre LP et du filtre de pondération des formants.

2.3.3 Prédiction à long terme

La prédiction linéaire à long terme est utilisée pour exploiter les corrélations à long terme existantes dans les signaux voisés. Ce filtre est appelé prédicteur de pitch, Il exploite la périodicité du signal. L'inverse du prédicteur de pitch est appelé filtre de pitch. Il modélise la fonction de la glotte et sa fonction de transfert décrit la structure harmonique du signal de parole.

Cette analyse a pour but de modéliser les vibrations des cordes vocales produisant les sons voisés qui possèdent des corrélations à long terme et se traduisent au niveau du signal de parole par une fréquence fondamentale F_0 ou hauteur de voix. Le prédicteur de pitch n'a aucun effet pour les signaux non voisés, puisque l'excitation non voisée est aléatoire et son spectre est monotone [26].

2.3.4 Modélisation du signal d'excitation

Dans un codage basé sur la prédiction linéaire et l'analyse par synthèse, une analyse de prédiction linéaire standard, calculée avec l'ordre, p , de coefficients, a_k , du filtre LP accompagnée d'une modélisation du pitch sont utilisées dans cet ordre pour déterminer l'excitation idéale du filtre de synthèse, et les paramètres de reconstruction de la parole. Tout en tenant compte d'une pondération perceptive, tel qu'il a été mentionné dans la section (2.3.3). Le signal de parole présente des corrélations à long terme qui sont causées par les vibrations des cordes vocales. Pour cela la périodicité à long terme est modélisée en représentant l'excitation courante par une version passée de celle-ci.

a. Estimation du délai tonal en boucle ouverte

Dans une analyse en boucle ouverte, l'estimation du pitch T_{bo} est calculée à partir du signal vocal pondéré perceptivement en procédant comme suit, calculer deux estimations de la tonie pour chaque trame, l'une pour les deux premières sous trames et la seconde pour les deux dernières. Pour extraire ces valeurs de pitch, nous devons trouver la période fondamentale. D'où une fonction d'intercorrélation sera appliquée au signal de parole pondéré [27].

b. Mise en forme du bruit harmonique

En utilisant un filtre de mise en forme du bruit harmonique, cela permet d'améliorer la qualité des signaux vocaux codés, sa formulation est donnée par [1], comme suit,

$$F_{bh}(z) = 1 - \beta z^{-T} \quad (2.13)$$

c. Estimation de la tonie en boucle fermée

Dans l'analyse en boucle fermée, un filtre combiné, $S(z)$, est utilisé, tel que,

$$S(z) = A(z) \cdot W(z) \cdot F_{bh} \quad (2.14)$$

La réponse impulsionnelle de ce filtre est désignée par, $s_i[n]$, avec, n , allant de 0 à 59 et $i = 0, 1, 2, 3$. Cette réponse impulsionnelle est nécessaire pour l'exploration des répertoires adaptatifs et fixes.

2.3.5 Le répertoire adaptatif conventionnel

Les paramètres du répertoire codent de séquences adaptatives, ACB, ou paramètres tonals sont, le délai tonal et le gain tonal. Dans la méthode du filtrage tonal par le répertoire adaptatif, le signal d'excitation est répété pendant des intervalles inférieurs à la longueur de la sous trame

[29]. Dans l'étape d'exploration du répertoire, l'excitation est complétée par le signal résiduel de prédiction linéaire, $e_{rs}(n)$, afin de simplifier la recherche en boucle fermée.

L'exploration du répertoire adaptatif est effectuée à chaque sous trame. La période fondamentale ou délai tonal, en boucle ouverte, T_{OL} est calculée pour tous les deux sous trames, c'est-à-dire pour chaque 120 échantillons. La période fondamentale est estimée dans l'étendue de 18 à 148 échantillons. La plage de retard effectif est de 18 à 141. Seules ces valeurs de retard sont générées par le codeur. Si un retard interdit est détecté par le décodeur, cette trame est signalée comme reçue par erreur [29].

2.3.6 Le répertoire fixe des excitations

La contribution du répertoire de codes fixes, FCB, provient d'un codage à impulsions multiples, MP, ou d'une procédure de codage, ACELP. Ces deux procédures sont presque similaires, avec une légère différence dans la quantification des impulsions du FCB. De même, les deux options de codage d'impulsions FCB placent un nombre limité d'impulsions dans une trame. L'option Multi-Pulse utilise, 6 ou 5 impulsions par sous trame, tandis que l'option ACELP, utilise 4 impulsions par sous trame. Toutefois les deux options considèrent deux grilles distinctes pour placer les impulsions.

Pour le cas du MP, l'une des grilles ne contient que des positions impaires et l'autre grille ne contient que des positions paires. Les deux grilles utilisent la même amplitude pour toutes les impulsions, mais chaque impulsion peut prendre un signe arbitraire. La procédure de codage, MP, place les impulsions de manière séquentielle dans de quelconques positions possibles mais sur une grille particulière, c'est-à-dire, soit celles de la grille 0 ou la grille 1.

2.4 Prédiction linéaire à excitation multi impulsionsnelle (MP-LPC)

Le codeur MP-LPC, ou LPC à excitation multi-impulsionnelle a été mise au point en 1982 par Atal et Remde [30]. C'est un modèle du conduit vocal, utilise l'enveloppe spectrale à court terme avec un filtre LPC pour produire une parole naturelle à bas débit, et de corriger les limites constatées sur la qualité de la parole obtenue par le codage LPC classique [31]. Cette technique proposé améliore l'excitation trop simplicité (mono-impulsion), par la recherche d'un autre modèle d'excitation (multi-impulsion) avec des impulsions multiple non uniforme espacées de façon non uniforme, et approchant au mieux le résiduel de prédiction. Les positions des impulsions sont déterminées séquentiellement par une procédure d'analyse par synthèse qui est pour la plupart des codeurs à bas débits.

L'avantage de cette méthode, est d'éviter la classification des sons de la parole en sons voisés et non-voisés. De plus, les erreurs de mesure du pitch inévitables aussi dans la LPC se trouvent ainsi éliminées. En effet, on doit trouver, $\check{s}(n)$, qui doit s'approcher de, $s(n)$, de façon à avoir $e_{rs}(n)$ minimal. La figure 2.4, illustre le principe de fonctionnement de ce type de codec.

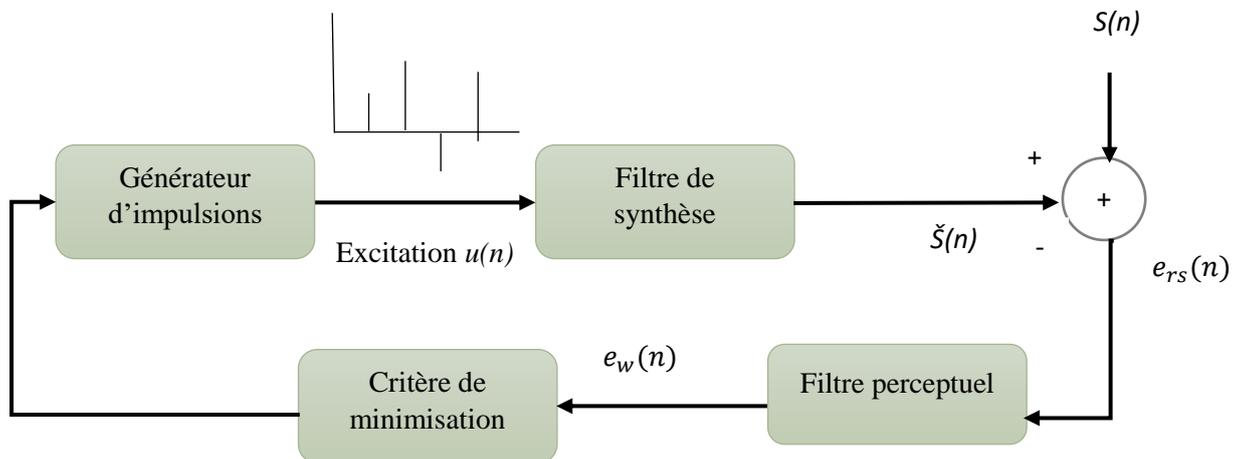


Fig. 2.4. Système d'analyse par Synthèse à excitation multi-impulsions, MP-LPC.

Où, $s(n)$, et $\check{s}(n)$, sont le signal original et son estimation, et $e_{rs}(n)$, est l'erreur de prédiction qui sera pondérée et filtrée par un filtre dit perceptuel qui utilise l'effet de masquage connu en acoustique, pour améliorer sensiblement la qualité de la parole décodée [30]. D'où sa fonction de transfert est définie dans l'équation (2.12).

2.5 Prédiction linéaire excitée par code algébrique (ACELP)

Les modèles de codage temporel à forme d'onde représentent un groupe important pour la compression des signaux de parole. Les grandes familles des modèles de type, CELP, utilisent les techniques suivantes, une boucle d'analyse par synthèse, un filtre de prédiction à long-terme et un filtre perceptuel. Ce codage est une variété des codeurs MP-LPC. Il permet d'obtenir des débits encore plus faibles (<5 kbps). Le modèle MP-LPC nécessite un débit minimum afin d'encoder son excitation. La qualité du signal de synthèse se détériore rapidement lorsque le modèle limite le débit attribué aux impulsions ou que le nombre d'impulsions diminue. Ainsi, pour augmenter la qualité du signal de synthèse tout en réduisant le débit, Schroeder et Atal proposent de changer le générateur d'impulsions du modèle MP-LPC pour un dictionnaire stochastique [32]. Le modèle, ACELP, qui est contenu dans plusieurs standards appartient à la famille CELP.

Le modèle ACELP remplace le dictionnaire stochastique du modèle CELP par un dictionnaire algébrique [28], qui propose une modification du schéma de fonctionnement à l'encodeur du modèle CELP. Dans un dictionnaire algébrique, les excitations possèdent peu d'impulsions et leur valeur d'amplitude vaut +1 ou -1. Ces deux caractéristiques d'un dictionnaire algébrique, peu d'impulsions et un nombre restreint de valeurs possibles pour l'amplitude permettent de réduire la complexité de la recherche dans le dictionnaire. La figure (2.5), montre les modifications apportées par le modèle ACELP, le modèle change l'emplacement du filtre perceptuel, $w(z)$, afin que le filtrage s'effectue avant le calcul de l'erreur quadratique entre le signal original, $s(n)$, et le signal de synthèse, $\check{s}(n)$.

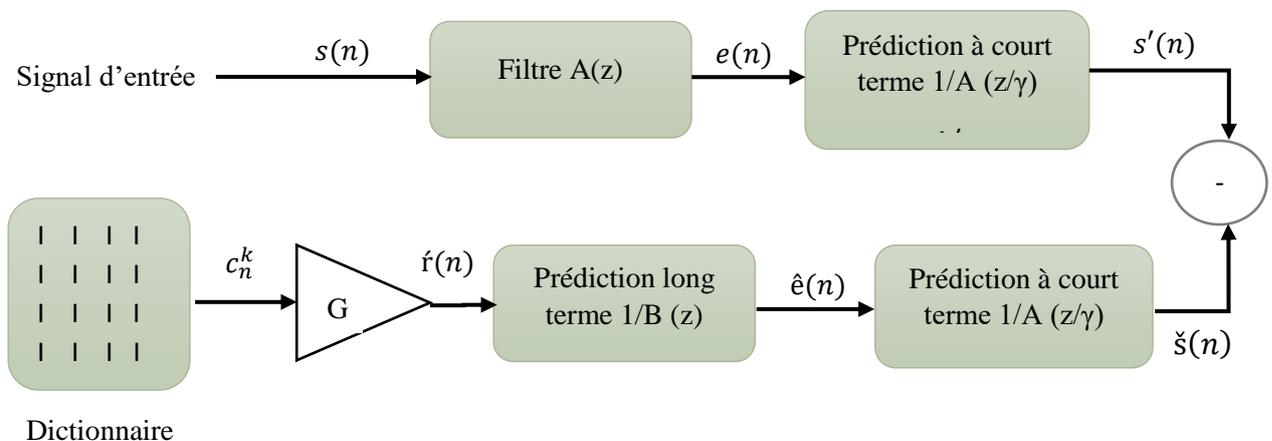


Fig. 2.5. Principe de fonctionnement dans un encodeur du modèle ACELP.

Dans ce codeur, la trame de parole est de 10 ms, de 80 échantillons, chaque trame est synthétisée en appliquant l'innovation précédente pondérée par un gain, G , à l'entrée de deux filtres variables mis en cascade. Le premier est un filtre à prédiction long terme, filtre de pitch de la forme, $1/B(z)$, ayant pour but de mieux modéliser la pseudo-périodicité du signal de parole et le second filtre est un filtre court terme, $1/A(z/\gamma)$, dont le but est de modéliser l'enveloppe spectrale du signal de parole. De même, un filtre de pondération de l'erreur dans les régions formantiques du spectre de la parole est aussi appliqué. Sa fonction de transfert est, $W(z)$, utilise des coefficients de prédiction non quantifiés, alors que le filtre LP utilise des coefficients quantifiés.

2.6 Codeurs de la parole basés sur CELP utilisés en VoIP

Il existe différents algorithmes de codage/décodage appelés aussi, codec. En téléphonie, c'est la branche ITU-T de l'organisme international télécommunications qui standardise les codecs

dans sa série de recommandations G.xxx. Nous citons ci-dessous quelques codeurs basés sur le modèle CELP.

a. G.711

Approuvé en 1965, le G.711 est le plus anciens des codeurs audio. Il est utilisé par le réseau d'accès RNIS et dans tous les cœurs de réseaux téléphoniques modernes. Un flux audio codé en ce codeur produit un flux de données de *64 kbps*, où chaque échantillon est codé sur un octet indépendant, en conséquence la taille de trame est de seulement *125 μs*. Bien que, toutes les applications VoIP mettent plus d'un échantillon par paquet, en général l'équivalent de *10 ms* de parole, soit 80 échantillons, [34].

b. G.722

Le G.722 fournit un codage large bande de meilleure qualité incluant le spectre jusqu'à 7kHz. Les débits disponibles sont de *48*, *56* ou *64 kbps*. Ce codeur permet de commuter à tout instant entre ces débits. Le G.722 est un codeur tout à fait adapté pour les applications de conférence professionnelles.

c. G.722.1

Le G.722.1 plus récent, également large bande, fonctionne à *24 kbps* ou *32 kbps*. Il en existe aussi une version à *16 kbps* qui est supportée par Windows Messenger. Ce codeur génère des trames de *20 ms*. Le G.722.1 est un compresseur à base de transformation qui est optimisé pour la parole et la musique [13].

d. G.729

Le G.729 est un algorithme de compression de données qui décompose la voix en paquets de *10ms*. Dans sa version standard, il opère à un débit de *8 kbps* mais il existe des variantes beaucoup moins utilisées fonctionnant à des débits de *6.4 kbps* et *11.8 kbps* [33]. Il est devenu, avec le G.723.1, l'un des codeurs les plus utilisés surtout pour transporter de la VoIP sur les WAN ; et pour le codage de la partie audio d'une visioconférence. Le G.729 peu adapté à la musique, au fax et aux signaux de modem.

e. G.723.1

Le G.723.1 est un codec assurant la compression de la voix dans des trames de *30ms*. Ce codec peut fonctionner selon deux débits :

- 6.3 kbps sur trames de 24 octets en utilisant l’algorithme MPC-MLQ avec un MOS de 3.9.
- 5.3 kbps sur trames de 20 octets en utilisant l’algorithme ACELP avec un score MOS de 3.7.

