

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE
UNIVERSITE A.MIRA-BEJAIA
FACULTE DE TECHNOLOGIE
DEPARTEMENT GENIE ELECTRIQUE



MEMOIRE
EN VUE DE L'OBTENTION DU DIPLOME DE
MASTER

Domaine : Sciences et Technologies
Filière : Télécommunications
Spécialité : Système des Télécommunications

Présenté par

DJOUADI Fouad
CHALAL Djafar

Thème

*Etude des techniques de masquage de pertes de
paquets (PLC) pour améliorer la robustesse du
codage CELP face aux erreurs de transmission
VoIP*

Devant le Jury

Président	M. TOUNSLIM
Examineur	M. ALLICHE.A
Encadreur	M. BENAMIROUCHE.N

Année Universitaire : 2018/2019

Remerciements

Nous remercions tous ceux qui ont contribué de près ou de loin à la réalisation de ce modeste travail.

Nous saurons remercier assez notre encadreur Mr Benamirouche pour son soutien à la fois pédagogique et moral et pour sa disponibilité, ce qu'il nous a permis d'achever ce projet.

Dédicaces

Nous dédions cet humble travail, à nos familles et à tous nos amis.

Djafar , Fouad

Table des matières

Remerciements

Dédicaces

Table des matières.....I

Liste des figures et tableaux.....V

Introduction générale.....1

CHAPITRE 1 : Transmission de la voix sur IP

1.1 Introduction : 3

1.2 Le signal de parole 4

1.2.1 Modèles du signal de parole 5

1.2.1.1 Modèle source-filtre :..... 5

1.3 la voix sur IP 6

1.3.1 définition..... 6

1.3.2 Architecture 6

1.3.3 Mode de fonctionnement 8

1.4 Protocoles de la VoIP 8

1.4.1 Protocole H.323 : 8

1.4.1.1 Avantages et inconvénients du protocole H.323 :..... 9

1.4.2 Protocole SIP 10

1.4.2.1 Avantages et inconvénients du protocole SIP :..... 10

1.4.3 Protocole RTP..... 11

1.4.3.1 Avantages et inconvénients..... 12

1.4.4 Protocole RTCP..... 12

1.4.4.1 Avantages et inconvénients..... 12

1.4.5 Protocole UDP..... 12

1.5 Conclusion.....13

CHAPITRE 2 : Modélisation LPC et codage de la parole

2.1. Introduction.....	14
2.2. Codage numérique de la parole.....	14
2.2.1 Les codecs à forme d'onde.....	15
2.2.2 Les codecs paramétriques (vocodeurs).....	15
2.2.3 Les codecs hybrides.....	15
2.3. Modélisation acoustique de la parole.....	Erreur ! Signet non défini.
2.3.1. Modélisation de l'enveloppe spectrale :.....	Erreur ! Signet non défini.
2.3.2. Fenêtrage d'un signal de parole.....	Erreur ! Signet non défini.
2.4 Analyse par synthèse (ABS).....	Erreur ! Signet non défini.
2.4.1. Analyse LP.....	Erreur ! Signet non défini.
2.4.1.1 Détermination des coefficients LP par la méthode d'autocorrélation :.....	19
2.4.2 Quantification des paires de raies spectrales (LSF).....	Erreur ! Signet non défini.
2.4.3 Filtre de pondération perceptive des formants.....	Erreur ! Signet non défini.
2.5. Modélisation de la corrélation à long terme.....	Erreur ! Signet non défini.
2.5.1. Estimation de la tonie.....	Erreur ! Signet non défini.
2.5.2 Mise en forme du bruit harmonique.....	Erreur ! Signet non défini.

2.5.3. Dictionnaire adaptatif conventionnel (ACB).....	Erreur ! Signet non défini.
2.5.4 Dictionnaire fixe (FCB).....	Erreur ! Signet non défini.
2.6 Codeur	
G.723.1.....	Erreur ! Signet non défini.
2.6.	
Conclusion.....	Erreur ! Signet non défini.

CHAPITRE 3 : Techniques de masquage des pertes de trames

3.1. Introduction	27
3.2. Perte de paquets.....	27
3.2.1 Les différentes causes de perte de paquets	27
3.2.1.1 Réseaux congestionnés	27
3.2.1.2 La gigue	28
3.2.1.3 Le délai de transit (latence).....	30
3.2.1.4 Perte due au protocole.....	31
3.3. Méthodes de masquage de pertes de paquets	31
3.3.1. Méthodes basées à la source	31
3.3.1.1. Priorisation et hiérarchisation	32
3.3.1.2. Entrelacement de trames	32
3.3.2. Méthodes basées au récepteur	33
3.3.2.1. Substitution	33
3.3.2.2. Masquage de pertes.....	33
3.3.2.3. PLC du décodeur G.723.1.....	34
3.4. Conclusion.....	35

CHAPITRE 4 : Simulations et interprétations des résultats

4.1 Introduction :	36
4.2 Simulation	36
4.2.1 Base de données TIMIT :	36
4.2.2 Evaluation de la qualité sonore perçue :	36
4.2.3 Modèle de Gilbert-Elliott	37
4.3 Partie pratique :	38
4.3.1 PLC du décodeur G.723.1 :	38
4.3.2 Utilisation de la trame précédente :	39
4.3.3 Utilisation de l'excitation nulle (zéros) :	40
4.3.4 Technique (voisée=PLC- G.723.1) et (non-voisée=zéros) :	41
4.3.5 Technique (voisée= PLC-G.723.1) (non-voisée=PLC-G.723.1 pour voisée) :	42
4.3.6 Utilisation du filtre LMS :	45
4.3.6.1 Principe :	45
4.3.6.2. filtre LMS seul :	46
4.3.6.3. Filtre LMS + PLC-G.7.23.1 :	47
4.4 Récapitulatif des methodes :	48
4.5 Conclusion :	49
Conclusion générale.....	50

Bibliographie

Liste des figures

CHAPITRE I

Fig.1.1 : appareil phonatoire.....	4
Fig.1.2 : exemple d'un son voisé.....	4
Fig.1.3 : exemple d'un son non-voisé.....	5
Fig.1.4 : modèle source-filtre du signal de la parole.....	6
Fig.1.5 : Architecture générale de la VoIP.....	7

CHAPITRE II

Fig.2.1. Enveloppe spectrale d'un signal de parole.....	17
Fig.2.2 Diagramme d'un modèle simplifié pour la production de la parole.....	18
Fig.2.3. Exemple d'un spectre de filtre LP, et du filtre de pondération des formants.....	22

CHAPITRE III

Fig.3.1. L'effet de la congestion du réseau.....	28
Fig.3.2. la gigue dans le réseau.....	29
Fig.3.3. Delai variable et le tampon d'elimination de gigue.....	29
Fig. 3.4 Source de délais.....	30
Fig.3.5. Délai de transit.....	31
Fig. 3.6. Priorisation et hiérarchisation.....	32
Fig. 3.7. Méthode d'entrelacement des paquets de données.....	33

CHAPITRE IV

Fig 4.1 Modèle de Gilbert-Elliot.....	37
Fig.4.2. Excitation générée par le PLC de G.723.1.....	38

Fig.4.3. excitation avec la trame passée.....	39
Fig.4.4. excitation avec des zéros.....	40
Fig.4.5.a. excitation (voisée=PLC- G.723.1).....	41
Fig.4.5.b. excitation (non-voisée=zéros).....	42
Fig.4.6.Technique (voisée= PLC-G.723.1) (non-voisée=PLC-G.723.1 pour voisée) pour différents L.....	44
Fig.4.7.a.excitation générée par LMS lors de l'apprentissage.....	46
Fig.4.7.b.excitation générée par LMS lors de la perte.....	46
Fig.4.8.excitation(en rouge) générée par LMS+PLC-G.723.1.....	47

Liste des tableaux

Tab.4.1. valeurs du PESQ pour G.723.1.....	39
Tab.4.2. valeurs du PESQ pour génération de la trame passée.....	40
Tab.4.3 valeurs du PESQ pour génération de l'excitation nulle.....	41
Tab.4.4. valeurs du PESQ pour excitation :(voisée= PLC-G.723.1) (non-voisée=zéros).....	42
Tab.4.5. Valeurs du PSQ pour différents L.....	45
Tab.4.6. valeurs du PESQ pour excitation générée par LMS.....	47
Tab.4.8. Valeurs du PESQ pour les différentes méthodes.....	48
Tab.4.7. valeurs du PESQ pour excitation générée par LMS+PLC-G.723.1.....	48

Liste des abréviations

AbS	Analysis by Synthesis
ACB	Adaptative Code-Book
ACELP	Algebraic Code Excited Linear Prediction
ADSL	Asymmetric Digital Subscriber Line
AMR-WB	Adaptive Multi-Rate Wideband
AR	Auto Regressive
CELP	Code Excited Linear Prediction
ECC	Error Correction Codes
FCB	Fixed Code-Book
GSM	Global System for Mobile Communications
IP	Internet Protocol
ITU	International Telecommunication Union
LP	Linear Prediction
LMS	Least Mean Square
LPC	Linear Prediction Coefficients
LSF	Line Spectral Frequencies
LTE	Long Term Evolution
MP-MLQ	Multi-Pulse Maximum Likelihood Quantification
OLA	OverLapp-Add
PCM	Pulse Code Modulation
PESQ	Perceptual Evaluation of Speech Quality
PLC	Packet Loss Concealment
PSVQ	Predictive Split Vector Quantizer
QoS	Quality of Servi
RS	Reed Selomon
RTCP	Real Time Control Protocol
RTP	Real-time Transport Protocol
SCCP	Skinny Client Control Protocol
SIP	Session Initiation Protocol

STN	Switched Telephone Network
TCP	Transmission Control Protocol
TIMIT	Texas Instruments and Massachusetts Institut of Technology
UDP	User Datagram Protocol
UMTS	Universal Mobile Telecommunication System
VoIP	Voice over Internet Protocol
Wi-Fi	Wireless Fidelity
WSOLA	Waveform Similarity based Overlap-Add

INTRODUCTION GENERALE

Introduction générale :

Dans les systèmes de la téléphonie filaire classique, 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. Ces derniers sont indispensables pour réduire le débit nécessaire à la représentation des signaux audio, pour plus d'efficacité, de souplesse de stockage et de diffusion à travers les réseaux. Le codage de parole est un domaine qui a fait l'objet de nombreux travaux ces dernières années. Malgré tout, ce champ de recherche reste relativement actif et de nouveaux formats audionumériques continuent à voir le jour. La tendance actuelle est de créer des codeurs universels, capable de coder aussi efficacement la parole et la musique, et pour certains, de gérer tous types de formats multicanaux.

A partir du début des années 70 apparaît la technique de Prédiction Linéaire LPC (Linear Predictive Coding), supposant que le signal de parole peut être considéré comme la sortie d'un filtre tout-pôle. Un codeur LPC particulièrement performant utilise des techniques d'analyse par synthèse LPAS (Linear Prediction Analysis-by-Synthesis). Dans un système de codage LPAS, le signal de parole à transmettre est codé en recherchant la meilleure excitation possible d'un filtre de synthèse, dont les coefficients sont déterminés par une analyse de Prédiction Linéaire. Le plus connu de ces codeurs LPAS est le codeur de type CELP (Code-Excited Linear Prediction) conçu pour le codage de la parole à des débits allant de 4 à 16 kbits/s. Pour des débits inférieurs à 4 kbps, les codeurs paramétriques sont plus efficaces. Pour atteindre ces taux de compression, les systèmes de codage paramétriques se basent sur la connaissance du processus de production de la parole. La technique consiste à extraire du signal de parole les paramètres les plus pertinents permettant au décodeur de le synthétiser. Les performances des codeurs paramétriques, appelés aussi vocodeurs, dépendent de la précision des modèles de production de parole. La plupart des codeurs paramétriques sont basés sur le codage prédictif linéaire (LPC), connus sous le nom de vocodeurs prédictifs. Dans ce type de vocodeurs, le conduit vocal est modélisé par un filtre tout-pôle. Grâce au développement de processeurs DSP (Digital Signal Processing) performants, des modèles mathématiques très proches du système phonatoire humain ont pu être mis au point et utilisés dans des algorithmes de codage de plus en plus complexes. La plupart des codeurs de parole récemment standardisés utilisent la quantification vectorielle pour la transmission de certains paramètres, notamment les coefficients de prédiction linéaire. Afin de réduire la complexité de calcul exigé durant le processus d'encodage dans une quantification vectorielle, des méthodes de recherche rapides et efficaces sont généralement utilisées.

Avec le déploiement de la transmission de la parole sur les réseaux à paquets (voix sur IP), où la congestion du trafic est chose commune, il est devenu une exigence de développer des algorithmes de compression de parole qui doivent être capables d'adapter dynamiquement le débit binaire, de la parole compressée, à la bande passante du réseau de transmission produisant une variation lisse et graduelle de la qualité subjective de la parole synthétisée.

Dans la rédaction de ce mémoire, la méthodologie est présentée comme suit ;

Dans le premier chapitre, nous Avons fait brièvement une présentation de la transmission de la voix sur la voie IP son architecture globale et ses principaux protocoles avec leurs avantages et inconvénients.

Dans le deuxième chapitre, nous avons expliqué le fonctionnement d'un codeur et les différents codecs de signal de parole et les différentes plages de débit ainsi que l'analyse du signal de parole par une analyse par synthèse type CELP avec une description générale du codec G723.1.

Dans le troisième chapitre, nous avons présenté les différentes causes de perte de paquets dans le réseau IP, et après nous avons présenté les différentes méthodes utilisées pour le masquage de ces pertes basées à l'émission et à la réception et une description générale du décodeur G723.1.

Le quatrième chapitre, la partie de simulation, ou nous avons implémenté des idées de masquage de pertes avec interprétation des résultats obtenus et description des outils utilisés.

CHAPITRE

I

1.1. Introduction

La voix sur IP (Voice over IP) est une technologie de communication vocale en pleine émergence. Elle fait partie d'un tournant dans le monde de la communication. Ce fût en 1996 la naissance de la première version voix sur IP appelée H.323 [1], issue de l'organisation de standardisation européenne ITU-T, cette technologie permet de transmettre de la voix sur un réseau IP qui transporte les données sous formes de paquets.

Avant l'envoi du signal de la parole dans le réseau IP, il subit plusieurs modifications en l'occurrence la numérisation et la compression, ensuite le récepteur tache à ce que le signal reçu soit audible en procédant à la décompression et à la conversion numérique-analogique.

Le signal de la parole est numérisé et compressé à l'aide d'un vocodeur ; par exemple, le G.723.1, puis, encapsulé dans un niveau transport. Le protocole de transport utilisé pour la VoIP est généralement UDP qui fonctionne sur le modèle BE (Best Effort ou meilleur effort). C'est-à-dire chaque paquet est transporté du mieux possible de l'expéditeur au destinataire. Le modèle BE ne garantit pas le contrôle et n'assure pas que les paquets sont bien arrivés à destination, ainsi intervient le protocole RTP/RTCP qui évalue et contrôle la qualité de service et les pertes de paquets.

Dans la commutation de paquet qui est utilisée dans la VoIP, les paquets n'empruntent pas le même trajet, ce qui engendre un temps de parcours différent d'un paquet à l'autre, et des paquets désordonnés à l'arrivée. Pour assurer une bonne synchronisation entre l'émetteur et le récepteur et retrouver l'ordre correct des paquets, les protocoles H.323, MGCP et SIP [2], sont déployés afin de prendre en charge la signalisation, et de garantir une meilleur QoS aux utilisateurs.

1.2. Le signal de parole

Le signal de parole est le résultat de l'enregistrement de la voix à l'aide d'un microphone, et résulte de la production de la voix par l'être humain [3].

La génération de l'onde acoustique se fait au niveau de des cordes vocales à l'intérieur du larynx, tel qu'il est illustré sur la figure (1.1).

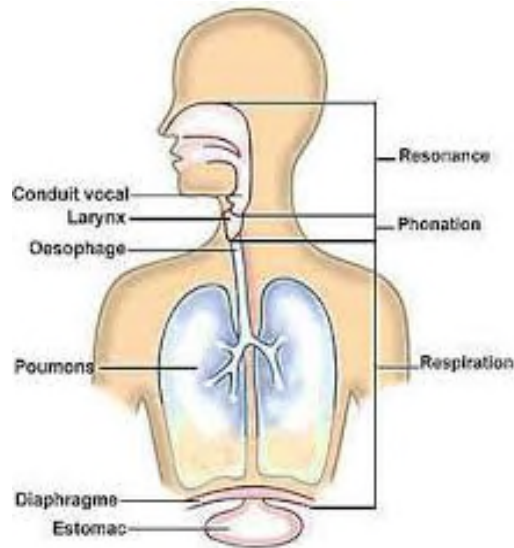


Fig.1.1 : appareil phonatoire.

