**Université ABBES LAGHROUR Khenchela**
**Faculté des Sciences et de la Technologie**
**Département de Génie Industriel**
**جامـــعة عبـــاس لـــــغرور خـنـشـلة**
**كليـــة العلـــوم والتكنـــولـــوجيـا**
**قسـم الهنـدسة الصنـاعية**

**N° Série :**.....

# Mémoire de fin d'étude

*Pour l'obtention du diplôme de Master*

## Filière : Télécommunications
## Spécialité : Télécommunications Avancées

*Réalisé par :*

Melle. Benachi Houria

Melle. Rihi Asma

**THEME**

# Effet de la perte des trames sur la qualité de la parole, cas de codeur CELP

*Soutenu le 2*/06/2016 *devant la commission d'examen composée de :*

| | | |
|---|---|---|
| **M.** | Bediaf aziz | Président |
| **M.** | Khalfaoui Mahiou | Directeur du Mémoire |
| **M.** | Bouden Soufian | Examinateur |

# *Dédicace*

*Je rends grâce à Dieu de m'avoir donné le courage et la volonté. Ainsi que la conscience d'avoir pu terminer mes études.*

*Je dédie ce modeste travail aux étoiles de ma vie* ***Ma mère*** *et* ***mon père*** *et* ***mes sœurs*** *et* ***mon frère Mohamed*** *A tous les membres de la famille* ***Rihi*** *et* ***Ben-Talha.***

*A toutes mes amies Et à tous ceux qui m'aiment.*

***ASMA.R***

# DEDICACE

*Au meilleur des pères*

*A ma très chère maman*

*Qu'ils trouvent en moi la source de leur fierté*
*A qui je dois tout*

*A ma sœur et mes frères*

*A qui je souhaite un avenir radieux plein de réussite*

*A mes Amies*

*A tous ceux qui me sont chers*

*(Houria )*

# *Remerciements*

*Tout d'abord, nous remercions Dieu, notre créateur de nos avoir donné les forces, la volonté et le courage afin d'accomplir ce travail modeste.*

*Nous adressons un grand remerciement à notre encadreur M.Khalfaoui Mahiou qui a proposé le thème de ce mémoire, pour ses conseils et son investissement du début à la fin de ce travail.*

*Nous tenons également à remercier messieurs les membres de jury pour l'honneur qu'ilsnous ont fait en acceptant de siéger à notre soutenance.*

*Enfin, nous tenons à exprimer notre profonde gratitude à nos familles qui nous ont toujours soutenues et à tout ce qui ont participé à la réalisation de ce mémoire. Ainsi qu'à l'ensemble des enseignants qui ont contribué à notre formation.*

# الملخص

ثمة مشكلة رئيسية في نظم الاتصالات الرقمية وهي الوقوع في أخطاء في البيانات المرسلة من خلال قناة مشوشة .
الهدف من هذا العمل هو دراسة خصائص FS-1016 وكذلك نوعية الصوت المشفر والمفك شفره من طرف هذا الأخير بدلالة أخطاء الإرسال.

إرسال الصوت مفهوم تولد من التكامل ، هذا الانتقال من حد إلى آخر، بعض الحزم تضيع ولن تصل إلى المستقبل الذي كانت متوجه اليه فقدان الإطارات بالنسبة إلى عتبة 5،10 حتى 3 من الفقدان لا يؤثر في نوعية الصوت واستبدال الإطارات الضائعة لن يكون ضروريا هذا ما يسمح لنا بتقليص زمن مرور الشبكات.

إدماج الإطارات المزيفة لن يؤثر على نوعية الصوت بالنسبة إلى عتبة $5.10^{-3}$

على العكس زمن مرور الشبكة يرتفع.

بالنسبة إلى استبدال الإطارات الضائعة يفضل استبدالها بالإطارات السابقة بدل من استبدالها بالإطارات المزيفة.

نستطيع القول أن المخبر يتوفر حاليا على المشفر وكل المعلومات اللازمة لاستخدامه.

**مفاتيح:**

قناة صاخبة, FS-1016 , نوعية الصوت المشفر والمفك شفره

إرسال الصوت عن طريق الحزم,الإطارات الضائعة

# Résumé

Un des problèmes majeurs dans les systèmes de communication numérique est l'apparition des erreurs sur les données transmises à travers un canal bruité.

L'objectif de ce travail consiste à étudier des caractéristiques du FS-1016 ainsi que la qualité de la parole codée et décodée par ce dernier en fonction des erreurs de transmission.

La transmission de la parole par paquet est un concept né de l'intégration, Lors de cette transmission d'une extrémité à autre, certains paquets seront perdus et n'arrive pas au récepteur auquel il était destinés.

La perte des trames pour un seuil de taux de l'ordre $5.10^{-3}$ de perte ne dégrade pas sensiblement la qualité de la parole, et le remplacement des trames perdues ne sera pas nécessaire, cela nous permet de minimiser le temps de traversée des réseaux.

L'insertion des fausses trames n'affecte pas la qualité de la parole pour un seuil de taux d'insertion de l'ordre $10^{-3}$, par contre le temps de traversé du réseau augmente. Cependant pour le

remplacement des trames perdues, il vaut mieux, remplacer les trames perdues par les précédentes que de les remplacer par des fausses trames.

On peut dire que le laboratoire dispose maintenant d'un codeur avec toutes les informations nécessaires pour son utilisation.

***MOTS-CLES :***

un canal bruité, FS-1016, la qualité de la parole codée et décodée, La transmission de la parole par paquet, La perte des trames, des réseaux, un codeur.

# Abstract

A major problem in digital communication systems is the occurrence of errors on data transmitted through a noisy channel.

The objective of this work is to study the characteristics of the FS-1016 and the quality of the encoded speech and decoded by the latter in terms of transmission errors.

The speech transmission packet is a concept born of integration At this transmission from one end to another, some packets will be lost and will not reach the receptor to which it was intended.

The loss of frames to a law-rate threshold 5.10-3 loss does not significantly degrade the quality of speech, and replacement of lost frames will not be necessary, it allows us to minimize network crossing time.

The insertion of false frames does not affect the speech quality for an insertion rate threshold of 10-3 order, against time through the network increases. However for the replacement of lost frames, it is better to replace the lost frames by earlier than replacing them with false frames.

We can say that the lab now has an encoder with all the information necessary for its use.

**Keywords:**

Noisy channel, FS-1016, the quality of the encoded speech and decoded, The speech transmission packet, lost frames, network, encoder.

# SOMMAIRE

# Liste des figures :

# *Introduction*

# INTRODUCTION GENERALE

Le moyen de communication que l'homme utilise le plus est bien la parole : il est simple, et le développement des communications lui ont donnée un aspect privilégié. L'usage du téléphone aussi bien le fixe que le mobile de dernière génération en est une bonne illustration. Ainsi, pour permettre à un grand nombre d'usagers 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 de la transmettre.

En effet, dans les premiers systèmes de téléphonie numérique, la parole est numérisée à 64 Kbps. De nombreux algotithmes ont été proposés pour diminuer ce débit tout en essayant de conserver une qualité subjective donnée en fonction des exigences de l'application à laquelle le codeur est destiné. 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 l'archivage ou la transmission. Le décodeur utilise ces paramètres pour reconstruire un signal de parole synthétique.

La plupart des algorithmes de codage mettent à profit un modèle linéaire simple de production de parole. Le signal de parole n'étant pas stationnaire, les codeurs le découpent généralement en trames quasi-stationnaires de durée entre 5 et 30 ms pour extraire les paires de raies spectrales ou coefficients LSF (Line Spectral Frequencies) qui sont déduites des coefficients de prédiction linéaire et qui possèdent de bonnes propriétés pour la quantification.
Le standard américain FS1016 est un codeur de parole à 4800 bits/s basé sur la technique CELP (Code Excited Linear Prediction). Ce standard a été développé par le département de la défense des Etats Unis d'Amérique (DoD) et le laboratoire AT&T Bell. Dans ce système de transmission 10 coefficients LSF (Line Spectral Frequencies) représentant les coefficients de prédiction sont codés sur 34 bits pour chaque trame de parole de 30 ms. Ce codage est effectué après une quantification scalaire.

L'objet de ce mémoire est l'étude des caractéristiques du FS-1016 ainsi que la qualité de la parole codée et décodée par ce dernier en fonction des erreurs de transmission. Ces erreurs interviennent soit au niveau des trames de paramètres fournies pa le codeur toutes les 30 ms et qui

forment le champ d'information du paquet de parole à transmettre, soit dans les autres champs du paquet et en particulier le champ d'adresse.

Ce mémoire est divisé en quatre chapitres.

- Le chapitre un représente une introduction aux différentes techniques de codage, il décrit brièvement les principaux de codage.
- Le chapitre deux présente une étude du codeur CELP d'une manière générale.
- Le chapitre troix décrit la mise en paquets de la parole, l'architecture des réseaux et la transmission de la parole par paquets.
- Le dernier chapitre présente l'ensemble des traveaux effectués, qui consistent en l'étude de la qualité de la parole en fonction des erreurs de transmission.

# *Chapitre 01*

# *Généralité sur les systèmes de codage*

## 1.1. Introduction

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

## 1.2. Description du signal vocal

Le signal de parole est issu d'un organe phonatoire comprenant les poumons, les cordes vocales, les cavités buccales, les cavités nasales, les lèvres et les narines (figure 1.1).

Au niveau acoustique, le signal est constitué par des variations de la pression de l'air engendrées par le conduit vocal **[1].**

Le système respiratoire fournit l'énergie nécessaire à la production de la parole, et se comporte comme une source continue, modulable de plusieurs manières.

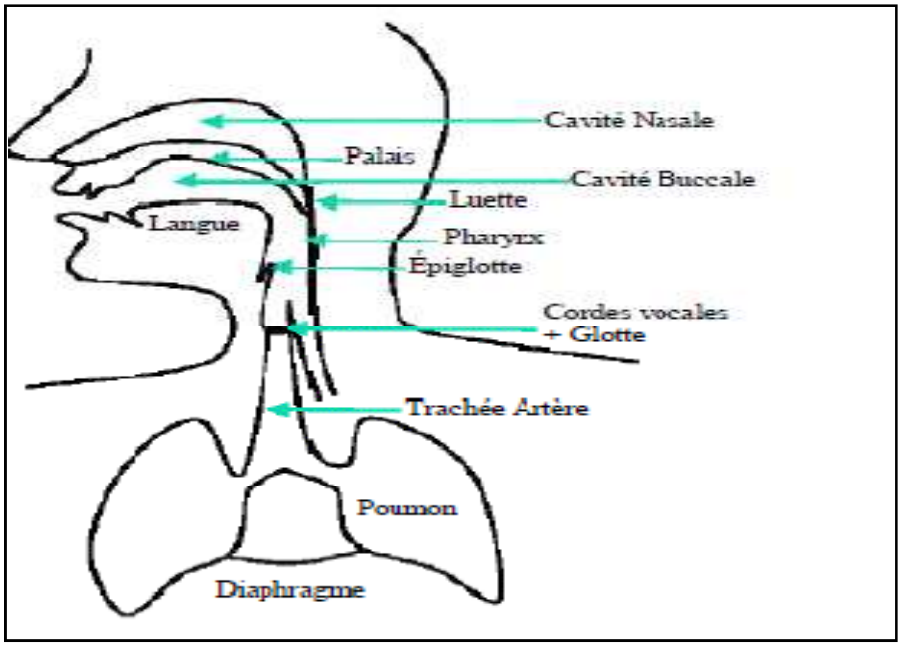


***Figure1.1:*** *Système phonatoire*

## 1.3. Paramètres du signal de parole

Le signal vocal est généralement caractérisé par trois paramètres: sa fréquence fondamentale, son énergie et son spectre.

### 1.3.1. La fréquence fondamentale

Elle représente la fréquence du cycle d'ouverture/fermeture des cordes vocales. Cette fréquence caractérise seulement les sons voisés, elle peut varier :

- De 80Hz à 200Hz pour une voix masculine,
- De 150Hz à 450Hz pour une voix féminine,
- De 200Hz à 600Hz pour une voix d'enfant.

### 1.3.2. L'énergie

Elle est représentée par l'intensité du son qui est liée à la pression de l'air en amont du larynx. L'amplitude du signal de la parole varie au cours du temps selon le type de son, et son énergie dans une trame est donnée par :

$$E = \sum_{n=0}^{N-1} s^2(\mathrm{n}) \qquad (1.1)$$

Avec N : la taille de la trame.

### 1.3.3. Le spectre

L'enveloppe spectrale ou spectre représente l'intensité de la voix selon la fréquence, elle est généralement obtenue par une analyse de Fourier à court terme. La quasi stationnarité du signal de parole permet de mettre en œuvre des méthodes efficaces d'analyse et de modélisation utilisées pour le traitement à court terme du signal vocal sur des fenêtres de durée généralement comprise entre *20ms* et *30ms* appelées trames, avec un recouvrement entre ces fenêtres qui assure la continuité temporelle des caractéristiques de l'analyse. La transformée de Fourier à court terme (*TFCT*) d'un signal échantillonné est par définition la transformée du signal pondéré.

$$\hat{\mathrm{S}}(k) = \hat{\mathrm{S}}\left(f = \frac{k}{N}\right) = \sum_{n=0}^{N-1} s(n).w(n).exp\,(-j2\pi nk/N) 0 \leq \mathrm{k} \leq \text{-}1 \qquad (1.2)$$

Où : N : Le nombre de points prélevés.

*S(k)* : Spectre complexe.

*s(n)*: Segment analysé.

W(n): Fenêtre de temps.

Le spectre de puissance (appelé aussi densité spectrale de puissance de la transformé de Fourier) est donné par:

$$\left|\hat{S}(k)\right|^2, 0 \leq k \leq \frac{N}{2} \quad (1.3)$$

## 1.4. Différentes techniques de codage

De nombreux travaux relatifs aux algorithmes de codage tendent à maximiser le compromis entre l'efficacité, le coût et la qualité de transmission des systèmes de communication en fonction des débits disponibles.

Les algorithmes de codage de parole peuvent être divisés en trois classes distinctes : les codeurs temporels, les codeurs paramétriques et les codeurs hybrides.

### 1.4.1. Les codeurs temporels

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 **[2].**

#### 1.4.1.1. Le codage MIC (modulation par impulsion et codage)

Le codage MIC est reposent exclusivement sur le théorème d'échantillonnage de Shannon et une quantification fixe, sont apparus dans les années 60. Etant donnée la distribution d'amplitudes des échantillons de parole, un quantificateur non-uniforme apporterait une meilleure qualité pour le même débit. Ainsi, l'Union Internationale des télécommunications a normalisé le codeur G.711 en 1972, un codeur logarithmique de parole de type PCM pour la transmission téléphonique avec un débit de 64 Kbits/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, 3400] Hz. Avec une complexité plus élevée, le codage de parole peut être obtenu avec des débits inférieurs.

Ce codage se fait en deux étapes, la conversion du continu en discret ou échantillonnage qui transforme une forme d'onde analogique ou continue dans le temps en une forme d'onde définie a des instant discrets dans le temps, la quantification qui transforme l'amplitude du signal à un instant discret dans le temps en un nombre (figure 1.2).

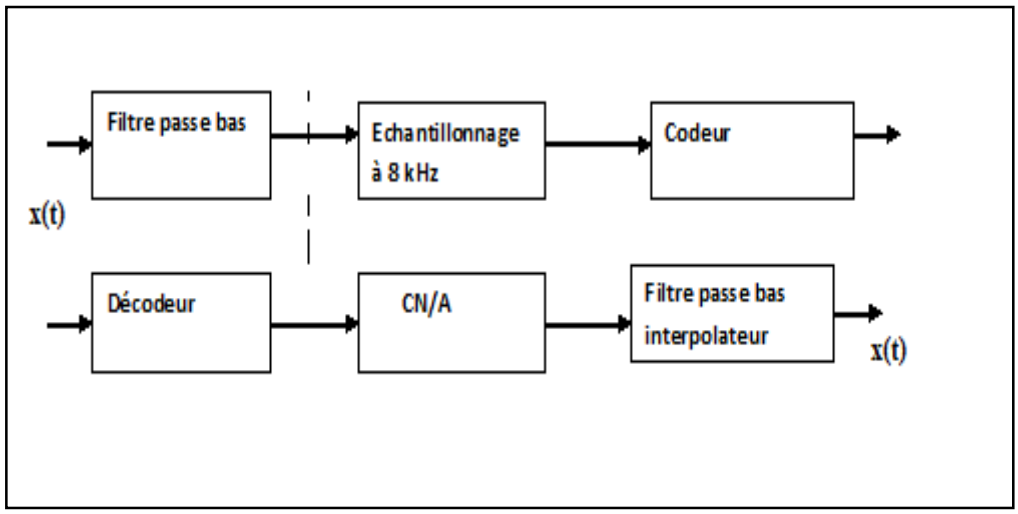


***Figure 1.2.*** *Système de transmission MIC*

### 1.4.1.2. Codage MIC différentiel adaptatif (MICDA ou ADPCM)

Dont le principe est de quantifier non plus la valeur d'un échantillon à un instant donné mais la différence avec une valeur prédite à partir d'échantillons précédents, a été introduite. Le signal à coder se limitant aux coefficients d'un filtre, le nombre de bits nécessaires à sa quantification est diminue. Cette codage (ADPCM) émerge au début des années 80 et permet de réduire le débit de moitié par rapport à loi PCM sans détériorer la qualité de parole et des services annexes. Le codeur G.721, normalisé par l'UIT depuis 1984, est un exemple de système ADPCM qui fonctionne à 32 Kbits/s.