Le codec G.723.1 est avant tout un codeur de voix. Son algorithme fait l’hypothèse que le signal traité correspond à une voix humaine occidentale. En particulier, il se révèle de qualité médiocre pour le codage de signaux non vocaux (musique, tonalités DTMF, Fax ...etc.

La norme ITU-T G.723.1 est une annexe qui définit un comportement différent du codec lorsqu’un silence est détecté. Cet algorithme supplémentaire est dénommé VAD/CNG, voice activity detection/ domfort noise generation, ou détection d’activité/bruit de confort. Dans ce cas seul le bruit de fond est modélisé et permet d’atteindre un débit extrêmement faible, avec 1 octet toute les 30 ms, tout en restituant l’ambiance sonore. 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. En raison de sa faible consommation en bande passante, ce codeur est particulièrement adapté à la VoIP, en l’absence de musique.

2.6.1 Le choix du codec G.723.1

Le choix de codec est un compromis entre la qualité de service souhaité et la capacité de l’infrastructure IP à délivrer une bande passante et des paramètres de QoS qui vont impacter cette qualité. Le paramètre le plus déterminant auquel on s’intéresse pour commencer est la bande passante que l’on met en regard du nombre de communications simultanées à écouter. Le tableau suivant permet d’effectuer rapidement le bilan de débit (bande passante) en fonction du codec choisi :

Codec	Débit	Durée de paquetage	Technique de paquetage
G.711	64 Kbps	5 ms	PCM
G.711	64 Kbps	10 ms	PCM
G.722	64 Kbps	20 ms	PCM
G.722.1	16 Kbps	30 ms	PCM
G.729	8 Kbps	10 ms	CS-ACELP
G.729	8 Kbps	20 ms	CS-ACELP
G.723.1	5.3 Kbps	30 ms	CELP
G.723.1	6.3 Kbps	30 ms	MP-MLQ

Tableau 2.1. Le bilan de débit des différents codecs

2.6.2 Le score MOS

Les codecs utilisés en VoIP présentent la particularité de compresser le signal original et dégradent par conséquent la qualité de la voix. Le score MOS fournit une indication quantifiée de la qualité après transmission et décompression [35]. L'un des intérêts de réaliser des simulations de trafic en avant-projet de déploiement de voix sur IP est de pouvoir calculer une valeur estimée de score MOS ; qui est une valeur numérique de 1 à 5, tel qu'il est présenté sur le tableau 2.2.

Score MOS	Qualité perçue
5	Excellente
4	Bonne
3	Moyenne
2	Dégradée
1	Mauvais

Tableau 2.2. La variation du score MOS et leurs indication

La méthode MOS est une mesure subjective de la QoS voix par des opérateurs humains. Elle n'a pas pour objet de fournir des données objectives mais d'obtenir une appréciation de la clarté de la voix reçue grâce à une enquête auprès d'un panel d'utilisateurs ou d'opérateurs représentatifs. Cette méthode est définie par la spécification ITU P800 (MOS pour note moyenne d'appréciation). Son application est coûteuse et délicate, mais elle reste le « juge de paix » en dernier recours lors de problèmes de plainte des utilisateurs du service de téléphonie [36].

2.6.3 Calcul du facteur R de l'E-model

L'ETSI a développé un modèle de calcul de la qualité de transport de la voix de bout en bout, de la bouche de l'émetteur à l'oreille du récepteur, connu sous le nom d'E-model. Ce modèle a été standardisé par l'ITU sous la référence G.107. Le principe de l'E-model consiste à calculer une grandeur unique R en fonction des paramètres que nous avons précédemment décrits dans ce document. La formule simplifiée du calcul de, R, est la suivante,

$$R = R_0 - L_s - L_d - L_e + A \quad (2.15)$$

Le principe de la formule est de partir d'un certain capital de QoS, égale à R_0 , et de lui imputer les dommages causés par les différents aspects de la transmission.

R_0 : Coefficient initial signal/ bruit, « capital initial de QoS », égal à 94.3 en VoIP.

L_s : Coefficient de dommages simultanés avec l'émission de la voix (bruit de fond).

L_d : Coefficient de dommages dus au délai de transmission et de transport.

L_e : Coefficient de dommage de distorsion causé par les équipements.

A : Coefficient d'amélioration.

Le facteur, R , ainsi est calculé entre, 0 à 100, et il permet de déduire directement un coefficient MOS, allant de 1 à 5. Dans la réalité les valeurs de R oscillent entre 50 et 93.2, soit la limite basse acceptable pour le récepteur, et la limite haute liée aux possibilités techniques de transformation de la voix humaine en signal.

2.6.4 Avantages et inconvénients

a. Les Avantages

- Une des raisons de la pérennité du G.723.1 est 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.
- En raison de sa faible consommation en bande passante, le G.723.1 est particulièrement adapté à la VoIP.
- Une meilleure technique de masquage de perte des paquets (PLC).

b. Les Inconvénients

- Non adapté à la musique, au fax, à la signalisation DTMF, pour Dual-Tone Multi-Frequency.
- Le codec G.711 permet de bénéficier au réseau constant de la meilleure qualité de service, tandis que la compression G.723.1 apporte avec elle des diminutions initiales de la QoS, immédiatement reflétées dans le score MOS de mesure de la qualité que nous avons déjà étudié dans la partie précédente.
- Les autres codecs ont bénéficié d'avancées par rapport au G723.1 en mathématiques et en puissance de calcul qui permettent d'améliorer le compromis qualité/débit.

2.7 Les méthodes de masquages de pertes de paquets (PLC)

Dans un system VoIP, au niveau du récepteur, certain paquets peuvent manquer, à cause des délais, à l'encombrement, ou aux erreurs de transfert. Dans les réseaux de communications, ces pertes sont causées par plusieurs facteurs liés aux différentes étapes de la chaine de transmission en VoIP, en particulier la congestion des nœuds (routeurs). Nous savons par ailleurs que la perte

des paquets dégrade la qualité de la voix et se traduit par des ruptures au niveau de la conversation. Il est par conséquent, indispensable de mettre en place un mécanisme de masquage de ces pertes au niveau de décodeur.

Plusieurs algorithmes de masquage des pertes de paquets appelés aussi PLC, packet loss concealment, existants ; qui consiste à générer un signal vocal synthétique destiné à remplacer les données manquantes, en assurant une transition douce entre le signal réel et le signal synthétique. Ces mécanismes de masquage de pertes se divisent en deux parties complémentaires :

- Réparation basée au niveau de l'émetteur (Sender-Based repair)
- Réparation basée au niveau de récepteur (Receiver-Based repair)

2.7.1 Techniques basés sur la réception

Cette catégorie permet d'offrir une récupération de données perdues au niveau de décodeur, soit par réplication d'onde ou par techniques d'estimation du signal avec utilisation d'un modèles de parole.

2.7.1.1 Réplication de forme d'onde

Ces méthodes de masquage de pertes de paquets reposent sur l'insertion directe d'une forme d'onde dans le trou. L'écart manquant est reconstruit en répétant une partie de la parole déjà reçue, La forme la plus simple consiste à répéter le dernier élément reçu. Plusieurs études proposent néanmoins des versions plus élaborées [37].

Cependant, le système de masquage de pertes de trames audio proposé pour le codeur G.723.1 est certainement le plus utilisé. En raison de leur simplicité à comprendre et à mettre en œuvre.

Deux méthodes ont été proposées qui sont : la première avec une approche à la tonie indépendante et la deuxième est basée sur une approche à la tonie dépendante. Dans la première approche, les ν échantillons du signal de parole les plus récents des trames reçues sont utilisés comme un gabarit de correspondance du modèle. Pour réduire la complexité de la méthode, la valeur de ν est constante et inférieur à la longueur de la trame N . Ainsi, les ν échantillons des N échantillons de la trame reçue sont utilisés pour remplacer la trame manquante avec un chevauchement-et-ajout (OLA, overlap and add) sur les deux côtés du vide pour l'approche de réplication à la tonie dépendante, les échantillons de la tonie la plus récente sont répétés jusqu'à ce que le vide soit complété [4].

La technique de masquage de trames manquantes basée sur réplcation de formes d'onde la plus raffinée est particulièrement celle proposé par Sanneck et al dans [38]. Cette technique est basée sur la superposition/addition avec correspondance de forme d'onde.

2.7.1.2 Utilisation de modèles de parole

Le signal parole, d'un côté ou des deux côtés d'une perte, est ajusté à un modèle, qui est utilisé pour régénérer le signal de parole qui comblera la période de pertes. Et dans notre cas on a utilisé un modèle de parole qui est basé sur le codage CELP de codeur G.723.1. En générale ce système de masquage de perte se base principalement sur l'extrapolation de l'enveloppe spectrale et les paramètres d'excitation. Et celle-ci se fait à base des paramètres LSF [4].

L'équation d'extrapolation s'écrit comme suit :

$$LSF_{ex} = \rho(LSF_{ap} - LSF_{vm}) + LSF_{vm} \quad (2.16)$$

Tel que, LSF_{ap} c'est le vecteur des coefficients LSF quantifiés de la dernière trame reçue avant la perte. Et le LSF_{vm} est la valeur LSF moyenne fixe, pour le ρ c'est le coefficient de prédiction utilisé dans le codage différentiel.

2.7.2 Techniques basés sur l'émission

La figure (2.6) rassemble les techniques de réparations basées au niveau de l'émetteur. Pour mieux comprendre les exemples qui suivent expliquant ces techniques, nous avons distingué une trame de donnée d'un paquet de donnée. Une trame représente un intervalle du flux sonore. Un paquet peut contenir une ou plusieurs trames encapsulées afin d'être envoyées sur le réseau.

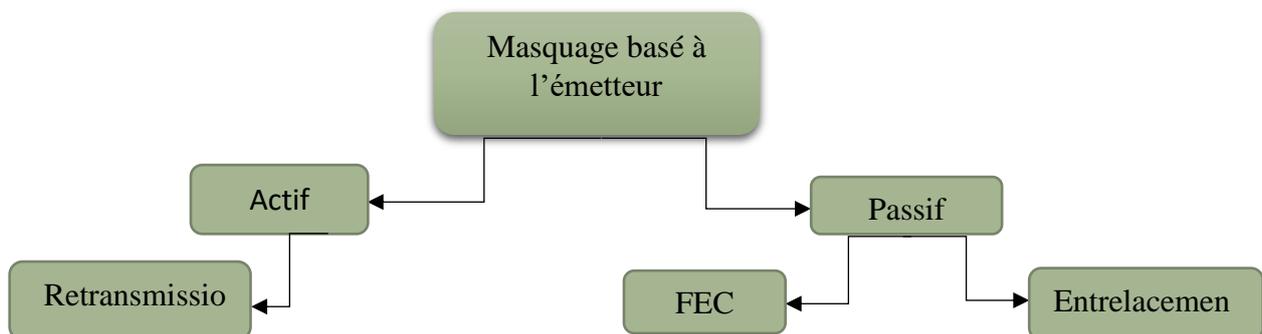


Fig. 2.6. Classification de techniques de masquage basées à l'émission.

2.7.2.1 Correction d'erreur en avale

La correction d'erreur par transmission est la plus connue, où des trames de parole de redondance sont enchainées, avec un retard, avec les paquets sélectionnés. Si une trame est

perdue, la version redondante retardée de cette trame peut être reçue correctement pour la décoder. Les méthodes FEC sont les plus efficaces, si la perte dans le réseau est prévisible, et si une bande passante supplémentaire est disponible. Le schéma de recouvrement repose sur l'addition de donnée de réparation au flux sortant. De ces données, les paquets manquants peuvent être réparés le cas de perte.

Le principe est illustré dans la figure (2.7), ci-dessous.

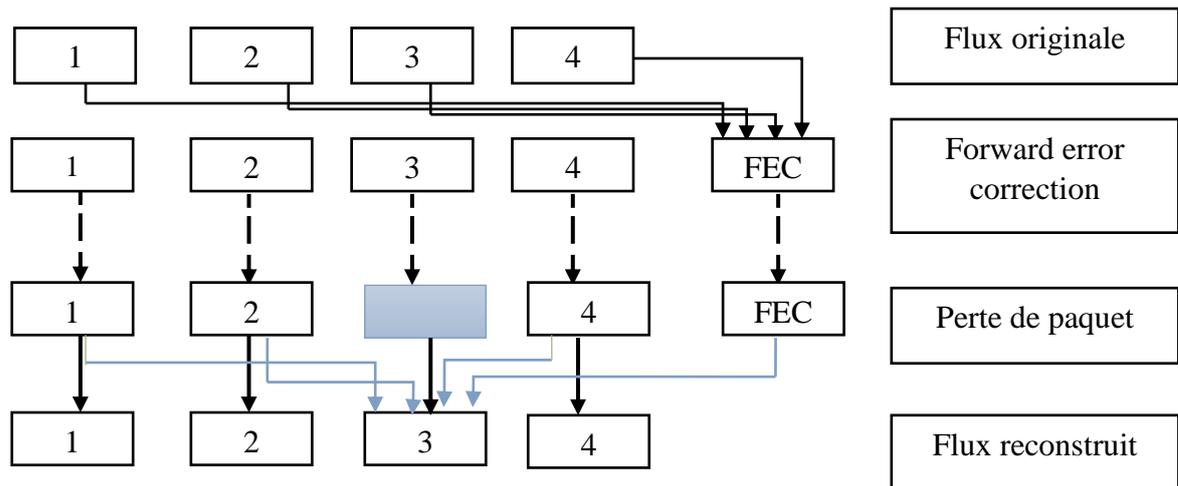


Fig.2.7. Exemple du principe FEC.

Plusieurs avantages découlent de cette méthode, nous pouvons citer la faible demande en ressource de calcul et la simplicité de l'implémentation. En contrepartie, cette technique impose un retard supplémentaire, une augmentation de la bande passante et une difficile implémentation au niveau du décodeur.