L'expulsion de l'air des poumons fait vibrer les cordes vocales si elles sont fermées (les muscles du larynx placent les cordes cote à cote) ; et ce qui produit une onde quasi périodique voir figure (1.2), dont la période fondamentale de cette onde est appelée pitch, et ce son est appelé son *voisé*.

Et si les cordes vocales sont ouvertes lors de l'expulsion de l'air des poumons (c'est-à-dire il passe sans les faire vibrer), cet air passe à travers un rétrécissement au niveau de la glotte ce qui entraîne des turbulences, il en résulte un son apériodique dit *non-voisé*, illustré par la figure (1.3), généralement considéré comme du bruit.

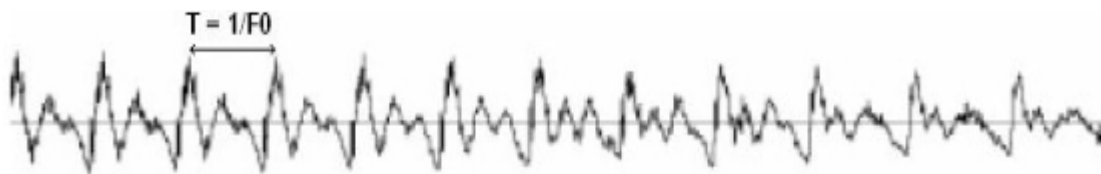


Fig.1.2 : exemple d'un son voisé



Fig.1.3 : exemple d'un son non-voisé

Ainsi, après avoir passé le larynx, l'onde acoustique excite le pharynx les cavités buccale et nasale qui forment un résonateur pour être diffractée par les lèvres.

1.2.1. Modèles du signal de parole

Afin de mieux représenter notre signal de parole pour des fins de codage, de transcription orthographique, phonétique ou pour la synthèse à partir de texte écrit ; il faut tout d'abord définir un modèle du signal de parole. En effet, plusieurs types de modèles de parole ont été proposés dans la littérature [4][5], on cite parmi eux :

- Le modèle source-filtre,
- Modèle harmonique plus bruit,
- Aspects linguistiques,

Etant donné que le vocodeur G.723.1, [6], utilisé dans la suite de notre travail ; est basé sur le modèle source-filtre, les notions que nous allons développer seront basées sur ce modèle.

1.2.1.1. modèle source-filtre :

Dans ce modèle, le signal de la parole est modélisé comme étant la sortie d'un filtre tout pole (filtre de synthèse) excité par une suite d'impulsions régulières (*voisement*) ou par un bruit blanc (*non-voisement*). La source correspond à la vibration des cordes vocales ou la production du bruit, et le filtre étant une modélisation du conduit vocal (tout le système vocal de la partie supra glottique jusqu'aux lèvres)[7]. Le signal $s(t)$, ainsi produit est le résultat de la convolution du signal de la source $e(t)$, par la fonction de transfert du filtre $h(t)$, comme l'illustre sur la figure (1.4).

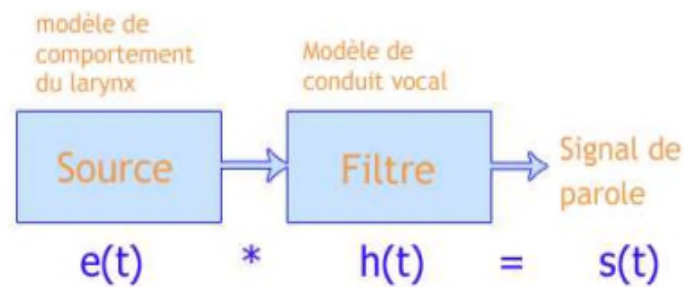


Fig.1.4 : modèle source-filtre du signal de la parole

1.3. la voix sur IP

1.3.1. définition

L'acronyme, VoIP signifie Voice over Internet Protocol ou Voix sur IP. Comme son nom l'indique, la VoIP permet de transmettre du son (en particulier de la voix) dans des paquets IP circulant sur Internet [2].

1.3.2. Architecture

La VoIP n'a pas encore de standard unique. De ce fait, chaque constructeur apporte ses normes et ses fonctionnalités à ses solutions. Les trois principaux protocoles sont H.323, SIP (Session Initiation Protocol) et MGCP/MEGACO. Il existe donc plusieurs approches pour offrir des services de téléphonie et de visiophonie sur des réseaux IP.

La figure (1.5), décrit de façon générale la topologie d'un réseau de téléphonie IP. Elle comprend toujours des terminaux, un serveur de communication et une passerelle vers les autres réseaux. Chaque norme a ensuite ses propres caractéristiques, pour garantir une plus ou moins grande qualité de service (QoS). Le réseau est aussi déportée soit sur les terminaux, soit sur les passerelles/contrôleur de commutation, appelées Gatekeeper. Nous retrouvons les éléments communs suivants :

- Le routeur : Permet d'aiguiller les données et de router les paquets entre deux réseaux. Certains routeurs permettent de simuler un Gatekeeper, grâce à l'ajout de cartes spécialisées supportant les protocoles VoIP.
- La passerelle : Permet d'interfacer le réseau commuté et le réseau IP.

- Le PABX : Est le commutateur du réseau téléphonique classique. Il permet de faire le lien entre la passerelle ou le routeur, et le RTC (Réseau Téléphonique Commuté). Toutefois, si tout le réseau devient IP, ce matériel devient obsolète.
- Les Terminaux : Sont généralement de type logiciel (software phone) ou matériel (hardphone), le softphone est installé dans le PC de l'utilisateur. L'interface audio peut être un microphone et des haut-parleurs branchés sur la carte son, même si un casque est recommandé.
- Le hardphone : est un téléphone IP qui utilise la technologie de la Voix sur IP, pour permettre des appels téléphoniques sur un réseau IP, tel que l'Internet, au lieu de l'ordinaire système PSTN (Public Switched Telephone Network). Les appels peuvent parcourir par le réseau internet comme par un réseau privé. Un terminal utilise des protocoles comme le SIP ou l'un des protocoles propriétaires tel que celui utilisée par Skype .

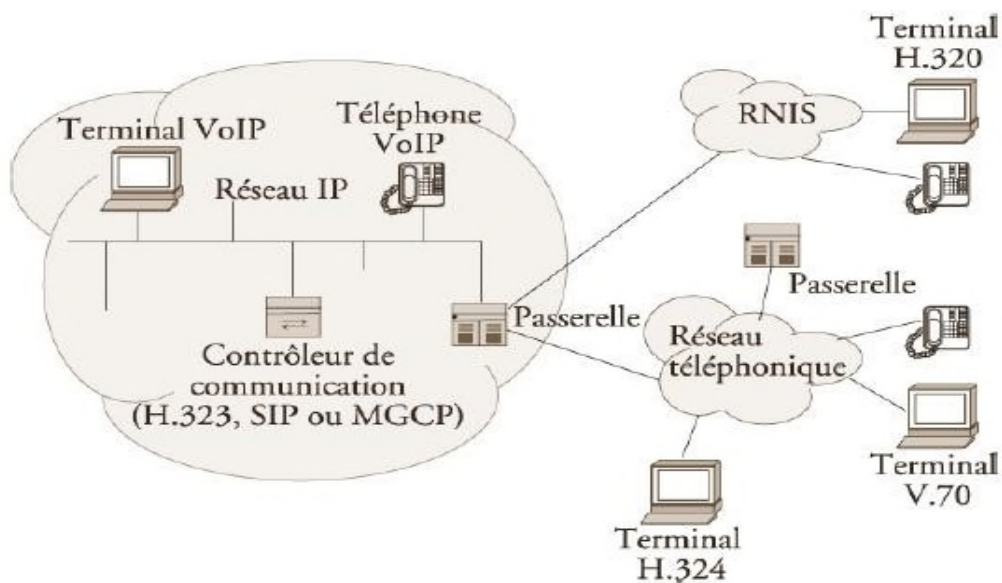


Fig.1.5 : Architecture générale de la VoIP

1.3.3. Mode de fonctionnement

La VoIP fonctionne par numérisation de la voix, puis par reconversion des paquets numériques en voix à l'arrivée. Le format numérique est plus facile à contrôler, il peut être compressé, routé et converti en un nouveau format de meilleure qualité. Le signal numérique est plus tolérant au bruit que l'analogique.

Les réseaux TCP/IP sont des supports de circulation de paquets IP, contenant un en-tête (pour contrôler la communication) et une charge utile, pour transporter les données.

Il existe plusieurs protocoles qui peuvent supporter la voix sur IP tel que : H.323, SIP et MGCP. Les deux protocoles les plus utilisés actuellement dans les solutions VoIP présents sur le marché sont le H.323 et le SIP [8].

1.4. Protocoles de la VoIP

Dans cette section, nous allons présenter les principaux protocoles qui interviennent pour l'établissement d'une communication sur la VoIP.

1.4.1. Protocole H.323 :

Le standard H.323[1], fournit depuis son approbation en 1996, un cadre pour les communications audio, vidéo et de données sur les réseaux IP. Il a été développé par l'ITU pour les réseaux qui ne garantissent pas une QoS (Quality of service), tels que FastEthernet et Token Ring. Il est présent dans plus de 30 produits. H.323 concerne le contrôle des appels, la gestion multimédia, la gestion de la bande passante pour les conférences point-à-point et multipoints. Il traite également de l'interfaçage entre le LAN (Local Area Network) et les autres réseaux.

Le protocole H.323 fait partie de la série H.32x qui traite de la vidéoconférence à travers les différents réseaux. Il inclut H.320 et H.324 liés aux réseaux ISDN (Integrated Service Data Network) et PSTN (Public Switched Telephone Network).

Plus qu'un protocole, H.323 crée une association de plusieurs protocoles différents et qui peuvent être regroupés en trois catégories : la signalisation, la négociation de codec, et le transport de l'information.

- Les messages de signalisation sont ceux envoyés pour demander la mise en relation de deux clients, qui indique que la ligne est occupée ou que le téléphone sonne, etc. Avec H.323, la signalisation s'appuie sur le protocole RAS (Registration/Admission/Status) pour l'enregistrement et l'authentification, et le protocole Q.931 pour l'initialisation et le contrôle d'appel.
- La négociation est utilisée pour se mettre d'accord sur la façon de coder les informations à échanger. Il est important que les téléphones (ou systèmes) utilisent un langage commun s'ils veulent se comprendre. Il s'agit du codec le moins gourmand en bande passante ou de celui qui offre la meilleure qualité. Il serait aussi préférable d'avoir plusieurs alternatives de langages. Le protocole utilisé pour la négociation de codec est le H.245.
- Le transport de l'information s'appuie sur le protocole RTP (Real-time Transport Protocol) qui transporte la voix, la vidéo ou les données numérisées par les codecs. Les messages RTCP (Real-time Transport Control Protocol) peuvent être utilisés pour le contrôle de la qualité, ou la renégociation des codecs si, par exemple, la bande passante diminue.

Une communication H.323 se déroule en cinq phases : l'établissement d'appel, l'échange de capacité et réservation éventuelle de la bande passante à travers le protocole RSVP (Ressource RéserveProtocol), l'établissement de la communication audio-visuelle, l'invocation éventuelle de services en phase d'appel (par exemple, transfert d'appel, changement de bande passante, etc.) et enfin la libération de l'appel [9].

1.4.1.1. Avantages et inconvénients du protocole H.323 :

La technologie H.323 possède des avantages et des inconvénients [9]. Parmi les avantages, nous citons :

- Gestion de la bande passante : H.323 permet une bonne gestion de la bande passante en posant des limites au flux audio/vidéo afin d'assurer le bon fonctionnement des applications critiques sur le LAN. Chaque terminal H.323 peut procéder à l'ajustement de la bande passante et la modification du débit en fonction du comportement du réseau en temps réel (latence, perte de paquets et gigue).
- Support Multipoint : H.323 permet de faire des conférences multipoint via une structure centralisée de type MCU (Multipoint Control Unit) ou en mode ad-hoc.

- Support Multicast : H.323 permet également de faire des transmissions en multicast.
- Interopérabilité : H.323 permet aux utilisateurs de ne pas se préoccuper de la manière dont se font les communications, les paramètres (les codecs, le débit...) sont négociés de manière transparente.
- Flexibilité : une conférence H.323 peut inclure des terminaux hétérogènes (studio de visioconférence, PC, téléphones...) qui peuvent partager selon le cas, de la voix de la vidéo et même des données.

Les inconvénients de la technologie H.323 sont :

- La complexité de mise en œuvre et les problèmes d'architecture en ce qui concerne la convergence des services de téléphonie et d'Internet, ainsi qu'un manque de modularité et de souplesse.
- Comprend de nombreuses options susceptibles d'être implémentées de façon différentes par les constructeurs et donc de poser des problèmes d'interopérabilité.

1.4.2. Protocole SIP

Le protocole SIP (Session Initiation Protocol),[10],est un protocole normalisé et standardisé par l'IETF (Internet Engineering Task Force) qui a été conçu pour établir, modifier et terminer des sessions multimédia. Il se charge de l'authentification et de la localisation des multiples participants. Il se charge également de la négociation sur les types de média utilisables par les différents participants, en encapsulant des messages SDP (Session Description Protocol). SIP ne transporte pas les données échangées durant la session comme la voix ou la vidéo. SIP étant indépendant de la transmission des données, tout type de données et de protocoles peut être utilisé pour cet échange. Cependant le protocole RTP (Real-time Transport Protocol) assure le plus souvent les sessions audio et vidéo. SIP remplace progressivement H323.

1.4.2.1. Avantages et inconvénients du protocole SIP :

Les principaux avantages du protocole SIP sont :

- Ouvert : Les protocoles et documents officiels sont détaillés et accessibles à tous en téléchargement.

- Standard : L'IETF a normalisé le protocole et son évolution continue par la création ou l'évolution d'autres protocoles qui fonctionnent avec SIP.
- Simple : SIP est simple et très similaire à http.
- Flexible : SIP est également utilisé pour tout type de sessions multimédia (voix, vidéo, mais aussi musique, réalité virtuelle, etc.).
- Téléphonie sur réseaux publics : Il existe de nombreuses passerelles (services payants) vers le réseau public de téléphonie (RTC, GSM, etc.) permettant d'émettre ou de recevoir des appels vocaux.
- Points communs avec H323 : L'utilisation du protocole RTP et quelques codecs son et vidéo sont en commun.

Par contre une mauvaise implémentation ou une implémentation incomplète du protocole SIP dans les User Agents peut perturber le fonctionnement ou générer du trafic superflu sur le réseau. Un autre inconvénient est le faible nombre d'utilisateurs ; en effet SIP est encore peu connu et peu utilisé par le grand public, n'ayant pas atteint une masse critique, il ne bénéficie pas de l'effet réseau.

1.4.3. Protocole RTP

RTP (Real time Transport Protocol), standardisé en 1996[11], est un protocole qui a été développé par l'IETF, afin de faciliter le transport temps réel de bout en bout, des flots donnés audio et vidéo, sur les réseaux IP. RTP est un protocole qui se situe au niveau de l'application et qui utilise les protocoles sous-jacents de transport TCP ou UDP. Mais l'utilisation de RTP se fait généralement au-dessus d'UDP, ce qui permet d'atteindre plus facilement le temps réel. Les applications en temps réels, comme la parole numérique ou la visioconférence constitue un véritable problème pour Internet. Une application en temps réel, exige une certaine qualité de service (QoS) que RTP ne garantit pas du fait qu'il fonctionne au niveau Applicatif.

De plus RTP est un protocole qui se trouve dans un environnement multipoint, donc on peut dire que RTP possède à sa charge, la gestion du temps réel, mais aussi l'administration de la session multipoint.

1.4.3.1. Avantages et inconvénients

Le protocole RTP permet de reconstituer la base de temps des différents flux multimédia (audio, vidéo, etc.); de détecter les pertes de paquets; et d'identifier le contenu des paquets pour leur transmission sécurisée.

Par contre, il ne permet pas de réserver des ressources dans le réseau ou d'apporter une fiabilité dans le réseau. Ainsi il ne garantit pas le délai de livraison.

1.4.4. Protocole RTCP

Le protocole RTCP (Real-time Transport Control Protocol) est fondé sur la transmission périodique de paquets de contrôle à tous les participants d'une session. C'est le protocole UDP (par exemple) qui permet le multiplexage des paquets de données RTP et des paquets de contrôle RTCP [9].

Le protocole RTP utilise le protocole RTCP, qui transporte les informations supplémentaires suivantes pour la gestion de la session.