### 1.4.1.3. Codage en bande élargie 64 Kbits/s par seconde

Certaines applications exigent une qualité supérieure à celle offerte par le MIC. Le codage en sous-bandes associé au codage MICDA, peut satisfaire ces exigences.

Le signal de parole limité à 8 kHz est échantillonné à 16 kHz. Il est ensuite divisé en deux sous bandes, une bande inférieure [0 :400Hz] et une bande supérieure [4000 :8000 Hz], par l'intermédiaire de deux filtres miroirs en quadrature (Figure 1.2).

Le signal issu de la sous-bande inférieure est comparé à sa valeur estimée et le signal résultant est envoyé dans un quantificateur adaptatif non linéaire à 60 niveaux. Dans la boucle de retour, les deux bits de poids faible du signal quantifié sont supprimés. Le signal ainsi obtenu est utilisé pour l'adaptation d'un quantificateur inverse de quinze niveaux pour la production d'un signal résiduel quantifié auquel on ajoute la valeur estimée. Le tout est utilisé pour la prédiction de

la parole. Le codage du signal de la sous-bande supérieure se fait de la même manière sauf qu'ici il n'y a pas suppression des bits. Dans ce type de codage, le canal à 64Kbits/s peut être partagé entre les données et la parole.

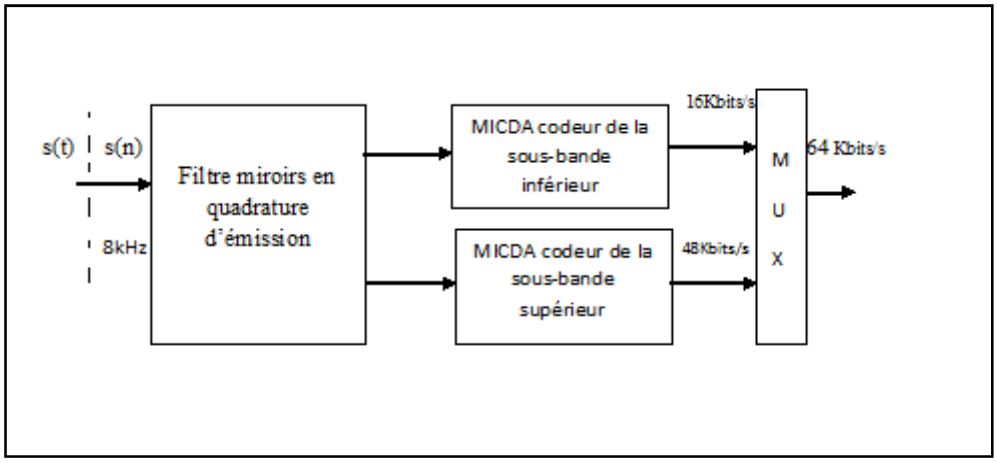


***Figure 1.3.****Principe du codeur sous-bandes*

## 1.4.2. Les codeurs paramétriques

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 **[3].**

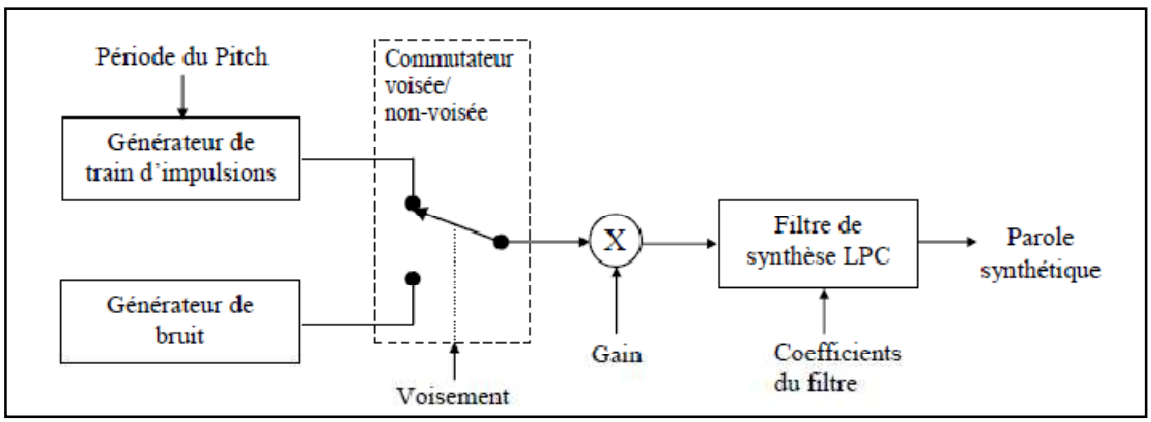


***Figure 1.4.*** *Le modèle LPC de production de la parole.*

Dans les vocodeurs prédictifs, le conduit vocal est modélisé par un filtre tout-pôle de fonction de transfert $H(z)$ :

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

$$A(z) = +\sum_{k=1}^{p} a_{\mathrm{k}}\ z^{-\mathrm{k}} \tag{1.5}$$

Le signal de parole est découpé en trames de courte durée. Pour une trame de 20 à 30 ms, le signale de parole est considéré comme stationnaire. Chaque trame de parole est analysée afin de déterminer les paramètres suivants :

- Les coefficients de prédiction $\{a_k\}_{\mathrm{k=1,\ldots,p}}$ pour le filtre de synthèse.
- Si la trame de parole est une trame voisée ou non voisée.
- La période du pitch dans le cas d'une trame voisée.
- Le gain principalement lié à l'énergie de la trame de parole.

Ces paramètres sont transmis au décodeur. À la réception, le signal de parole est reconstitué par le passage d'un signal d'excitation à travers le filtre de synthèse LPC1/ $A(z)$ (figure1.4). Le signal d'excitation est un train d'impulsions périodiques (à la cadence du pitch) ou un bruit blanc selon que le signal à reproduire est voisé ou non voisé. Ce signal d'excitation modélise le signal résiduel de prédiction obtenu par le passage du signal de parole à travers le filtre inverse $A(z)$. Le modèle LPC de production de la parole est donné à la figure 2.4. Le standard fédéral FS1015 (LPC10E) fonctionnant à un débit de 2.4 kbit/s est un exemple de codeur de parole paramétrique basé sur le modèle de production LPC **[3].**

### 1.4.3. Les codeurs 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. 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).

La structure de base de codeur hybride se compose d'une étape d'analyse qui consiste à déterminer les coefficients du filtre de synthèse et d'une étape d'analyse par synthèse qui consiste à trouver une séquence d'excitation qui minimise un critère d'erreur (*Figure 1.5)*

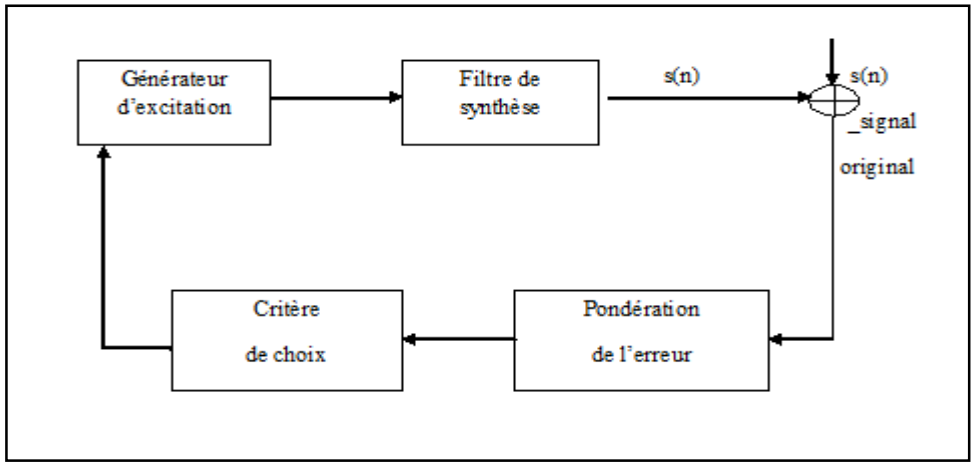


***Figure 1.5*** *codeur hybride (analyse par synthèse)*

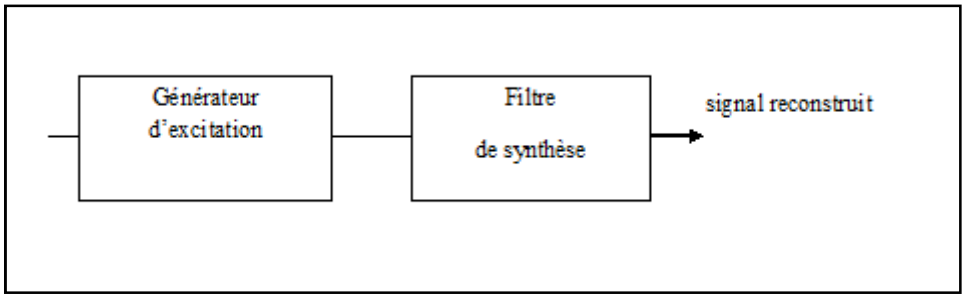


***Figure 1.6*** *Décodeur hybride*

### 1.4.3.1. Codeur prédictif adaptatif

Ce codeur utilise la méthode d'analyse par prédiction linéaire, il procède tout d'abord à une déconvolution du signal vocal par filtrage inverse, fournit ainsi un signal résiduel codé par un codage temporel classique [6proph]. Le signal transmis se compose du signal résiduel et des coefficients du filtre qui servent pour la déconvolution. La structure d'un tel codeur est très voisine

de celle d'un codeur différentiel. Elle nécessite simplement la transmission des paramètres (énergie et les valeurs des coefficients) pour des tranches de signal de 10 à 20 ms (Figure 1.6).

Le signal est analysé par bloc avec les avantages suivants :

- les coefficients du filtre sont calculés à partir du signal original
- La transmission de ces coefficients, qui contiennent une part importante de l'information ne demande qu'un débit faible. La corrélation à long terme peut être aussi exploitée par ce type de codage

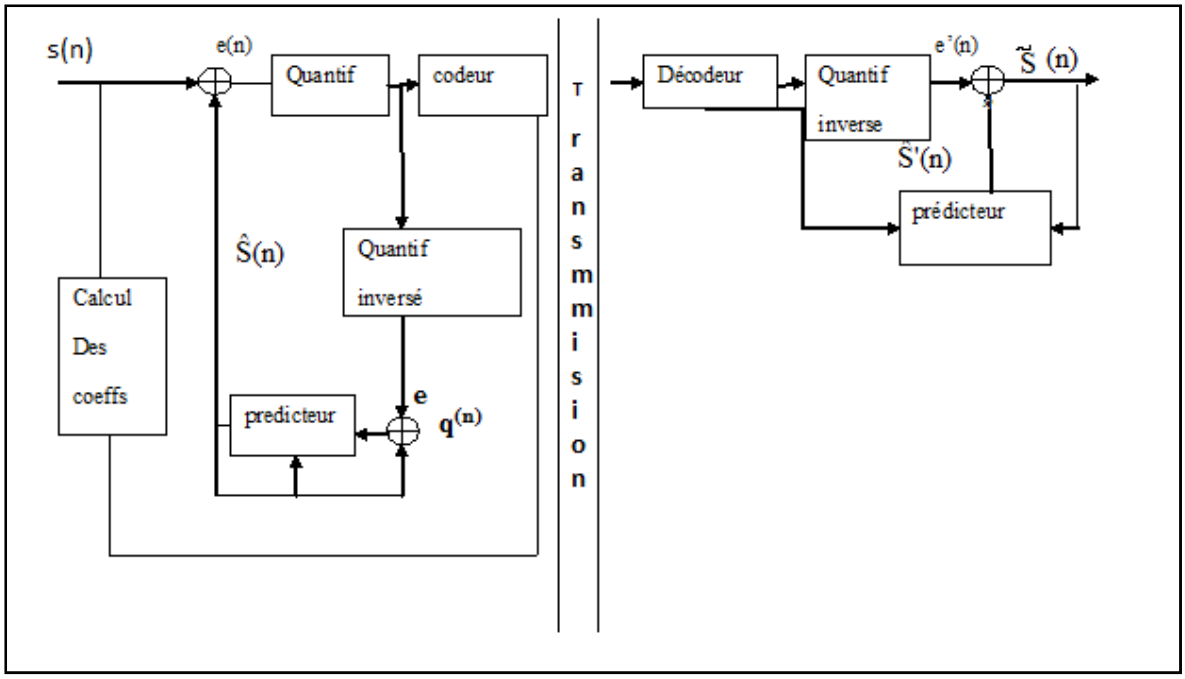


***Figure 1.7.*** *Schéma de principe d'un codeur prédictif adaptatif*

## 1.4.3.2. Codeur excité par code

Le code Excited Prédictive Codec (CELP) cherche à coder le signal résiduel d'une façon vectorielle. Dans le codeur, le signal de parole est découpé par exemple en trames de 160 échantillons. Les coefficients du filtre de synthèse à court-trame sont calculés toute les trames. Chaque trame est ensuite découpé en quatre sous-trames de longueur 40 échantillons. Le signal de parole est comparé à un signal synthétique, qui résulte du filtrage d'un vecteur bruit issu d'un dictionnaire. Généralement, ce dictionnaire contient 1024 vecteurs connus en termes identiques aussi bien par l'émetteur que par le récepteur.

Chaque vecteur est filtré à long-terme et un filtre à court-terme pour produire un signal synthétique et on choisit de garder l'excitation qui produit le signal de parole le plus proche du signal original. Une fois le numéro de cette séquence d'excitation déterminé, il suffit de le transmettre.

### 1.4.3.3. Prédictifs excités par impulsions multiples

Dans ce codeur l'excitation est représenté comme une suite d'impulsions multiples dont la position et l'amplitude sont calculées de manière à donner un signal synthétique qui ressemble le plus possible au signal original **[4].**

L'algorithme cherche à déterminer l'excitation à fournir au filtre de synthèse pour minimiser l'énergie du signal résiduel. L'amplitude et la position de ces impulsions sont transmises en même temps que les coefficients du filtre linéaire ayant servi à la déconvolution.

De nombreuses recherches ont montré qu'il suffisait de quelques impulsions, de l'ordre de 10 par tranche de 10 ms de signal, pour modéliser correctement le signal de parole.

### 1.4.3.4. Codeur hybrides en sous-bandes

Le codage en sous-bandes consiste à filtre le signal de parole par un banc de filtres en quadrature couvrant toute la bande téléphonique, puis sous-échantillonner à la cadence de Nyquist et finalement quantifier séparément. Le signal de chaque sous-bande est codé par un codeur hybrident**[4].**

Pour reconstituer le signal, on somme les signaux issus de chaque sous bande, décodés, reéchantillonnés et filtrés par des filtres identiques à ceux de l'analyse. Le reéchantillonnage du signal des sous-bandes est important, car il permet de conserver la même quantité d'information.

### 1.4.4. Mesure de la qualité

Pour mesurer la qualité du codage, on utilise des mesures objectives qui sont purement mathématiques, dont la plus couramment utilisés par les codeurs qui essayent de préserver la forme du signal est le rapport signal à bruit. Les autres mesures de la qualité sont des mesures subjectives qui évaluent la qualité du codage à l'écoute. On distingue la technique Diagnostic Rythme Test (DRT) qui mesure l'intelligibilité sur un grand nombre de mots, la technique Diagnostic Acceptability Measure (DAM) qui mesure le naturel perçu de la parole, la technique Mean Opinion Score (MOS) ou l'auditeur évalue un codeur sur une échelle d'écoute allant de 1 à 5 (5 : excellent, 4 : bon, 3 : passable, 2 : mauvais, 1 : médiocre).

Cette technique a été appliquée aux codeurs les plus courants, pour une bande passante téléphonique de 3.4khz et on a trouvé :

- Pour le MIC (G711) à 64 Kbits/s 4.3.
- Pour ADPCM (G721) à 32 Kbits/s 4.1.
- Pour LD.CELP (G728) à 16 Kbits/s 4.
- Pour le CELP à 8 Kbits/s 3.7 Kbits/s 3.7.
- Pour CELP (FS-1016) 3.

## I.5. Conclusion

Dans ce chapitre nous avons présenté les différents types de codeurs paroles et nous sommes intéressés aux codeurs hybrides qui utilisent de façon complémentaire les avantages des techniques de codage temporelles et paramétriques pour permettre un codage efficace du signal de parole.

Le but du codage est de réduire le nombre d'information à envoyer chaque seconde tout en gardant une qualité adéquate au besoin.

# *Chapitre 02*

# *Codeur CELP*

## 2.1. Introduction

Les techniques de codage à bas débit tendent à réduire le débit d'information, lorsqu'on transmet ou lorsqu'on mémorise le signal de parole. Actuellement la technique la plus utilisée est le codage par prédiction linéaire à excitation par codes (CELP: Code ExitedLinearPrediction), où l'on peut atteindre des débits de transmission aussi bas que 4 800 bits/s.

