الجمهورية الجزائرية الديمقراطية الشعبية

*République Algérienne Démocratique et Populaire*

وزارة التعليم العالي والبحث العلمي

*Ministère de l'enseignement supérieur et de la recherche scientifique*

جامعة الشهيد حمة لخضر الوادي

*Université EchahidHammaLakdhar-EL OUED*

كلية التكنولوجيا

*Faculté de technologie*

قسم الهندسة الكهربائية

**Département de génie électrique**

# MEMOIRE DE FIN D'ETUDE

**En vue de l'obtention du diplôme de Master Académique en génie électrique**

**Spécialité : systéme et télècommunication**

## THEME

# Etude comparative des techniques de codage de la parole à travers un canal bruité.

**Présenté Par :**

**M$^{elle}$ : Ben Yaia Idris**

**M$^{elle}$ :Salem Mohamed Mouloud**

**Devant le jury composé de :**

| | |
|---|---|
| **Président** | U. El-Oued |
| **Examinateur** | U. El-Oued |
| **Encadreur** | U. El-Oued |

**Année universitaire :** 2016/2017

# *Remerciements*

*Tout d'abord, tout remerciement et louange à notre Dieu `ALLAH' le Tout-puissant de nous avoir donné le courage, la volonté et la patience de mener à terme ce présent travail.*

*Ce travail a été effectué sous l'encadrement de* ***Dr. AJGOU Riadh****, à qui nous voudrons témoigner toute notre reconnaissance, pour nous avoir fait bénéficier de ses compétences scientifiques*

*Nous remercions vivement nos familles pour leur aide morale et matérielle durant toute la période de préparation.*

*Nous tenons à remercier aussi tous les enseignants qui ont contribué à notre formation au département de Technologie*

# Résumé

La parole a une place centrale dans la communication humaine. Les dernières années ont connu l'apparition de nouveaux systèmes de communication numériques pour lesquels le signal parole (voix) est codé. Le plus récemment, le besoin pour le codage de la parole est dû à l'augmentation rapide dans les systèmes sans fil, et dans la voix sur Internet Protocole (VoIP) où la parole est juste un type de données transportés sur le réseau de l'IP. Ainsi le type de codage de la parole est essentiel dans les réseaux mobiles (GSM, UMTS…). Le but de codage de la parole est de produire la plus haute qualité de la parole possible pour un objectif d'avoir un faible BER. Le codeur de la parole consiste d'un encodeur et un décodeur. L'encodeur prend le signal parole numérique pour le compresser et éliminer la redondance. Le rôle de décodeur est de reconstruire le signal original. Dans ce travail on va explorer quelque types de codage comme: PCM, DPCM, ADPCM…… qui correspondent au International Telecommunications Union – Telecoms (ITU-T) pour un objectif de faire une étude comparative, à travers un canal numérique bruité.

**Mots clés :** BER, BPSK, AWGN, PCM, DPCM, ADPCM .

# ملخص

للخطاب مكانة مركزية في التواصل الإنساني .في الأعوام الأخيرة عرفنا المظهر الجديد لنظم الاتصالات الرقمية التي من اجلها تم تشفير إشارة الخطاب , في الآونة الأخيرة أصبحت الحاجة إلى ترميز الكلام سببها الزيادة السريعة في الأنظمة اللاسلكية ونقل الصوت عبر بروتوكولات الانترنت VOIP حيث أن الكلمة فيه مجرد نوع واحد من البيانات المنقولة عبر شبكةIP إضافة إلى أن هكذا نوع من ترميز الكلام أساسي في الشبكات المتنقلةGSM ) , ( … UMTS.
الهدف من ترميز الكلام هو إنتاج اعلي مستوى ممكن من الجودة من اجل أن يكون BER ضعيف , يتكون مبرمج تشفير الكلام من وحدة فك الترميز : تعمل على إعادة بناء الإشارة الأصلية مشفر : يأخذ إشارة الكلام الرقمية للضغط و القضاء على التكرار. في هذا العمل سوف نستكشف بعض أنواع الترميز PCM, DPCM, ADPCM…. الذي يتوافق مع الاتحاد الدولي للاتصالات لهدف دراسة مُقارنة من خلال قناة رقمية صاخبة.

**الكلمات المفتاحية :** ADPCM , DPCM , PCM AWGN, BPSK , BER

# Abstract

Speech has a central place in human communication.In recent years we have introduced the new appearance of digital communication systems for which the speech signal is encrypted Recently, the need for speech coding has been caused by the rapid increase in wireless systems and voice over VOIP Where the word is just one type of data transmitted over the IP network In addition, this type of speech encoding is essential in( GSM, UMTS ….)

Target networks The purpose of speech coding is to produce the highest quality possible in order to have a weak BER The speech encryption programmer consists of: Decoder: Works to rebuild the original signal Encode: Take the digital speech signal to press and eliminate repetition In this work we will explore some coding types PCM, DPCM , ADPCM ……. Which corresponding to the International Telecommunications Union – Telecoms (ITU-T) For the purpose of comparative study through digital channel a noisy .