Les récepteurs utilisent RTCP pour renvoyer vers les émetteurs un rapport sur la QoS. Ces rapports comprennent le nombre de paquets perdus, le paramètre indiquant la variance d'une distribution (plus communément appelé la gigue : c'est à dire les paquets qui arrivent régulièrement ou irrégulièrement) et le délai aller-retour. Ces informations permettent à la source de s'adapter, par exemple, de modifier le niveau de compression pour maintenir une QoS.

1.4.4.1. Avantages et inconvénients

Le protocole de RTCP est adapté pour la transmission de données temps réel. Il permet d'effectuer un contrôle permanent sur une session et ces participants. Par contre il fonctionne en stratégie bout à bout. Et il ne peut pas contrôler l'élément principal de la communication "le réseau "[12]

1.4.5. Protocole UDP

Le protocole UDP est un protocole de transport orienté « non connexion » sans confirmation. À l'opposé du protocole TCP, il permet aux applications d'échanger des datagrammes sans accusé de réception ni remise garantie [13].

L'avantage du protocole UDP est son temps d'exécution court qui permet de tenir en compte les contraintes de temps réel que la VoIP doit assurer.

1.5. Conclusion

Après avoir défini et expliqué le processus de génération du signal de la parole, nous avons présenté les mécanismes de la transmission de ce signal sur un réseau IP que nous avons défini et donné son architecture. De plus, l'énumération des différents traitements et protocoles qui entrent en vigueur ont été ajoutés, afin de comprendre le mécanisme adopté pour l'élaboration d'une communication sur VoIP. Ainsi que les paramètres qui nous renseignent sur la QoS de ce réseau exploité.

Par sa simplicité, la VoIP représente la solution la plus efficace pour effectuer des conversations, ce qui explique son évolution rapide.

Pour mieux comprendre les pertes de paquets qui surviennent lors de la réception, nous allons d'abord aborder dans le chapitre suivant le codage de la parole qui constitue un élément majeur pour la reconstitution du signal à la réception.

CHAPITRE

II

2.1. Introduction

Le codage de parole permet la réduction de débit de transmission du signal dans des canaux à largeur de bande limitée. La largeur de bande du canal de transmission doit être minimisée tout en préservant la qualité du signal vocal reconstruit. Dans le cas de la transmission de la voix sur IP, la réduction du débit limite le nombre ou la taille des paquets à envoyer sur le réseau. Dans les systèmes de téléphonie filaire classiques, la parole est numérisée à 64 kbps.

De nombreux algorithmes ont été proposés, [14] [15] [16] [17][6]....etc., pour diminuer ce débit tout en essayant de conserver une qualité subjective suffisante. Un système de codage de la parole comprend deux parties, le codeur et le décodeur. Le codeur analyse le signal pour en extraire un nombre réduit de paramètres pertinents qui sont représentés par un nombre restreint de bits pour archivage ou transmission, le décodeur utilise ces paramètres pour reconstruire un signal de parole synthétique.

L'analyse de la parole est une étape indispensable à toute application de synthèse, de codage, ou de reconnaissance. Elle repose en général sur un modèle. Celui-ci possède un ensemble de paramètres numériques, dont les plages de variation définissent l'ensemble des signaux couverts par le modèle.

2.2. Codage numérique de la parole

Pour permettre à un plus grand nombre d'utilisateurs de communiquer, il faut que les équipements qui véhiculent ces énormes flux d'informations ne soient pas trop encombrés et saturés ; d'où l'utilité de compresser l'information avant sa transmission. Ainsi des méthodes de traitement du 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. Plusieurs algorithmes de plus en plus sophistiqués pour la compression du signal de parole ont été développés[18] [19] [20]...etc.

Les codecs de la parole peuvent être répartis en trois classes : les codecs à forme d'onde, les vocodeurs, et les codecs hybrides. On distingue en général 3 plages de débits :

- Les hauts débits, supérieurs à 16 kbps, correspondant à des algorithmes de codage de la forme d'onde non spécifiques à la parole [21],

- 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[21].
- 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[21].

La plupart des algorithmes de codage mettent à profit un modèle linéaire simple de production de la parole. Le signal de parole est souvent modélisé (modèle « source-filtre ») comme la sortie d'un filtre tout pôle (appelé filtre de synthèse) dont la fonction de transfert représente l'enveloppe spectrale, excité par une entrée dont les caractéristiques (en particulier la fréquence fondamentale) déterminent la structure fine du spectre. Plusieurs techniques de modélisation et de codage de parole ont été proposées, comme le modèle de l'analyse par synthèse (AbS) qui se base sur la modélisation d'enveloppe spectrale par le modèle autorégressif (AR) et la prédiction linéaire (LP), que nous allons voir dans ce chapitre.

2.2.1 Les codecs à forme d'onde

Les codeurs temporels utilisent des techniques de codage qui cherchent avant tout à préserver la forme temporelle du signal de parole, ce qui les rend robustes aux différents types d'entrée et ne sont donc pas spécifiques au signal de parole. Ce type de codeur offre des débits supérieurs à 16 kbit/s. La qualité du signal synthétisé obtenue est excellente pour un débit relativement élevé. Le plus ancien procédé de transmission numérique de la parole est la modulation par impulsion codée MIC ou encore PCM (Pulse Code Modulation). L'Union Internationale des Télécommunications (UIT) a normalisé le codeur G.711 en 1972[22], un codeur logarithmique de parole de type PCM pour la transmission téléphonique avec un débit de 64 kbit/s. Ce type de codage échantillonne le signal de parole à une fréquence de 8 kHz et opère une quantification sur 8 bits du signal de parole dans la bande de fréquences 300 Hz à 3400 Hz.

2.2.2 Les codecs paramétriques (vocodeurs)

Ces codeurs ont été conçus pour des applications à très bas débit (inférieur à 4 kbits/s) et sont principalement prévus pour maintenir l'intelligibilité du signal vocal. Pour atteindre ces taux de compression, les systèmes de codage paramétriques se basent sur la connaissance

du processus de production de la parole. La technique consiste à extraire du signal de parole les paramètres les plus pertinents permettant au décodeur de le synthétiser. Les performances des codeurs paramétriques, appelés aussi vocodeurs, dépendent de la précision des modèles de production de parole. La plupart des codeurs paramétriques sont basés sur le codage prédictif linéaire (LPC), connus sous le nom de vocodeurs prédictifs. Ces codeurs fournissent, à faibles débits, des performances supérieures à celles des codeurs temporels (codeurs de forme d'onde).

2.2.3 Les codecs hybrides

La fréquence d'échantillonnage étant fixe, la réduction de débit des codeurs d'onde fait chuter rapidement la qualité d'écoute pour des débits inférieurs à 16 kbit/s. Les codeurs hybrides utilisent les deux méthodes temporelle et paramétrique de façon complémentaire, ce qui permet un codage de parole de bonne qualité à des débits relativement faibles [16,17] [23,24]. Ces codeurs sont basés sur des techniques de codage temporel auxquelles des modèles de production de parole sont associés pour améliorer leur efficacité. Cependant, ce type de codage nécessite des coûts de calculs plus importants. Tous les codeurs hybrides s'appuient sur une analyse LPC pour obtenir les modèles de synthèse de parole. Les deux techniques paramétrique et temporelle modélisent respectivement le conduit vocal et le signal résiduel ou erreur de prédiction (l'excitation).

2.3. Modélisation acoustique de la parole

Pour réduire le débit tout en conservant une qualité suffisante ou pour améliorer la qualité, pour un débit imposé, il faut chercher les paramètres pertinents qui constituent le signal de parole.

Pour cela un modèle simplifié du système phonatoire ne retenant que les paramètres les plus significatifs du signal est recherché.

Une modélisation précise du conduit vocal et la source d'excitation est nécessaire pour produire un signal intelligible et naturel.

2.3.1. Modélisation de l'enveloppe spectrale :

La production du signal de parole est un processus aléatoire, l'enveloppe spectrale correspond au spectre d'amplitude du filtre modélisant le conduit vocal et le spectre de la source glottique. Ce filtre contient les informations de phase et d'amplitude des composants spectrales du signal de parole. L'enveloppe spectrale sert alors à coder les amplitudes.

L'enveloppe spectrale du signal de parole est le facteur le plus déterminant en termes d'identification du locuteur par l'audition.

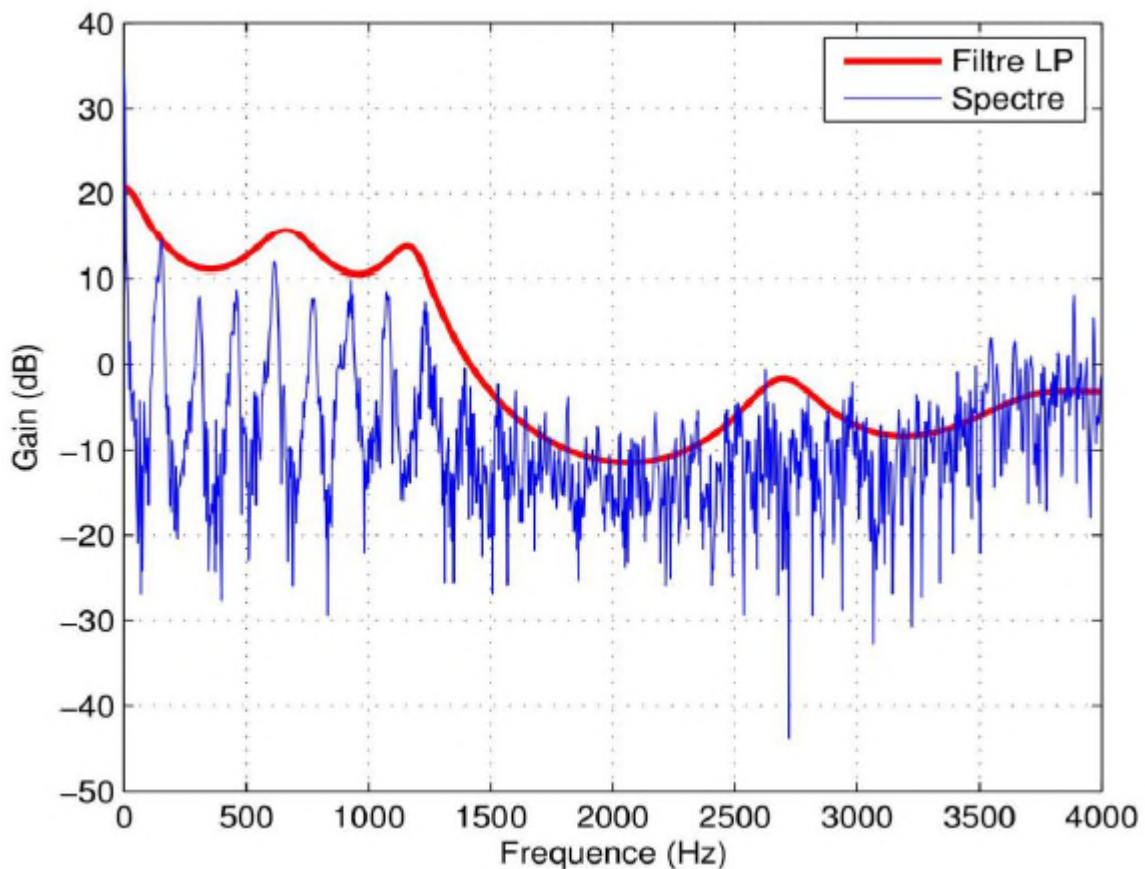


Fig.2.1. Enveloppe spectrale d'un signal de parole.

2.3.2. Fenêtrage d'un signal de parole

Le fenêtrage est utilisé dès qu'on s'intéresse à un signal de longueur volontairement limité. En effet un signal réel ne peut qu'avoir une durée limitée dans le temps, que plus est, un calcul ne peut se faire que sur nombre de points finis.

Le but du fenêtrage du signal de parole est de le découper en trames quasi-stationnaires de durée de 10ms à 30ms, étant donné qu'il n'est pas stationnaire.

2.4 Analyse par synthèse (ABS)

Le principe de l'analyse par synthèse se base sur la génération des signaux de sortie qui correspondent aux différents choix de paramètres des signaux d'excitation, chaque signal d'excitation candidat est passé à travers le filtre de synthèse (LP) et comparé à la parole d'entrée. La combinaison de paramètres qui génère le meilleur signal sera choisie.

2.4.1. Analyse LP

Pour l'analyse LP (prédiction linéaire), On peut voir le diagramme du modèle de base pour la production de paroles, appropriée à l'analyse de prédiction linéaire dans la figure (2.2) ci-dessous :

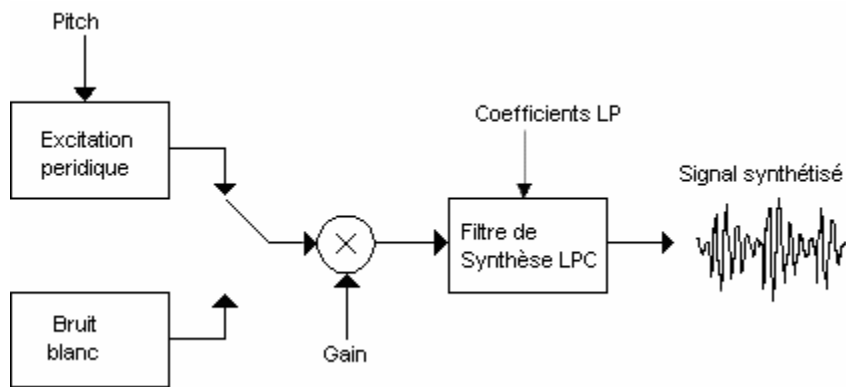


Fig.2.2diagramme d'un modèle simplifié pour la production de la parole.

Dans ce modèle, le système vocal et l'excitation glottale sont représentés par un filtre numérique variant dans le temps dont la fonction du système d'état stable est de la forme :

$$H(z) = \frac{G}{1 - \sum_{k=1}^P \alpha_k z^{-k}} \quad (2.1)$$

Le système peut être excité soit par un train d'impulsion pour la parole voisée, soit par une séquence aléatoire pour la parole non voisée. On peut donner les échantillons de parole $s(n)$ utilisant l'équation de différence simple :

$$s(n) = \sum_{k=1}^P \alpha_k s(n - k) + Gu(n) \quad (2.2)$$

Le prédicteur linéaire avec des coefficients de prédiction α_k et d'ordre p est défini comme un système dont la sortie est :

$$\hat{s}(n) = \sum_{k=1}^P \alpha_k s(n - k) \quad (2.3)$$

La fonction du système de ce prédicteur linéaire est :

$$P(z) = \sum_{k=1}^P \alpha_k z^{-k} \quad (2.4)$$

L'erreur de prédiction est définie comme :

$$e(n) = s(n) - \hat{s}(n) \quad (2.5)$$

$$e(n) = s(n) - \sum_{k=1}^P \alpha_k s(n-k) \quad (2.6)$$

Ainsi le filtre d'erreur de prédiction est la sortie du système dont la fonction de transfert est :

$$A(z) = 1 - \sum_{k=1}^P \alpha_k z^{-k} \quad (2.7)$$

La comparaison des équations (2.1) et (2.5) suggère que si $\alpha_k = a_k$, donc $e(n) = G.u(n)$ et dans une telle condition, le filtre d'erreur de prédiction $A(z)$ sera un filtre inverse pour le système $H(z)$.

$$H(z) = \frac{G}{A(z)} \quad (2.8)$$

Le problème de base de prédiction linéaire est de déterminer un ensemble de coefficients de prédiction $\{a_k\}$ directement du signal de paroles et de cette façon on obtient une bonne estimation des propriétés spectrales de signal de parole à l'aide de l'équation (2.8).

Donc le but est de minimiser l'erreur de prédiction $e(n)$, entre le signal de parole $s(n)$ et son estimation $\hat{s}(n)$, en adaptant les coefficients $\{a_k\}$ $k=0..p$ pour chaque trame de N échantillons. Pour cela le critère des moindres carrés est utilisé :

$$\min(\sum_{n=0}^{N-1} e^2(n)) \quad (2.9)$$

2.4.1.1 Détermination des coefficients LP par la méthode d'autocorrélation :

A cause de la contrainte de l'utilisation des trames de N échantillons, il est important dans la méthode d'autocorrélation d'appliquer une fenêtre. Normalement, une fenêtre de Hamming est employée. L'application de la fenêtre élimine les problèmes causés par des changements rapides du signal aux bords de la fenêtre, qui assure une transition lisse d'un bloc à un autre des paramètres estimés dans l'analyse.