## 2.2. Définition d'un codeur CELP idéal

Un codeur CELP idéal est composé d'un dictionnaire contenant un ensemble de vecteurs d'excitation. Le nombre de ces vecteurs est généralement de 1024 et chaque vecteur porte un numérode séquence connu par l'émetteur et par le récepteur. Chaque vecteurs d'excitation est multiplié par un gain g, qui a pour objet d'optimiser la ressemblance entre le signal ainsi syntétisé et le signal de parole original et d'un modèle H(z) généralement un filtre tous pôles qui sert á supprimer les redondances ou á décorréler le signal de parole par filtrage inverse **[5].**

Dans ce type de codage on transmet au décodeur les paramètres du filtre ainsi que le gain et l'index de la séquence d'excitation optimale, pour synthéser un signal de parole qui ressemble au signal original.

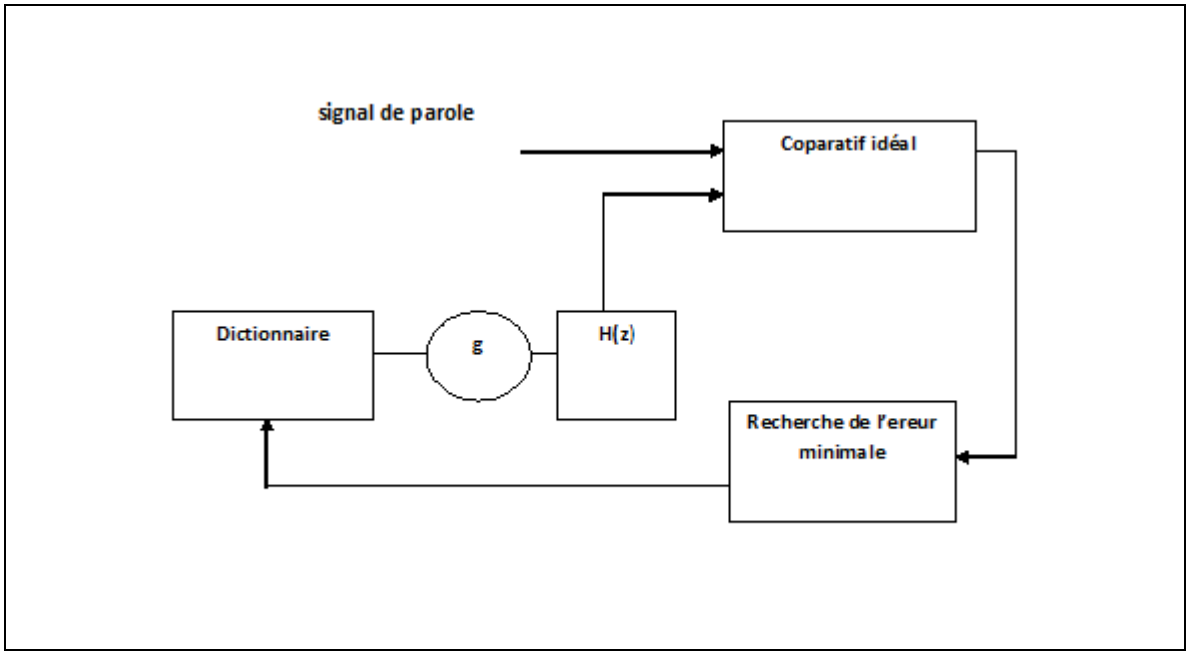


***Figure2.1****Codeur CELP idéal*

### 2.2.1. Modèle

Le modèle H(z) est un filtre tous pôles pour lequel on doit déterminer des coefficients à l'aide de la prédiction linéaire sur des fenêtres d'analyse de l'ordre de 80 à 240 échantillons pour une fréquence d'échantillonnage de 8 kHz. Ces fenêtres d'analyse ont été choisies en fonction du signal de parole qui est stationnaire sur des intervalles de temps de l'ordre de 10 à 30 ms. Generalement ce modèle comporte de huit à dix coefficients.

### 2.2.2. Le critère de minimisation choisi

Pour trouver l'excitation optimale qui nous fournit le signal le plus proche du signal original, on doit filtrer l'ensemble des séquences du dictionnaire par le filtre perceptuel W(z) et on choisit celle qui minimiser l'erreur quadratique moyenne$\sum e_n^2$.

$$W(z) = \frac{A(z)}{A(z/\gamma)} \text{ Avec } \qquad A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + - - - + a_p z^{-1-p} \tag{2.1}$$

La trasformée en Z inverse de l'erreur s'écrit :

$$E(z) = (S(z) - \tilde{S}_j(z)) / W(z) \tag{2.2}$$

$\tilde{S}_j(z)$ : représente la trasformée en Z de la réponse du filtre de synthèse à la $j_{ième}$ séquence d'excitation.

Le filtrage perceptuel a pour objet, de réduire la distortion du signal de parole perçue à l'écoute **[5,6].**

Il s'applique au signal d'erreur avant quantification, pour modéliser le spectre de bruit du quantificateur, de façon à qu'il reste suffisamment en dessous du spectre du signal utile.

Le diagramme du codeur idéal devient :

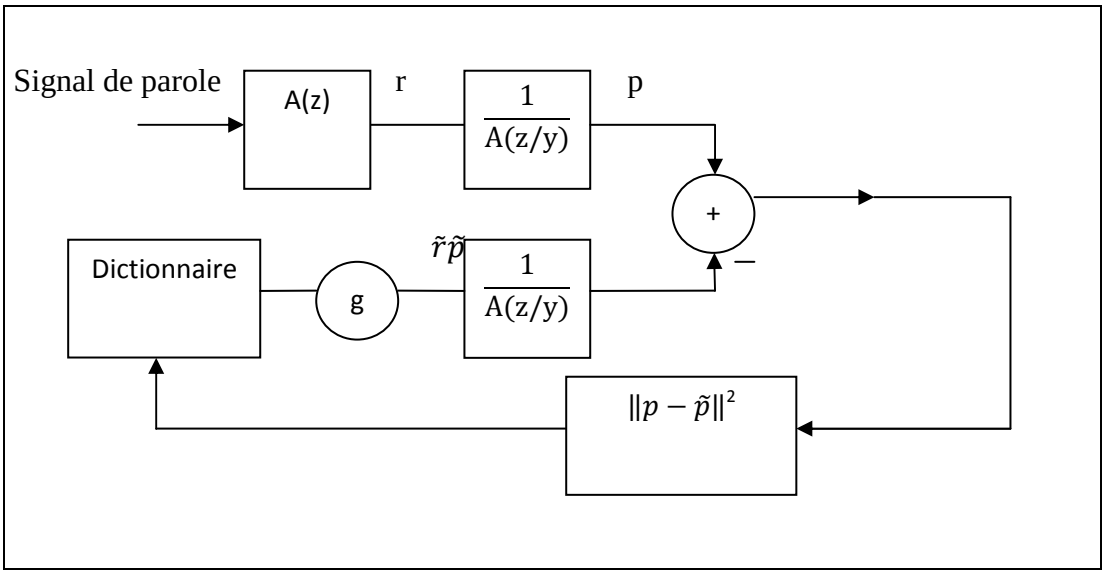


***Figure2.2.*** *Codeur CELP clasique*

$\tilde{p}$ : représente le signal synthétique.

P : représente le signal percpetuel dû au filtrage du signal de parole par le filtre perceptuel.

## 2.2.3. Générateur de code

Le générateur d'excitation est connu aussi bien par l'émetteur que par le récepteur. Il est composé d'un ou plusieurs modules d'excitation. Chaque module comporte un dictionnaire composé par un certain nombre de vecteurs d'excitation, chaque vecteur est multiplié par un gain de façon à ce que le vecteur filtré par le filtreperceptuel donne le signal perceptuel le plus ressemblant possible au signal perceptuel original.**[5,6]**

La forme la plus simple d'excitation est donnée par :

$$\tilde{r}_n = g_j e_n^j \qquad \text{pour n = 0, 1,2 .....,  N-1} \tag{2.3}$$

N : représente le nombre d'échontillons. $g_j$ : représente le gain optimal au sens du critère choisi.

$e_n^j$ : représente la séquence d'excitation d'index j.

Dans le cas ou on utilise un nombre K de dictionnaire, l'excitation composée est :

$$\tilde{r}_n = \sum_{k=1}^{K} g_{j(k)}\, e_n^{j(k)} \tag{2.4}$$

### 2.2.4. Définition du dictionnaire

Plusieurs types de dictionnaires ont été définis : des dictionnaires lacunaires ou la majorité des composantes des vecteurs sont nuls ; des dictionnaires d'impulsions régulièrement espacées, contenant la mémoire des excitations passées ; des dictionnaires algébriques structurées ; des dictionnaires construits avec des séquences multipulses;des dictionnaires binaire ou les échantillons prennent les valeurs -1 et +1; des dictionnaires ternaires ou les valeurs possibles des échantillons sont -1, 0 et +1**[6].**

### 2.2.5. Mémoire du filtre et influence desfenêtres d'analyse précédentes

La séquence d'excitation $\tilde{r}$ est choisis si le signal$\tilde{p}$, réponce du filtre 1/ A (z /$\gamma$) à $\tilde{r}$ , Minimise$\|p - \tilde{p}\|^2$.Soit h la réponce impulsionnelle infinie du filtre 1/ A (z/$\gamma$), alors p à pour expression :

$$\tilde{p}_n = \sum_{i=0}^{\infty} h_i \, \tilde{r}_{n-i} \tag{2.5}$$

Cette expression peut se décomposer en deux termes :

$$\tilde{p}_n = \sum_{i=0}^{n} h_i \, \tilde{r}_{n-i} + \sum_{i=n+1}^{\infty} h_i \, \tilde{r}_{n-i} \qquad \text{n = 0, 1,2,...., N-1} \tag{2.6}$$

Le premier terme a priori inconnu.

Le second terme$\tilde{p}_n^0 = \sum_{i=n+1}^{\infty} h_i \, \tilde{r}_{n-1}$ correspond aux conditions initiales du filtre de synthèse 1/A (z/$\gamma$).Ces conditions initiales sont dues aux excitations précédentes. Comme la séquence $\tilde{p}^0$ est constante et ne dépend pas du mot de code choisis, elle peut être soustraite du signal de parole perceptuelisép. Soit $\tilde{p}^{-1,}$ l'excitation tel que $p^1$ = p - $\tilde{p}^0$. Cette séquence est l'excitation pour la sous-trame qui permet une reconstruction parfaite de l'original p.

Le critére à minimiser est donc :

$$\|p^1 - \tilde{p}^1\|^2$$

Plus généralement

$$\left\|P^1 - \sum_{k=1}^{K} \tilde{P}^k\right\|^2 \tag{2.7}$$

$\tilde{P}^k$ : représente les contributions des K vecteurs issus des K dictionnaires.

Le filtre de synthése est donc un filtre sans conditions initiale.

## 2.2.6. Modélisation optimale au sens des moindres carrés

Partant du signal original perceptualisé auquel on a enlevé la contribution des fenêtres précédentes. On cherche à construire une estimation $\tilde{p}$ fonction du signal de parole original perceptualisé et optimalisé et optimsant le critère du moindre carrés.
Soit K le nombre de déctionnaires utilisés pour modéliser la séquence d'excitation. On cherche donc à déterminer les indices j(1),....., j(K) des vecteurs d'excitation et les gains$g_{j(1)}$,.., $g_{j(k)}$ de façon à minimiser l'erreur perceptuelle quadratique moyenne E =$\|p - \tilde{p}\|^2$.
E peut sécrire aussi de la maniére suivante :

$$E = \sum_{n=0}^{N-1}\left(p - \sum_{k=1}^{K} g_{j(k)} f_n^{j(k)}\right)^2 \quad (2.8)$$

N : représente le nombre d'échontillons dans la fenêtre d'analyse courante

$$\tilde{p} = \sum_{k=1}^{K} g_{j(k)}\, f^{j(k)} \quad (2.9)$$

$$f_n^j = \sum_{i=0}^{n} h_i\, e_{n-1}^j \text{ avec } f^j = \left[f_0^j, f_1^j, \ldots\ldots, f_{N-1}^j\right] \text{pour j =1,...,}\Delta \quad (2.10)$$

Δ : représente le nombre de vecteurs contenus dans un dictionnaire

$f_n^j$ : représente le résultat du filtrage de $\left[e_0^j, e_1^j, \ldots, e_{N-1}^j\right]$ par le filtre perceptuel.

Le développement de l'expression 2.9 donne :

$$\begin{aligned}
\tilde{p}_0 &= g_{j(1)} f_0^{j(1)} + g_{j(2)} f_1^{j(2)} + \cdots + g_{j(K)} f_0^{j(K)} \\
\tilde{p}_1 &= g_{j(1)} f_1^{j(1)} + g_{j(2)} f_1^{j(2)} + \cdots + g_{j(K)} f_1^{j(K)} \\
&\vdots \\
\tilde{P}_{N-1} &= g_{j(1)} f_{N-1}^{j(1)} + g_{j(2)} f_{N-1}^{j(2)} + \cdots + g_{j(K)} f_{N-1}^{j(K)}
\end{aligned} \quad (2.11)$$

En notation matricielle $\tilde{P}$ devient égal à :

$$\tilde{P} = \text{Ag} \quad (2.12)$$
$$\tilde{P} = [\tilde{P}_0 \tilde{P}_1 \ldots \tilde{P}_{N-1}]^T$$

$$A = \begin{bmatrix} f_0^{j(1)} & . & . & . & f_0^{j(K)} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ f_0^{j(1)} & . & . & . & f_0^{j(K)} \end{bmatrix}$$

$$G = [p - AG]^T [p - AG] \quad (2.13)$$

Avec

$$P = [p_0 \ . \ . \ . \ p_{N-1}]^T$$

$\tilde{p}$Serala meilleure estimation si E est minimale. Pour que E soit minimale, il faut que ces dirivés partielles par rapport aux $g_{j(k)}$ soient nulles. Aprés dérivition on obtient :

$$A^T[p - A\ g] = 0 \quad (2.14)$$

Qui peut s'écrire aussi : $A^T A\ g = A^T p$ (2.15)

On a ainsig = $[A^T A]^{-1} A^T p$ (2.16)

On remplace g par sa valeur dans 2.10 alors p devient égal :

$$p = A\ [A^T A]^{-1} A^T p \quad (2.17)$$

Minimiser E revient à maximiser l'expression suivante :

$$\|\tilde{p}\|^2 = p^T A\ (A^T A)^{-1} A^T p \quad (2.18)$$

Cette expression est obtenu en développemant les exprissions 2.8 et 2.13.

La recherche de la solution optimale consiste à essayer les kuplets formés de K séquences issues des K dictionnaires. Ceci est pratiquement difficile si les K vecteurs sont issus des K dictionnaires différents de tailles $T_1 ... T_K$ , le nombre de possibilités est n =$T_1 . T_2 ... T_K$. Si les vecteurs sont issus du même dictionnaire le nombre de possibilités est : $n_T^K = \frac{T!}{K!(T-K)!}$oùT la taille du dictionnaire.

On utilise le plus généralement un algotithme sous-optimal. Les vecteurs filtrés sont cherchés itérativement.

## 2.2.7. Algorithme itératif standard

A la première itération, un seul vecteur $f_n^j$ du premier dictionnaire est sélectionné on a :

$$A^T A = \langle f^j , f^j \rangle \tag{2.19}$$

$$A^T p = \langle f^j , p \rangle \tag{2.20}$$

$\langle x, y \rangle$ : représente le produit scalaire de deux vecteurs x et y.