2.7.2.2 Entrelacement de trames

La technique d'entrelacement, ou interleaving en anglais, est très utile lorsque, les paquets contiennent plusieurs trames et le délai de bout-en-bout n'est pas important. Avant transmission du flux, les trames sont réarrangées de telle manière que celles, initialement, adjacentes se retrouvent séparées dans le flux transmis, puis remises dans leur ordre original au niveau de récepteur. La figure (2.8), illustre un exemple où chaque paquet contient 4 trames, détaillé dans [39].

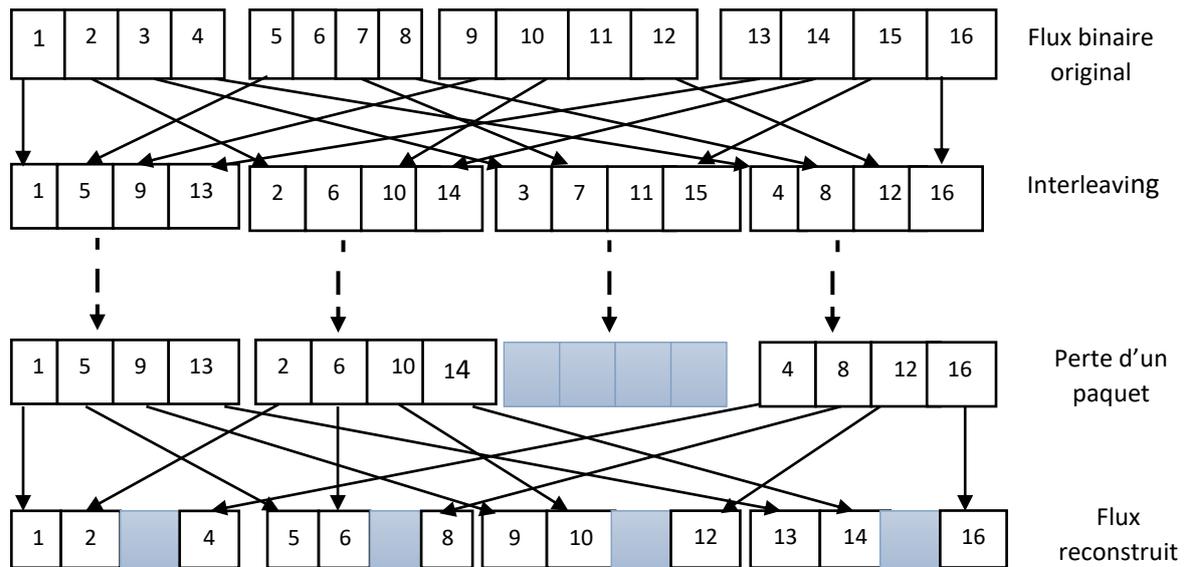


Figure 2.8. Exemple du principe d'entrelacement de trames.

L'augmentation de latence constitue un sérieux inconvénient à l'utilisation d'entrelacement dans ces applications interactives. Alors que le maintien d'une bande passante stable avant et après son implémentation représente son avantage majeur.

2.7.2.3 Retransmission

Cette technique retransmet simplement les paquets perdus, elle est difficilement applicable pour les applications interactives et pour lesquelles les délais de bout-en-bout sont réduits. Cependant pour des conditions de délai plus souple, cette méthode peut être implémentée.

2.8 Conclusion

Un aperçu global sur le codage de la parole basé sur le codage à prédiction linéaire a été présenté dans ce chapitre. En particulier, nous avons présenté d'abord la définition de codage de la parole. Puis, nous avons traité l'algorithme de prédiction linéaire basé sur l'analyse par synthèse. En effet, nous avons détaillé, des caractéristiques sur l'excitation totale qui est souvent modélisée en deux parties. Une prédiction à court terme sur une tranche de parole suivie par une prédiction à long terme. De plus nous avons détaillé la prédiction MP-LPC et le système de fonctionnement de codeur ACELP ainsi que les différents codeurs basés sur CELP. Ensuite, nous avons décortiqué l'utilisation et le fonctionnement du codec G723.1, qui est la base de notre étude, en citant quelques avantages et inconvénients de ce dernier. D'autre part une étude sur les différentes méthodes de masquage PLC a été faite.

Chapitre 2 – codage de la parole : analyse par synthèse

Au terme de notre étude sur les algorithmes de codage, nous avons fait un constat sur la qualité de codage qui permet une bonne compression de la voix et par conséquent l'optimisation de la bande passante nécessaire tout en offrant une augmentation du nombre d'appels simultanés possibles. En effet, les nombreuses méthodes de numérisation existantes ont pour objectif d'exploiter la grande variété de codecs CELP vu leur bonne qualité de compression, mais malheureusement, la forte intercorrélation utilisée pour le codage de deux trames successives, les rend très vulnérables face aux pertes de paquets. Donc pour cela, nous allons tenter de proposer des méthodes de masquage de pertes de paquets basées sur le codec G723.1, afin d'améliorer la qualité de service des systèmes exploitant ce codec à la présence de perte

Chapitre 3

Amélioration de la qualité de service (QoS) : Simulation

3.1 Introduction

La qualité de service est un ensemble de normes et de contraintes que le réseau doit respecter pour offrir un niveau de service approprié à la transmission avec une bonne perception, et ce, dans un délai raisonnable. Un des paramètres de base qui caractérise la QoS dans un trafic IP est le taux de perte de paquets. En effet, dans un codec, il existe des algorithmes intégrés de correction de perte de trames, tel que, la technique de masquage de pertes (PLC).

Dans le cas de la mauvaise estimation du signal de parole, des trames manquantes sont affectées et une propagation d'erreur est introduite sur les trames bien reçues. Pour cela une amélioration des algorithmes de masquage de perte (PLC) est nécessaire pour une bonne estimation du signal de la parole en cas de pertes. Dans ce qui suit, nous allons proposer des méthodes alternatives pour améliorer l'algorithme de masquage de pertes de paquets, PLC. Ces approches se basent sur le déploiement d'un algorithme d'optimisation pour générer une meilleure excitation à la place de celle offerte par l'algorithme PLC original du codec G723.1 [1].

3.2 Qualité de service

Actuellement, les différents flux sont regroupés sur un seul et même support que constituent les réseaux IP. Les débits ne cessent donc de grimper et les réseaux sont désormais multiservices. Cette centralisation simplifie grandement les tâches des administrateurs et utilisateurs car il n'y a plus qu'une seule infrastructure à gérer. Cependant, sans régulation du trafic, et pour cela les réseaux IP se retrouvent vite saturés. La multiplication des flux provoque l'engorgement des liaisons, il faut donc faire la police pour fluidifier la circulation. Afin de pouvoir assurer un niveau de service satisfaisant pour les utilisateurs et améliorer les performances d'un réseau. Il convient d'envisager une notion nouvelle qui est la qualité de service.

La qualité de service est la capacité à véhiculer dans de bonnes conditions un type de trafic donné, en terme de disponibilité, débit, délai de transmission, gigue, taux de perte de paquets... et aussi c'est le mécanisme utilisé afin d'assurer la priorisation du trafic de la connexion et garantir une bande passante minimale. Donc, la QoS est essentielle pour la VoIP. Si notre réseau est saturé, la communication en temps réel est particulièrement atteinte, car il n'importe pas s'il s'agit de vidéo, d'audio ou des deux à la fois, il devient impossible de maintenir la conversation.

La téléphonie IP peut garantir une qualité d'appel élevée si la signalisation et les canaux audio ont la priorité sur les autres utilisations du réseau, grâce au QoS. Pour que le client se bénéficie d'une bonne qualité de voix, principaux problèmes susceptibles d'apparaître dans ce type de service doivent être résolus : la latence, la gigue, l'écho et la perte de paquets. Pour mesurer la qualité du signal, il existe deux types de mesure, la mesure objective et la mesure subjective. Les mesures objectives de la qualité de la parole sont purement des mesures mathématiques évaluées en utilisant des distances euclidiennes. Les mesures subjectives de qualité évaluent la qualité de codage par des tests d'écoute par exemple, MOS, qu'on a vu auparavant.

3.3 Types de signaux

La production d'énergie de signal de parole a deux processus qui peuvent intervenir simultanément :

3.3.1 Le son voisé

Dans ce cas, les cordes vocales sont placées cote à cote. Sous l'influence de l'air expulsé par les poumons, ces cordes vocales vibrent et produisent ainsi une onde quasi-périodique. La fréquence fondamentale de cette onde sonore est dite « pitch ».

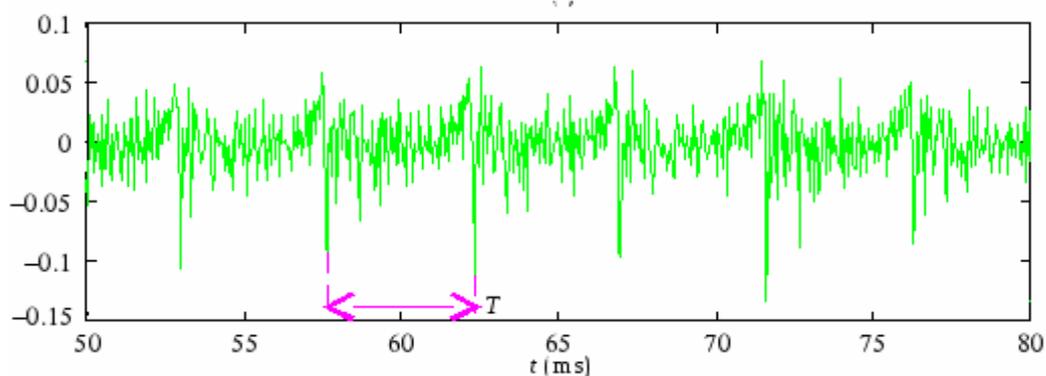


Fig. 3.1 Signal de parole d'un son voisé

3.3.2 Le son non voisé

L'air passe à travers les cordes vocales sans les faire entrer en vibration. Il en résulte un son aperiodique.

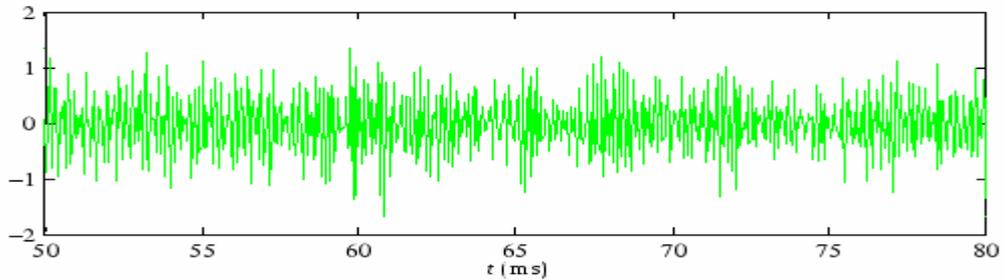


Fig. 3.2 Signal de parole d'un son non voisé

3.4 Model et outils de mesure

3.4.1 Model Gilbert Elliot

Dans notre étude, nous avons adopté le modèle Gilbert Elliot basé sur une chaîne de Markov. Ce modèle a deux états reflétés, si le paquet précédent est reçu (état 0) ou perdu (état 1). Ainsi il est défini par les probabilités de transition p et q comme sont représentés sur la figure (3.3). Soit cpl dénote la probabilité pour que le modèle de réseau abandonne un paquet sachant que le paquet précédent est abandonné, c'est-à-dire la probabilité pour que le modèle reste dans l'état 1, cette probabilité est également connue comme la probabilité conditionnelle de perte (cpl). Et le taux de perte de paquets (PLR) peut être calculé comme suit [4],

$$PLR = \frac{p}{p+q} \quad (3.1)$$

$$cpl = 1 - q \quad (3.2)$$

A partir de ces probabilités p et q , La probabilité pour qu'un paquet soit abandonné sans connaître si le paquet précédent est livré ou abandonné. C'est-à-dire la probabilité de perte sans conditions (upl) est exactement la probabilité pour que le modèle de réseau soit dans l'état 1.

$$upl = \frac{p}{p+q} \quad (3.3)$$

La longueur moyenne d'une rafale de perte, ABL , qui correspond au modèle (cpl) peut être déterminée par la formule suivante :

$$ABL = \sum_{i=1}^{\infty} i \cdot (1 - cpl) \cdot cpl^{i-1} = \frac{1}{q} \quad (3.4)$$

La figure 3.3 présente le modèle de Markov avec ses probabilités de transition.

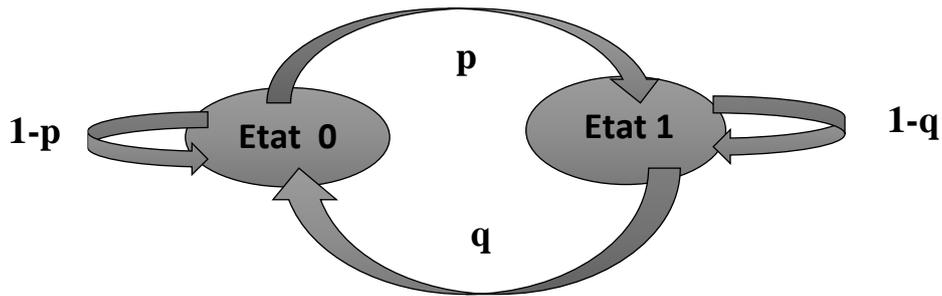


Fig. 3.3 Modèle de Gilbert-Elliott

3.4.2 Outils de mesure

a- PESQ

Pour évaluer la qualité de la voix transmise par un système de télécommunication, nous utilisons une méthode objective de prévision de la qualité de la téléphonie et des codes vocaux de bout en bout. Il s'agit de l'outil PESQ, normalisé par la recommandation P.862 de l'ITU-T [40] et qui permet une évaluation de la qualité vocale perçue. Dans le cadre de cette évaluation, deux paramètres d'erreur sont calculés dans le modèle cognitif, lesquels seront combinés pour obtenir une note MOS objective concernant la qualité d'écoute.

b- La base de donnéé :

Dans notre travail, on va utiliser une base de données de parole TIMIT [41], [42], est une base de données acoustique et phonétique dédiée à la reconnaissance de la parole indépendamment du locuteur. Elle contient des enregistrements à large bande de 630 locuteurs américains (438 hommes et 192 femmes), répartis en huit dialectes régionaux. Les séquences de cette base sont sous-échantillonnées à 8khz avec une longueur de 8s à 19s donc 15s à la moyenne pour chaque séquence, pour une mesure de qualité en bande étroite. Cette base sert pour la mesure de qualité objective des deux catégories signal de test et signal de référence (test and training), en prenant un ensemble de 40 séquences de parole. Afin d'obtenir une note d'évaluation globale du PESQ moyen pour chaque condition de canal (taux de pertes). La valeur du PESQ moyen pour un seul canal de transmission avec pertes est calculée comme suit [4] :

$$PESQ_{moy} = \frac{1}{N_{sq}} \cdot \frac{\sum_{j=1}^{N_{sq}} PESQ_{j.Lsqj}}{\sum_{j=1}^{N_{sq}} L_{sqj}} \quad (3.5)$$

Ou N_{sq} est le nombre de séquence, pour notre cas ça sera 40, et L_{sqj} la longueur d'une séquence avec j le nombre de trames de chaque séquence.