Les paramètres optimaux $\{a_k\}$, $k=0\dots p$, avec $a_0 = 1$, sont obtenus en annulant la dérivée de l'équation

$$\min(\sum_{n=0}^{N-1} e^2(n)) \quad (2.10)$$

Par rapport aux coefficients a_j :

$$\frac{\delta(\sum_{n=0}^{N-1} e^2(n))}{\delta a_j} = 2 \sum_{n=0}^{N-1} e(n) \frac{\delta e(n)}{\delta a_j} \quad (2.11)$$

$$= 2 \sum_{n=0}^{N-1} e(n) \left[\frac{\delta s(n)}{\delta a_j} - \frac{\delta \hat{s}(n)}{\delta a_j} \right] \quad (2.12)$$

$$= 2 \sum_{n=0}^{N-1} e(n) \frac{\delta \sum_{k=1}^p s(n-k)}{\delta a_j} \quad (2.13)$$

$$= 2 \sum_{n=0}^{N-1} e(n) s(n-j) \quad (2.14)$$

$$= 2 \sum_{n=0}^{N-1} [s(n) - \hat{s}(n)] s(n-j) \quad (2.15)$$

$$= 2 \sum_{n=0}^{N-1} s(n) s(n-j) - 2 \sum_{n=0}^{N-1} s(n) \hat{s}(n-j) \quad (2.16)$$

$$= 2 \sum_{n=0}^{N-1} s(n) s(n-j) - 2 \sum_{n=0}^{N-1} \sum_{k=1}^p a_k s(n-j) s(n-j) \quad (2.17)$$

En posant les coefficients d'autocorrélation $\{r_j\}$ $j=0\dots p$ obtenus par:

$$r_j = \sum_{k=1}^p s(n) s(n-j) \quad \text{avec } j=0,\dots,p \quad (2.18)$$

L'annulation de l'équation (2.11) permet d'avoir les équations de Yule-Walker :

$$r_j - \sum_{k=1}^p a_k r(|j-k|) = 0 \quad \text{avec } j = 0, \dots, p \quad (2.19)$$

Ces équations peuvent être mises sous forme matricielle :

$$\begin{bmatrix} r_0 & r_1 & \dots & r_{p-1} \\ \dots & \dots & \dots & \dots \\ r_{p-1} & r_{p-2} & \dots & r_0 \end{bmatrix} \begin{bmatrix} a_1 \\ \dots \\ a_p \end{bmatrix} = \begin{bmatrix} r_1 \\ \dots \\ r_p \end{bmatrix} \quad (2.20)$$

Or :

$$R a = r$$

Cette matrice symétrique, dont les éléments d'une même diagonale sont égaux, est dite de Toeplitz.

Le vecteur des coefficients LP $\{a_k\}$ $k = 0\dots p$ est obtenu par l'inversion de la matrice de corrélation :

$$\begin{bmatrix} a_1 \\ \dots \\ a_p \end{bmatrix} = \begin{bmatrix} r_0 & \dots & r_{p-1} \\ \dots & \dots & \dots \\ r_{p-1} & \dots & r_0 \end{bmatrix}^{-1} \begin{bmatrix} r_1 \\ \dots \\ r_p \end{bmatrix} \quad (2.21)$$

Ce système d'équation linéaire peut être résolu avec de multiples algorithmes, mais le plus efficace c'est celui de Levinson-Durbin qui est un algorithme récursif qui exploite la forme Toeplitz et la symétrie du système d'équation.

2.4.2 Quantification des paires de raies spectrales (LSF)

Afin de transmettre les coefficients de prédiction linéaire LPC au décodeur, ces coefficients doivent être quantifiés, mais leur vulnérabilité au bruit de quantification, il faut d'abord les transformer.

Les paramètres LSF sont obtenus en calculant les racines des polynômes $D(z)$ et $Q(z)$ de l'ordre de $(p+1)$ par :

$$D(z) = A(z) - Z^{-(p+1)} A(Z^{-1}) \quad (2.22)$$

$$Q(z) = A(z) + Z^{-(p+1)} A(Z^{-1}) \quad (2.23)$$

Où $A(z)$ est le filtre de synthèse, et si l'ordre de p est pair, les deux polynômes peuvent être exprimés par :

$$D(z) = (1 - z^{-1}) \prod_{i=2,4,\dots,p} (1 - 2 \cos(2\pi f_i) z^{-1} + z^{-2}). \quad (2.24)$$

$$Q(z) = (1 + z^{-1}) \prod_{i=1,3,\dots,p-1} (1 - 2 \cos(2\pi f_i) z^{-1} + z^{-2}). \quad (2.25)$$

Et on sait que les paramètres LSFs ont la particularité d'ordre ascendant :

$$f_0 < f_1 < f_2 < \dots < f_{p-1} < f_p < f_{p+1}.$$

La stabilité du filtre de synthèse est mieux garantie si l'ordre des paramètres LSFs est respecté. De plus une autre propriété s'ajoute, telle que $f_0 = 0$ et $f_{p+1} = 0.5$, imposée par les racines fixes $z=1$ et $z=-1$, des polynomes $D(z)$ et $Q(z)$ respectivement. Les paramètres LSF sont très bien corrélés entre deux trames successives ce qui permet une réduction dans l'allocation de bits en utilisant une quantification prédictive.

2.4.3 Filtre de pondération perceptive des formants

Pour atténuer les pics (formants) du spectre LP et d'amplifier les régions entre les pics (vallées) ; un filtre qui contient des pôles et des zéros avec des coefficients qui changent pour chaque sous trame formé par les coefficients LP non-quantifiés est appliqué au spectre LP, appelé filtre de pondération perceptive des formants, son rôle est de masquer le bruit dans les zones de fréquences vallées ; et s'écrit comme suit :

$$W_{PP}(z) = \frac{A(\gamma_1 z)}{A(\gamma_2 z)} \quad (2.27)$$

$$\gamma_1 = 0.9, \quad \gamma_2 = 0.5 .$$

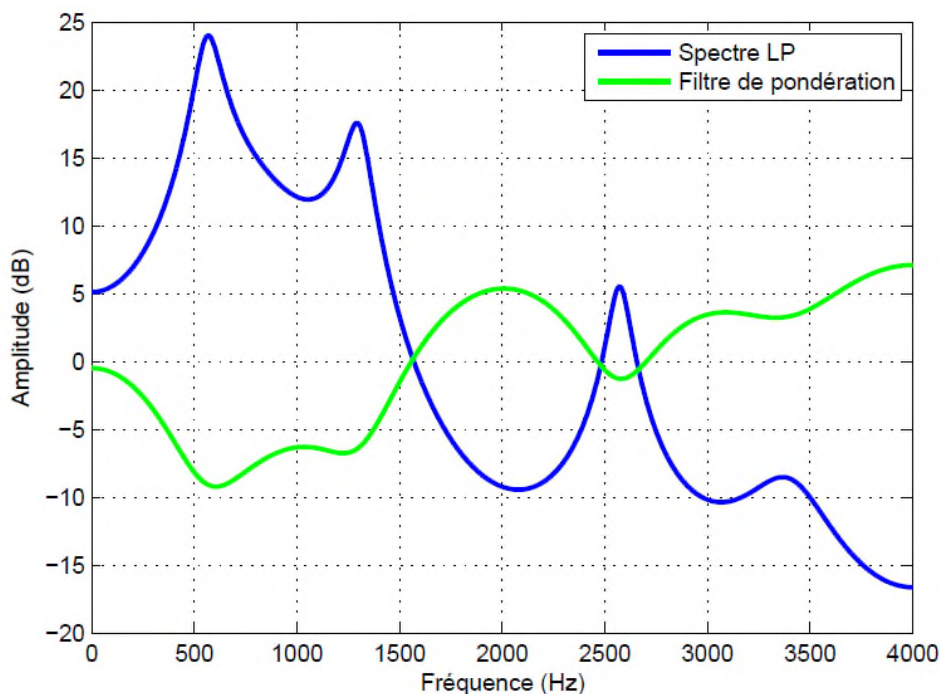


Fig.2.3. Exemple d'un spectre de filtre LP, et du filtre de pondération des formants

2.5. Modélisation de la corrélation à long terme

Le signal de parole présente des corrélations à long terme qui sont causées par les vibrations des cordes vocales. Pour les signaux voisés, les corrélations avec des distances égales à la période fondamentale, caractérisent le voisement, d'où les redondances fréquemment trouvées dans le signal de parole, de plus ces redondances seront modélisées puis utilisées afin de mieux coder ou compresser les segments du signal de parole.

2.5.1. Estimation de la tonie

Pour chaque trame On calcule deux estimations de tonie : l'une pour les deux premières sous-frames et une pour les deux dernières. L'estimation de la période fondamentale en boucle ouverte, LOL , est calculée à partir du signal vocal pondéré perceptivement, $f[n]$. Pour déterminer la période fondamentale, on fait appel à une méthode de maximisation du critère d'inter-corrélation, $COL(j)$, exprimé comme suit :

$$C_{OL}(j) = \frac{\left(\sum_{n=0}^{t_{max}} f[n] \cdot f[n-j] \right)^2}{\sum_{n=0}^{t_{max}} f[n-j] \cdot f[n-j]} \quad t_{min} \leq j \leq t_{max} \quad (2.28)$$

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

2.5.2 Mise en forme du bruit harmonique

Pour améliorer la qualité des signaux vocaux codés, on construit un filtre de mise en forme du bruit harmonique exprimé comme suit ;

$$P_i = 1 - \beta z^{-l} \quad (2.29)$$

2.5.3. Dictionnaire adaptatif conventionnel (ACB)

Le nombre de périodes de pitch à examiner, en boucle fermée, est réduit en effectuant une procédure à deux étapes :

- Une analyse en boucle ouverte (effectuée sur chaque trame de 10 ms) estime, dans un premier temps, la période de pitch optimale T_{op} en effectuant une recherche sur des

valeurs entières favorisant les petites valeurs du pitch (pour éviter le choix d'un multiple du pitch).

- La recherche au sein du dictionnaire adaptatif (la recherche du pitch en boucle fermée) est effectuée pour chaque sous-trame de 5 ms. La complexité de la recherche en boucle fermée est réduite grâce à la réduction de l'intervalle des valeurs de pitch possible autour de la valeur optimale T_{op} obtenue en boucle ouverte. Pour la première sous-trame, le retard T_1 est trouvé en recherchant autour du pitch en boucle ouverte T_{op} . Pour la seconde sous-trame, la recherche de pitch en boucle fermée est faite sur 9 délais entiers autour de la valeur (T_1). Où (T_2) est la partie entière du pitch T_1 de la première sous-trame. Les délais entiers de T_1 et T_2 , sont affinés en faisant intervenir des délais fractionnaires (autour de la valeur entière déjà trouvée).

2.5.4 Dictionnaire fixe (FCB)

Le dictionnaire fixe est basé sur un dictionnaire à structure algébrique où chaque vecteur de dimension $L = 40$ échantillons contient quatre impulsions signées, chacune d'elles pouvant occuper 8 à 16 positions possibles. La structure du dictionnaire fixe est donnée au tableau (2.2). Le vecteur issu du dictionnaire est construit en prenant un vecteur nul de dimension $L = 40$, et en plaçant les quatre impulsions unité aux positions $\{m_k\}_{k=0,\dots,3}$, multipliés avec leur signe correspondant

2.6 Codeur G.723.1[6],

Ce codeur est conçu de façon à fonctionner avec un signal numérique que l'on obtient en filtrant d'abord la bande téléphonique (Recommandation G.712) de l'entrée analogique, puis en l'échantillonnant à 8000 Hz puis en le convertissant linéairement en mots MIC de 16 éléments binaires pour l'entrée du codeur. Il convient d'effectuer la conversion inverse vers le mode analogique par des moyens similaires à la sortie du décodeur. D'autres caractéristiques d'entrée/sortie, comme celles qui sont spécifiées par la Recommandation G.711[22], pour les données MIC à 64 kbit/s, seront converties linéairement en mots MIC de 16 bits avant codage, ou seront reconverties dans leur format approprié après décodage linéaire des mots MIC de 16 bits. Le flux binaire allant du codeur au décodeur est défini dans le cadre de la présente Recommandation.

Le codeur est fondé sur les principes du codage prédictif linéaire (LPC) par analyse et synthèse, en vue de minimiser un signal d'erreur pondéré par une courbe de perception. Le

codeur travaille sur des blocs (trames) contenant chacun 240 échantillons, ce qui correspond à une longueur de trame de 30 ms à 8 kHz. Chaque bloc passe d'abord par un filtre passe-haut afin d'éliminer la composante aperiodique. Il est ensuite subdivisé en quatre sous-trames de 60 échantillons chacune. Au moyen du signal d'entrée non traité, on calcule pour chaque sous-trame un filtre de codage prédictif linéaire (LPC) (*linear prediction coder*) du 10^{ème} ordre. Le filtre LPC pour la dernière sous-trame est calculé au moyen d'un quantificateur des vecteurs prédictifs dédoublés (PSVQ) (*predictive split vector quantizer*). Les coefficients de codage LPC non quantifiés servent à construire le filtre de pondération perceptive à court terme qui est utilisé pour filtrer la trame complète afin de donner le signal vocal pondéré par la perception.

Le signal vocal ainsi pondéré sert à calculer, toutes les deux sous-trames (120 échantillons), la période fondamentale en boucle ouverte, L_{OL} . Cette estimation de la hauteur tonale est effectuée sur des blocs de 120 échantillons. La période fondamentale (inverse de la fréquence fondamentale) est recherchée dans l'étendue de 18 à 142 échantillons.

A partir de ce point, le signal vocal est traité à raison de 60 échantillons par sous-trame.

Au moyen de la période fondamentale estimée, on construit un filtre de mise en forme du bruit harmonique. La combinaison du filtre de synthèse à codage LPC, du filtre de pondération perceptive des formants, et du filtre de mise en forme du bruit harmonique sert à créer une réponse impulsionnelle qui est utilisée pour d'autres calculs.

La réponse impulsionnelle et l'estimation de la période fondamentale, L_{OL} , permettent de calculer un prédicteur de tonie en boucle ouverte. On fait appel à un prédicteur de tonie du cinquième ordre. La période fondamentale est calculée sous la forme d'une petite valeur différentielle de part et d'autre de l'estimation L_{OL} . La contribution du prédicteur de tonie est ensuite soustraite du vecteur cible initial. La période fondamentale et la valeur différentielle sont toutes deux transmises au décodeur.

Finalement, la composante aperiodique de l'excitation est approchée. Pour le débit supérieur, on fait appel à l'excitation par quantification d'impulsions multiples selon le critère du maximum de vraisemblance (MP-MLQ). Pour le débit inférieur, on fait appel à l'excitation par séquences codées à structure algébrique (ACELP).

2.6. Conclusion

Dans ce chapitre, nous avons essayé d'expliquer d'une façon explicite le fonctionnement d'un codeur de parole, basé sur l'analyse par synthèse qui à la fois permet le codage de la parole à des débits bas et au même temps réalise une compression sans l'altération du signal de la parole. Nous avons aussi présenté comment déterminer les coefficients de prédiction linéaire LP par la méthode d'autocorrélation et leur quantification afin de les transmettre au décodeur en les transformant en paramètres LSF. On a aussi vu le rôle du filtre de pondération perceptuelle de formant dans l'atténuation de bruit et comment estimer le délai tonal. Nous avons présenté l'optimisation des deux registres ACB et FCB qui permettent la création des signaux candidats pour avoir un signal synthétisé qui se rapproche le maximum possible au signal original. Enfin On a donné une description générale du codec G.723.1-basé CELP. Malgré de tous ces paramètres optimisés, dans la transmission des données sur la VOIP il y a toujours des pertes de paquets qui provoquent des erreurs sur la reconstruction du signal de parole au niveau du décodeur.

Après avoir compris le fonctionnement d'un codeur de parole, notamment le G.723.1, nous allons dans le prochain chapitre étaler les différentes causes de ces pertes de paquets, afin d'y remédier et garantir une bonne qualité de service.

CHAPITRE

III

3.1. Introduction

En téléphonie sur IP, la voix est transmise sous forme de paquet IP sur le réseau. Un paquet correspond au codage d'une trame du signal (de 10ms à 30 ms de parole). En raison de l'architecture Best-Effort des réseaux IP, les paquets de voix peuvent être détruits lors d'une congestion, ou arriver suffisamment en retard pour être considérés par le système comme perdus. Le problème est donc de masquer les pertes. Dans ce chapitre nous allons voir les différentes causes de pertes de paquets et les différentes méthodes de masquage (PLC) qui ont été élaborées afin d'arriver à une qualité de service(QoS) la meilleure possible.