Ensuite on selectione le deuxiéme vecteur du premier dictionnaire et ainsi de suite **[5].** Aprés avoir essayé tout les vecteurs du premier dictionnaire. On choisit l'index j(1) du vecteur qui maximise l'expression suivante :

$$\langle p, f^j \rangle \langle f^j , f^j \rangle^{-1} \langle f^j , p \rangle = \langle f^j , p \rangle^2 / \langle f^j , f^j \rangle \tag{2.21}$$

Une fois l'index est trouvé on doit alors calculer le gain qui lui est associé par :

$$g^{j(1)} = \frac{\langle f^{j(1)}, p \rangle}{\langle f^{j(1)}, f^{j(1)} \rangle} \tag{2.22}$$

A la kième itération, la contribution des k-1 premiers vecteurs $f^{j(1)}$ est retirée de p.

$$p^k = p - \sum_{i=1}^{k-1} g_{j(i)} f^{j(i)} \tag{2.23}$$

Cette opération a pour objet de donner un nouveau signal perceptuel p, qui permet de calculer un nouvel index et un nouveau gain dans un autre dictionnaire, qui vérifient les expressions suivantes :

$$G(k) = \text{argmax} \frac{\langle f^{j(k)}, p^k \rangle}{\langle f^{j(k)}, f^{j(k)} \rangle} \tag{2.24}$$

$$g_{j(k)} = \frac{\langle f^{j(k)}, p^{k} \rangle}{\langle f^{j(k)}, f^{j(k)} \rangle} \quad (2.25)$$

Etant donné que $\langle f^{j}, p^{k+1} \rangle = \langle f^{j}, f^{j} \rangle - g_{j(k)} \langle f^{j}, f^{j} \rangle$ , l'inetercorrélation nécessaire à l'étape k+1, peut être calculée à partir de l'intercorrélation à l'étape k, on obtient alors l'algorithme standard :

Pour k = 1,....., K

Pour j = 1,.....,Δ

. $\boldsymbol{\alpha}^{j(k)} = \langle f^{j(k)}, f^{j(k)} \rangle$et $\boldsymbol{\beta}^{j(k)} = \langle f^{j(k)}, p \rangle$

. j(k) = argmax($\alpha^{j(k)}/\beta^{j(k)}$) et $g_{j(k)} = (\alpha^{j(k)}/\beta^{j(k)})$

. $p^{k+1} = p^{k} - g_{j(k)} f^{j(k)}$

## 2.2.8. Introduction d'un dictionnaire prédictif adaptatif

Pour produire la structure périodique du signal de parole lorsque le retard M est inférieur à la durée de la sous-fenêtre d'analyse. On utilise un dictionnaire prédictif adaptatif. Ainsi la séquenced'excitation est composée de deux termes **[5,6].**

$$\tilde{r} = g_j e_n^j + b_{\tilde{r}\mathrm{n-M}} \quad (2.26)$$

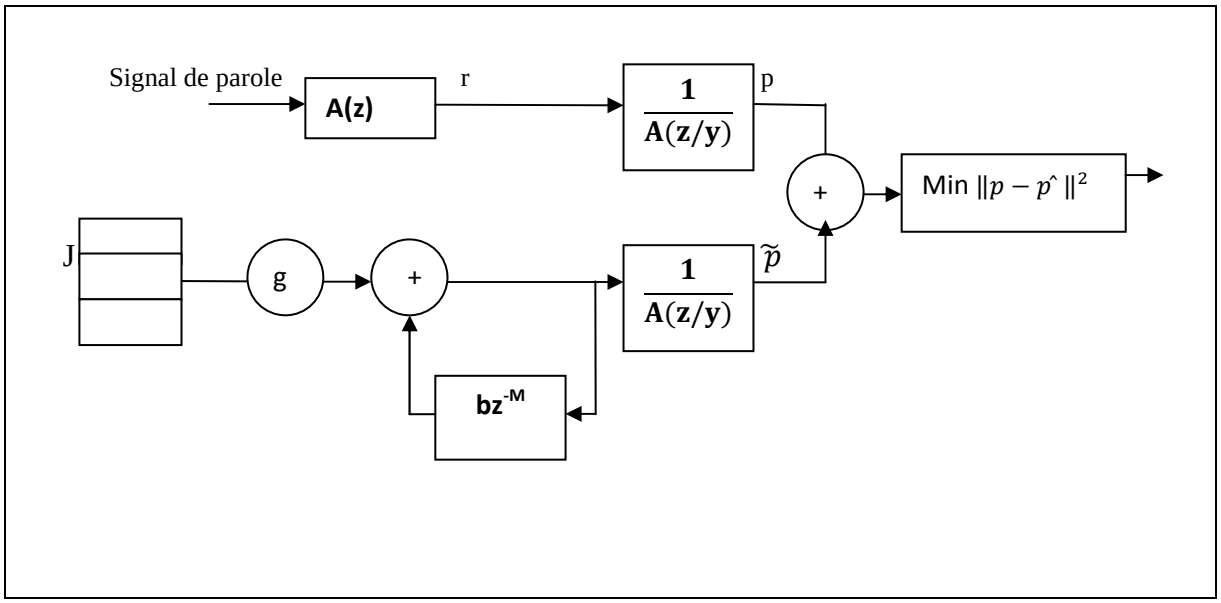


***Figure2.3****Modélisation signal perceptuel avcec prédicteur à long-terme*

La valeur de décalage M doit être supérieure ou égale à la taille de la fenêtre d'analyse. Il est nécessaire alors de faire un traitement par sous-fenêtre. La fréquence du fendamental moyenne est

de l'ordre de 100 Hz pour un locuteur masculin, et de 250 Hz pour un locuteur féminin ; cela veut dire que les valeurs probables de M sont respectivement égales à 80 et 32 échantillons. Habituellement la fenêtre d'analyse prend 160 échantillons, c'est pourquoi on divise cette dernière en quatre sous-fenêtres et on détermine le modèle d'excitation pour chaque sous-fenêtre. La ditermination de M et b peuvent être calculés de la même maniére que l'index j et le gain $g_j$ à la condition de construire un conditionnaire prédictif comportant l'excitation passée voir matrice si desous :

$$\begin{bmatrix} \tilde{r}_{-N} & \tilde{r}_{-N+1} & \cdot & \cdot & \tilde{r}_{-1} \\ \tilde{r}_{-N-1} & \tilde{r}_{-N} & \cdot & \cdot & \tilde{r}_{-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \tilde{r}_{-2N+1} & \cdot & \cdot & \cdot & \tilde{r}_{-N} \\ \tilde{r}_{-2N} & \cdot & \cdot & & \tilde{r}_{-N-1} \\ \cdot & \cdot & \cdot & \cdot & \\ \tilde{r}_{-M\,max} & \cdot & \cdot & \cdot & \tilde{r}_{-M\,max+N-1} \end{bmatrix}$$

Ce dictionnaire posséde une proprité intéressante. Lorsqu'on passe d'une fenêtre d'analyses à la suivante, l'ensemble du dictionnaire n'est pas remis en cause. Uniquement N vecteurs doivent être actualiés. Les autres se déduisent par translation vers le bas.Ce dictionnaire peut être construit on le prolongant vers le haut de la façon suivante : on utilise les N-1 échantillons disponibles$\tilde{r}_{-N+1}$,...,$\tilde{r}_{-1}$ et on les complète par $\tilde{r}_{-N+1}$, ensu on utilise les N-2 échantillons et on les complète par $\tilde{r}_{-N+2}, \tilde{r}_{-N+3}$ etc. On obtient :

$$\begin{bmatrix} \tilde{r}_1 & \cdot & \cdot & \cdot & \cdot & \tilde{r}_{-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \tilde{r}_{-Mmin} & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \tilde{r}_{-N+1} & \cdot & \cdot & \tilde{r}_{-1} & \tilde{r}_{-N+2} & \tilde{r}_{-N+3} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \tilde{r}_{-Qmax} & \cdot & \cdot & \cdot & \cdot & \tilde{r}_{-Qmax+N-1} \end{bmatrix}$$

## 2.3. Quelques codeurs hybrides récent

Plusieurs codeurs hybrides sont actuellement normalisés, on peut citer :

- le standard européen du GSM (group spécial mobile) pour la téléphonie cellulaire : C'est un codeur à 13 kbits/s. Il utilise un prédicteur à long-terme à un pas et un filtre à court-terme à 8 coéfficients, calculées sur des trames de 160 échantillons. Chaque trame est divisée en 4 sous trames qui sont codées séparément. L'excitation reconstruite est une séquence de 13 impulsions obtenue à partir du signal résiduel. La séquence d'excitation est codée par codeur MIC adaptatif.
- Le fédéral standard 1016 : C'est un codeur qui fonctionne à 4.8 kbits/s, il utilise un prédicteur à long-terme et un prédicteur à court-terme à 10 coéfficient, calculés sur des trames de 240 échantillons, et qui sont codées sur 34 bits. Chaque trame est divisée en 4 sous-trames de 60 échantillons chacune. L'excitation reconstruite est la somme d'un dictionnaire stochastique de 512 vecteurs.
- Le VSELP de Motorola : Motorola àdéveloppédeux codeurs VSELP, chacun d'eux utilise dix coefficients de parcor pour le filtre à court-terme ; la trame utilisée est de 30 ms et est codée sur 38 bits par une quantification scalaire. Chaque sous-trame est obtenue par une excitatation modélisée par un vecteur issu du dictionnaire de 2048 vecteurs pour le codeur à 4.8 kbits/s ou de deux vecteurs d'excitation issus de dictionnaires de 128 vecteurs pour le codeur à 8 kbits/s.
- Le codeur LD-CELP qui est n avec un normalisé par le CCITT pour un débit de 16 kbits/s avec un délai de 5 ms. Ce codeur utilise un prédicteur à court-terme d'ordre 50 et un vecteur d'excitation de longueur 50 et codé sur 10 bits.

## 2.4. Conclusion

Le codeur CELP à beaucoup évolué depuis le modèle de Schreoder et Atal avec des modifications de structure et de dictionnaire. Il permet de coder la parole d'une manière vectorielle à des débits très bas. Sa caractéristique principale c'est qu'il utilise des prédicteurs adaptatifs pour éliminer les redondances à court-terme et à long-terme du signal de parole. Il met en oeuvre aussi une procédure de recherche de l'excitation optimale, qui consiste à minimiser un critère pertinent d'un point de vue perceptif, faisant intervenir l'erreur entre le signal original et le signal reconstruit.

# *Chapitre 03*

# *Généralité sur la transmission de la parole par paquet*

## 3.1. Introduction

La transmission de la parole par paquet est un concept né de l'intégration des services. Elle a fait, depuis un certain temps, l'objet de réseaux expérimentaux à des débits divers. Elle exploite les techniques MIC connues sous la recommandation G.711 du CCITT, Le MICDA selon la recommandation G.721, G.726 et G.727 du CCITT et aussi les algorithmes CELP, dont les débits s'étendent de 4 à 16 Kbits/s. La parole est transmise avec succès sur les canaux de RNIS (Réseau Numérique à intégration de service) selon un ensemble de protocoles intégrés, découlant de la connexité numérique de ce réseau. Les données d'images et de parole vont donc partager d'une façon transparente les services protocolaires offerts par le réseau. La transmission de la parole par paquets intéresse aussi le domaine des réseaux locaux qui offrent des débits de transmission élevés, de plusieurs mégabits par seconde.

## 3.2. Définition d'un réseau

Un réseau et un moyen de communication qui permet à deux correspondants de communiquer indépendamment de la distance qui les sépare **[7].**

Pour que deux correspondants communiquent, il faut qu'ils utilisent les mêmes mécanismes de communication (même interface électrique et même langage), à travers une liaison physique. Pour que plusieurs correspondants communiquent, il faut pouvoir commuter la liaison de l'un à l'autre, en utilisant les mêmes mécanismes de communication **[6,7].**

Le réseau communication comporte alors des nœuds de communication capable de faire progresser la communication jusqu'au destinataire.

## 3.3. Différents types de commutations

Il existe trois types de réseaux, les réseaux à commutation de circuit, les réseaux à commutation de messages et les réseaux à commutation de paquet. Les réseaux à commutation de circuits ont été les premiers à apparaître.

### 3.3.1. Commutation de circuits

C'est historiquement la première à avoir été utilisé, elle consiste à créer dans le réseau un circuit particulier entre l'émetteur et le récepteur avant que ceux-ci ne commencent à échange désinformations. Ce circuit sera propre aux deux entités communiquant et il sera libéré lorsque l'un des deux coupera sa communication.

Par contre, si pendant un certain temps les deux entités ne s'échangent rien le circuit leur reste quand même attribué. C'est pourquoi, un même circuit (ou portion de circuit) pourra être attribué à plusieurs communications en même temps. Cela améliore le fonctionnement global du réseau mais pose des problèmes de gestion (files d'attente, mémorisation,...).

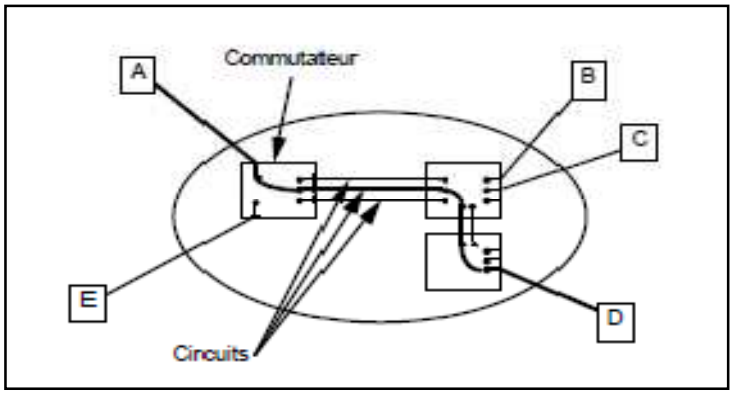


***Figure3.1.*** *Principe de la commutation de circuits*

## 3.3.2. Commutation de message

Elle consiste à envoyer un message de l'émetteur jusqu'au récepteur en passant de nœud de commutation en nœud de commutation. Chaque nœud attend d'avoir reçu complètement le message avant de le réexpédier au nœud suivant.

Cette technique nécessite de prévoir de grandes zones tampon dans chaque nœud du réseau, mais comme ces zones ne sont pas illimitées il faut aussi prévoir un contrôle de flux des messages pour éviter la saturation du réseau. Dans cette approche il devient très difficile de transmettre de longs messages. En effet, comme un message doit être reçu entièrement à chaque étape si la ligne a un taux d'erreur de 10-5 par bit (1 bit sur 105 est erroné) alors un message de 100000 octets n'a qu'une probabilité de 0,0003 d'être transmis sans erreur.

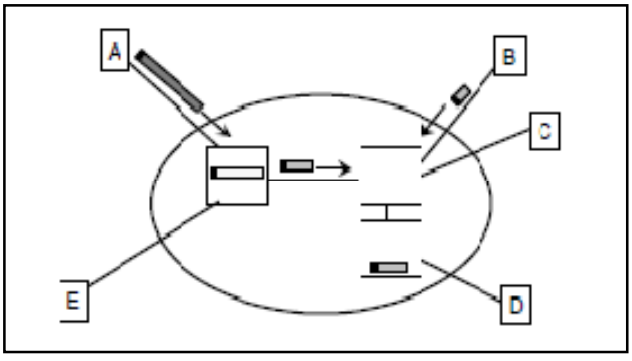


***Figure3.2.*** *Principe de la commutation de messages*

### 3.3.3. Commutation par paquets

Elle est apparue au début des années 70 pour résoudre les problèmes d'erreur de la commutation de messages. Un message émis est découpé en paquets et par la suite chaque paquet est commuté à travers le réseau comme dans le cas des messages. Les paquets sont envoyés indépendamment les uns des autres et sur une même liaison on pourra trouver les uns derrière les autres des paquets appartenant à différents messages.

Chaque nœud redirige chaque paquet vers la bonne liaison grâce à une table de routage. La reprise sur erreur est donc ici plus simple que dans la commutation de messages, par contre le récepteur final doit être capable de reconstituer le message émis en rassemblant les paquets. Ceci nécessitera un protocole particulier car les paquets peuvent ne pas arriver dans l'ordre initial, soit parce qu'ils ont emprunté des routes différentes, soit parce que l'un d'eux a du être réémis suite à une erreur de transmission **[8].**

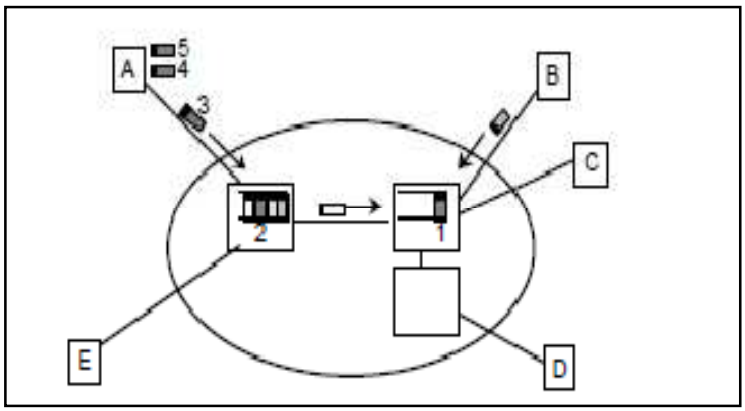


***Figure3.3.*** *Principe de la commutation par paquets*

## 3.4. Architecture du réseau

Au début des années 70, chaque constructeur a développé sa propre solution réseau autour d'architecture et de protocole privés et il s'est vite avéré qu'il serait impossible d'interconnecter ces différents réseaux si une norme internationale n'était pas établie.

Cette norme établie par l'internationale standard organisation (ISO) est la norme open system interconnexion (OSI, interconnexion de systèmes ouverts).

Un système ouvert est un ordinateur, un terminal, un réseau, n'importe quel équipement respectant cette norme et donc apte à échanger des l'information avec d'autres équipement hétérogènes et issus de constructeurs différents. La première objectif de la norme OSI a été de

définir un modèle de toute architecture de réseau base sur découpage en sept couches chacun de ces couches correspondant à une fonctionnalité particulière d'un réseau.

### 3.4.1. Rôle des différentes couches (OSI)

L'organisation internationale OSI a normalisé un modèle de sept couches. Ces couches se répartissent en deux classes :

- Les couches 1 à 4 dites couches basses, sont chargées de la mise en œuvre des fonctions de transport de l'information. Elles assurent le contrôler des erreurs de bonne réception, le contrôle de flux, etc.

-Les couches de 5 à 7 dites couches hautes, assurent les fonctions de traitement de l'information : la synchronisation des échanges, la représentation de l'information, etc.