3.5 Génération du signal d'excitation

Une procédure de masquage des erreurs a été incorporée dans le décodeur afin de réduire la dégradation dans le signal vocal reconstitué en raison d'effacements de trame dans le flux binaire. Ce processus de masquage des erreurs est fonctionnel lorsque la trame des paramètres du codeur, correspondant à une trame de 10 à 20 ms, a été identifiée comme étant perdue. La stratégie de masquage consiste à reconstruire la trame actuelle sur la base de l'information déjà reçue. Cette méthode remplace le signal d'excitation manquant par un signal de caractéristiques similaires. Pour cela nous avons tenté de proposer des méthodes ou des modifications du PLC du G.723.1 afin de mieux estimer l'excitation de la trame perdue. Pour ce faire, à chaque fois, la qualité de la méthode proposée sera jugée. Pour ce, nous avons fait appel au métrique PESQ pour comparer la robustesse de nos méthode face au PLC du codec G.723.1 original.

3.5.1 Génération du signal d'excitation avec des zéro

Dans cette partie nous avons essayé d'implémenter l'idée de masquage de perte avec le remplacement de la trame perdue avec une suite de zéros. Cette idée permettra au décodeur de considérer la trame perdue comme une trame qui ne contient aucune information. L'excitation générée comparée à l'excitation originale sont données par la figure (3.4).

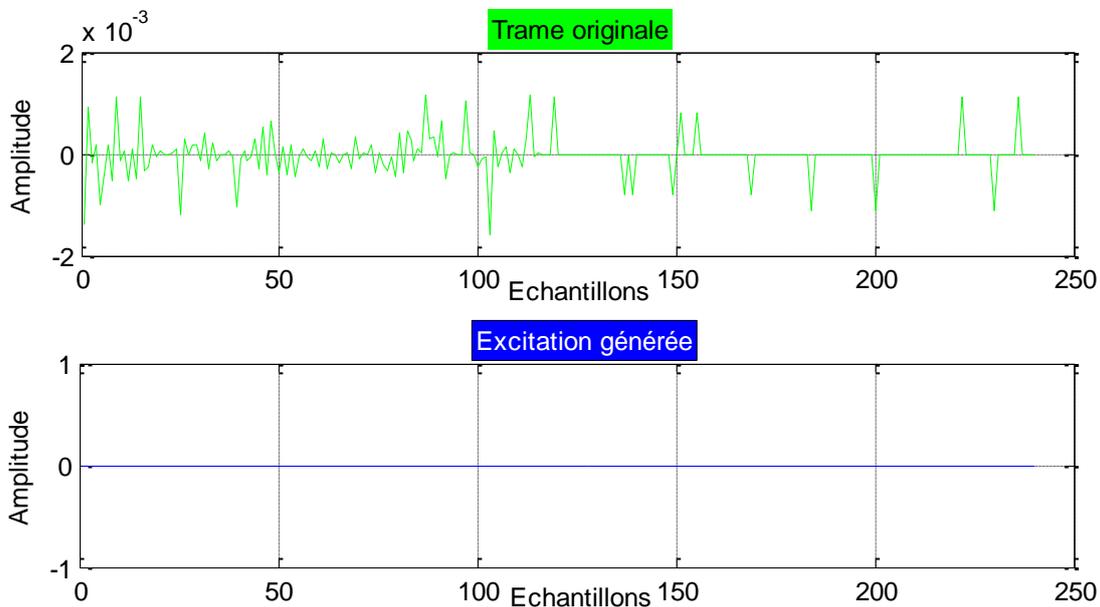


Fig. 3.4 L'excitation générée par implémentation des zéros

A partir de la figure (3.4), nous pouvons voir que la qualité subjective du signal est totalement absente. Mais il est tout de même nécessaire de calculer les résultats de mesures objectives du PESQ pour mieux juger. Pour tester l'efficacité de toutes les méthodes proposées, nous avons calculé le PESQ sur le PLC original et pour toutes les méthodes en question. Pour

cela, nous avons utilisé un ensemble de masques avec des taux de pertes de, 8%, 10%, 13%, 16%, 18%, 20%, contenant chacune, une rafale de pertes de 2 trames perdues successivement. Les valeurs du PESQ pour la méthode d'implémentation de zéros est données par le tableau (3.1).

Les résultats du PESQ en fonction des taux de pertes	
Taux de pertes	PESQ Moyen
2%	3.4147
4%	3.0932
6%	2.8480
8%	2.4847
10%	2.3148
13%	2.0727
16%	1.8568
18%	1.7383
20%	1.6503

Tableau 3.1 Les résultats du PESQ avec implémentation des zéros

D'après le tableau (3.1), pour un taux de perte de 8%, nous avons un résultat du PESQ 2.4847. Mais à chaque fois que le taux de perte est plus important, la valeur de PESQ se dégrade rapidement et fortement. Pour un taux de perte de 20%, nous avons obtenu une valeur de PESQ égale à 1.6503, chose qui confirme davantage l'inefficacité de remplacement par la trame passée.

3.5.2 Génération du signal d'excitation de la trame passée

Dans cette méthode, nous avons proposé un PLC qui utilise la totalité de l'excitation de la trame passée pour remplacer l'excitation de la trame actuelle perdue. Cette méthode est aussi nommée réplique de forme d'ondes qu'on a étudié dans chapitre 2, exploitée dans certains codecs standardisés à l'occurrence le G.729. La figure (3.5) illustre, l'excitation donnée par cette méthode en comparaison avec la bonne excitation sans perte.

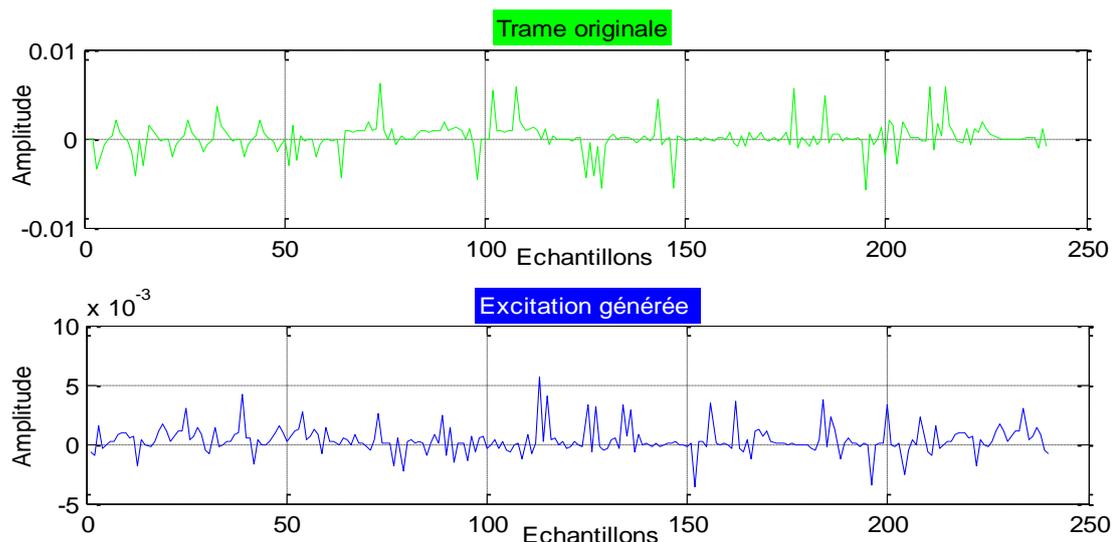


Fig. 3.5 Excitation générée par la trame passée

D’après la figure (3.5), les deux excitations ne se ressemblent vraiment pas, puisque dans cette méthode le type (voisé/non-voisé) de la trame passée n’est pas pris en compte. Ce qui nous donne des résultats moyens. Pour mieux juger et montrer l’efficacité de la méthode proposée, nous avons utilisé 6 taux de pertes, et les résultats PESQ obtenus sont donnés dans le tableau (3.2).

Les résultats du PESQ en fonction des taux de pertes	
Taux de pertes	PESQ Moyen
2%	3.5597
4%	3.1537
6%	2.9566
8%	3.0120
10%	2.9075
13%	2.6804
16%	2.5601
18%	2.4648
20%	2.4824

Tableau 3.2 Les résultats du PESQ avec la trame passée.

Avec cette méthode de masquage, les résultats du PESQ sont supérieurs aux résultats de la méthode d’implémentation par des zéros. Pour un taux de perte 8%, par exemple, le résultat de PESQ est de 3.0120, par contre, pour l’implémentation par des zéros est de 2.4847 pour le même taux de pertes. Et pour un taux de 20%, le PESQ est égal à 2.4824, pour la trame passée, et il est de 1.6503, pour l’implémentation par des zéros.

3.5.3 Génération du signal d'excitation PLC du codec G723.1

La méthode PLC originale implémentée dans le G.723.1, se base sur la génération des excitations. L'excitation régénérée est utilisée en tant qu'entrée dans filtre de synthèse LP sur la base des coefficients LP antérieurs. D'une manière plus facile, elle repose sur la dernière trame qui a été bien reçue pour générer une nouvelle trame qui va remplacer la trame perdue. Les résultats de l'excitation PLC régénérée et la bonne excitation originale de la trame perdue sont donnés par la figure (3.6) ci-dessous.

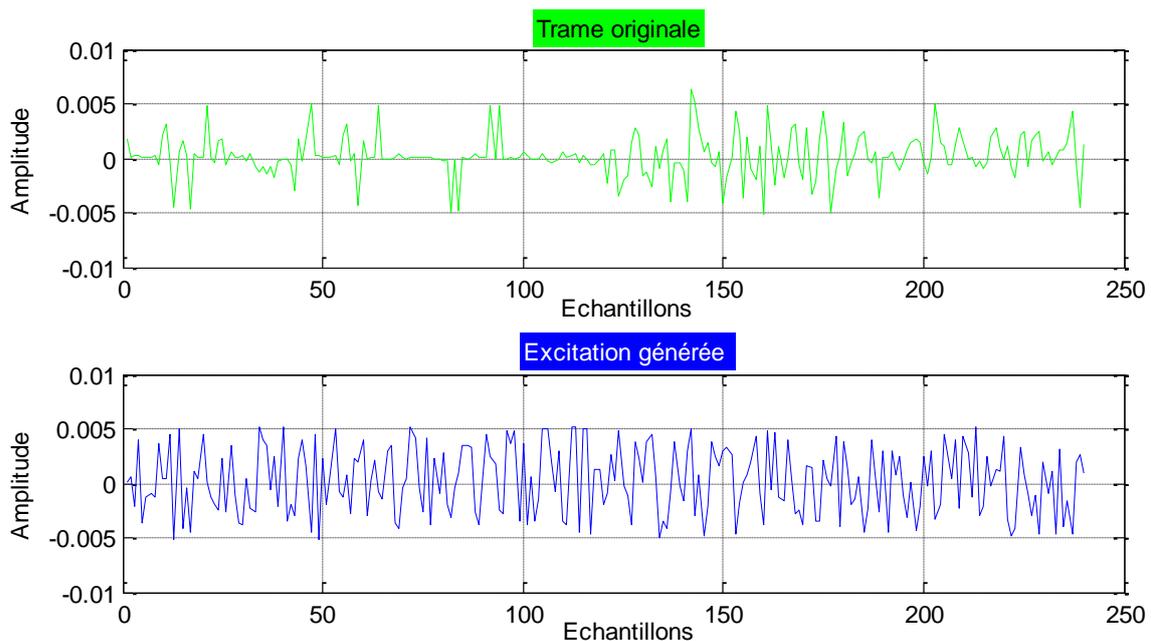


Fig. 3.6 Excitation générée par le PLC du G723.1, trame passée non voisée.

D'après la figure (3.6), nous pouvons bien voir que l'excitation générée dans le cas où la trame passée est jugée non voisée, ne correspond pas à l'excitation originale. Vu que l'excitation PLC donnée est basée sur le non-voisement modélisée par un bruit, or que, le voisement est basé sur la périodicité ou redondance existante dans le signal de la parole.

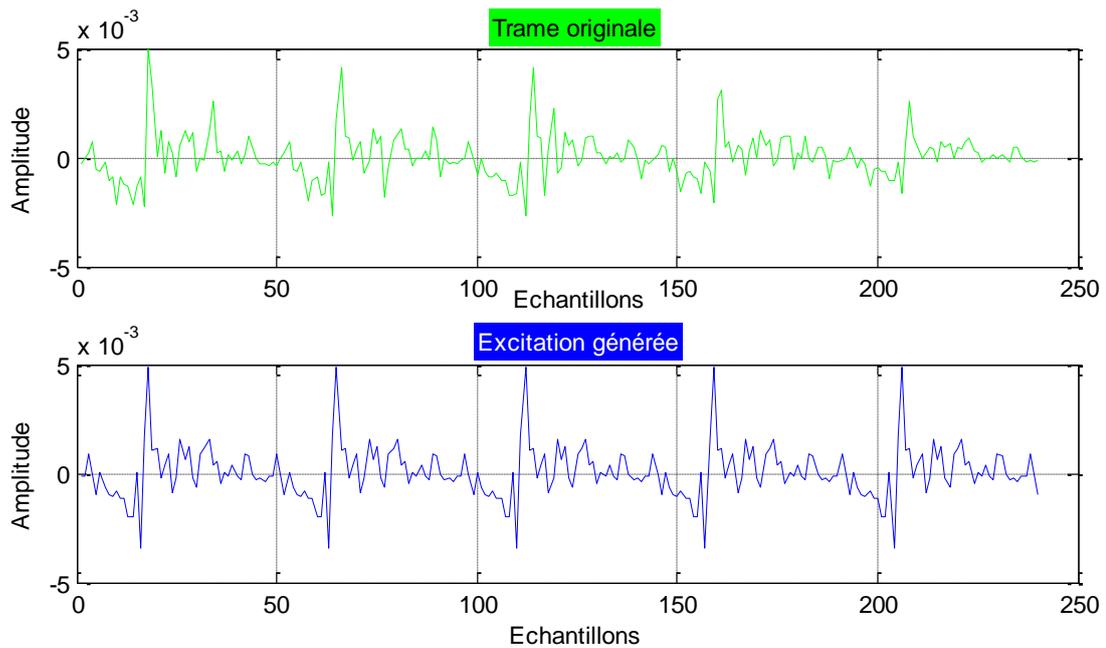


Fig. 3.7 Excitation générée par le PLC du G723.1, trame passée voisée

D’après la figure (3.7), nous pouvons bien voir que l’excitation générée dans le cas où la trame passée détectée voisée, l’excitation PLC de la trame perdue correspond vraiment à l’excitation de la bonne trame originale sans perte.