3.2. Perte de paquets

Les utilisateurs qui naviguent sur le internet ou accèdent à d'autres réseaux en ligne sont capables de faire l'envoi et la réception d'unités de données (paquets). Tout au long de ce processus ; un ou plusieurs paquets peuvent être perdus lors du transfert et ne peuvent pas atteindre l'adresse de destination. C'est ce qu'on appelle la perte de paquets. Elle peut se manifester sous forme de lenteur des services, de perturbation de la connexion au réseau et de perte totale de la connectivité du réseau. La perte de paquet peut affecter n'importe quelle application, mais elle est plus susceptible de perturber celles qui dépendent de transfert de données en temps réel comme la téléphonie ainsi que d'autres plateformes de système numérique.

3.2.1 Les différentes causes de perte de paquets

3.2.1.1 Réseaux congestionnés

Les réseaux qui atteignent leur capacité maximale sont considérés comme un réseau congestionné et sont plus susceptibles de subir des pertes de paquets en raison de l'augmentation du trafic .Comme le processus de transfert de paquet suit des étapes spécifiques les échecs de connexion peuvent entraîner la perte de certains paquets. La congestion se traduit généralement par des dépassements du tampon du routeur, alors que les nœuds envoient plus de paquets que le réseau ne peut en gérer. Deux types de solution ; La prévention (ou anticipation), ou la régulation :

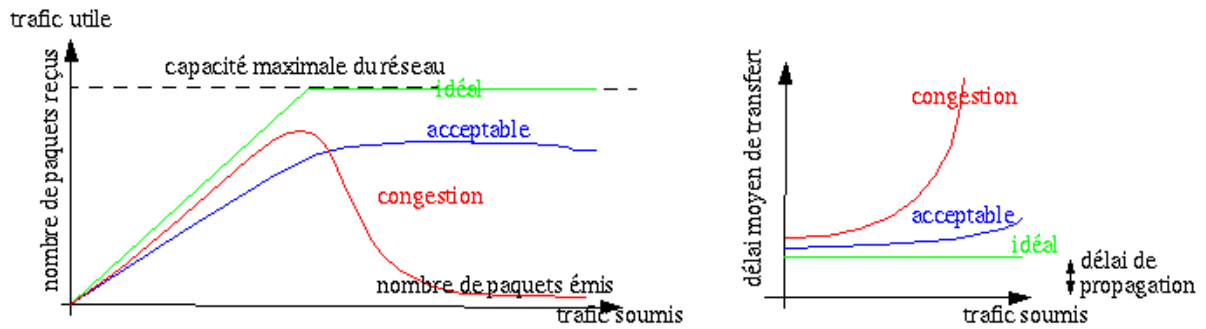


Fig.3.1. L'effet de la congestion du réseau

-**Pré-allocation des ressources** ; réservation des ressources, l'espace de stockage, portion de bande passante, utilisés uniquement dans le cadre de la connexion (monopolisation inutile des ressources)

- **Le contrôle de congestion isarithmique** ; conservation d'un nombre constant de paquets en circulation dans le réseau.

- **Rétro-contrôle** ; contrôle du débit des émetteurs, chaque nœud surveille le taux d'utilisation de ses liaisons. Envoi de paquet d'engorgement vers les nœuds émetteurs de paquet devant être acheminés sur les liaisons chargées

-**Priorité** ; nécessité de définir des critères pour choisir les paquets à détruire ou à acheminer en priorité.

3.2.1.2 La gigue

La gigue (Jitter) est une variation, ou un écart des impulsions de commutation dans un signal numérique haute fréquence, cette déviation peut être exprimée en termes d'amplitude, de temps de phase ou de largeur de l'impulsion.

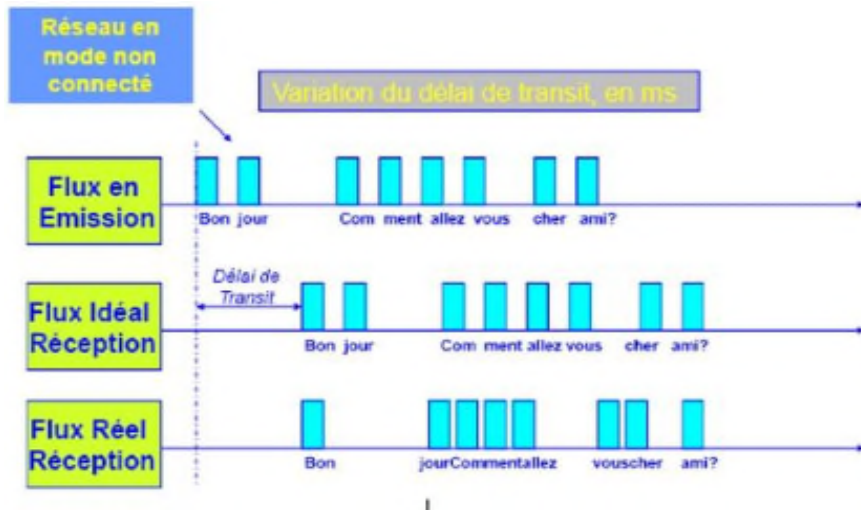


Fig.3.2. la gigue dans le réseau.

Dans les réseaux IP, la gigue représente la variation de la latence d'un flux de paquets entre deux systèmes lorsque certains paquets prennent davantage de temps pour passer d'un système à l'autre. Ce phénomène peut être le résultat d'un encombrement du réseau, d'un glissement de phase, ou d'un changement d'acheminement ; la figure (3.2), illustre l'effet de la gigue.

Un tampon (buffer) peut être atténué les effets de la gigue, soit sur le réseau d'un routeur ou un commutateur. L'application reçoit ainsi essentiellement les paquets du tampon plutôt que directement du réseau. Ils sont envoyés par le tampon à la vitesse régulière, de façon à atténuer les variations temporelles des paquets qui arrivent dans le tampon, voir figure (3.3).

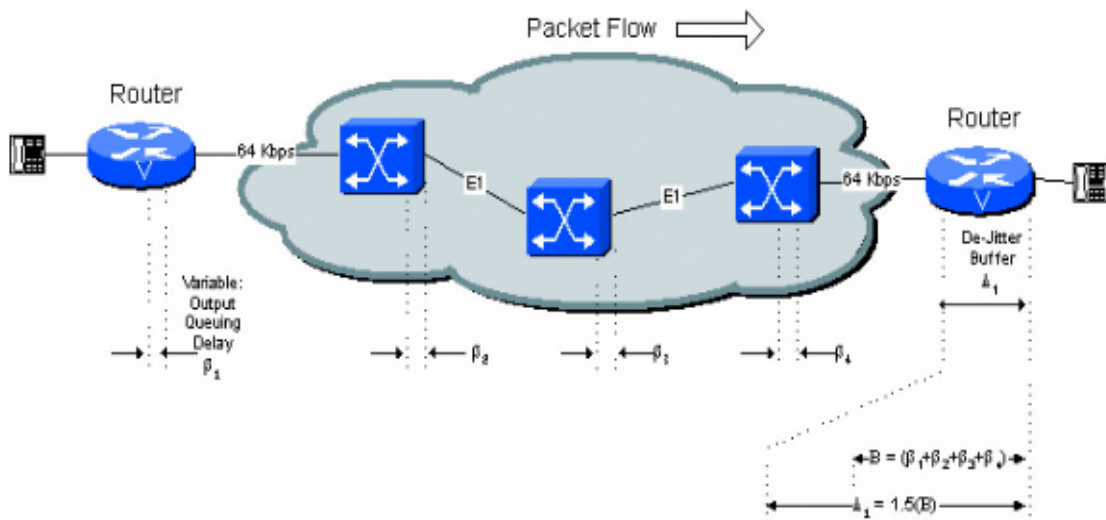


Fig.3.3. Delai variable et le tampon d'élimination de gigue.

Il existe d'autres techniques permettant de réduire les effets de gigue lorsque plusieurs voies de trafic sont disponibles. Il est ainsi possible d'acheminer le trafic de manière sélective vers la voie la plus stable, ou de choisir systématiquement la voie la plus proche de la vitesse de livraison du paquet cible.

3.2.1.3 Le délai de transit (latence)

Le délai de transit est le temps mis par un paquet pour aller d'un hôte émetteur à un hôte récepteur, tel qu'il est illustré sur la figure (3.4).

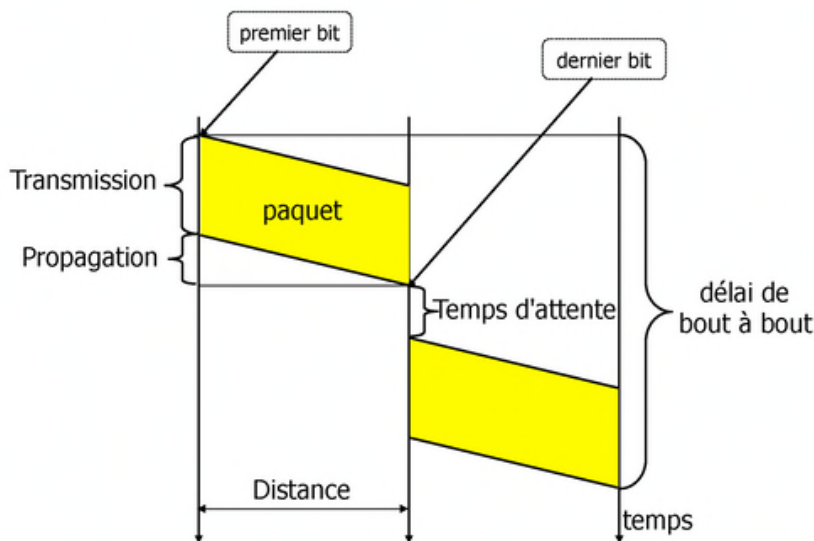


Fig. 3.4 Source de délais.

Il est la somme de plusieurs petits, ou moins petits délais ; tel que :

- délai dû au codage,
- délai de mise en trame sur le support (négligeable),
- délai de traversée du réseau (entrée, sortie, transit, fils d'attente),
- temps de service (négligeable),
- la bufférisation destinée à corriger la gigue,
- délai de décodage.

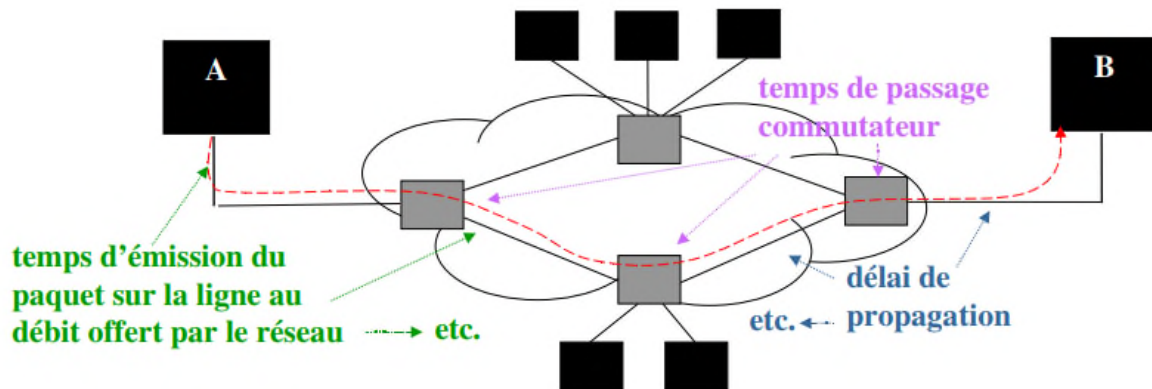


Fig.3.5. Délai de transit.

3.2.1.4 Perte due au protocole

Dans le cas de la voix IP, le protocole de transport utilisé est généralement UDP. L'UDP n'offre aucune garantie quant à la sécurité et à l'authenticité des données. Le fait de renoncer à l'authentification mutuelle de l'expéditeur et du destinataire permet au protocole UDP d'assurer une vitesse de transmission exceptionnelle. Toutefois, le protocole UDP ne peut garantir l'intégrité et la sécurité des paquets de données. L'ordre dans lequel les paquets ont été envoyés n'est pas non plus garantie.

3.3. Méthodes de masquage de pertes de paquets

La qualité audio d'un appel utilisant la technologie Voix sur IP est assujettie à plusieurs facteurs, entre autres les pertes de paquets. En effet, ce dernier est un facteur déterminant de la qualité d'un appel voix sur IP ; un taux de pertes élevé peut causer une dégradation énorme de la qualité audio qui peut même rendre la communication impossible.

Des techniques de masquage de pertes interviennent pour pallier au problème de perte de paquets, certaines sont utilisées pour protéger le signal avant la transmission, d'autres permettent de reconstruire au mieux le signal au niveau du récepteur.

3.3.1. Méthodes basées à la source

Ces méthodes se basent sur l'ajout de l'information (redondance), et de codes correcteurs d'erreurs, comme par exemple d'envoyer deux flux de données, le premier à fort débit et haute qualité de description ; le deuxième avec une description sommaire du signal mais à un débit beaucoup plus réduit.

3.3.1.1. Priorisation et hiérarchisation

Les méthodes de priorisation ressemblent aux méthodes hiérarchiques et prennent souvent en compte la manière dont est encodé le signal. Des priorités sur les trames sont posées en fonction de l'énergie du signal, de la différence avec le paquet précédent ou encore des indicateurs de transition dans la trame. De cette manière une priorité est accordée selon qu'un paquet peut facilement être prédit ou non. De la même manière, les paramètres de codage peuvent être hiérarchisés, en accordant une priorité supérieure aux bits les plus significatifs lors de la quantification des paramètres les plus sensibles (pitch, l'énergie court terme, les coefficients LPC).

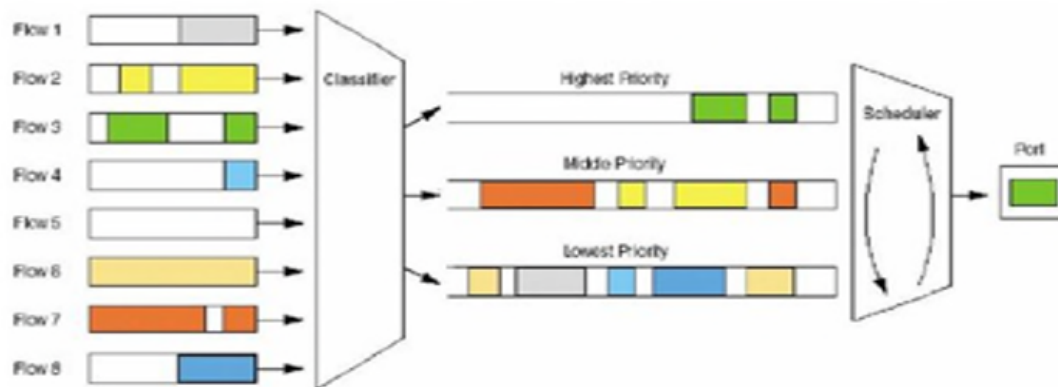


Fig. 3.6. Priorisation et hiérarchisation de paquets

Comme illustré dans la figure (3.6) :

- Les paquets de priorité élevée sont servis d'abord
- Risque de famine pour les paquets de priorité faible

3.3.1.2. Entrelacement de trames

Le principe est de mettre les paquets d'une manière désordonnée par rapport à leur succession chronologique, ce qui minimise l'effet de perte de paquets si la perte est isolée ; en revanche ça engendre un retard additionnel du fait que le récepteur est appelé à remettre les paquets dans leur ordre chronologique.

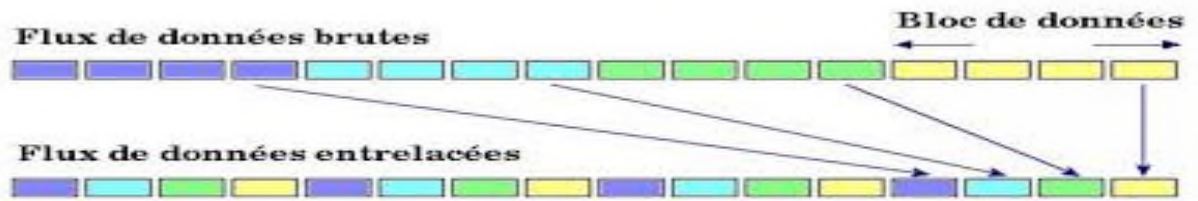


Fig. 3.7.Méthode d'entrelacement des paquets de données.

3.3.2. Méthodes basées au récepteur

Sont les méthodes qui nécessitent l'intervention du récepteur, parmi ces méthodes on trouve les techniques de :

3.3.2.1. Substitution

Les techniques de substitution sont très efficaces car elles ne nécessitent pas de calculs lourds et elles sont généralement rapides. Dans les cas où les pertes sont peu fréquentes et avec des paquets de petites tailles.