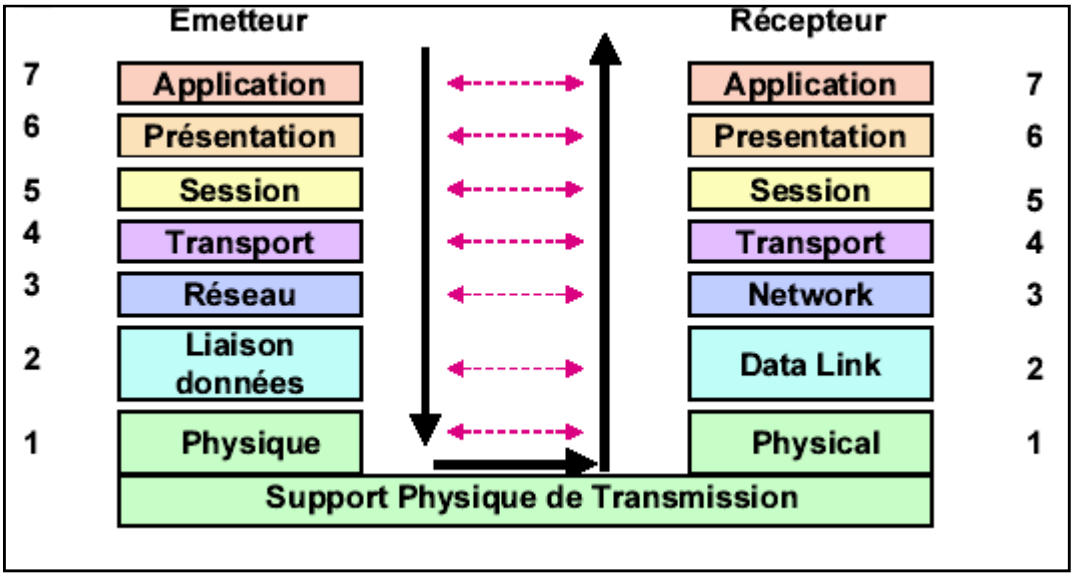


***Figure3.4.*** *Le model OSI en détail.*

- **La couche physique**

Cette couche définit les caractéristiques techniques, électriques, fonctionnelles et procédure les nécessaires à l'activation et à la désactivation des connexions physiques destinées à la transmission de bits entres deux entités de la couche liaisons de données.

- **La couche liaison**

Cette couche définit les moyens fonctionnels et procéduraux nécessaires à l'activation et à l'établissement ainsi qu'au maintien et à la libération des connexions de liaisons de donnes entre les entités du réseau.

Cette couche détecte et corrige, quand cela est possible, les erreurs de la coche physique et signale à la couche réseau les erreurs irrécupérables.

- **La couche réseau**

Cette couche assure toutes les fonctionnalités de services entre les entités du réseau, c'est à dire : l'adressage, le routage, le contrôle de flux, la détection et la correction d'erreurs non résolues par la couche liaison pour préparer le travail de la couche transport.

- **La couche transport**

Cette douche définit un transfert de données entre les entités en les déchargeant des détails d'exécution (contrôle entre l'OSI et le support de transmission).

Son rôle est d'optimiser l'utilisation des services de réseau disponibles afin d'assurer à moindre coût les performances requise par la couche session.

- **La couche session**

Cette couche fournit aux entités de la couche présentation les moyens d'organiser et de synchroniser les dialogues et les échanges de données.

Il s'agit de la gestion d'accès, de sécurité et d'identification des services.

- **La couche présentation**

La couche présentation s'occupe de la syntaxe et de la sémantique de l'information transmise, pour que les machines hétérogènes reliées par un même réseau utilisent le même langage et pour que la décompression des données soit dans une même norme agréée.

- **La couche application**

Cette couche assure aux processus d'application le moyen d'accès à l'environnement OSI et fournit tout les services directement utilisables par l'application (transfert e données, allocation de ressources, intégrité et cohérence des informations, synchronisation des applications).

## 3.5. Mise en paquet de la parole (protocole G.764)

Avant la mise en paquet de la parole, les échantillons de la parole peuvent être codés à l'extrémité d'origine; côte émission par l'une des méthodes existantes G.711, G721, G726 ou autres, si possible avec l'élimination des intervalles de silence. Les échantillons SONT recueillis sur une période de 16 ms et divisés en blocs de 128 bits. Les blocs sont organisés de manière à faciliter la suppression des blocs en cas de congestion du réseau. Le protocole G764 propose deux types de trames UIH et UI. Le paquet UI est utilisé pour le transport de la signalisation canal par canal. Il ne comporte pas de numéro de séquence et peut donc être perdu sans modification. Par

contre le paquet UIH est utilisé pour le transport de la parole décodée. Sa différence avec le paquet UI et que la séquence de contrôle et obtenue sur l'en-tête de la trame (les huit premières octets à l'exclusion du fanion) et non sur la trame totale.

En éclaire le données de parole ne sont pas protégées car l'information est plus sensible aux délais imposer par des erreurs de transmission bit.

### 3.5.1. Description des différents champs du paquet UIH

Le paquet UIH représenté par la figure 3.5, est utilisé pour le transport des données de laparole **[9].** Il est formé par un certain nombre de champs qui seront explicités comme suit.

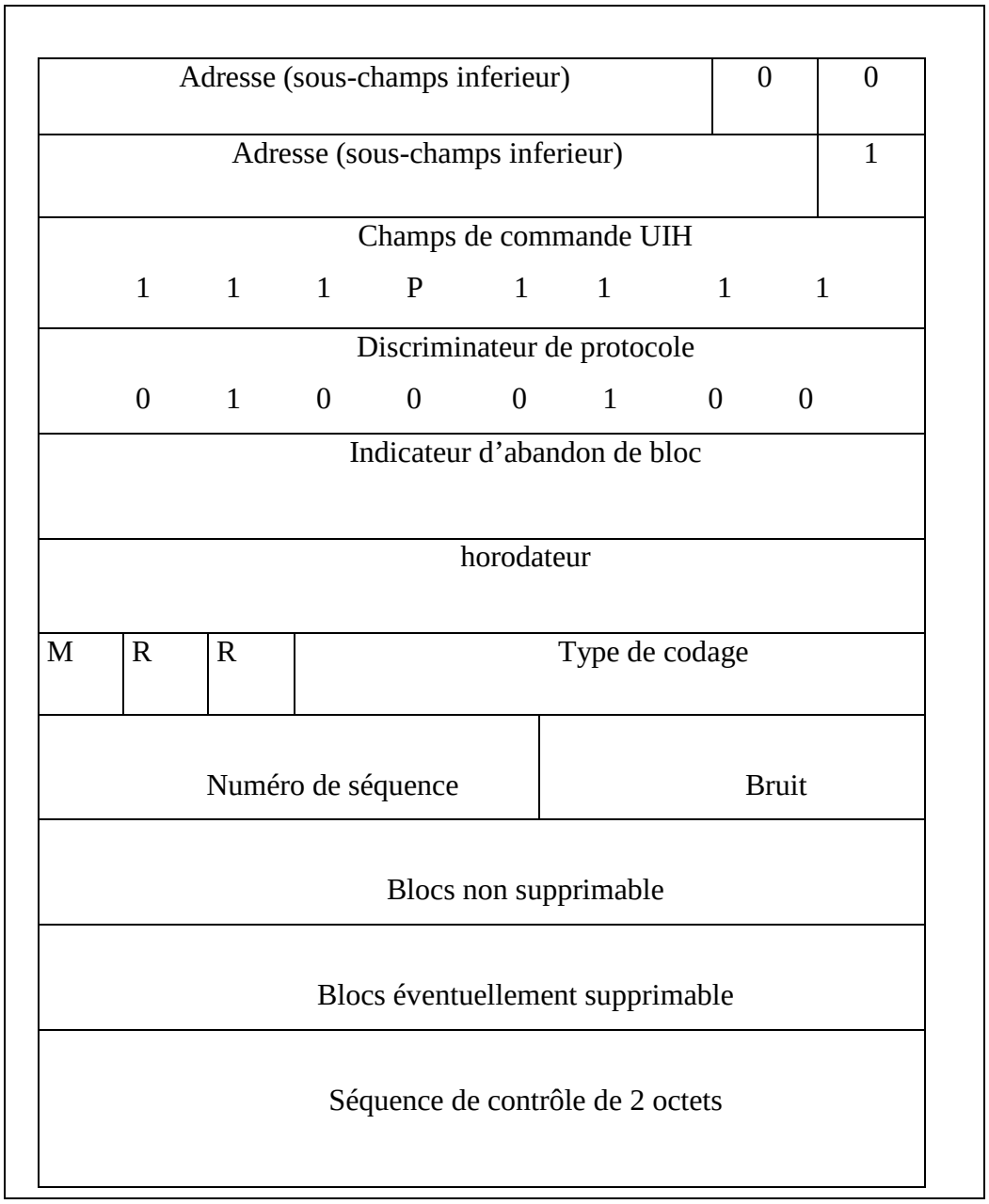


***Figure3.5.****Format de la trame UIH*

### 3.5.1.1. Discriminateur de protocole

Le champ de discriminateur de protocole (PD) et le premier octet de l'en-tête du paquet (octet 4 de la trame), il indique d'une manière unique le type de service transporté dans la trame UIH. Sa valeur est indiquée sur la figure 3.5.

### 3.5.1.2 Indicateur d'abondant

L'indicateur d'abondant de bloc (BDI) recherche l'état de suppression de blocs dans le paquet de parole. Un bloc est formé de bits de même poids recueillis dans tous les échantillons de parole mis en paquet **[9].**

Un bloc contient 128 bits, ce qui correspond à un intervalle de mise en paquet de 16 ms pour une fréquence d'échantillonnage de 8 kHz. Les blocs formés sont rangés décroissant de poids (figure 3.6).

| Numéro de bit | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| R | R | 2 | 1 | R | R | 2 | 1 |

***Figure 3.6.****Format de* l'indicateur *de suppression de bloc(BDI)*

C1 et C2 représentent le sous-champ C qui indique le nombre de blocs que l'on peut encore supprimer à tout nœud intermédiaire du réseau. La valeur de C varie de zéro à trois, cette valeur est modifiée par rapport à la valeur initiale, elle diminue à chaque fois que des blocs sont supprimés du paquet pour indiquer le nombre de blocs qu'il reste encore à supprimer.

**Tableau1**

| C1 | C2 | Nombre de blocs supprimable |
|---|---|---|
| 0 | 0 | 0 blocs |
| 0 | 1 | 1 bloc |
| 1 | 0 | 2 blocs |
| 1 | 1 | 3 blocs |

M1 et M2 forment le sous champ M, qui indique le nombre total de blocs à supprimer du paquet de parole, en période d'encombrement quand il traverse tout le réseau. La valeur du sous-champ M varie de zéro à trois, cette valeur n'est pas modifiée par rapport à sa valeur initiale.

**Tableau2**

| M1 | M2 | Nombre de blocs supprimable |
|---|---|---|
| 0 | 0 | 0 blocs |
| 0 | 1 | 1 bloc |
| 1 | 0 | 2 blocs |
| 1 | 1 | 3 blocs |

Si les échantillons de parole sont codés avec un débit fixe à l'extrémité d'origine, les sous-champs M et C sont mis à zéro.

### 3.5.1.3. L'horodateur

L'horodateur (TS) est codé sur huit bits, il enregistre le retard variable cumulatif dans les files d'attente rencontrées par un paquet traversant le réseau avec une résolution de 1ms. La valeur maximale valable dans le champ (TS) ne doit pas dépasser 20 ms. Si après la mise à jour, le retard variable dépasse 200 ms, sa valeur est fixée à 200 ms.

### 3.5.1.4. Type de codage

Le champ de type de codage (CT) indique la technique de codage employée pour coder les échantillons de parole au point extrémité d'origine avant la mise à jour en paquet, ce champ est codé sur 5 bits ce qui permet d'utiliser 32 type de codage.

Le bit M est mis à 1 pour tous les paquets, à l'exception du dernier du dernier paquet d'une salve ou il est mis à 0.

### 3.5.1.5 Numéro de séquence et champ de bruit

Le numéro de la séquence (SEQ**)** est employé par les points extrémités dans le processus de formation de paquet pour déterminer le premier paquet d'une salve de parole et aussi détecter la perte de paquet. Le SEQ, une fois associé à l'horodateur(TS), permet la suppression de la variable de retard dans les réseaux.

La numérotation se fait modulo 16 pour un cycle de 1 à 15 pour les paquets suivants le premier paquet d'une salve de parole.

Le champ de bruit indique le niveau de bruit de fond présent à l'extrémité d'origine et qui est utilisé par l'extrémité de terminaison pour déterminer le niveau de bruit de fond qui peut être régénéré en cas d'absence de paquet de parole.

### 3.5.1.6. Champ d'information

Le champ d'information contient des blocs de 128 bits chacun, ces blocs sont organisés selon les poids des bits, le premier bloc contient les bits des poids forts (MSB) de tous les échantillons, le second les bits MSB suivants, etc.…. Les bits d'un bloc sont ordonnés selon leurs numéros d'échantillons, la figure 6 représente le format des bits avant leur mise en paquets.

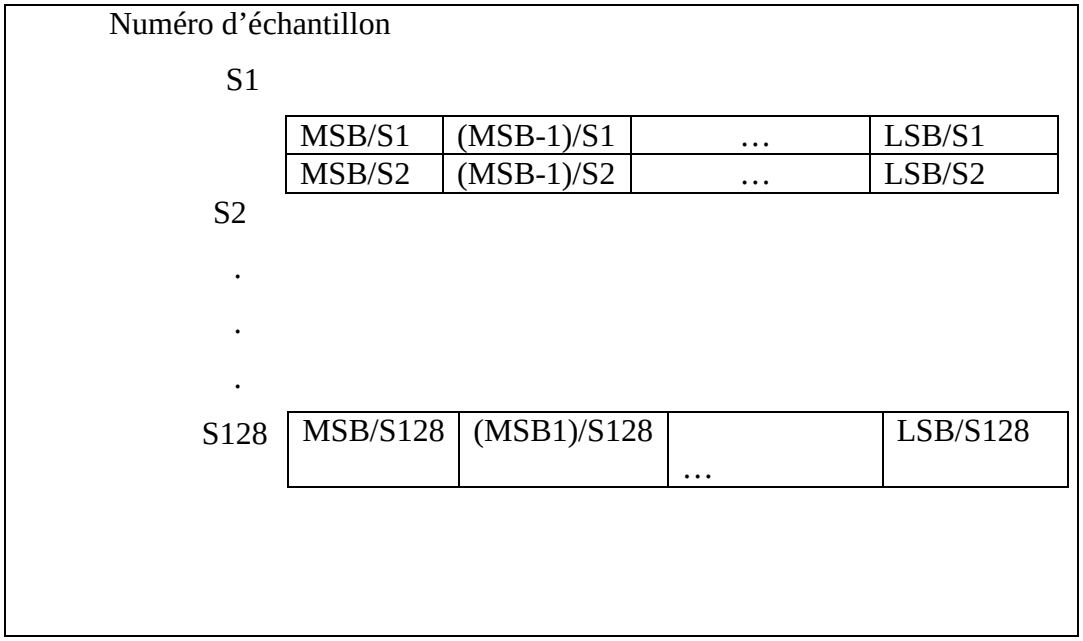


***Figure 3.7.*** *Format des bits avant leur mise en paquet*

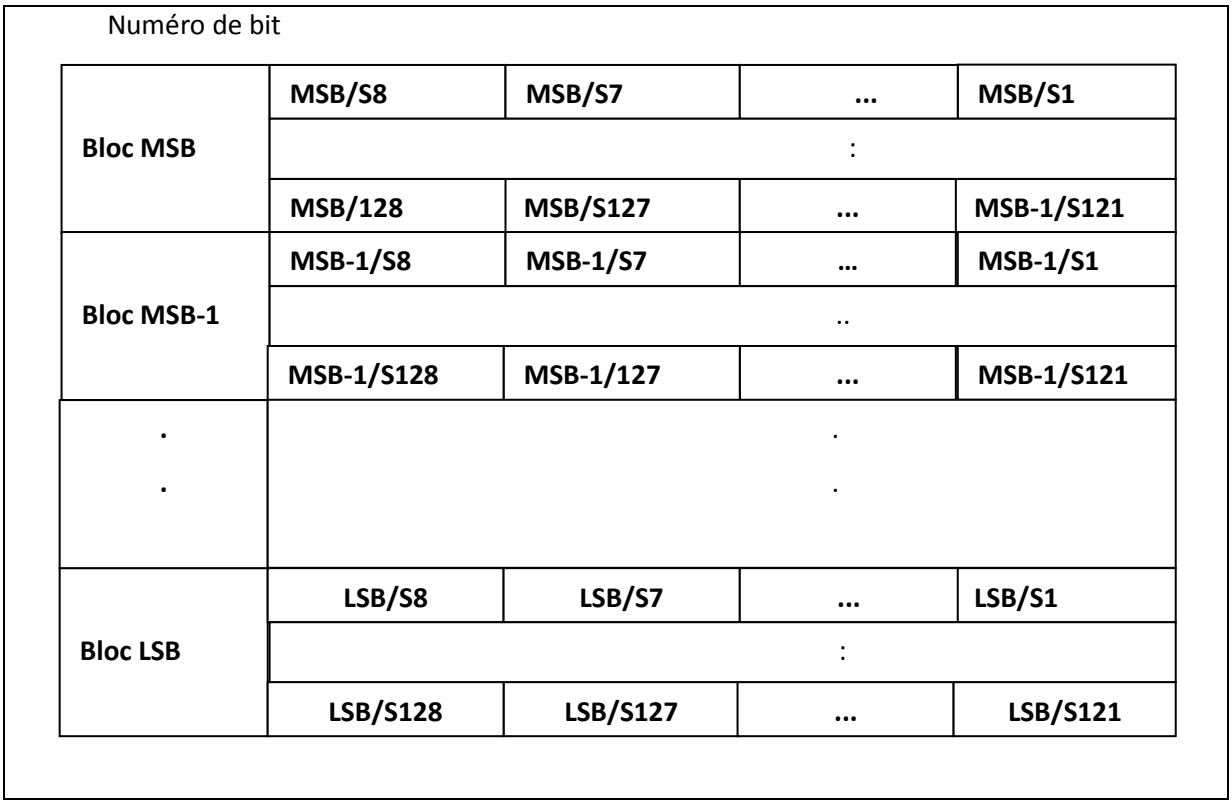


***Figure 3.8.*** *Champ d'information des paquets de parole*

## 3.6. Conclusion

La mise en paquet de la parole permet de partager un canal de transmission entre les données et la parole. Elle se base sur la norme du CCITT (recommandation G.764), qui est aussi la base pour le reliage des trames et la commutation des paquets pour le X.25. Un paquet de parole comprend plusieurs champs. Dont le champ d'information représente de la parole codée. Pour les codeurs hybrides, le champ d'information contient les paramètres qui servent pour la synthèse de la parole. Généralement ces paramètres sont les coefficients du filtre de synthèse, la période du fondamental, son gain et enfin le numéro de la séquence d'excitation.