Les valeurs du PESQ PLC de G.723.1 obtenues sont données par le tableau (3.3).

Les résultats du PESQ en fonction des taux de pertes	
Taux de pertes	PESQ Moyen
2%	3.5454
4%	3.4213
6%	3.3255
8%	3.1695
10%	3.1006
13%	3.0108
16%	2.9021
18%	2.8233
20%	2.7484

Tableau 3.3 Les résultats du PESQ avec le PLC de G.723.1

Pour le codeur G.723.1, un taux de perte de 8%, donne une valeur du PESQ 3.1695 et la dégradation est plus forte lorsque le taux de pertes augmente. Lorsque le taux de perte est égal à 20%, nous avons obtenu un résultat du PESQ égal à 2.7484. Ces résultats du PESQ seront un repère de comparaison pour les méthodes proposées. Donc, comme premier constat, le PLC

original du codec G.723.1 est mieux que les deux méthodes précédentes. Et pour cela, nous avons proposé une nouvelle méthode alternative pour mieux estimé le signal d'excitation.

3.6 Régénération du signal d'excitation par un filtrage adaptatif

Le filtrage adaptatif est présent en traitement du signal dans de nombreuses applications, tels que, l'identification de système, la réduction d'interférence, l'annulation d'écho ou encore l'égalisation et la réduction de propagation d'erreur [43]. Pour réduire davantage cette dernière, nous allons introduire au niveau du décodeur un filtre des moindres carrés récursifs (RLS), afin d'offrir une meilleure estimation du signal d'excitation avec une forte ressemblance en forme d'onde avec celle de la trame perdue.

3.6.1 Algorithme RLS

Le RLS est un algorithme fournit une convergence rapide, une erreur asymptotique faible et est peu sensible aux variations des valeurs propres du système. Dans le but d'offrir un meilleur estimé de signal d'excitation, nous définissons, α comme filtre RLS avec L coefficients, tel que, $\alpha_{j=1,\dots,L}$, et $e_{PLC}(n)$, étant le vecteur contenant les 240 échantillons de la trame PLC courante. Les étapes de sa réalisation est comme suit :

a. Étape de filtrage

Implique le calcul des données de sortie du filtre, tel que :

$$\ddot{e}(n) = \alpha(n) \cdot e_{PLC}(n) \quad (3.6)$$

L'erreur à minimiser est définie comme la différence entre le signal estimé, $\ddot{e}(n)$ et le signal de l'excitation courante $\hat{e}(n)$,

$$\varepsilon(n) = \ddot{e}(n) - \hat{e}(n) \quad (3.7)$$

b. Étape d'adaptation

Procède à la mise à jour du filtre en fonction de l'estimation de l'erreur, pour cela la fonction du coût est définit par [44] :

$$J(n) = \sum_{i=1}^n \lambda^{n-1} |\varepsilon(i)|^2 \quad (3.8)$$

Où, λ ($0 < \lambda < 1$) est le facteur d'oubli, qui permet à l'algorithme d'oublier le passé trop lointain et de poursuivre les non stationnarités intervenantes dans les signaux.

$$\alpha(n) = \alpha(n - 1) + G \cdot [\hat{e}(n) - e_{PLC}^t(n)\alpha(n - 1)] \quad (3.9)$$

Avec le gain G qui se définit comme suit :

$$G(n) = \frac{\lambda^{-1} P(n-1) e_{plc}}{1 + \lambda^{-1} e_{plc}^t(n) P(n-1) e_{plc}(n)} \quad (3.10)$$

Le calcul de G est effectué à l'aide de la matrice de covariance qui se définit selon :

$$P(n) = \lambda^{-1} P(n - 1) - \lambda^{-1} G(n) e_{plc}^t(n) P(n - 1) \quad (3.11)$$

Et cette matrice initialisée de la manière suivante :

$$P(0) = \sigma^{-1} I \quad (3.12)$$

Avec I , la matrice identité et σ une constante positive très faible.

3.6.2 Principe de base des méthodes proposées

a. Expérience 1 : RLS-PLC

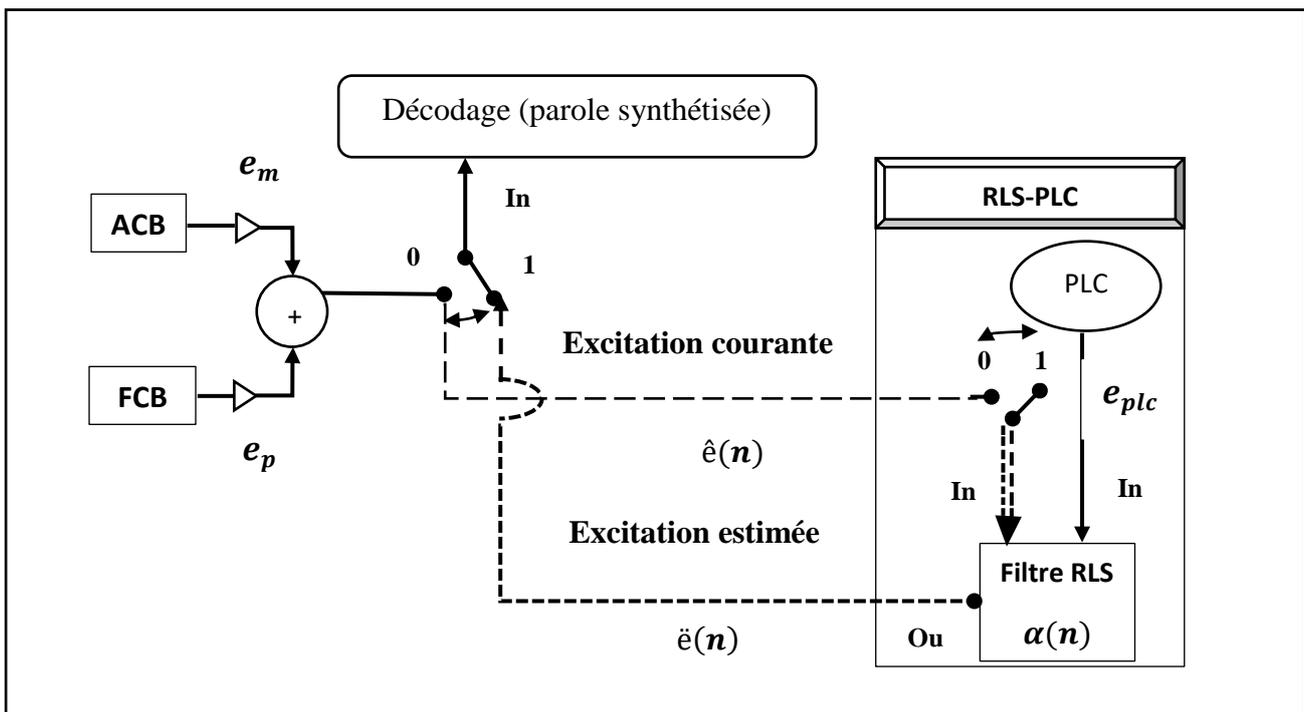


Fig. 3.8 Schéma bloc de décodeur modifié pour la méthode proposée (RLS-PLC)

Le schéma de la figure (3.8), montre le module PLC du codec G.723.1 complété par un filtre RLS pour produire une excitation basée sur les composantes de l'excitation qui sort du module PLC. Ce filtre effectue une estimation du signal d'excitation sur l'ensemble d'une trame. Il en

résulte une séquence d'excitation de 40 échantillons. Pour ce faire, nous avons proposé un système de filtrage adaptatif qui intervient sur deux étapes principales :

- Apprentissage : état '0' sans perte.

Le filtre RLS est considéré comme un égaliseur avec deux entrées en minimisant l'erreur quadratique entre ces deux vecteurs, pour ajuster ses coefficients et les utilise dans l'état '1' avec perte. Le premier vecteur contient les 240 échantillons disponibles de l'excitation de la trame PLC $e_{plc} = [e_{plc1}(n), \dots, e_{plc240}(n)]$. Par ailleurs, le second est constitué des 240 échantillons provenant de l'excitation de la trame courante $\hat{e}(n) = [\hat{e}_1(n), \dots, \hat{e}_{240}(n)]$.

Le PLC du codec G.723.1 travaille avec nombre d'échantillon égal à 240, cela justifie notre choix de prendre directement les 240 échantillons requis dans cette approche.

- Filtrage : état '1' Avec perte.

La fonction RLS dans ce cas, est vue comme étant un simple filtre avec une entrée, une sortie et une réponse impulsionnelle d'une longueur de 240 échantillons. En effet, le signal sortant de ce filtre $\hat{e}(n)$ est excité par le vecteur de 240 échantillons de l'excitation $e_{plc}(n)$ et remplace la trame perdue une fois il y'a perte, qui donne une régénération d'un bon estimé du signal $\hat{e}(n)$.

Pour initialiser les paramètres de notre filtre adapté, nous avons effectué des variations sur sa longueur L et le facteur d'oubli λ . Au premier lieu, nous avons effectué des tests exhaustifs sur un ensemble de 8 séquences afin de choisir les bonnes valeurs de λ et L qui donne une bonne qualité de parole objective maximale, par rapport au PLC original tel que : $PESQ_{plc}(8séq) = [3.5626 \ 3.4061 \ 3.2645 \ 3.1278 \ 3.0828 \ 2.9320 \ 2.8222 \ 2.7459 \ 2.7441]$

Par ailleurs, nous présenterons les résultats du PESQ sur un ensemble de 40 séquences (10 minutes de discussions). Les tableaux qui suivent montrent les résultats des tests obtenus.

		Facteur d'oubli $\lambda = 0.9$							
PLR \ L	2%	4%	6%	8%	%10	%13	%16	%18	%20
L=2	3.4900	3.2006	3.0703	2.7601	2.6423	2.6275	2.4229	2.1341	2.1840
L=4	3.5140	3.1559	3.0641	2.7660	2.6203	2.5927	2.3643	1.9816	2.1432
L=6	1.6430	3.4362	2.9615	2.9034	2.7735	2.4573	2.5405	2.2025	2.0370
L=8	3.4271	2.7638	2.7452	2.6679	2.4643	2.4959	2.2873	1.9864	2.0177
L=10	3.2174	2.7782	2.7102	2.4566	2.3104	2.3731	2.2053	1.8205	1.9658
L=12	3.3266	2.8553	2.8028	2.5856	2.3797	2.3453	2.1861	1.9125	1.8849

		Facteur d'oubli $\lambda = 0.95$							
PLR \ L	2%	4%	6%	8%	%10	%13	%16	%18	%20
L=2	3.3384	3.2046	3.0492	2.7598	2.6562	2.6683	2.4817	2.2362	2.1980
L=4	3.4998	3.2437	3.0917	3.8063	2.7411	2.6931	2.4906	2.3064	2.1448
L=6	3.4581	3.2333	3.0671	2.8059	2.7043	2.7097	2.5062	2.3690	2.1498
L=8	3.4723	3.2074	3.0779	2.7379	2.6123	2.7087	2.5443	2.2963	2.1868
L=10	3.4861	3.1557	2.9758	2.7280	2.5550	2.7074	2.5322	2.2940	2.1221
L=12	3.4287	3.0853	2.9273	2.7359	2.5900	2.6923	2.4920	2.3018	2.0808

		Facteur d'oubli $\lambda = 0.96$							
PLR \ L	2%	4%	6%	8%	%10	%13	%16	%18	%20
L=2	3.4637	3.1787	3.0241	2.6914	2.6020	2.5797	2.3680	2.2420	2.0893
L=4	3.4998	3.2352	3.0426	2.7144	2.6427	2.6173	2.4092	2.2818	2.0531
L=6	3.3482	3.2105	3.0180	2.7528	2.6897	2.6445	2.4433	2.3199	2.0882
L=8	3.4856	3.2243	3.0317	2.7155	2.6547	2.6433	2.4693	2.3433	2.1825
L=10	3.3482	3.2231	3.0314	2.7456	2.6448	2.6696	2.4814	2.3312	2.1395
L=12	3.4816	3.1516	2.9867	2.7298	2.6568	2.6709	2.4305	2.2790	2.1031

Tableau 3.4. Les résultats du PESQ en fonction du facteur d'oubli, λ , et de la longueur, L, de RLS-PLC associés aux différents taux de pertes de paquets (PLR).

Le tableau ci-dessus, montre les résultats de PESQ pour différentes valeurs de λ (0.9 ; 0.95 ; 0.96) en fonction de différents taux de pertes PLR (2%, 4%, 6%, 8%, 10%, 13%, 16%, 18%, et 20%). Après l'analyse des résultats obtenus pour ces tests, nous avons remarqué qu'avec une longueur du filtre égale à $L = 6$, et $\lambda = 0.9$ le filtre arrive à offrir un bon estimé du signal d'excitation (régénérer par PLC) qui se traduit par de bonnes valeurs du PESQ particulièrement pour le taux de pertes 4%. Donc nous sommes basés sur ces résultats pour choisir les bonnes valeurs de λ et L, afin de les utiliser pour l'ajustement des variantes de la méthode proposée RLS-PLC. Le tableau (3.5) ci-dessous montre les résultats obtenus pour un test effectué sur un sous-ensemble de 40 séquences de parole, extrait de la base de données TIMIT.

Les résultats du PESQ en fonction des taux de pertes	
$\lambda = 0.9$	$L = 6$
Taux de pertes	PESQ Moyen
2%	1.2713
4%	3.4633
6%	3.1321
8%	2.9544
10%	2.7367
13%	2.5557
16%	2.3735
18%	2.1366
20%	1.9746

Tableau 3.5 Les résultats du PESQ avec RLS-PLC

D’après le tableau (3.5), pour une longueur du filtre $L = 6$, un facteur d’oubli $\lambda = 0.9$ et un taux de perte de 4% on remarque une valeur améliorée par rapport aux résultats du PESQ avec le PLC original de codeur G.723.1 qui a été présenté par le tableau (3.3).