En général, les paquets perdus sont recréés en remplissant les trous par du silence, du bruit blanc, en répétant le précédent paquet reçu ou en appliquant une méthode plus évoluée de pattern matching. Le modèle de pattern matching le plus usuel utilise des portions d'échantillons précédentes ou suivantes les plus corrélés à la portion perdue. D'autres techniques estiment la période du pitch avec le pitch du paquet précédent, ou effectuent une substitution en se basant sur les paramètres de prédiction linéaire des trames précédentes.

Ces méthodes s'avèrent efficaces si les pertes ne sont pas très fréquentes ou si elles n'interviennent pas sur des trames successives. Les réseaux mobiles étant très perturbés il est préférables de ne pas envisager l'utilisation des techniques de substitution pour pallier aux pertes de paquet.

3.3.2.2. Masquage de pertes

Les processus de masquage reconstruisent les paquets manquants de diverses manières. Les plus simples mais peu efficaces pour des applications requérant de la haute qualité sonore, se basent sur les méthodes connues de substitution ou d'insertions. Telles qu'abordées précédemment, elles consistent à remplacer le paquet perdu par un silence ou du bruit, à trouver des correspondances ou à effectuer des interpolations avec les informations contenues

dans les paquets correctement reçus (substitution par le paquet précédant dans le pire des cas, réplication du pitch, modification de l'échelle temporelle, etc.).

Une autre catégorie de techniques regroupe les méthodes de régénération. Elles utilisent l'information des paquets précédents, et parfois des suivants pour générer l'information manquante (interpolation de l'état, reconstruction basée sur le modèle de la voix,...etc.). Une trame est alors approximée en utilisant les informations des trames adjacentes puisqu'un signal de parole est considéré localement comme stationnaire [25].

.

Ces méthodes sont plus efficaces mais aussi plus complexes, elles sont souvent utilisées seules ou combinées dans les décodeurs de parole. Elles nécessitent une implémentation plus importante que la simple substitution avec souvent une logique de plus haut niveau permettant de gérer plusieurs cas de figures (lorsqu'une trame est perdue mais pas la précédente, lorsque plusieurs trames consécutives sont perdues, lorsque la dernière trame reçue est voisée ou non-voisée, etc.)

3.3.2.3. PLC du décodeur G.723.1

Le système de masquage de pertes implémenté dans le codec G.723.1, se base principalement sur l'extrapolation de l'enveloppe spectrale et les paramètres d'excitation. L'extrapolation de l'enveloppe spectrale se fait à base des coefficients LSF de telle sorte que l'équation d'extrapolation s'écrit comme suit,

$$LSF_{ep} = (LSF_{vc} - LSF_{mf}) + LSF_{mf} \quad (3.1)$$

Tel que, LSF_{vc} est le vecteur des coefficients LSF quantifiés de la dernière trame reçue avant la perte. LSF_{mf} est la valeur LSF moyenne fixe, ρ est le coefficient de prédiction utilisé dans le codage différentiel qui est augmenté de la normale $12/32$ à $23/32$. Avant la conversion des LSF en coefficients LP, une séparation minimale accrue des valeurs LSF est imposée.

Si la trame actuelle était marquée comme perdue et que la trame précédente a été classée comme non voisée, l'excitation de la trame actuelle est produite au moyen d'un générateur de nombres aléatoires uniformes.

$$e(n) = A_{plc} * \eta(n) \quad (3.2)$$

Sachant que, $\eta(n)$ est une valeur aléatoire dans l'intervalle $[-1,1]$.

Ou A_{plc} est le gain d'excitation calculé comme suit,

$$A_{plc} = \beta A_{pl} \quad (3.3)$$

Ou $\beta = 3/4$. Le gain des excitations est calculé à base des gains des deux dernières sous-trames de la trame reçue comme suit :

$$A_{plc} = \frac{A_g + (i_2 + i_3)}{2} \quad (3.4)$$

Sachant que, i_2 et i_3 , sont les indices des tables de gains de quantification, si le nombre de trames perdues successivement dépasse trois la valeur des échantillons de l'excitation seront mises à zéro.

A_g représente l'amplitude du gain des deux sous-trames.

Si la trame précédente a été marquée comme voisée, l'excitation est calculée comme suit :

$$e(n) = \beta e'(n) \quad (3.5)$$

Tel que :

$$e'(n) = \begin{cases} e(n) & n < 0 \\ e[\text{mod}(n, Lp) - Lp] & n \geq 0 \end{cases} \quad (3.6)$$

3.4. Conclusion

Nous avons présenté dans ce chapitre les différentes causes de pertes de paquets, en l'occurrence la congestion, la gigue, le délai de transit et les pertes liées aux protocoles. Ainsi nous avons présenté les principales méthodes de masquage de pertes et leurs différentes techniques qui sont réparties en deux catégories selon leur positionnement et on a vu leurs avantages et inconvénients ; à la source, méthode priorisation et hiérarchisation (risque d'exclusion mutuelle) et méthode d'entrelacement (retard additionnel pour les remettre en ordre). Au récepteur ; méthode de substitution efficaces seulement si les pertes ne sont pas très fréquentes ou si elles n'interviennent pas sur des trames successives, et nous avons fait une description sur le PLC du décodeur G.723.1.

Le chapitre suivant traite le côté pratique de masquage de pertes, ou nous allons simuler sous Matlab quelques programmes issus de la littérature et essayer de nouvelles techniques.

CHAPITRE

IV

4.1 Introduction

Comme précisé au chapitre précédent, différentes techniques de masquage de pertes de paquets existent, et fait l'objet d'un important domaine de recherche.

De nos jours, le volume de data échangé via IP ne cesse de croître, chose qui mène les scientifiques et chercheurs à optimiser au mieux les paquets échangés sans altérer l'information.

Dans ce chapitre, nous allons simuler différentes méthodes de masquage de pertes de paquets ; commençant par les méthodes basiques, et le PLC intégré au décodeur G.723.1 et proposant des modifications sur le PLC original afin de l'améliorer.

4.2 Simulation

4.2.1 Base de données TIMIT :

TIMIT [27][28] est en fait une base de données regroupant pas moins de 630 voix enregistrées. Cette base a été la première à être largement utilisée, de par son caractère complet. Son principe est en effet très simple, puisque pour chaque phrase lue par un locuteur, trois fichiers permettent de regrouper l'ensemble des données : un fichier texte avec la phrase, un fichier avec la transcription phonologique de la phrase, puis un dernier avec la description acoustique du signal de parole.

Dans nos simulations, nous avons extrait 40 séquences sonores de la base de données TIMIT échantillonnées à 8 KHz, correspondant à 10 minutes de parole.

4.2.2 Evaluation de la qualité sonore perçue :

L'évaluation de la qualité vocale perçue (anglais : *Perceptual Evaluation of Speech Quality*, PESQ)[29] est un outil permettant d'évaluer la qualité de la voix transmise par un système de télécommunication.

Donnant des valeurs allant de -0.5 à 4.5 (de qualité médiocre à excellente), les valeurs du PESQ nous renseignent sur la qualité sonore transmise ainsi, en simulant des pertes elles nous renseignent sur l'efficacité de la technique de masquage de pertes de paquets introduite.

Lors de la simulation les 40 séquences de parole contiennent plusieurs longueurs ainsi, pour chaque séquence on aura son PESQ, à la fin nous allons calculer le $PESQ_{moyen}$ qui est la

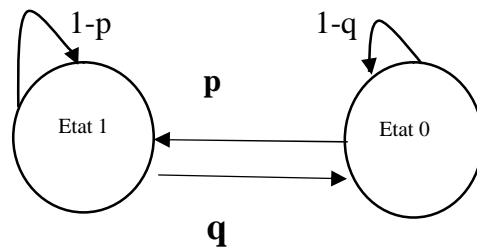
somme des valeurs du PESQ pour différentes séquences pondérée par sa longueur divisée sur la somme des longueur des séquences (nombre de trames L_i).[25]

$$PESQ_{moyen} = \frac{\sum_{i=1}^N PESQ_i * L_i}{\sum_{i=1}^N L_i} \quad (4.1)$$

4.2.3. Modèle de Gilbert-Elliott

Les systèmes de transmission de données par paquets sont soumis au phénomène de pertes. Comme nous l'avons vu précédemment, ces pertes sont dues soit à une congestion à un nœud du réseau, soit, dans le cadre de la téléphonie sur IP à un trop long retard d'un paquet qui est alors détruit, un lien réseau coupé. Dans les réseaux IP, la perte de paquets est par nature en rafale. On distingue la perte en rafale de la perte de paquets consécutifs dans le sens où la perte en rafale est une période de temps où une grande majorité de paquets sont perdus alors que lors d'une perte de paquets consécutifs, tous les paquets sont perdus.

Les pertes de paquets sont modélisées à l'aide d'une chaîne de Markov à deux états selon l'approche proposée par Gilbert et Elliott [30] . La figure (4.1) présente ce modèle.



Modèle de Gilbert-Elliott.

Ainsi, ce modèle est basé sur deux états, l'état **1**, le paquet est perdu alors que dans l'état **0** le paquet est reçu, avec **p** la probabilité de passer de l'état où le paquet est reçu à l'état de pertes, **q** est la probabilité de passer de l'état **1** à l'état **0**, c'est à dire, considérer que le paquet est perdu alors qu'il est reçu. Et les probabilités **(1-p)** et **(1-q)** sont associées au maintien de chaque état dans sa situation original.

Le taux de paquets perdus moyen, TPQ est alors donné par la relation suivante :

$$TPQ = \frac{p}{p+q} \quad (4.2)$$

4.3 Partie pratique :

Nous allons simuler différentes méthodes que nous allons évaluer après afin de mieux reproduire l'excitation perdue, pour cela nous avons proposé cinq différentes idées que nous allons comparer, tout en se référant au PLC intégré dans le décodeur G.723.1.

4.3.1 PLC du décodeur G.723.1 :

Le PLC implémenté dans le décodeur G.723.1, tient son excitation à générer sur la dernière trame reçue correctement ; c'est-à-dire, il exploite les coefficients LP précédents pour générer l'excitation désirée.

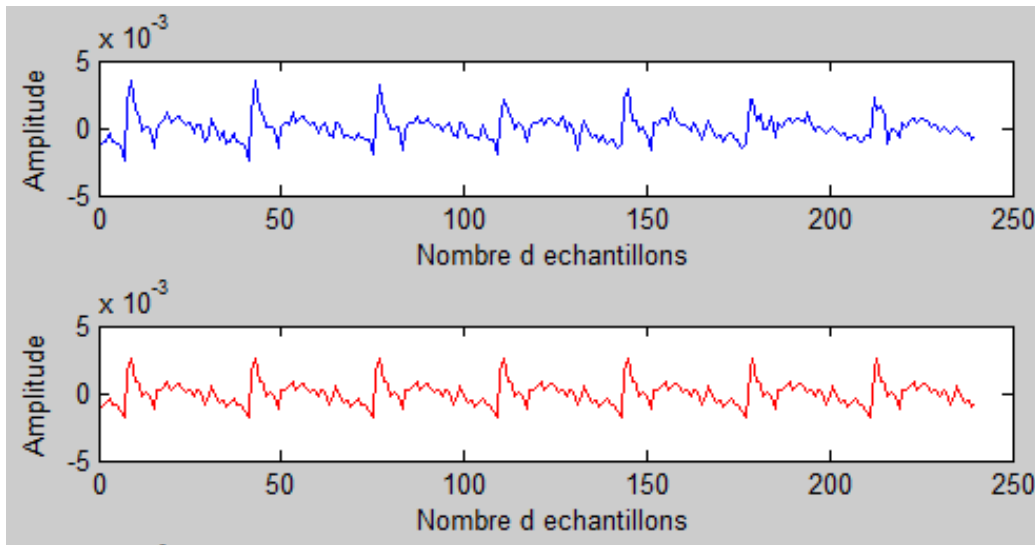


Fig.4.2. Excitation générée par le PLC de G.723.1

Dans la figure (4.2) nous distinguons deux signaux, celui en bleu représente l'excitation originale (sans perte), et le rouge l'excitation générée par le PLC du décodeur G.723.1.

Nous remarquons qu'il y a bien une légère différence entre les deux signaux (surtout au niveau temporel).

Le tableau (Tab 4.1) dresse les différentes valeurs du PESQ_{moy} obtenues pour différents masques (2%, 4%, 6%, 8%, 10%, 20%), ces masques représentent le taux de pertes injecté dans les 40 séquences avec une rafale de perte de 2 pertes successives.

Taux de pertes	PESQ moyen
2%	3.5441
4%	3.2876
6%	3.0855
8%	2.9081
10%	2.761
20%	2.2403

Tab.4.1. valeurs du PESQ pour G.723.1

Comme indiqué dans le tableau (Tab 4.1), les valeurs du PESQ diminuent plus le taux de perte augmente.

4.3.1.1 Utilisation de la trame précédente :

On remplace la trame perdue par la dernière trame reçue, cette technique est dite de substitution.

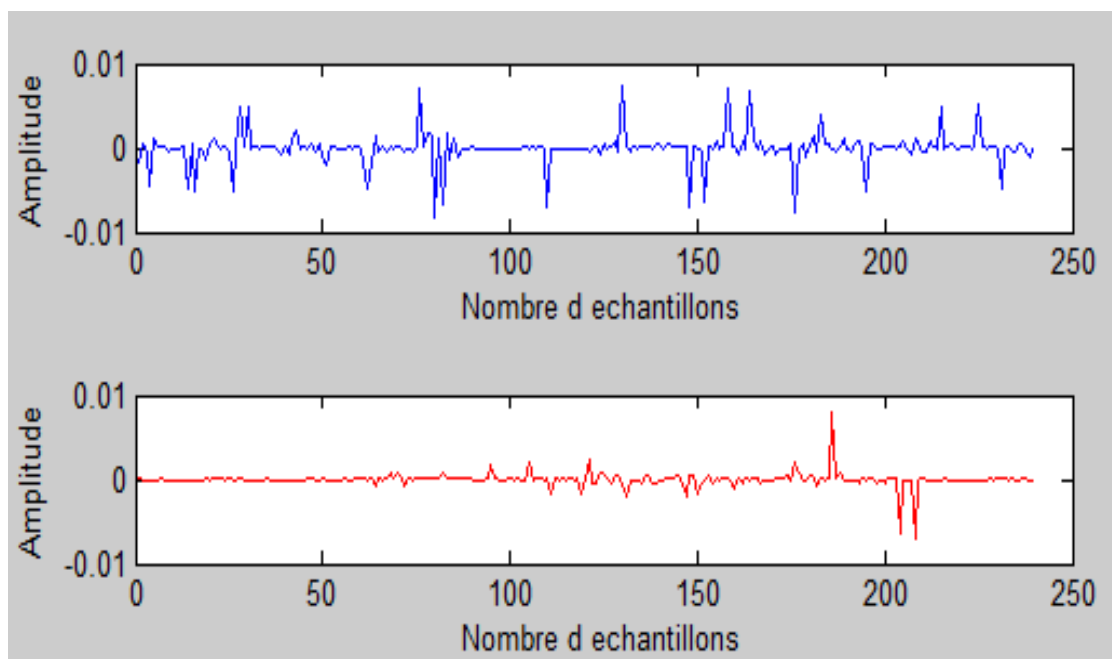


Fig.4.3. excitation avec la trame passée

Dans la figure 4.3. L'excitation en bleu est l'originale, et la rouge est la précédente ; on remarque pas une ressemblance dans les deux excitations.

Taux de pertes	PESQ moyen
2%	3.4199
4%	3.127
6%	2.8968
8%	2.6677
10%	2.4934
20%	1.9612

Tab.4.2. valeurs du PESQ pour génération de la trame passée

Des valeurs du PESQ obtenues, nous confirmons que la technique de génération de la trame passée, est efficace pour des taux de pertes faibles, mais de qualité moins efficace pour des taux élevés.