# *Chapitre 04*

# *Etude de la qualité de la parole codée par le codeur CELP en fonction des erreurs de transmission*

## 4.1. Introduction

Le codeur CELP utilisé au laboratoire est le FS-1016, pour pouvoir l'utliliser comme un outil pou nos expériences, on a voulu connaître ses caractéristiques : les commandes avec lesquelles on peut le lancer, les fichiers qu'il utilise en lecteur et en écriture, le format de ces fichiers ainsi que le problème de son adaptation pour le pont audio de notre machine, pour avoir une idée sur la qualité de la parole codée et décodée. Le but de ce chapitre et d'exposer l'enssemble des travaux qui ont été effectués pour connaître les performances du codeur et pour résoudre les problèmes exposés au début de ce chapitre.

## 4.2. Le codeur CELP FS-1016

Le codeur CELP FS-1016 a été choisé par le ministère de la défense américain pour le remplacement du vocodeur LPC à 2400 bits/s et il proposé comme standard pour l'OTAN ; ses principales caractéristiques sont :

- Une trame de 240 échantillons prélevés à 8 kHz et divisée en quatre sous trame de 60 échantillons chacune
- Une prédicteur à court-terme d'ordre 10 dont les coefficients sont calculés une fois chaque trame et quantifiés d'une manière scalaire.
- La méthode utilisée pour le calcul est l'autocorrélation avec une fenêtre de Hamming.
- Les dictionnaires utilisés sont deux dictionnaires, l'un est adaptatif et contient 250 vecteurs, pour une analyse en boucle fermée et l'autre est un dictionnaire stochastique contenant 512 vecteurs gaussiens, dont 75% des éléments sont nuls.

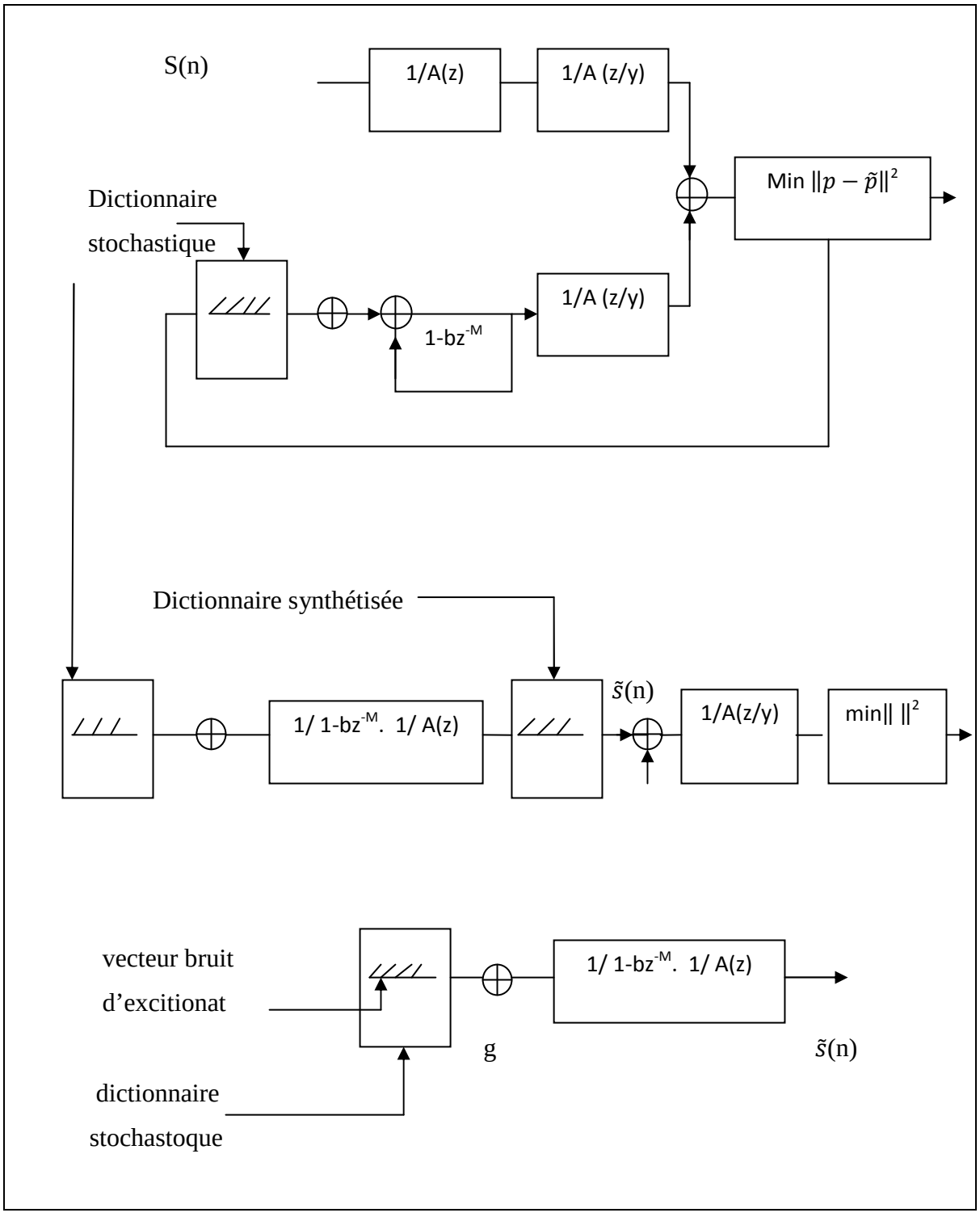


***Figure 4.1.*** *Principe du CELP*

## 4.3. Codage et décodage des échantillons de parole par le CELP

Le codeur CELP utilise en entrée un fichier de parole binaire de format 16 bits. Ce fichier doit porter l'extension (.spd) c'est à dire un nom suivit par .spd. La commande pour lancer le programme CELP est celp –i [nom du fichier à lire en entrée avec l'extension .spd] -0 [nom du fichier en sortie sans l'extension .spd]. A la fin de l'exécution du programme en aura deux fichiers en sortie, un fichier de paramètres qui contient les coefficients du filtre court-terme, l'index de la séquence d'excitation, son gain, le délai et son gain, tous quantifiés, ce fichier porte l'extension ( .chan ) et un fichier des échantillons de parole écrit en binnaire, de format 16 bits qui porte l'extension ( npf.spd ). On peut donner un exemple d'utilisation du codeur.

- Le fichier de lecteur ou d'entée a pour nom parole1 écrit en binnaire avec un format 16 bits. Pour qu'il soit lu par le codeur il faut qu'il porte l'extension (.spd). Alors le nom du fichier devient parole1.spd.

-Le fichier d'écriture ou de sortie aura pour nom parole1_out.

- Le lancement du CELP se fait de la manière suivante : celp –i parole1.spd -o parole1_out. En sortie on aura deux fichiers, le fichier de paramètres qui aura pour nom parole_out.chan et le fichier de parole qui aura pour nom parole1_outnpf.spd.

Le problème consiste alors à transformer les formats des fichiers de parole en format CELP et transformer ensuite le format CELP du fichier de parole fourni par le codeur après décodage, en format audio pour écouter les échantillons de parole décodés et avoir une idée sur la qualité de décodage. Pour résoudre ce problème, on a réalisé deux procédures. La première transforme le format de n'importe quel fichier de parole en format CELP. La seconde transforme le format CELP en fotrmat audio.

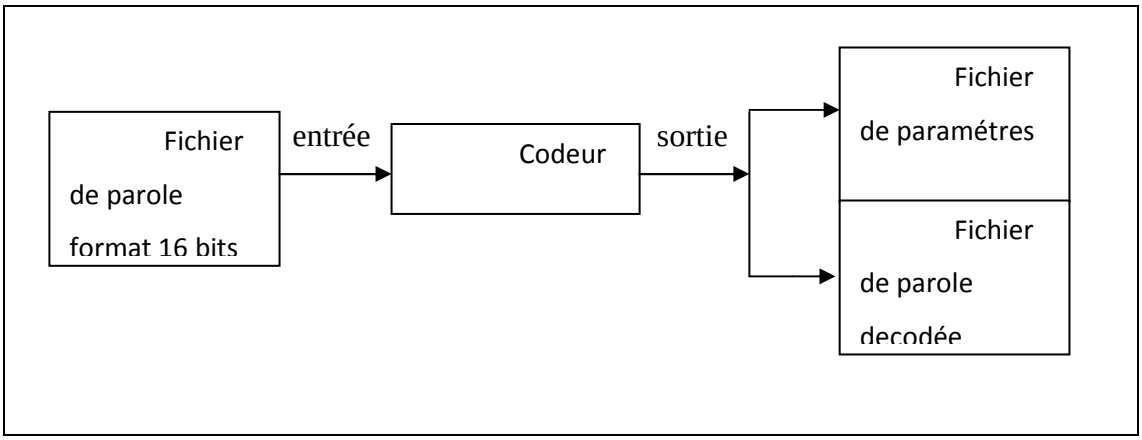


***Figure 4.2.*** *Principe d'utilisation du codeur CELP*

### 4.3.1. Procédure d'acquisition des échantillons de parole

La procédure d'acquisition des échantillons de parole s'appelle record2. Elle permet à travers le pont audio à un locuteur de parler et d'aquérir des échantillons de parole par l'intermédiare d'un micro. Les échantillons acquis sont formatés en format audio 8 bits. La procédure record2 transforme le format audio 8 bits en format CELP 16 bits. Elle permet aussi transformer le format d'un fichier de parole quelconque en format CELP. Une fois ces échantillons enregistrés et transformés, ils sont mis dans un fichier à part.

### 4.3.2. Procédure de réception des échantillons de parole

Les échantillons acquis et mis dans un fihier de parole au format 16 bits sont envoyés au CELP. Une fois codés et décodés, ils sont ensuite mis dans un autre fichier de même format. Pour connaître l'efficacité du codage, on avait besoin d'écouter le signal décodé. Pour cette raison on a réalisé une procédure, qui permet d'écouter des fichiers de parole. Cette procédure s'appelle play2. Elle permet de convertir le format CELP en format audio 8 bits et d'envoyer le fichier ainsi trouvé sur le pont audio et enfin 1 écouter. Elle permet aussi d'écouter des fichiers de parole quelconque en transformant leurs formats en format audio.

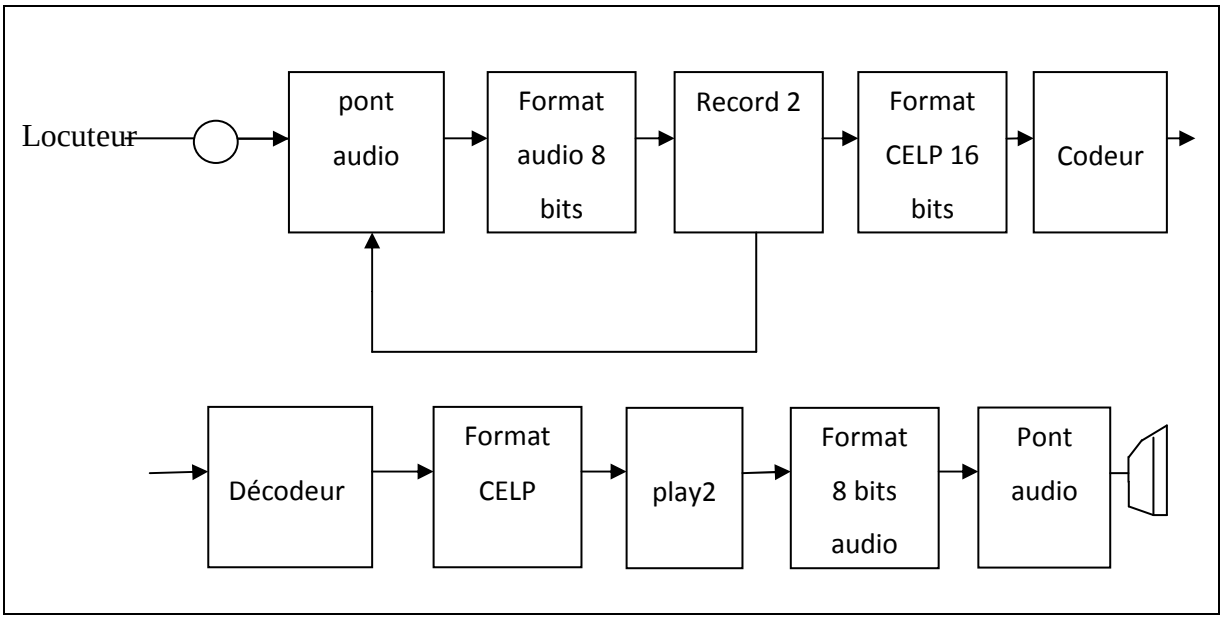


***Figure 4.3.*** *Adaptation du pont au codeur CELP*

## 4.4. Séparation de l'algorithme du FS-1016

Dans le codeur CELP le décodeur est inclus dans le programme principal. Il faut alors le séparer du codeur.

Cette séparation permet d'obtenir deux parties :

Une partie codage ou partie analyse, qui permet de produire un fichier de paramètres écrit en Ascii. Ce fichier contient les coefficients du filtre à court-terme, la valeur du délai, son gain, l'index de la séquence d'excitation et le gain par le quel doivent être multipliés les vecteurs de cette séquence d'excitation.

Ces paramètres sont codés comme suit :

- L'index est codé sur 9 bits, chaque sous-trame.
- Le gain est codé sur 5 bits, chaque sous-trame.
- Le délai est une fois codé sur 8 bits pour une sous-trame et une fois sur 6 bits pour la sous-trame suivante et ainsi de suite.
- Les coefficients du filtre sont codés chaque trame [3 4 4 4 4 3 3 3 3 3].
-

Le programme exécutable du codeur est nomé celp2. La commande de son lancement se fait de la même manière que celle du programme celp.

Une partie décodage ou partie la synthèse qui permet d'utiliser le fichier de paramètres pour la prédiction de la parole. Le programme exécutable du décodeur est appelé celp1. Pour le lancer, on utilise la commande suivante : celp1 –c 3m3f_out.chan –o 1am. A la fin de son exécution on aura un fichier de parole de format 16 bit écrit en binnaire, qui a pour nom lamnsf.spd.

Pour que le fichier de paramètre soit lu par celp1, il doit toujours porter l'extension (.chan).

3m3f_out.chan : représente un exemple de fichier de lecteur utilisé par celp1.

Lam : représente un exemple de fichier écriture utilisé par celp1.

Lamnpf.spd : représente lefichier lam au quel celp1 ajoute l'extension npf.spdaprés avoir mis dedans les échantillons de parole décodés.

Le but de cette séparation est de préparer le celp pour une éventuelle utilisation pour la transmission de la parole dans un réseau et pour permettre aussi d'estimer ses performances en fonction des erreurs de transmission.