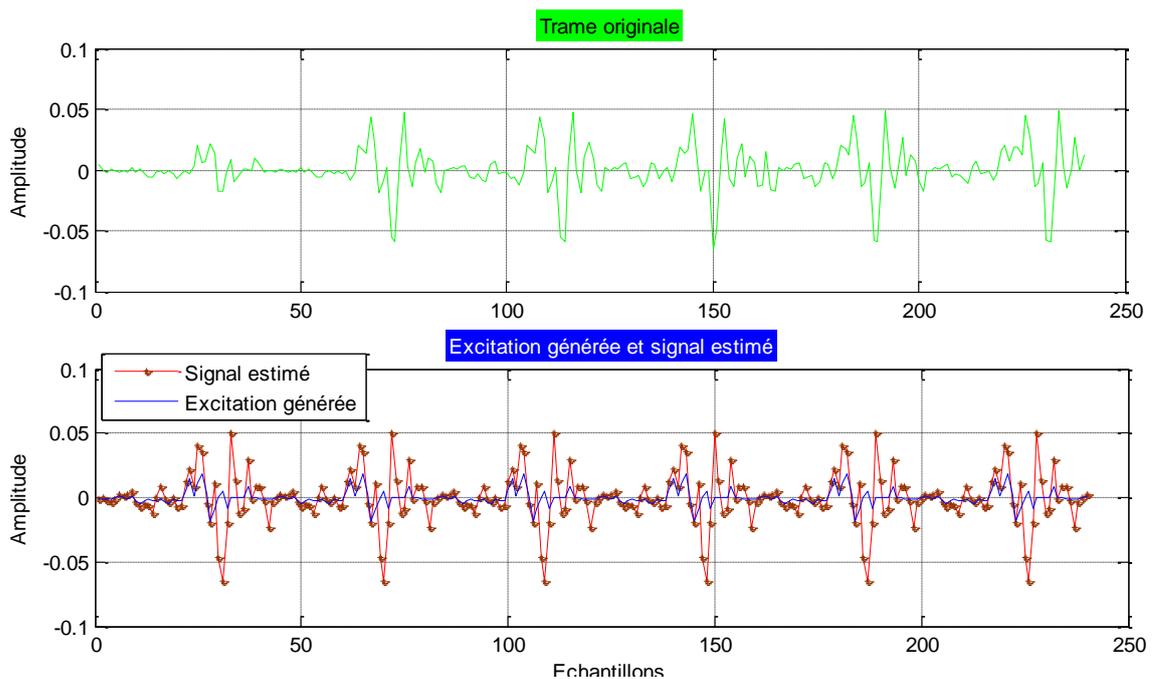


Fig. 3.9 Signal estimé par RLS-PLC

D’après la figure (3.9), nous pouvons bien voir que l’excitation régénérée dans le cas de RLS-PLC, se rapproche vraiment de l’excitation de la bonne trame originale sans perte. On constate que la convergence de filtre RLS commence à partir de 60 échantillons, d’où la différence entre la trame originale et l’excitation estimée dans la première partie du signal.

b. Expérience 2 : RLS-TP

Dans cette expérience, nous avons remplacé l’excitation PLC par la trame passée (les derniers 240 échantillons), afin de trouver une bonne estimation de cette dernière comparant aux valeurs de PESQ de l’excitation de la trame passée. Pour ce faire, nous avons effectué les mêmes étapes présentées précédemment. Le tableau (3.6), illustre les résultats du PESQ (sur 8 séquences) de la méthode proposée en fonction du facteur d’oubli, λ , et la longueur, L , du filtre RLS associés au différents taux de pertes de paquets (PLR). Sachant que, $PESQ_{TP}(8 \text{ séq}) = [3.5254 \ 3.3283 \ 3.1625 \ 2.9245 \ 2.8614 \ 2.7001 \ 2.5547 \ 2.4940 \ 2.5239]$.

		Facteur d’oubli $\lambda = 0.90$								
PLR \ L		2%	4%	6%	8%	10%	13%	16%	18%	20%
L=2		3.3875	3.1312	2.9079	2.4071	2.3545	2.4548	2.2954	2.1277	1.9696
L=4		3.3333	3.0675	2.9270	2.4105	2.3211	2.4889	2.3015	2.1161	1.9245
L=6		3.3973	2.9874	2.7799	2.3843	2.2587	2.4429	2.2622	2.0611	1.8855
L=8		3.3738	2.9057	2.5312	2.3697	2.2722	2.4047	2.1947	2.0089	1.7549
L=10		3.3796	2.9102	2.6445	2.3689	2.1825	2.3565	2.1035	1.9660	1.7948
L=12		3.3659	2.9033	2.6013	2.1693	2.0854	2.3145	2.0797	1.9334	1.6321

		Facteur d’oubli $\lambda = 0.95$								
PLR \ L		2%	4%	6%	8%	%10	%13	%16	%18	%20
L=2		3.4414	3.1187	2.9104	2.6014	2.4999	2.4543	2.2025	2.0994	1.9783
L=4		3.4414	3.1450	2.9492	2.5350	2.4626	2.4522	2.2384	2.0982	1.9860
L=6		3.4751	3.1775	2.9838	2.6871	2.6378	2.6541	2.4676	2.3185	2.1727
L=8		3.4751	3.1982	3.0167	2.7231	2.6377	2.6455	2.4700	2.3306	2.1431
L=10		3.4041	3.4881	3.2119	3.0159	2.7378	2.6156	2.6553	2.4700	2.1803
L=12		3.3903	3.4849	3.2022	2.9786	2.6597	2.6005	2.6455	2.4330	2.1531

		Facteur d’oubli $\lambda = 0.98$								
PLR \ L		2%	4%	6%	8%	10%	13%	16%	18%	20%
L=2		3.4025	3.0940	2.8343	2.4990	2.3861	2.2708	2.0243	1.8817	1.8161
L=4		3.4162	3.0936	2.8469	2.5237	2.4141	2.2915	2.0473	1.9069	1.8488
L=6		3.4162	3.0936	2.8753	2.5378	2.4518	2.3036	2.0976	1.9582	1.8488
L=8		3.4162	3.0936	2.8732	2.5378	2.4627	2.3428	2.1106	1.9582	1.8618
L=10		3.4382	3.1034	2.8972	2.5598	2.4627	2.3428	2.1106	1.9712	1.8760
L=12		3.4382	3.1034	2.8939	2.5598	2.4627	2.3708	2.1335	1.9941	1.8618

Tableau 3.6 Les résultats du PESQ en fonction du facteur d’oubli, λ , et de la longueur, L , de RLS-TP associés aux différents taux de pertes de paquets (PLR).

Le tableau (3.6), illustre les résultats de PESQ pour différentes valeurs de λ (0.9 ; 0.95 ; 0.98) en fonction de différents taux de pertes PLR . Les résultats obtenus pour cette méthode sont intéressants, car ils sont très proches de ceux de la trame passée. Nous avons remarqué qu’avec une longueur du filtre égale à, $L = 10$, et $\lambda = 0.95$, le filtre arrive à offrir un bon estimé du signal d’excitation, particulièrement pour les taux de pertes 4%, 6%, 8% et 16%. Le tableau (3.7) ci-dessous montre les résultats obtenus pour un test effectué sur un sous-ensemble de 40 séquences de parole pour la méthode RLS-TP.

Les résultats du PESQ en fonction des taux de pertes	
Taux de pertes	PESQ Moyen
$\lambda = 0.95$	$L = 10$
2%	3.4904
4%	3.2131
6%	2.9870
8%	2.8487
10%	2.6902
13%	2.5622
16%	2.4457
18%	2.3467
20%	2.1579

Tableau .3.7 Les résultats du PESQ pour la méthode RLS-TP.

Pour cette deuxième expérience, nous pouvons remarquer que la méthode RLS-TP proposée offre de meilleures valeurs PESQ par rapport aux résultats de la trame passée. En particulier, lorsque PLR prends les valeurs 4%, et 6% de pertes. En effet, pour un taux de 4%, nous avons un $pesq = 3.2131$, contre, $pesq = 3.1537$ et pour un taux de 6%, nous avons une valeur de PESQ égale à $pesq = 2.9870$ contre $pesq = 2.9566$ de la trame passée.

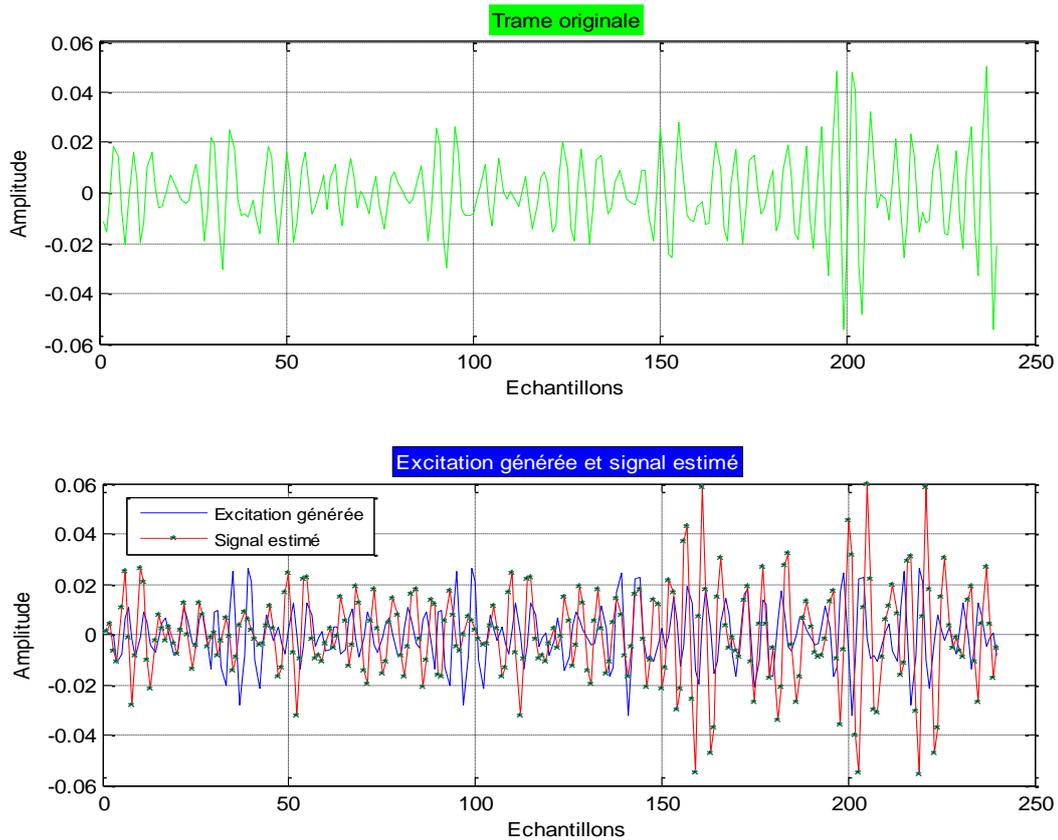


Fig. 3.10 Signal estimé par RLS-TP

D'après la figure (3.10), nous remarquons que le signal estimé approche en forme d'onde du signal original mais avec des amplitudes différentes, cela est dû à l'ajustement des coefficients de filtre selon la forme de la trame originale. Par contre, lorsque l'excitation générée ressemble cette dernière, le filtre n'intervient pas.

c. Expérience 3 : RLS comme FEC

➤ La méthode RLS-PLC comme FEC

Dans La figure (3.11), nous décrivons comment intégrer la méthode utilisée précédemment (RLS-PLC) au niveau du codeur CELP de la norme G.723.1, afin d'obtenir les informations supplémentaires FEC. Pour ce faire, l'excitation régénérée $\hat{e}(n)$, est utilisée pour la mise à jour de la mémoire ACB désynchronisée.

Par ailleurs, la figure (3.12) montre la resynchronisation de la mémoire ACB au niveau du décodeur. Cette procédure aura lieu une fois la trame précédente considérée comme perdue. Donc, l'excitation donnée par le répertoire ACB sera remplacée par les paramètres FEC ajoutés.

Pour tester l'efficacité de cette méthode proposée, nous avons suivi les mêmes étapes citées précédemment. Commenant par des tests sur 8 séquences en variant les paramètres de filtre λ et L . Ensuite, nous avons choisi les bonnes valeurs des paramètres RLS pour trouver les résultats PESQ sur un ensemble de 40 séquences.

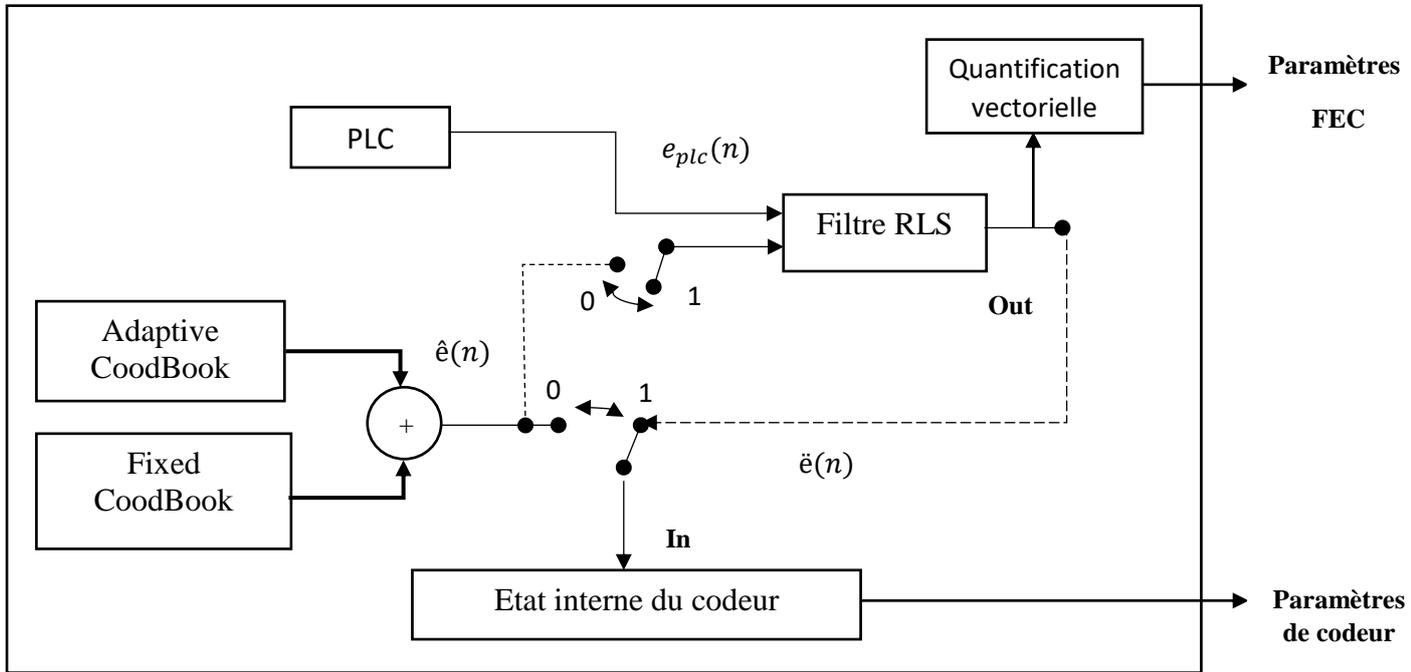


Fig. 3.11 Schéma du codeur modifié pour la méthode RLS-PLC comme FEC.