1.1. Utilisation de l'excitation nulle (zéros) :

Dans cette méthode, on remplace la trame perdue non par sa précédente, mais par des zéros (du silence) comme montré dans la figure 4.4 :

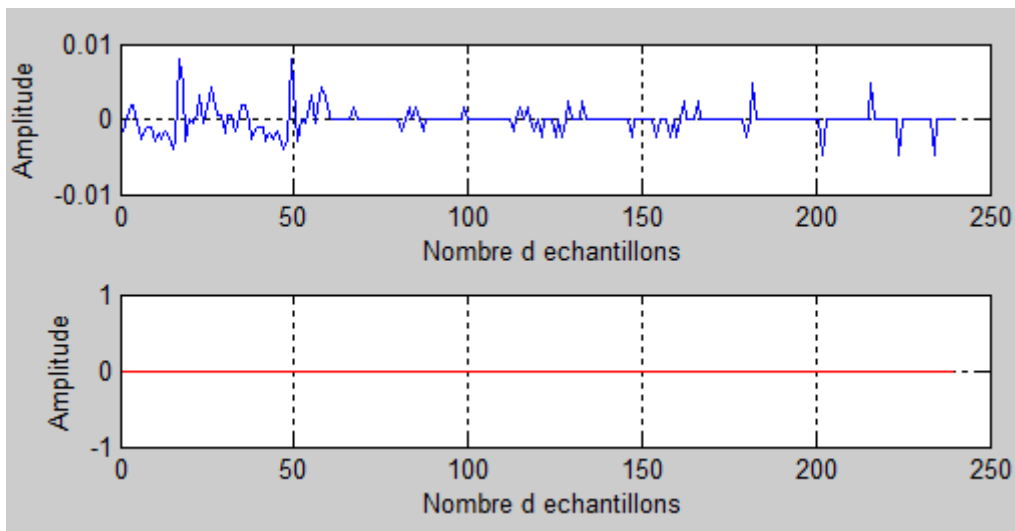


Fig.4.4. excitation avec des zéros

Et nous avons relevé les valeurs suivantes :

Taux de pertes	PESQ moyen
2%	3.2134
4%	2.768
6%	2.4376
8%	2.1579
10%	1.9517
20%	1.6351

Fig.4.4. excitation avec des zéros

Comme la technique de génération de la trame passée, la génération de l'excitation nulle ne déroge pas de la loi des techniques de substitution (voir chapitre 3) ; c'est-à-dire efficaces seulement si les pertes ne sont pas très fréquentes.

4.3.2 Technique (voisée=PLC- G.723.1) et (non-voisée=zéros) :

Dans ce cas, nous avons pris en compte la notion de voisement et non-voisement de la dernière trame reçue, et l'excitation générée dépend de cette notion.

Si la trame précédente est voisée, on génère le PLC G.723.1, sinon une excitation nulle, comme illustré sur la figure (4.5.a) et (4.5.b).

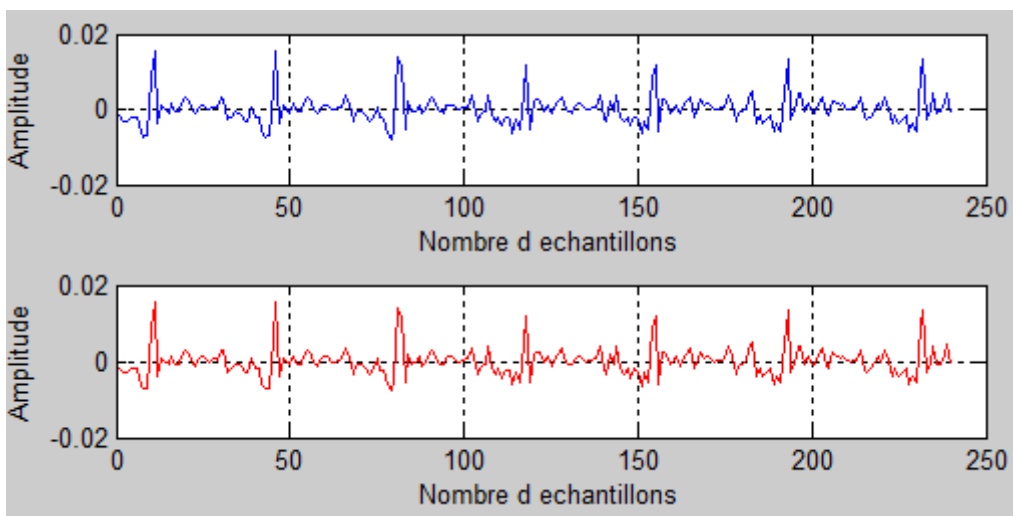


Fig.4.5.a. excitation (voisée=PLC- G.723.1)

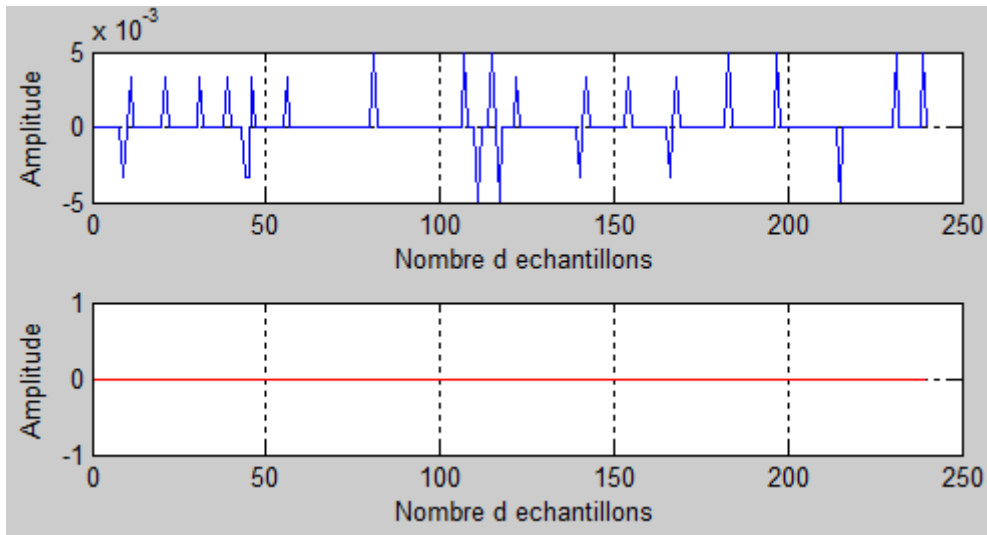


Fig.4.5.b. excitation (non-voisée=zéros)

Taux de pertes	PESQ moyen
2%	3.3812
4%	3.0818
6%	2.7911
8%	2.5675
10%	2.379
20%	1.799

Tab.4.4. valeurs du PESQ pour excitation :(voisée= PLC-G.723.1) (non-voisée=zéros)

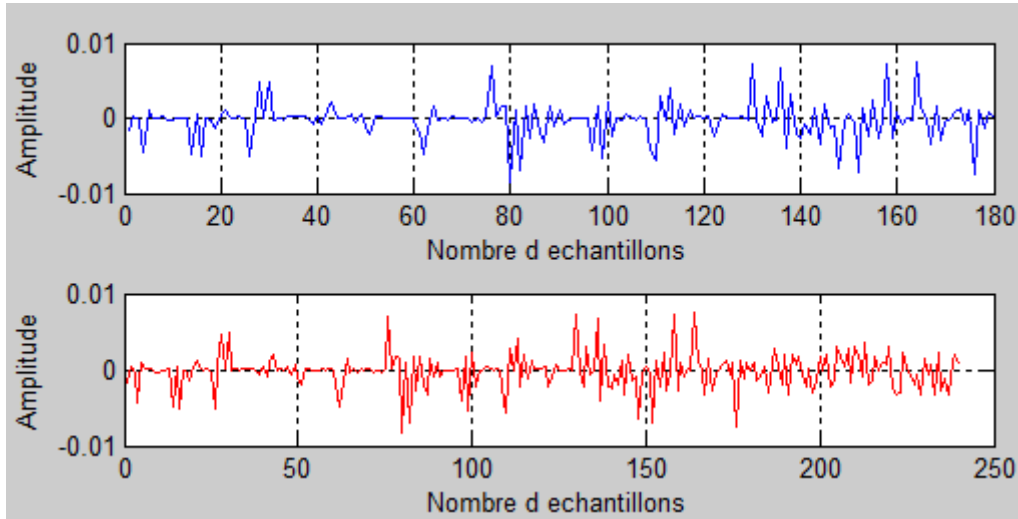
Les valeurs du PESQ tableau (4.4) montrent que cette méthode est meilleur que la précédente (excitation nulle), mais moins performante par rapport aux deux premières (à savoir le PLC du G.723.1, et la génération de l'excitation précédente).

4.3.3 Technique (voisée= PLC-G.723.1) (non-voisée=PLC-G.723.1 pour voisée) :

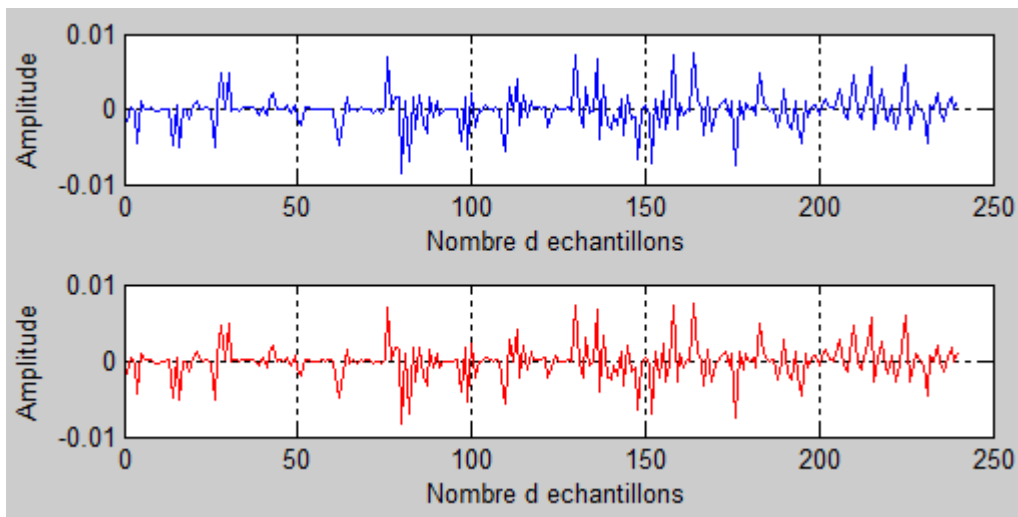
Le PLC de G.723.1 alloue deux traitements dépendant de la trame précédente si elle est voisée ou pas, on a remarqué [26] que pour les trames précédentes voisées le PLC masque bien les pertes, et moins bien pour les trames précédentes non-voisées.

Pour les trames précédentes voisées, on utilise le PLC de G.723.1, pour les non-voisées, on force le décodeur à les considérer comme voisées et leur allouer le même traitement fait pour les trames voisées.

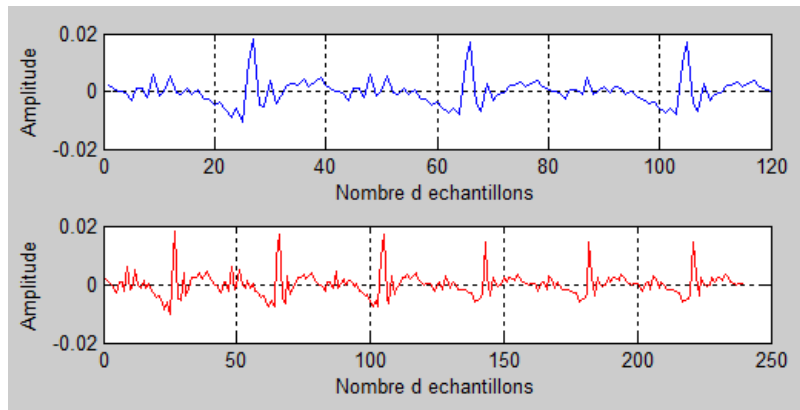
En considérant la trame précédente non-voisée comme voisée, on doit lui attribuer un délai tonal L (fréquence fondamentale), on a pris $L=30, 35, 45, 50, 90$.



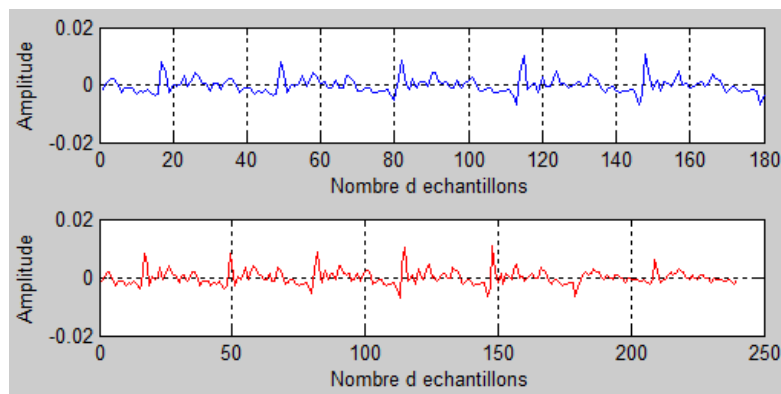
$L=30$



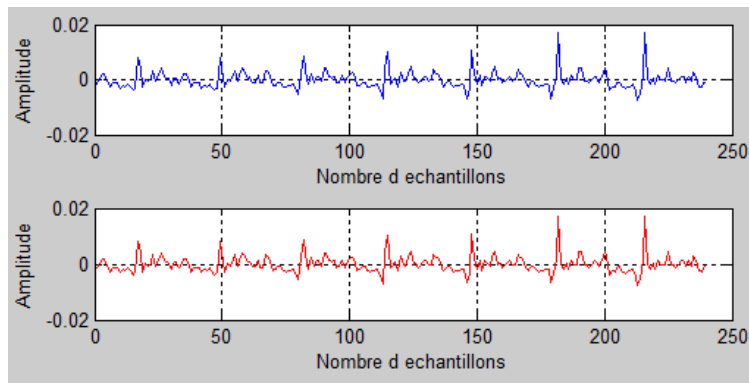
$L=35$



L=45



L=50



L=90

Fig.4.6. Technique (voisée= PLC-G.723.1) (non-voisée=PLC-G.723.1 pour voisée) pour différents L.

Ainsi nous avons relevé les résultats suivants :

	2%	4%	6%	8%	10%	20%
L=30	3.456	3.1992	2.9616	2.7427	2.5733	2.024
L=35	3.4776	3.231	3.0169	2.8116	2.6483	2.1145
L=45	3.4784	3.2326	3.0162	2.8093	2.6478	2.1148
L=50	3.482	3.2375	3.0242	2.8175	2.6540	2.1272
L=90	3.482	3.2354	3.0208	2.8166	2.6537	2.1234

Tab.4.5. Valeurs du PSQ pour différents L

Le choix des L autour de L=45 (30, 35, 45,50) n'est pas aléatoire, [25] a montré que les délais tonaux (fréquence fondamentales) des séquences de la voix gravitent beaucoup plus autour de 45 qu'ailleurs; et pour L=90 ce n'est que le double de la valeur 45.

D'après le tableau (4.5), les meilleurs valeurs sont pour un délai tonal L=50.

4.3.4. Utilisation du filtre LMS :

Le filtre LMS (least mean square) est un filtre égaliseur adaptatif ; dans la figure (4.2) nous avons remarqué un léger décalage temporel entre l'excitation du PLC G.723.1, et la bonne excitation ; nous optons pour le LMS pour égaliser les deux signaux.

➤ **Principe :**

Cette technique se résume en deux étapes :

- Une étape d'apprentissage : son utilité est pour ajuster les paramètres (coefficients) du filtre LMS ; cette étape intervient lorsque la trame est bien reçue.
- La deuxième étape : c'est lorsqu'il y a perte, on injecte la dernière trame reçue dans le filtre qui rappelle le, a des coefficients réajustés pendant la période d'apprentissage, ainsi nous limiterons la propagation d'erreur.