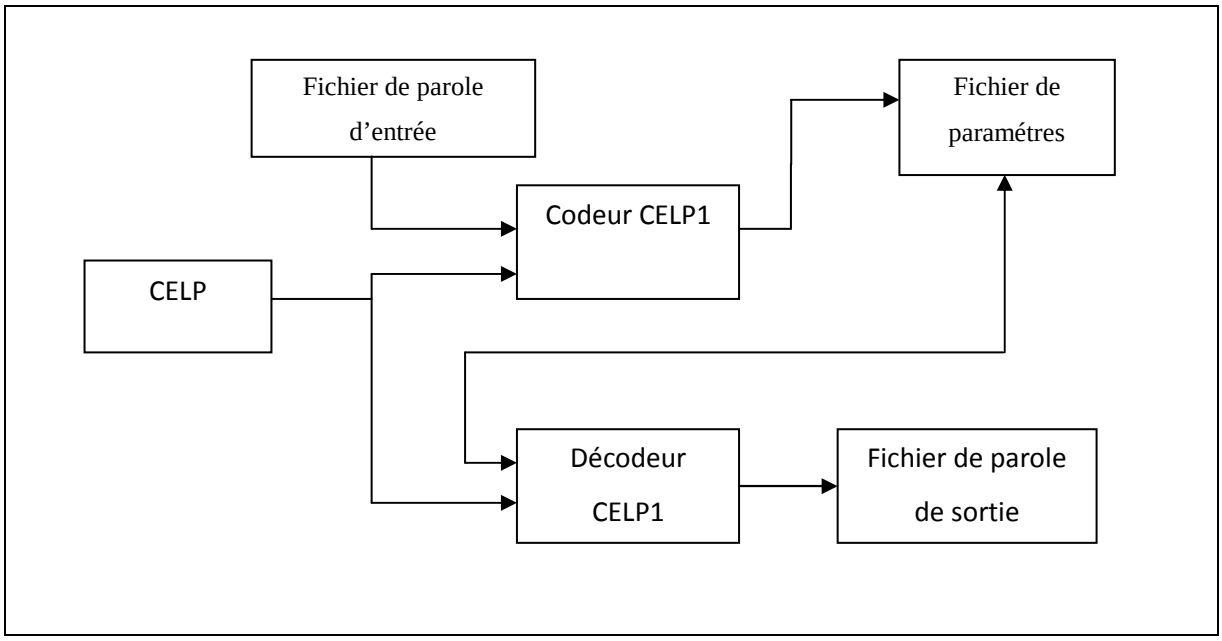


***Figure 4.4.*** *Séparation du codeur CELP en partie codage et partie décodage*

## 4.5. Mesure du rapport signal à btuit

La mesure de la qualité la plus couramment utilisée pour les codeurs qui essayent le préserver la forme du signal reste le rapport signal à bruit.

S : signal de la parole original.

$\tilde{S}$ : signal de la parole synthétisé.

Pour un enregistrement de N échantillons, on définit l'energie du signal par

$$E_s = \sum_{n=0}^{N-1} S \tag{4.1}$$

L'energie de l'erreur est donnée par

$$E_e = \sum_{n=0}^{N-1} e_n^2 = \sum_{n=0}^{N-1} (S_n^2 - \tilde{S}_n^2) \tag{4.2}$$

Le rapport signal à bruit est alors donné par

$$RSB = 10 \log \frac{E_s}{E_e} \tag{4.3}$$

Le signal de parole est non-stationnaire, certains segments du signal peuvent avoir une énergie plus ou moins grande en supposant que l'énergie de l'erreur soit à peu près constante ; le rapport signal à bruit peut être très important comme très faible.

On utilise de fait plutôt le rapport signal à bruit segmental. Le signal est découpé en L segment de 15 à 20ms et on calcule une moyenne de RSB de deux façons.

$$\text{RSB 1} = \frac{1}{L}\sum_{1}^{L} 10\log\frac{\sum_{n=0}^{N-1} S^2(n)}{\sum_{n=0}^{N-1} e^2(n)} \tag{4.4}$$

$$\text{RSB 2} = 10\log\frac{1}{L}\sum_{1}^{L}\frac{\sum_{n=0}^{N-1} S^2(n)}{\sum_{n=0}^{N-1} e^2(n)} \tag{4.5}$$

## 4.5.1. Calcul du rapport signal à bruit du signal codé par le CELP

Le calcul des coefficients se fait selon la norme définie dans le standard fédéral. Elle introduit un retard supplémentaire de 15 ms si l'on code les sous-trame de la trame$t_n$, le calcul se fait avec les sous-trames de la trame $t_n$ et les sous-trames de la trame$t_{n+1}$, c'est à dire qu'il prend les deux dernières sous-trames de la trame $t_n$et les deux premières sous-trames de la trame $t_{n+1}$ et ainsi de suite (figure 4.5).

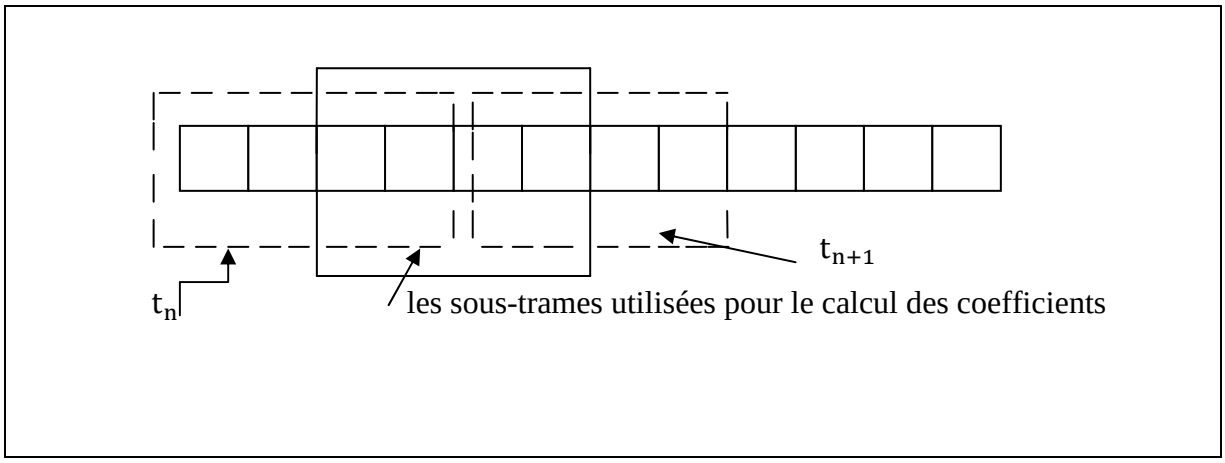


***Figure 4.5.*** *Principe du calcul des coefficients à court – terme*

Dans l'algorithme, le codage se fait de la manière suivante. On prend deux sous-trames de valeur nulle et on les combine avec les deux premières sous-trames de la trame$t_1$, qui est formée par les 240 premiers échantillons. Ensuite on prend les deux dernières sous-trames de la trame précédente avec les deux premières sous-trames de la trame $t_2$ et ainsi de suite.

D'aprés ce procédé à la fin du décodage on aura deus sous-trames perdues et cela parce que la lecture des échantillons de parole se fait par blocs de 240 échantillons.

### 4.5.1.1. Préaccentuation des valeurs avant le codage

Le CELP ne code pas directement les échantillons acquis, il leurs fait subir tout d'abord une préaccentuation, ensuite il les code et le calcul du rapport signal à bruit se fait entre les valeurs préaccentuées et les valeurs décodées. Pour récupérer ces valeurs, afin de les utiliser pour le calculde rapport signal à bruit, on a ouvert un fichier en écriture à l'intérieure du programme principal.

### 4.5.1.2. Calcul du rapport signal à bruit d'une sinusoïde

Pour avoir une preuve visuelle sur la qualité de codage et du décodage, on a appliqué sinusoïde. La reconstitution était quasi-parfaite. Son rapport signal à bruit segmental moyen est de 20.22dB. D'après la figure on voit bien le retard introduit par le CELP, un retard de 15 ms qui s'ajoute au retard de transmission. La sinusoïde utilisée a une période de 7.5 ms, qui correspond à 60 échantillons, c'est à dire égale a la taille d'un vecteur issu du dictionnaire adaptatif.

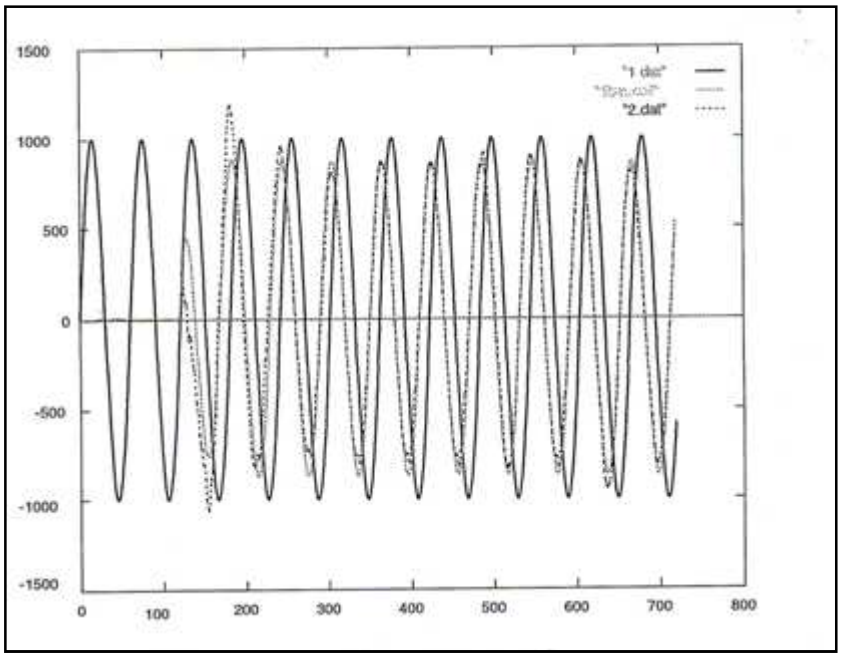


**Figure 4.6.** a: _____Schéma d'une sinusoide

b: …….Schéma de la sinusoïdepréaccentuée.

*c*: ……*Schéma de la sinusoïde après décodage*

### 4.5.1.3. Calcul du raopport signal à bruit pour un signal de parole

Pour évaluer la qualité du codage, le CELP était testé sur un ensemble de phrases et on a trouvé les résultats des rapports signal à bruit suivants.

**Tableau 1**

| Locuteurs | Sexe | Phrases | RSB segmental en dB |
|---|---|---|---|
| 1 | M | 34 à 42 | 9.77 |
| 2 | M | 43 à 56 | 11.34 |
| 1 et 2 | M | 34 à 51 | 10.54 |

Dans ces tests les locuteurs 1 et 2 sont différents.Pour essayer de trouver les mêmes résultats que ceux figurant dans le tableau 1, alors on a effectué un enregistrement avec la voix d'un locuteur sur station de travail avec les phrases 34 à 42 et les phrases 34 à 51 et un enregistrement avec la voix d'un autre locuteur avec les phrases 34 à 56. Après codage et décodage de ces phrases, on a trouvé les résultats des RSB suivant.

**Tableau 2**

| Locuteurs | Sexe | Phrases | RSB segmental en dB |
|---|---|---|---|
| 1 | M | 34 à 42 | 9.7 |
| 2 | M | 43 à 56 | 9.82 |
| 1 | M | 34 à 51 | 9.66 |

### 4.5.1.4. Interprétation

Les phrases décodées une première fois contiennent les caractéristiques du codeur CELP. Lorsqu'elles sont codées et décodées leursqualités se dégrade. Cette dégradation est duà la perte d'information sur les échantillons de la parole. Lorsque un signal de parole est codé et décodé une

première fois, il perd un peut de son intelligibilité, ce qui provoque une dégradation dans la qualité de la parole.

### 4.5.2. RSB et qualité de parole en fonction des erreurs de transmission

Lors de la transmission de la parole par paquetd'une extrémité à une autre, certains paquets seront perdus et n'arrivent pas au récepteur auquel ils étaient destinés. D'autres paquets en provenance de l'autre extrémité, peuvent arriver à ce même récepteur, suite à des erreurs en bits dans leurs champs d'adresse, c'est le phénomène d'insertion à tort des paquets. A-delà de certains taux de perte ou d'insertion de paquets, la qualité de la parole décodée est dégradée.Dans le cas de perte de paquets, la qualité de la parole peut être préservée en remplaçant les paquets perdus :

- Soit par la retransmission du paquet perdu.
- Soit par le remplacement du paquet perdu par le précèdent.
- Soit par le remplacement du paquet perdu par un bruit de fond.

Dans le cas d'insertion à tort des paquets, la qualité peut être protégée en éliminant les faux paquets.
Le codeur CELP FS-1016 fournit une trame de paramètres toute les 30 ms. Cette trame forme le champ d'information du paquet de parole à transmettre.
La perte des paquets dans le cas du codeur FS-1016 provoquent la perte des trames. Par contre l'insertion à tort des paquets provoque une insertion des fausses trames.

L'etude qui a été effectuée dans cette partie a deux objectifs, le premier consiste à estimer les taux de perte et d'insertion, au-delà desquels la qualité de la parole est dégradée. Le second consiste à proposer des solutions qui permettent de protéger la qualité de la parole contre ces phénomènes.

Pour arriver à ce but plusieurs simulations ont été réalisées :

Dans la première, on supprime des trames du fichier de paramètres. Ce procédé simule la perte des trames due à la perte de paquets. Le fichier de paramètres obtenu est appliqué au décodeur celp1. Aprés décodage la valeur de l'energie du signal fournit par celp1 est comparée à celle du signal décodé sans perte.

Dans la deuxième, on remplace les trames suprimées du fichier de paramètres par les précédentes. Cette duplication simule le remplacement des trames perdues dû aux pertes de paquet de parole. Après décodage du fichier de paramètres, on calcule le rapport signal à bruit.

Dans la troisième, on procède de la même manière que dans la première simulation sauf que dans cette dernière, au lieu de supprimer des trames du fichier de paramètres, on insert des fausse trames.

Dans la quatrième, on procède de la même manière que dans la deuxième simulation, sauf qu'ici la trame perdue est remplacée par une fausse trame.

Dans la cinquième, on introduit des erreurs en bits dans le fichier de paramètres ; ensuite ce fichier est appliqué au décodeur. Après décodage on calcule le rapport signal à bruit.

Ces simulation on été effectuée sur un fichier de 20000 trames. Ce nombre était choisi pour optimiser l'espace sur le disque, ainsi que le temps de calcul. Il permet aussi d'atteindre un taux d'erreurs de l'ordre $10^{-5}$.

### 4.5.2.1. Mesure de la qualité de la parole en fonction de perte des trames sans remplacement

La perte des trames ou des paquets de parole est due à deux causes principales ; des erreurs sur le champ d'adresses et la saturation des files d'attentes. Le champ d'adresses contient une adresse de mise sur réseau et une adresse de destination. A chaque noeud de réseau le champ d'adresses est examiné, et toute erreur non corrigible entraîne l'élimination du paquet de parole.

La perte des paquets est due aussi à la saturation des files d'attente. Lors de son cheminement sur le réseau, le paquet de parole rencontre des files d'attente à chaque noeud. Une de ces files d'attente peut être complètement remplie par d'autres paquets en provenance d'autres sources et qui utilise le même chemin passant par ce noued. Alors le paquet sera perdu, c'est le phénomène de congestion.

Cette perte de paquets povoque des vides dans les séquences de parole, ce qui engendre une dégradation dans sa qualité.

Pour mesurer le taux de perte, au-delà duquel ces vides ou cette discontinuité dans la parole sera perceptible, on a procédé de la manière suivante : On élimine des trames dans le fichier de paramètre sans les remplacer, ensuite on applique ce dernier au décodeur celp1. Après décodage, le signal de parole obtenu subit un test d'écoute sur station de travail, afin de savoir s'il est compréhensible ou non.

Après cette étude on a remarqué que, la qualité de la parole demeure bonne jusq'à un taux de perte égale à $10^{-3}$. Au-delà de ce taux la qualité commence à se dégrader. Vu la nature de l'oreille humaine, cette dernière ne pas détecter le vide provoqué par une perte de signal de parole de l'ordre de 30 ms toutes les 30 secondes. Par contre, à partir d'une perte de l'ordre de 30 ms toutes les 15 secondes, ce vide ou cette discontinuité dans la parole deviennet perceptibles. C'est pourquoi à partir de ce taux le remplacement des trames perdues est indispensable.D'autre part on a voulu illustrer les résultats obtenus par une courbe. Après chaque test d'écoute du signal de parole décodé avec perte, on compare la valeur de son énergie à celle du signal de parole décodé sans perte. Notons que dans notre étude, la valeur de l'énergie du signal de parole décodé sas perte est considéré comme 100% de la qualité de la parole. Finalement on obtient une courbe de la qualité de la parole en fonction de taux de perte.

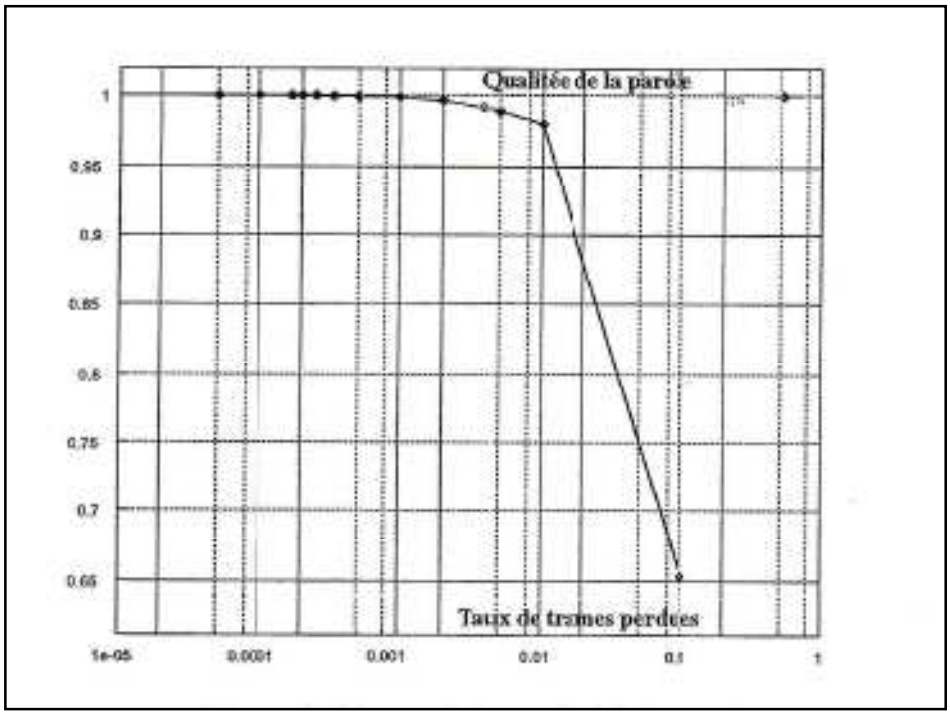


***Figure 4.7.*** *Mesure de la qualité de la parle en fonction des trames perdues*

## 4.5.2.2. Interprétation

Lorsque l'énergie d'un signal de parole obtenu après décodage est comprise entre 100% et 99% de sa valeur, sa qualité reste bonne. Ces deux valeurs correspondent respectivement au taux de perte 0 et $10^{-3}$. Par contre lorsque la valeur de l'énergie du signal de parole est comprise dans un intervalle inférieur à 99% de sa valeur, sa qualité est dégradée. Cela correspond à des taux de perte supérieurs ou égaux à $5.10^{-3}$.