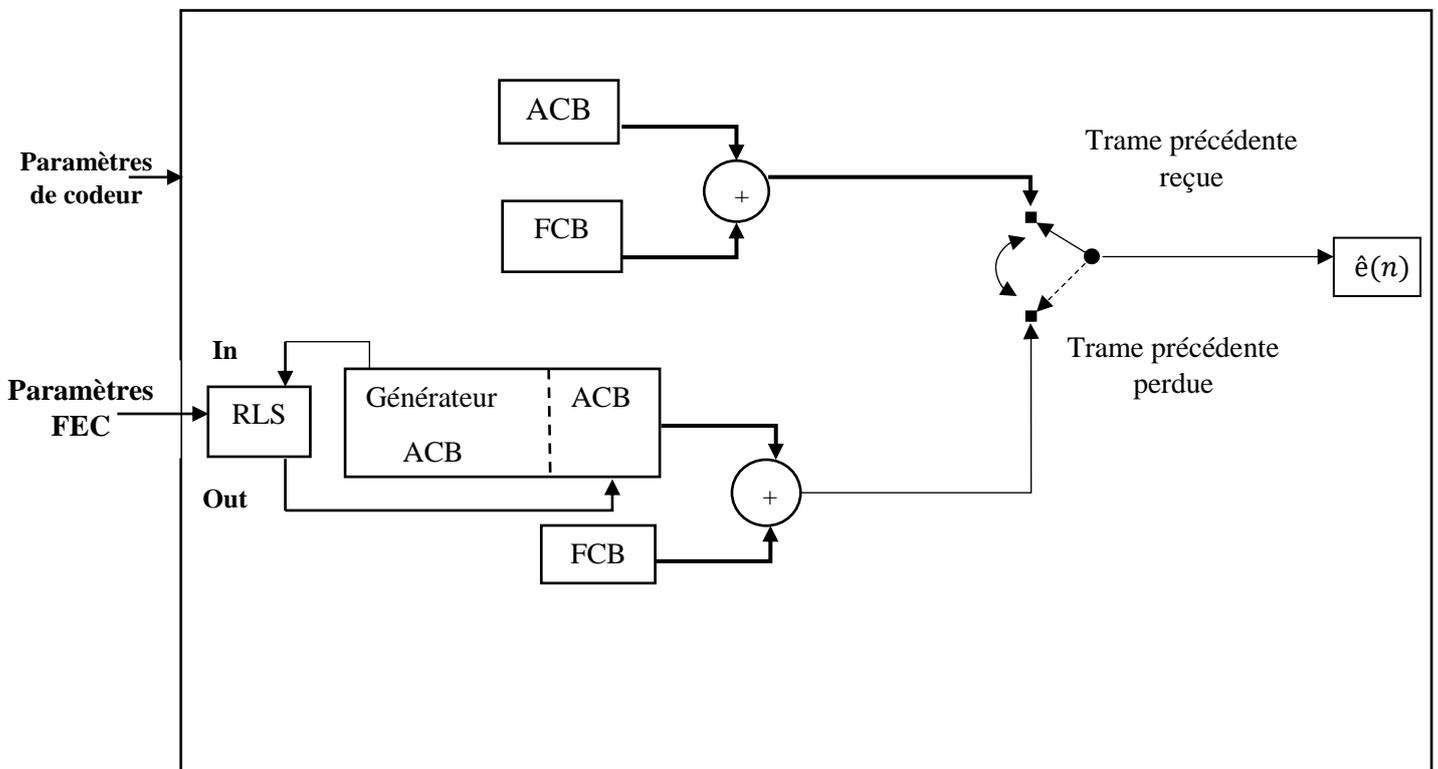


Fig. 3.12 Schéma du décodeur modifié pour la resynchronisation du ACB.

		Facteur d'oubli $\lambda = 0.68$							
PLR \ L	2%	4%	6%	8%	10%	13%	16%	18%	20%
L=2	3.4391	3.2895	3.1751	3.0761	3.0004	2.9659	2.8205	2.7271	2.4949
L=4	3.5286	3.4160	3.2842	3.0378	2.9933	2.7693	2.4428	2.3611	2.5977
L=6	3.4861	3.3748	3.2137	3.0374	2.9601	2.8312	2.5093	2.3892	2.5909
L=8	3.5339	3.3870	2.2542	2.9830	2.7666	2.6508	2.3409	2.3124	2.3358
L=10	3.4735	3.2635	3.1297	2.8192	2.5496	2.6442	2.3433	2.2902	2.1009
L=12	3.4598	3.2651	3.1616	2.9078	2.8354	2.5756	2.3001	2.2562	2.3424

		Facteur d'oubli $\lambda = 0.69$							
PLR \ L	2%	4%	6%	8%	10%	13%	16%	18%	20%
L=2	3.3296	3.4533	3.3048	3.1646	3.0138	2.9874	2.9409	2.8205	2.7792
L=4	3.4696	3.4051	3.2842	3.0171	2.9912	2.8438	2.5134	2.4746	2.5816
L=6	3.5290	3.3886	3.2267	3.0508	2.9860	2.8170	2.4858	2.3903	2.6168
L=8	3.5339	3.4070	3.2542	3.0179	2.9032	2.6759	2.3799	2.2822	2.4651
L=10	3.4614	3.2874	3.1600	2.8440	2.7792	2.6515	2.2906	2.2638	2.3154
L=12	3.5104	3.2481	3.1300	2.9709	2.7194	2.6136	2.3573	2.2529	2.1818

Tableau 3.8 Les résultats du PESQ en fonction du facteur d'oubli, λ et de la longueur, L.

Avec cette méthode, on remarque à travers ces tableaux que les résultats PESQ sont améliorés par rapport à celles du RLS-PLC. De plus, nous avons eu une amélioration des certains taux de pertes en comparant les valeurs avec PLC original. Par exemple, pour $\lambda = 0.68$, nous avons obtenu un PESQ amélioré pour des taux de pertes 13% et 4% pour des longueurs de $L = 2$ et $L = 4$, respectivement. Et une qualité de parole objective meilleure lorsque $\lambda = 0.69$ et $L = 2$ en particuliers pour les taux de pertes 4%, 8%, 13%, 16%, 18% et 20%. Le tableau (3.9) illustre les résultats de PESQ sur un ensemble de 40 séquences avec, $\lambda = 0.69$ et $L = 2$

Les résultats du PESQ en fonction des taux de pertes	
$\lambda = 0.69$ $L = 2$	
Taux de pertes	PESQ Moyen
2%	3.5806
4%	3.2794
6%	3.4140
8%	3.0776
10%	2.9445
13%	2.7916
16%	2.6723
18%	2.5661
20%	2.7959

Tableau 3.9 : Les résultats du PESQ avec RLS-PLC comme FEC.

Chapitre 3 –Amélioration de la qualité de service (QoS) : Simulation

A partir ces valeurs présentées dans le tableau (3.9), on note qu’avec la méthode RLS-PLC comme FEC lorsque $PLR = 2\%, 4\%, 20\%$, les résultats du PESQ ont augmenté par rapport à ceux de RLS-PLC et même de PLC original. Ce qui nous donne une qualité de service meilleur pour ces taux de pertes.

➤ La méthode RLS-TP comme FEC

C’est le même principe que l’expérience précédente, avec RLS-PLC comme FEC, en remplaçant le module PLC par la trame passée.

		Facteur d’oubli $\lambda = 0.68$							
PLR \ L	2%	4%	6%	8%	%10	%13	%16	%18	%20
L=2	3.5116	3.1557	2.9660	2.8290	2.7332	2.7776	2.5738	2.4842	2.3203
L=4	3.5116	3.1868	2.9823	2.8685	2.7694	2.6996	2.5485	2.4441	2.2661
L=6	3.4844	3.1170	2.8813	2.8027	2.6925	2.5903	2.5530	2.3973	2.1947
L=8	3.3559	2.6997	2.7763	2.7606	2.6825	2.5442	2.3491	2.2304	2.1628
L=10	3.4256	3.0817	2.8875	2.6481	2.5591	2.6271	2.4344	2.3244	2.0989
L=12	2.7372	2.1484	1.9655	1.6510	1.5777	1.7205	1.4981	1.4796	1.3468

		Facteur d’oubli $\lambda = 0.70$							
PLR \ L	2%	4%	6%	8%	%10	%13	%16	%18	%20
L=2	3.4617	3.1249	2.9213	2.7102	2.6213	2.7181	2.5385	2.4353	2.1552
L=4	3.4974	3.1661	2.9479	2.8558	2.7702	2.7605	2.5892	2.4852	2.2749
L=6	3.4979	3.1840	2.9826	2.8869	2.7768	2.7388	2.5914	2.4602	2.2218
L=8	3.1321	2.5438	2.3149	2.3574	2.2647	2.2362	2.0387	1.9231	1.8084
L=10	2.9506	2.4108	2.1205	1.9201	1.9349	2.0746	1.8497	1.7291	1.4858
L=12	2.8417	2.2556	2.0028	1.7962	1.8207	1.9212	1.6433	1.5198	1.3406

Tableau 3.10 Les résultats du PESQ en fonction du facteur d’oubli, λ et de la longueur, L.

Le tableau (3.10), montre les résultats de PESQ pour la méthode RLS-TP comme FEC. On remarque que ces tests sont améliorés par rapport à la méthode RLS-TP pour les petites longueurs de filtres ($L = 2,4,6$). Par contre, pour des longueurs plus grandes ($L = 8,10,12$) on visualise une diminution considérable presque pour tous les taux de pertes. Donc on a choisi $L = 2$ et $\lambda = 0.68$ pour un test de 40 séquences présenté par le tableau suivant.

Les résultats du PESQ en fonction des taux de pertes	
$\lambda = 0.68$	$L = 2$
Taux de pertes	PESQ Moyen
2%	3.4911
4%	3.1656
6%	2.9585
8%	2.7422
10%	2.5493
13%	2.5488
16%	2.3646
18%	2.2422
20%	2.0594

Tableau 3.11 Les résultats du PESQ avec la méthode RLS-TP comme FEC

Dans le tableau précédent (3.11), on observe une amélioration au niveau de masque 4% et 6% pour une longueur de $L = 2$ et un facteur d'oubli $\lambda = 0.68$ pour cette méthode (RLS-TP comme FEC).

3.6.3 Résultats généraux et discussion

PLR \ Méthodes	2%	4%	6%	8%	10%	13%	16%	18%	20%
Génération avec des zéros	3.4147	3.0932	2.8480	2.4847	2.3148	2.0727	1.8568	1.7383	1.6503
Génération avec la trame passée	3.5597	3.1537	2.9566	3.0120	2.9075	2.6804	2.5601	2.4648	2.4824
Génération avec PLC	3.5454	3.4213	3.3255	3.1695	3.1006	3.0108	2.9021	2.8233	2.7484
RLS-PLC	1.2713	3.4633	3.1321	2.9544	2.7367	2.5557	2.3735	2.1366	1.9746
RLS-TP	3.4904	3.2131	2.9870	2.8487	2.6902	2.5622	2.4457	2.3467	2.1579
RLS-PLC comme FEC	3.5806	3.3751	3.4140	3.0776	2.9445	2.7916	2.6723	2.5661	2.7959
RLS-TP comme FEC	3.4911	3.1656	2.9585	2.7422	2.5493	2.5488	2.3646	2.2422	2.0594

Tableau 3.12 Les résultats des méthodes utilisées

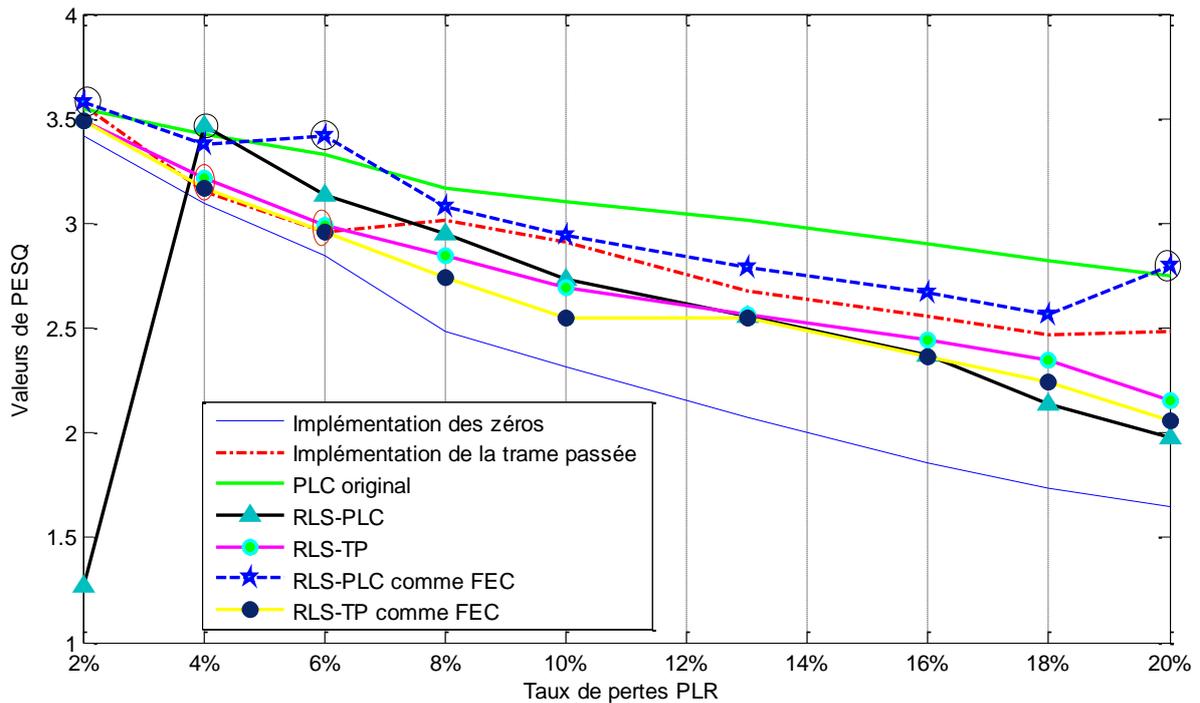


Fig. 3.15 L'évolution du PESQ pour chaque méthode

Le tableau (3.12) et la figure (3.15), présentent une comparaison entre les différentes méthodes proposées. D'après les résultats on constate que :

- ✓ PLC original du codec G.723.1 donne une excitation meilleure que la trame passée et celle d'implémentation des zéros.
- ✓ La méthode RLS-PLC est conseillée pour améliorer la qualité de service du codec G.723.1 avec un taux de pertes égale à 4%. Par contre, RLS-PLC comme FEC pour les taux de pertes de 2%, 6% et 20%.
- ✓ L'utilisation du filtre RLS avec la trame passée, améliore la robustesse du codec G.729 avec des taux de pertes de 4% et 6%. De même, pour la méthode RLS-TP comme FEC.