4.3.4.1. Filtre LMS seul :

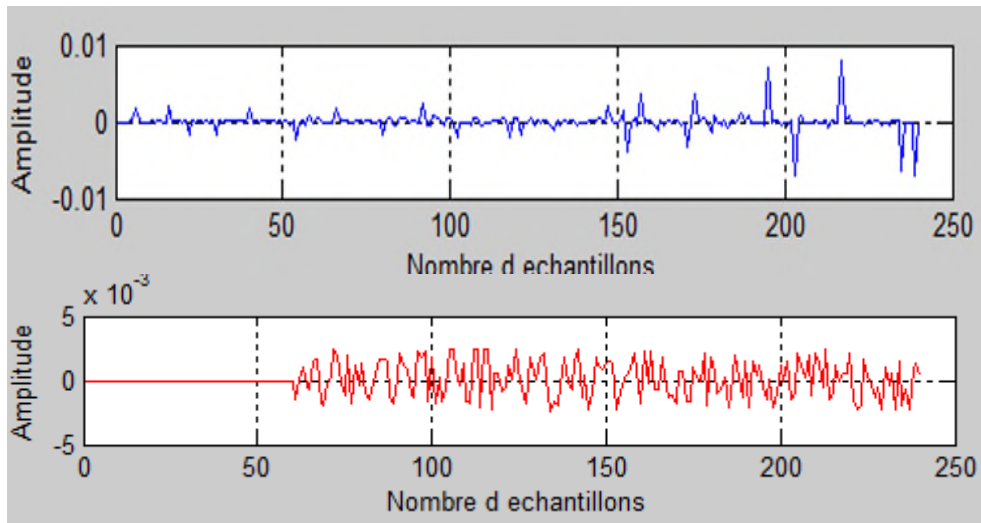


Fig.4.7.a.excitation générée par LMS lors de l'apprentissage

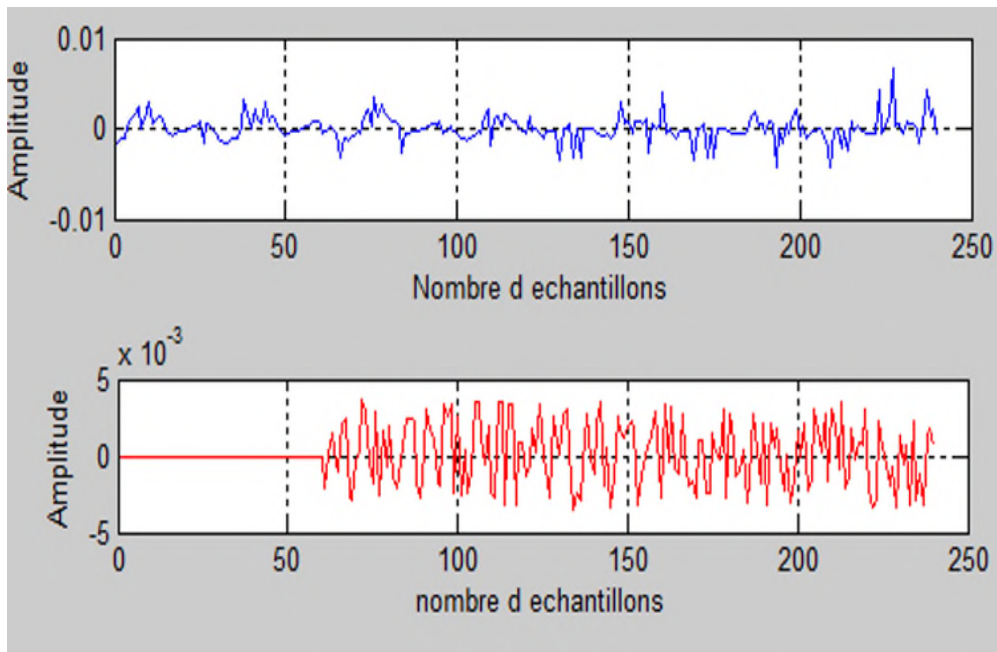


Fig.4.7.b.excitation générée par LMS lors de la perte

La figure (4.7.b) montre que les valeurs de l'excitation générée par LMS lors de la perte sont très inférieures aux valeurs de la bonne excitation.

Les valeurs du PESQ sont notées dans le tableau (4.6) ci-dessous :

Taux de pertes	PESQ moyen
2%	3.2134
4%	2.7687
6%	2.4376
8%	2.1579
10%	1.9517
20%	1.4399

Tab.4.6. valeurs du PESQ pour excitation générée par LMS

Du tableau (4.6), on remarque que le filtre LMS n'est pas efficace comparant aux autres méthodes.

4.3.4.2. Filtre LMS + PLC-G.723.1 :

Pour pallier aux faibles valeurs du signal généré par LMS, nous avons opté de leur additionner les valeurs du PLC-G.723.1, donc le filtre LMS sert comme un filtre de pondération pour l'excitation LPC-G.723.1.

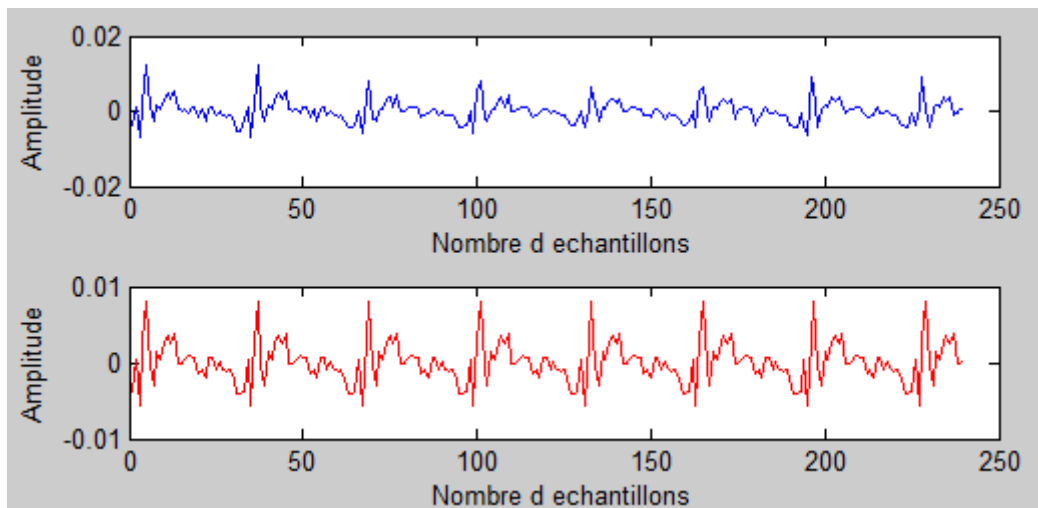


Fig.4.8.excitation(en rouge) générée par LMS+PLC-G.723.1

Dans la figure (4.8), le signal en bleu est celui de la bonne excitation, le rouge est celui généré par LMS + PLC-G.723.1. On remarque une légère différence entre les deux.

Taux de pertes	PESQ moyen
2%	3.5582
4%	3.3752
6%	3.1923
8%	3.0298
10%	2.89
20%	2.3876

Tab.4.7. valeurs du PESQ pour excitation générée par LMS+PLC-G.723.1

Du tableau (4.7), les valeurs du PESQ de cette méthode ont considérablement augmenté par rapport à la méthode LMS seul, atteignant une valeur de 3.5582 pour un taux de perte de 2%.

4.4 Récapitulatif des méthodes :

Des tableaux précédents, on résume les valeurs obtenues dans le tableau (4.8) :

	2%	4%	6%	8%	10%	20%
PLC	3.5441	3.2876	3.0855	2.9081	2.761	2.2403
Excitation précédente	3.4199	3.127	2.8968	2.6677	2.4934	1.9612
Excitation nulle	3.2134	2.768	2.4376	2.1579	1.9517	1.6351
V=PLC N.V=nulle	3.3812	3.0818	2.7911	2.5675	2.379	1.799
V=PLC NV= V-PLC	3.482	3.2375	3.0242	2.8175	2.6540	2.1272
LMS seul	3.2134	2.7687	2.4376	2.1579	1.9517	1.4399
LMS+PLC	3.5582	3.3752	3.1923	3.0298	2.89	2.3876

Tab.4.8. Valeurs du PESQ pour les différentes méthodes

Tq : - V= voisée

- N.V= non-voisée

En comparant les valeurs obtenues dans le tableau précédent, on peut constater que la méthode la moins appropriée pour le masquage des pertes est celle de LMS seul ; et la plus performante est celle du LMS + PLC-G.723.1.

La dernière technique proposée est légèrement meilleur que celle proposée par codeur G.723.1 (tout en se basant d'elle) ; technique qui peut être perfectionnée en trouvant les bons paramètres initiaux (le pas d'adaptation, la longueur du filtre...).

4.5 Conclusion :

Sachant les différentes causes de pertes, et les multiples méthodes de masquage de pertes de paquets, et le fonctionnement du codeur G.723.1 dans les deux chapitres précédents, nous avons pu proposer six techniques que nous avons testées dans ce dernier chapitre.

En présentant premièrement les outils de mesure que nous utilisons (à savoir la base de données TIMIT, le PESQ qui est l'outil d'évaluation objective de parole, et le canal de transmission Gilbert-Elliott), nous avons après calculer les valeurs du PESQ pour le PLC imbriqué dans le décodeur G.723.1 qui nous a servi de repère pour comparer nos méthodes proposées.

Nous avons servi aussi de la notion du voisement et non-voisement pour le traitement de génération de l'excitation afin de varier les expériences.

Enfin nous avons opté pour un filtre adaptatif simple et très répandu qui est le LMS, et nous avons pu obtenir des améliorations dans la qualité objective de la voix en le combinant avec le PLC d'origine.

CONCLUSION GENERALE

Conclusion générale ;

Dans un monde en plein essor technologique, le multimédia prend de plus en plus de place dans notre société, et la voix sur IP s'érige de place au détriment des autres systèmes de communication par sa simplicité d'utilisation, sa réduction des coûts, et son enrichissement du contenu. Mais elle doit relever certains défis techniques pour son amélioration. Parmi ces défis la gestion de pertes de paquets qui altèrent le signal, donc la qualité de service. Proposer des techniques capables d'augmenter la robustesse du codage de parole du type CELP face aux erreurs de transmission est le but de notre étude.

L'étude réalisée au cours de ce mémoire a porté sur quatre axes principaux :

Le premier axe, le signal de parole, son processus de génération et sur la transmission de la voix sur IP, son architecture, son mode de fonctionnement et ses différents protocoles.

Le deuxième, le codage numérique du signal de parole de type CELP basé sur l'analyse par synthèse, on a vu les différents codecs, leur différence de principe de fonctionnement et leur plage de débit. On a fait une modélisation acoustique détaillée de la parole, modélisation précise du conduit vocal et de la source d'excitation en suite une présentation générale du vocodeur G.723.1.

Le troisième, sur les pertes de paquets, leurs causes (réseaux congestionnés, gigue, latence et les pertes dues au protocole) et les différentes techniques de masquage qui sont réparties en deux méthodes, méthode basée à la source et méthode basée au récepteur et nous avons fait une description sur le PLC du décodeur G.723.1.

Le quatrième, est la partie d'application de techniques de masquage de pertes, commençant par une définition des outils de mesure à savoir la base de données TIMIT, le canal de transmission GILBERT-ELLIOTT basé sur le modèle markovien, et le PESQ qui est un outil d'évaluation de la qualité vocale transmise par un système de télécommunication. Enfin nous avons simulé différentes méthodes de masquage. D'abord par PLC du décodeur G.723.1, dont les valeurs du PESQ servaient de repère pour les autres méthodes que nous avons proposées et testées. En se basant sur la modification de l'excitation générée par le PLC du décodeur G.723.1 nous avons testé six techniques de masquage des pertes afin d'améliorer la robustesse du codeur G.723.1 face aux erreurs de transmission, et nous avons obtenu pour la technique où on a combiné le filtre adaptatif simple LMS avec le PLC du décodeur des

résultats encourageants qui peuvent être améliorés davantage, ou en utilisant d'autres filtres adaptatifs plus complexes comme le NLMS (LMS normalisé)

Dans le domaine de traitement de signal, qu'il soit pour le codage numérique, la compression des données ou le masquage de perte, peut être encore enrichi par d'autres travaux dans le but d'obtenir un meilleur résultat.

Références bibliographiques

- [1]. Thom, G. A. (1996). H. 323: the multimedia communications standard for local area networks. *IEEE Communications Magazine*, 34(12), 52-56.
- [2]. Site : <https://www.frameip.com/voip/>
- [3]. Douglas, D. M. O. S. Codage échelonnable à granularité fine de la parole: Application au codeur G. 729.
- [4].Koenig, L. (2011). *Masquage de pertes de paquets en voix sur IP* (Doctoral dissertation, Institut National Polytechnique de Toulouse).
- [5].Lachachi, N. (2006). Codage paramétrique de la parole en vue de sa transmission sur internet. Mémoire de Magister. Université d'Oran.
- [6]. UIT, S.D.L.N.D.T., Codeur vocal à double débit pour communications multimédias acheminées à 5.3 Kbits/s et à 6.3 Kbits/s. (2006).
- [7]. Ghio, A. (2006). Modélisation du conduit vocal.
- [8].Didi, S. (2014). *La Téléphonie sur IP (ToIP)* (Doctoral dissertation).
- [9]. Ouakil, L., &Pujolle, G. (2008). *Téléphonie sur IP*. Editions Eyrolles.
- [10]. Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., ...&Schooler, E. (2002). *SIP: session initiation protocol* (No. RFC 3261).
- [11]. Schulzrinne, H., Casner, S., Frederick, R., & Jacobson, V. (2003). *RTP: A transport protocol for real-time applications* (No. RFC 3550).
- [12]. Pujolle, G. (2014). *Les réseaux*. Editions Eyrolles.
- [13]. Postel, J. (1980). *User datagram protocol* (No. RFC 768).
- [1]. Thom, G. A. (1996). H. 323: the multimedia communications standard for local area networks. *IEEE Communications Magazine*, 34(12), 52-56.
- [2]. Site : <https://www.frameip.com/voip/>
- [3]. Douglas, D. M. O. S. Codage échelonnable à granularité fine de la parole: Application au codeur G. 729.
- [4].Koenig, L. (2011). *Masquage de pertes de paquets en voix sur IP* (Doctoral dissertation, Institut National Polytechnique de Toulouse).
- [5].Lachachi, N. (2006). Codage paramétrique de la parole en vue de sa transmission sur internet. Mémoire de Magister. Université d'Oran.
- [6]. UIT, S.D.L.N.D.T., Codeur vocal à double débit pour communications multimédias acheminées à 5.3 Kbits/s et à 6.3 Kbits/s. (2006).
- [7]. Ghio, A. (2006). Modélisation du conduit vocal.

- [8]. Didi, S. (2014). *La Téléphonie sur IP (ToIP)* (Doctoral dissertation).
- [9]. Ouakil, L., & Pujolle, G. (2008). *Téléphonie sur IP*. Editions Eyrolles.
- [10]. Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., ...& Schooler, E. (2002). *SIP: session initiation protocol* (No. RFC 3261).
- [11]. Schulzrinne, H., Casner, S., Frederick, R., & Jacobson, V. (2003). *RTP: A transport protocol for real-time applications* (No. RFC 3550).
- [12]. Pujolle, G. (2014). *Les réseaux*. Editions Eyrolles.
- [13]. Postel, J. (1980). *User datagram protocol* (No. RFC 768).
- [14]. Recommendation, C. C. I. T. T. (1988). Pulse code modulation (PCM) of voice frequencies. In *ITU*.
- [15]. Rec. ITU-T. G.722.2, (2003). Wideband coding of speech at around 16 Kbits/s using AM-RWB. Int Telecom Union. Geneva.
- [16]. Rec, I. T. U. T. (1996). G. 729, Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP). *International Telecommunication Union, Geneva*.
- [17]. Rec, I. T. U. T. (1994). G. 728, Coding of speech at 16 Kbits/s using a low-delay code excited linear prediction.
- [18]. Chu, W. C. (2003). Speech coding algorithms. *Foundation and evolution of standardized coders*.
- [19]. Spanias, A. S. (1994). Speech coding: A tutorial review. *Proceedings of the IEEE*, 82(10), 1541-1582.
- [20]. Goldberg, R., & Riek, L. (2000). *A practical handbook of speech coders*. CRC press.
- [21]. Baudoin, G., Cernocky, J., Gournay, P., & Chollet, G. (2000, September). Codage de la parole à bas et très bas débits. In *Annales des télécommunications* (Vol. 55, No. 9-10, p. 462). Springer-Verlag.
- [22]. CCITT III, F. (1972). 4-Recommendation G. 711.
- [23]. Schroeder, M., & Atal, B. S. (1985, April). Code-excited linear prediction (CELP): High-quality speech at very low bit rates. In *ICASSP'85. IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 10, pp. 937-940). IEEE.
- [24]. Salami, R., Laflamme, C., Adoul, J. P., Kataoka, A., Hayashi, S., Moriya, T., ...& Shoham, Y. (1998). Design and description of CS-ACELP: A toll quality 8 kb/s speech coder. *IEEE transactions on Speech and Audio Processing*, 6(2), 116-130.

- [25]. Benamirouche, N., Amélioration du codage de la parole de type celp face aux erreurs de transmission. Thèse de doctorat. Université Abderrehmane Mira Béjaia. Faculté de Sciences et Technologie.
- [26]. Ichalal, S., Bouaoun, S., & Benamirouche, N. (2018). *Etude des techniques de masquage de perte de paquets (PLC) pour améliorer la robustesse du codage de la parole face aux erreurs de transmission (VoIP)+ Application* (Doctoral dissertation, Université Abderrahmane Mira-Bejaia).
- [27]. Garofolo, J. S. (1988). The structure and format of the DARPA TIMIT CD-ROM prototype. *Documentation of DARPA TIMIT*.
- [28]. Lamel, L. F., Kassel, R. H., & Seneff, S. (1989). Speech database development: Design and analysis of the acoustic-phonetic corpus. In *Speech Input/Output Assessment and Speech Databases*.
- [29]. Recommendation, I. T. (2001). Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. *Rec. ITU-T P. 862*.
- [30]. Gilbert, E. N. (1960). Capacity of a burst-noise channel. *Bell system technical journal*, 39(5), 1253-1265.