### 4.5.2.3. Calcul du rapport signal à bruit : cas du remplacement des trames perdues

Dans l'étude précédente, on a remarqué qu'au-delà de certain taux de perte la qualité de la parole est dégradée et le remplacement des trames perdues est nécessaire. Plusieurs techniques peuvent être employées pour compenser les trames perdues. Permi eux on cite :

- Le remplacement des trames perdues par les précédentes.
- Le remplacement des trames perdues par des trames bruits.

Pour choisir le meilleur moyen pour le remplacement du paquet perdu. On a effectué une étude, qui consiste à mesurer le rapport signal à bruit pour le signal de parole en fonction de la duplication des trames perdues, par des fausses trames et par les trames précédentes. Pour réaliser cette étude, on a effectué deux simulations :

- La première consiste à perdre une trame de paramètre, ensuite la remplacer par la trame précédente.Après décodage et calcul du rapport signal sur bruit on a trouvé la courbe suivante

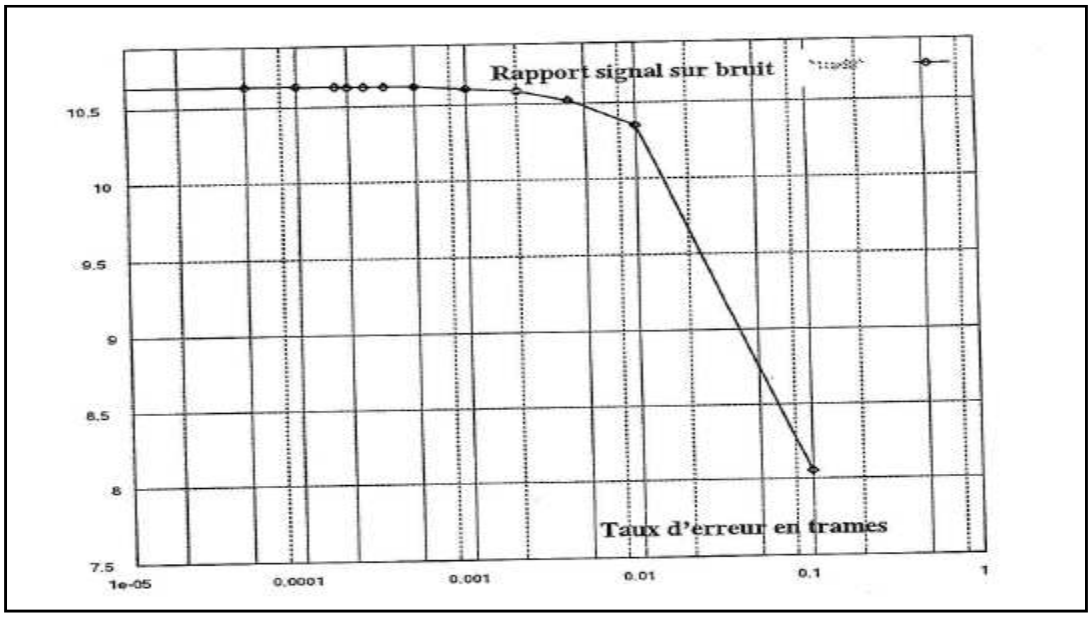


***Figure 4.8.*** *RSB cas du remplacement d'une trame perdue par la précédente*

- La deuxième consiste à remplacer une trame de paramètres par une fausse trame. Après décodage et calcul du rapport signal à bruit, on a trouvé la courbe suivante.

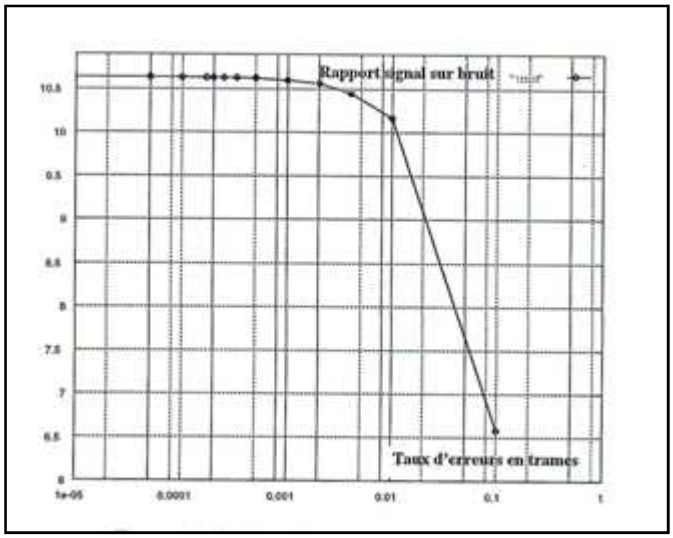


***Figure 4.9.*** *RSB cas du remplacement d'une trame perdue par une fausse trame*

### 4.5.2.4. Interprétation

Le signal de parole est quasi-périodique sur des intervalles de temps de l'ordre de 20 à 30 ms. Le fait de dupliquer une trame revient à remplacer une période par la précédente c'est pourquoi la variation du RSB est trés lente et la courbe est presque constante jusq'a un taux égal à $10^{-3}$. Ensuite elle commence à décroître et à partir de taux d'erreur $10^{-2}$, la descente est rapide ce qui nous permet de conclure qu'à partir d'un taux d'erreur de $10^{-2}$ la qualité est relativement dégradée, cela était vérifié par un test d'écoute.

Le RSB en fonction des fausses trames décroît plus rapidement que le RSB en fonction des trames dupliquées par les trames précédentes, car une fausse trame c'est du bruit ce qui engendre un mauvais décodage et par conséquent une qualité de parole dégradée.

### 4.5.2.5. Mesure de la qualité de la parole en fonction des trames insérées à tort

Lorsqu'il y a une erreur sur le champ d'adresse, l'adresse d'acheminement peut être modifée, et une trame qui n'est pas originalement destiné à un récepteur lui parvient tout de même. Au-delà d'un certain taux d'insertion la qualité de la parole est dégradée et la suppression des trames insérées est nécessaire.

Dans cette étude, on àessayé de mesurer les taux d'insertion au-delà desquels la qualité de la parole est dégradé. Pour arriver à cet objectif on a employé le même procédé que celui employé dans le paragraphe 4.5.2.1, sauf qu'ici au lieu de perdre des trames, on insert des fausses trames. Après cette étude, on a remarque que l'insertion des fausses trames dans des séquences de parole, peut être gênante pour la fidélité du signal de parole a partir d'un taux de l'ordre $3.3.10^{-4}$. Au-delà

de ce taux la qualité devient très dégradée. C'est pourquoi il faut prévoir un procédé d'élimination des fausses trames.

Pour illustrer ces observations par une courbe, on a utilisé le même raisonnement que celui utilisé dans le paragraphe 4.5.2.2, sauf qu'ici au lieu de calculer son énergie totale, on calcule sa valeur moyenne sur des segments de parole de 60 échantillons. Ce nombre correspond à la l durée de la sous-fenêtre d'analyse. La valeur obtenu est comparée à la valeur moyenne de l'énergie du signal calculé sans insertionde fausses trames. On a obtenu la courbe suivante :

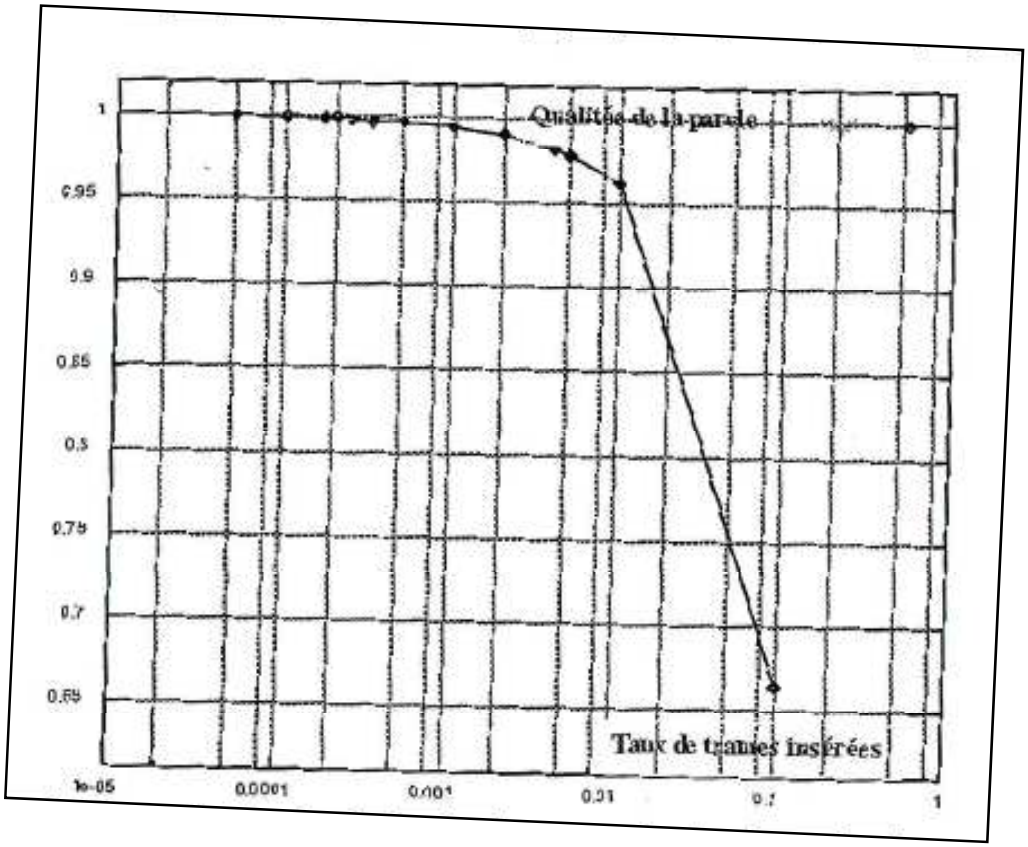


***Figure 4.10.****Mesure de la qualité de la parole en fonction des trames insérées à tort*

## 4.5.2.6. Interprétation

Lorsque énergie d'un signal de parole obtenu après décodage est compris entre 100% et 98% de sa valeur, sa qualité reste bonne. Ces deux valeurs correspondent respectivement au taux de perte 0 et $5.10^{-4}$. Par contre lorsque la valeur de l'énergie moyenne du signal de parole estcomprise dans un intervalle infèrieur à 98% de sa valeur, sa qualité est dégradée. Cela correspond à des taux de perte supèrieurs ou égaux à $10^{-3}$.

### 4.5.3. Conclusion

La perte des trames pour un seuil de taux de l'ordre $5.10^{-3}$ de perte ne dégrade pas sensiblement la qualité de la parole, et le remplacement des trames perdues ne sera pas nécessaire, cela nous permet de minimiser le temps de traversée des réseaux. L'insertion des fausses trames n'affecte pas la qualité de la parole pour un seuil de taux d'insertion de l'ordre $10^{-3}$, par contre le temps de traversé du réseau augmente. Cependant pour le remplacement des trames perdues, il vaut mieux, remplacer les trames perdues par les précédentes que de les remplacer par des fausses trames.

# *Conclusion*

# Conclusion générale

Le but de ce mémoire est de étudié la qualité de la parole en fonction des erreurs de transmission.

Dans une première étape on a résoudre le problème de format du fichier de parole que le codeur utilise en entrée. Après avoir résolu ce problème, on a fait une adaptation du codeur au point audio. Cette adaptation nous permet d'écouter les fichiers de parole décodée et avoir aussi une opinion subjective sur la qualité du décodage.

Dans une seconde étape, on a séparé le codeur du décodeur, pour pouvoir étudier la qualité de la parole en fonction des erreurs de transmission, car dans le programme principal le décodeur est inclus, Cette séparation permet d'obtenir un codeur, qui fournit un fichier de paramètres et un décodeur auquel on applique ce fichier de paramètres. Les erreurs de transmission interviennent normalement au niveau de ces paramètres.

L'étape suivante consiste à simuler des erreurs de transmission et à mesurer la qualité de la parole décodée. Ces erreurs ont été simulées de plusieurs manières. Après avoir fait ces simulations, on est arrivé à certains résultats qui permettent de conclure, que la perte des paquets de parole sans remplacement n'affecte pas la qualité de la parole qu'a partir d'un taux d'erreur de l'ordre de $5.10^{-3}$. Cette perte permet de minimiser le temps de traversée du réseau, car elle évite le temps de stockage du paquet perdu dans le buffer de stockage des paquets de parole, qui serviront au décodage. Au-delà de ce taux le remplacement des paquets perdus est indispensable.

Pour éviter la retransmission du paquet perdu, on le remplace par le paquet précédent. Le remplacement par le paquet précédent protège la qualité de la parole contre des taux de perte de l'ordre de 0.01.

L'étude concernant l'insertion à tort des paquets, a montré que la qualité de la parole se dégrade à partir d'un taux d'insertion de l'ordre $10^{-3}$, c'est pourquoi il faut prévoir un procédé d'élimination des faux paquets à partir de ce taux.

Les erreurs en bits n'affectent sensiblement la qualité qu'à partir d'un taux de l'ordre de $10^{-3}$. Ainsi la transmission de la parole sur des réseaux qui présentent ce taux d'erreurs en bits est déconseillé, car la qualité de la parole sera nettement dégradée.

Dans le manuel d'utilisation du CELP FS-1016, il manquait beaucoup d'information l'algorithme CELP. Cette étude a permis de combler beaucoup de ces lacunes et on peut dire que le laboratoire dispos maintenant d'un codeur avec toutes les informations nécessaires pour son utilisation.

# REFERENCES BIBLIOGRAPHIQUES

**[1]**Combescure P., *Mathieu M. (1985) : Codage numérique des signaux sonores, l'Echo des recherches vol. 121*

**[2]R. Goldberg**, **and L. Riek**, *a Practical Handbook of Speech Coders, CRC Press, Boca Raton London New York Washington, D.C., 2000.*

**[3] W. C. Chu**, Speech Coding Algorithms: Foundation and Evolution of Standardized Coders, John Wiley & Sons, 2003

**[4]Calliope.** *La parole et son traitement automatique. Collection technique et scientifique de télécommunication ; Masson1986*

**[5]Michel Mauc M.** *(1993), Réduction de la complexité des algorithmes de codage deparole de type CELP. Thèse de doctorat, Université Paris Sud (Orsay)*

**[6]Dymarski, P**. *Moreau N., VosW., Détermination et codage de l'excitation dans un codeur CELP, 13éme colloque GRETSI pp 725-728,1991*

**[7]Andrew T**., *Université d'Amsterdam, Réseaux Architecture, protocole, application Inter Edition, Paris(1991)*

**[8]Jean-Pierre R.** *Les réseaux locaux et compac 3G, DAC/RSX Décembre1990*

**[9]CCITT** *Recommandation G.764 Aspect Généraux Des systèmes de transmission du numérique équipement, Mise en paquet de la parole-protocole de transmission de la parole par paquet Genève 1990*

# Liste des Symboles

| | |
|---|---|
| IP | Internet protocoles. |
| LPC | Langage Parlé Complété |
| MIC | Modulation par impulsion et codage. |
| PCM | Pulse Code Modulation. |
| MICDA | Modulation par impulsion et codage Différentiel adaptatif. |
| L'UIT | L'Union internationale des télécommunications. |
| CELP | Le code Excited Prédictive Codec. |
| DRT | Diagnostic Rythme Test. |
| DAM | Diagnostic Acceptability Measure. |
| MOS | Mean Opinion Score. |
| GSM | Group special mobile. |
| VSELP | Vector sum exited prediction. |
| CCITT | Consultative Committeefor International Telephony and Telegraphy**.** |
| RNIS | Réseau Numérique à intégration de service. |
| ISO | L'internationale standard organisation. |
| OSI | Interconnexion de systèmes ouverts. |
| UIH | United Information Highway. |
| PD | Discriminateur de protocole. |
| BDI | L'indicateur d'abondant de bloc. |
| RSB | Rapport signal à bruît. |