3.7 Conclusion

Dans ce chapitre, on a défini la notion de qualité de service, le modèle et ces outils de mesures utilisés. Pour améliorer cette qualité, nous avons testé plusieurs techniques de génération du signal d'excitation pour les trames perdues. Commencant par une implémentation des zéros, puis la génération avec la trame passée, ensuite avec le PLC original du codec G.723.1. Afin d'améliorer ces types de méthodes de masquage, nous avons proposé d'introduire

Chapitre 3 –Amélioration de la qualité de service (QoS) : Simulation

un filtre adaptatif RLS pour trouver un bon estimé du signal perdu. Pour ce faire, le filtre ajuste ces coefficients par une phase d'apprentissage, pour ensuite, passer à la phase de filtrage.

En fait, la première expérience se base sur l'implémentation du filtre RLS avec PLC du codec G.723.1 au niveau du décodeur uniquement. Tandis que, la seconde consiste à améliorer l'excitation de la trame passée, en remplaçant le PLC dans la méthode RLS-PLC avec la trame passée. En ce qui concerne la dernière expérience, la trame perdue sera régénérée en utilisant le filtre RLS comme FEC basé au niveau du codeur. Au premier lieu, avec module PLC, ensuite avec la trame passée.

Enfin, chaque méthode proposée améliore le PESQ pour des taux de pertes particuliers pour des cas différents de longueurs du filtre et valeurs du facteur d'oubli. Donc chaque expérience dépend des paramètres choisis. Tout de même, les résultats obtenus prouvent l'efficacité des méthodes proposées pour des taux de pertes bien précis. Comme perspectives, il est nécessaire, d'effectuer d'autres choix et tester d'autres types de filtres adaptatifs.

Conclusion

Générale

Conclusion générale

Dans ce projet de fin d'études, nous avons fait une étude sur la qualité de service (QoS) associée à la transmission de la VoIP. Pour se faire, nous avons détaillé les différents mécanismes permettant d'atteindre une bonne qualité de service sur des environnements de communication de données sévères. Notre objectif en premier lieu, est de comprendre le principe des codeurs de parole CELP. Car ces derniers sont les plus utilisés pour le codage de la voix sur IP en mode de en commutation de paquets. En effet, la majorité des codecs utilisés dans la VoIP sont basés sur le CELP grâce à leur bonne qualité de compression. Parmi ceux, nous retrouvons le codeur G.723.1 qui fonctionne avec deux débits distincts, un à 5.3 kbps et le second à 6.3 kbps. Toutefois, dans le but de remédier aux pertes, ce dernier est implémenté d'un algorithme intégré de correction de perte de trames (PLC). Afin d'améliorer ses qualités objective et subjective, nous avons proposé des techniques pour mieux estimer le signal de la parole des segments perdus et être capable d'augmenter la robustesse du codeur G.723.1 en présence d'erreurs de transmission.

En premier lieu, nous avons présenté un aperçu global sur la transmission de la voix sur le réseau IP, son architecture, son principe de fonctionnement, les protocoles de transport et signalisation utilisés et les défis rencontrés notamment la perte de paquets. Nous avons contribué à cet axe en essayant de présenter des méthodes dites de masquage de perte de paquets (PLC) pour y remédier.

Par ailleurs, le deuxième chapitre est dédié au codage de la parole CELP basé sur analyse par synthèse. En commençant par exposer l'idée générale sur le codage de parole à prédiction linéaire. Par la suite nous avons illustré les différents codeurs basés sur CELP. En effet, la différence principale entre ces codeurs est la méthode qui est employée pour coder l'excitation, en particulier le nombre de bits employés à cette fin.

En dernier lieu, dans le chapitre trois, nous avons présenté de nouvelles techniques qui exploitent les propriétés de codage CELP, pour surpasser les problèmes liés aux erreurs de transmission. En effet, pour améliorer la qualité de service de la transmission de la voix sur IP, nous avons suggérés d'utiliser un filtre adaptatif des moindres carrés récursifs (RLS) comme modèle d'estimation sur deux volets distincts. Le premier, est basé au niveau du décodeur comme une amélioration du PLC original, tandis que le second, est un modèle basé au niveau du codeur comme technique de correction précoce d'erreur (FEC). Débutant par une combinaison de ce filtre avec le PLC en modifiant le décodeur G.723.1, puis une régénération de la trame passée a été faite avec la méthode RLS-TP. Ensuite, pour chaque expérimentation,

Conclusion générale

nous avons obtenu des améliorations des valeurs PESQ pour des taux de perte particuliers, par rapport à la technique de masquage originale.

Au terme de ce travail, nous estimons que les résultats obtenus sont assez intéressants qui nous ont permis de nous imprégner davantage par des connaissances techniques et nouvelles sur la technologie des réseaux de communication de paquets pour la transmission de la voix sur IP. Comme perspectives, nous souhaitons améliorer les méthodes mises en place afin d'attendre des améliorations sur les différents taux de pertes avec d'autres filtres adaptatifs, à savoir, le filtre de Wiener et le filtre de projection affine.

Bibliographie

Normes :

- [1] Rec. ITU-T G.723.1, (1996). Dual rate speech coder for multimedia communication transmitting at 5.3kbit/s and 6.3kbit/s.
- [13] Rec. ITU-T. G.722.1, (2005). Low-complexity coding at 24 and 32 Kbit/s for hands-free operation in systems without low frame loss.
- [14] Rec. ITU-T. G.722.2, (2003). Wideband Coding of Speech at Around 16 kbit/s Using AM-RWB. *Int Telecomm Union, Geneva.*
- [15] Rec. ITU-T. G.729, (1996). Coding of speech at 8 kbit/s using conjugate structure algebraic-code-excited linear prediction (cs-acelp). *tech. rep., International Telecommunication Union.*
- [28] L'UIT, S.D.L.N.D.T.D., (2006). Codeur vocal à double débit pour communications multimédias acheminées à 5,3 kbit/s et à 6,3 kbit/s.
- [40] Perceptual Evaluation of Speech Quality (PESQ), (2001). An Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs. ITU-T P.862 Recommendation.

Sites web :

- [2] FrameIP.com. (2017). *VoIP-la voix sur IP*. <https://www.frameip.com>
- [3] Wikipédia. (2016). *Itinérance dans l'Union européenne*. <https://fr.m.wikipedia.org/wiki/>
- [5] WikiMémoires. (2001, 25 mars). *Etude et mise ne place d'un centre d'appels via IP*.
<https://wikimemoires.net>
- [35] W3Tel. *Comprendre la VoIP*. <https://www.w3tel.com/>

Mémoires & thèses :

- [4] BENAMIROUCHE, N. (2017, janvier). *Amélioration du codage de la parole de type CELP face aux erreurs de transmission* (Thèse). Université Abderrahmane Mira Bejaia, faculté Sciences et Technologies.
- [6] BOUZAIDA, R. (2011). *Etude et Mise en place d'une Solution VoIP Sécurisée* (Mémoire). Université Virtuelle de Tunis.
- [7] DIDI, S., & GUERRICHE, M. (2014, juin). *La Téléphonie sur IP (ToIP)* (Mémoire). Université Abou Bakr Belkaid– Tlemcen, département d'Informatique
- [10] SAIDI, M. (2010, mai). *Codage Sinusoïdal de la Parole pour des Transmissions de la Voix sur IP (VoIP)* (Thèse). Université des Sciences et de la Technologie Houari Boumediene USTHB, faculté d'Electronique et Informatique.
- [12] RATSIMBAZAFY, S. (2007, décembre). *Le codage de la parole en GSM et les schémas de codage en GPRS* (Mémoire). Université d'antananrivo, département télécommunication.
- [16] DJAMAH, M. *codage échelonnable à granularité fine de la parole utilisant la quantification vectorielle arborescente* (Thèse). Université du Québec Centre Energie Matériaux Télécommunication (INRS).
- [17] Roy, G., & Eng, B. (1990). *Low-rate analysis-by-synthesis wideband speech coding*. Department of Electrical Engineering McGill University.
- [20] LACHACHI, N.-E. (2006, octobre). *Codage paramétrique de la parole en vue de sa transmission sur Internet* (Mémoire). Université d'Oran -Es-Senia-, département d'Informatique.
- [21] TAMANNA, I. (2000, avril). *Interpolation of Linear Prediction Coefficients for Speech Coding* (Mémoire). Department of Electrical Engineering McGill University.
- [22] NADIM, B. (1998, mai). *Robust Spectral Parameter Coding in Speech Processing* (Mémoire). Department of Electrical Engineering McGill University.
- [24] BOUDRAA, B. (2006, septembre). *Analyse et Synthèse Multi-Impulsionnelle* (Thèse). Université des Sciences et de la Technologie Houari Boumediene USTHB, faculté d'Electronique et Informatique.
- [26] GRASSI, S. (1998). *Optimized Implementation of Speech Processing Algorithms* (Thèse). Université de Neuchatel.

Bibliographie

- [29] KABAL, P. (2009). *ITU-T G.723.1 Speech Coder: A Matlab Implementation*. Department of Electrical & computer Engineering McGill University.
- [34] BOUZID, R., & CHABANA, D. (2019, juillet). *Etude d'un système de communication VoIP* (Mémoire). Université Abderrahmane Mira Bejaia, faculté Sciences et Technologies.
- [37] KOENIG, L. (2011, janvier). *Masquage de pertes de paquets en voix sur IP* (Thèse). Institut National Polytechnique de Toulouse (INP Toulouse).
- [39] CHAOUCH, H. (2005, juin). *Amélioration des performances du codec G.729* (Mémoire). Université des Sciences et de la Technologie Houari Boumediene USTHB, faculté d'Electronique et Informatique.
- [44] BOUNIF, A. (2014, octobre). *Etude de l'annulation d'écho acoustique stéréophonique par les algorithmes de filtrage adaptatif transverse* (Thèse). Université Blida 1, faculté des Sciences de l'Ingénieur.

Livres :

- [8] LAURENT, O., & GUY, P. (2008). *Téléphonie sur IP*. (EYROLLES) (2ème édition éd.). EYROLLES.
- [9] DI GALLO, F. (1999-2000). *Cours de réseaux et systèmes*. (CNAM BORDEAUX).CNAM.
- [18] Moreau, N. (2007). *Outils pour la compression application à la compression des signaux audio*. Edition collection technique et scientifique des télécommunications. PARIS-TECH.

Articles :

- [11] MURGIA, C. FENG, G., GUYADE, A. & QUINQUIS, C. (1995). Codage des signaux de parole (20HZ à 15KHZ) à très faible retard au débit de 64 Kbit/s. *institue de la communication parlée. URA CNR-368*, (pp. 345-348).
- [19] MAKHOUL, J. (1975). Linear Prediction: A Tutorial Review. *In Proc. IEEE*, (pp. 561-580).
- [23] ITAKURA, F. (1975). Line spectrum representation of linear predictor coefficients of speech signals. *The Journal of the Acoustical Society of America*, 57(S1), (pp. 35-35).
- [25] ATAL, B. & REMDE, J. (1982). A new model for LPC excitation for producing natural sounding speech at low bit rates. *In Proc. ICASSP-82*, (pp. 616-617).
- [27] LANDAU, I. D., N. M'SIRDI, & M. M'SAAD. (1986). Techniques de modélisation récursive pour l'analyse spectrale paramétrique adaptative. *Revue de Traitement du Signal* 3, (pp. 183-204).
- [32] SCHROEDER, M. & ATAL, B. (1985). Code-Excited Linear Prediction (CELP): High-quality speech at very low bit rates. *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-85*, volume Vol.10, (pp. 937-940).
- [33] LAFLAMME, C., ADOUL, J.-P., SU, H. & MORISSETTE, S. (1990). On reducing computational complexity of codebook search in CELP coder through the use of algebraic codes. *International Conference on Acoustics, Speech, and Signal Processing, 1990. ICASSP-90*, volume Vol. 1, (pp. 177-180).
- [38] SANNECK, H., STENGER, A., YOUNES, K. B., & GIROD, B. (1996). A new technique for audio packet loss concealment. *IEEE. In Global Telecommunications Conference, GLOBE-COM'96.*, Communications: The Key to Global Prosperity (pp. 48-52).

Bibliographie

[41] LAMEL, L., KASSEL, R. & SENEFF, S. (1986). Speech database development: Design and analysis of the acoustic-phonetic corpus. *In Proc Speech Recognition Workshop (DARPA)*, (pp. 100-110).

[43] BENAMIROUCHE, N., & BOUDARAA, B. (2013). Improved Frame Loss Recovery Using RLS for PLC Excitation Shape Reconstruction. *Proceedings Engineering & Technology*, Vol. 1, (pp. 42-45).

Rapport:

[32] Accellent. (2007). *La qualité de Service de la Voix sur IP (principes et assurance)*. Le Descartes. <https://wallu.pagesperso-orange.fr/VoIP.pdf>

CD-ROOM :

[42] GAROFOLO, J. S. The Structure and Format of the DARPA TIMIT, *CD-ROM Prototype, Documentation of DARPA TIMIT*.

Résumé

La qualité de service est un paramètre standard très important pour juger la VoIP. La QoS traite l'adéquation mentionnée en termes de paramètres de performances du réseau : débit, gigue, délai et perte de paquets. De toute évidence, la majorité des codecs de la parole utilisés dans la voix sur IP sont basés sur le CELP vu la qualité de compression offerte. Malheureusement, ils sont très vulnérables face aux pertes de paquets chose qui dégrade davantage la qualité de service. De ce fait, notre travail est basé sur l'amélioration de l'algorithme du masquage de perte de paquets afin de remédier aux problèmes de pertes et offrir une meilleure qualité de décodage en présence de pertes. Ainsi, nous avons proposé deux méthodes, une est basée au niveau du codeur, comme FEC et la seconde est basée au décodeur. Pour se faire, nous avons fait usage d'un filtrage adaptatif RLS comme estimateur des signaux de l'excitation pour les trames perdues. Enfin, les résultats expérimentaux ont prouvé l'efficacité des deux méthodes proposées et une amélioration de la qualité de service par rapport au codec original.

Abstract

Quality of service is a very important standard parameter in judging VoIP service. QoS addresses the mentioned fit in terms of network performance parameters: throughput, jitter and buffer delays and packet loss. Obviously, the majority of speech codecs used in VoIP are CELP-based due to the offered quality of compression. Unfortunately, they are very vulnerable to packet loss which further degrades the quality of service. Therefore, our work is based on packet loss concealment algorithm improvement in order to remedy the loss problems and provide better decoding quality in the presence of losses. Thus, we have proposed two methods, one is coder-based method, as FEC, and the second is decoder-based method. To do this, we made use of adaptive RLS filtering as an estimator of the excitation signal in case of lost frames. Finally, the experimental results proved the effectiveness of the two proposed methods and an improvement in the quality of service is obtained relatively to the original codec.