**Keyword:** BER, BPSK, AWGN, PCM, DPCM, ADPCM .

:

# *Sommaire*

# Sommaire

**Chapitr III : Simulation des différents codeurs**

# *Introduction générale*

# *Résumé*

La parole a une place centrale dans la communication humaine. Les dernières années ont connu l'apparition de nouveaux systèmes de communication numériques pour lesquels le signal parole (voix) est codé. Le plus récemment, le besoin pour le codage de la parole est dû à l'augmentation rapide dans les systèmes sans fil, et dans la voix sur Internet Protocole (VoIP) où la parole est juste un type de données transportés sur le réseau de l'IP. Ainsi le type de codage de la parole est essentiel dans les réseaux mobiles (GSM, UMTS…).

Le but de codage de la parole est de produire la plus haute qualité de la parole possible pour un objectif d'avoir un faible BER. Le codeur de la parole consiste d'un encodeur et un décodeur. L'encodeur prend le signal parole numérique pour le compresser et éliminer la redondance. Le rôle de décodeur est de reconstruire le signal original. Dans ce travail on va explorer quelque types de codage comme: PCM, DPCM, ADPCM…… qui correspondent au International Telecommunications Union – Telecoms (ITU-T) pour un objectif de faire une étude comparative, à travers un canal numérique bruité.

**Mots clés :** BER, BPSK, AWGN, PCM, DPCM, ADPCM .

# ملخص

للخطاب مكانة مركزية في التواصل الإنساني .في الأعوام الأخيرة عرفنا المظهر الجديد لنظم الاتصالات الرقمية التي من اجلها تم تشفير إشارة الخطاب , في الآونة الأخيرة أصبحت الحاجة إلى ترميز الكلام سببها الزيادة السريعة في الأنظمة اللاسلكية ونقل الصوت عبر بروتوكولات الانترنت VOIP حيث أن الكلمة فيه مجرد نوع واحد من البيانات المنقولة عبر شبكة IP إضافة إلى أن هكذا نوع من ترميز الكلام أساسي في الشبكات المتنقلة ( GSM,UMTS … ) .

الهدف من ترميز الكلام هو إنتاج اعلي مستوى ممكن من الجودة من اجل أن يكون BER ضعيف , يتكون مبرمج تشفير الكلام من وحدة فك الترميز : تعمل على إعادة بناء الإشارة الأصلية مشفر : يأخذ إشارة الكلام الرقمية للضغط و القضاء على التكرار. في هذا العمل سوف نستكشف بعض أنواع الترميز PCM, DPCM, ADPCM…. الذي يتوافق مع الاتحاد الدولي للاتصالات لهدف دراسة مُقارنة من خلال قناة رقمية صاخبة.

**الكلمات المفتاحية :** ADPCM , DPCM , PCM AWGN, BPSK , BER

# *Abstract*

Speech has a central place in human communication.In recent years we have introduced the new appearance of digital communication systems for which the speech signal is encrypted Recently, the need for speech coding has been caused by the rapid increase in wireless systems and voice over VOIP Where the word is just one type of data transmitted over the IP network In addition, this type of speech encoding is essential in( GSM, UMTS ….)

Target networks The purpose of speech coding is to produce the highest quality possible in order to have a weak BER The speech encryption programmer consists of: Decoder: Works to rebuild the original signal Encode: Take the digital speech signal to press and eliminate repetition In this work we will explore some coding types PCM, DPCM , ADPCM ……. Which corresponding to the International Telecommunications Union – Telecoms (ITU-T) For the purpose of comparative study through digital channel a noisy .

**Keyword:** BER, BPSK, AWGN, PCM, DPCM, ADPCM .

# *Introduction générale*

Le but principal d'un système de communication numérique est de transmettre des informations à partir d'une ou plusieurs sources à une ou plusieurs destinations avec une grande fiabilité et la bonne qualité de service.

Cependant, à cause de la présence du bruit dans les canaux de communication, l'information reçue, en général diffère de celle transmise. La tolérance aux erreurs de transmission à des débits élevés devient donc un critère important pour améliorer la qualité de service. En effet, la qualité d'une transmission numérique dépend principalement de la probabilité d'occurrence d'erreurs dans les symboles transmis. Cette probabilité d'erreur étant fonction du rapport signal à bruit, une des solutions pour améliorer la qualité de transmission est d'augmenter la puissance d'émission. Malheureusement, cette solution est souvent très couteuse en termes d'énergie et de technologie, ce qui rend sa réalisation souvent impossible. Une alternative pour protéger l'intégrité du message transmis, consiste à ajouter aux informations utiles, des informations redondantes selon une règle bien déterminée. La génération de la redondance par le codeur à l'émission permet à l'aide d'un algorithme de décodage à la réception de détecter ou corriger un nombre fini d'erreurs introduites par le bruit du canal de transmission.

Ce mémoire est structuré de trois chapitres, comme suit :

Le premier chapitre est une introduction générale au domaine du signal de la parole, décrit le signal de parole, ses caractéristiques, son système de production et quelques rappelles essentiels . On explore les différentes techniques d'évaluation de la qualité de la parole

Au deuxième chapitre, nous introduisons les éléments de base nécessaires de la chaîne de transmission numérique, de l'émetteur vers le récepteur . Ainsi, on explore les différents classes de codage de la parole et introduisons le principe de quelques codeurs d'un signal de la parole qui conforment à l'IUT (International Union Telecommunication).

Le dernier chapitre. Nous donnons le modèle de simulation et les différents paramètres utilisés, puis nous présentons tous les résultats obtenus dont, on a fait une étude comparative des codecs PCM, DPCM et ADPCM point de vue temps d'exécution, PESQ et BER.

A la fin et avec une conclusion générale et perspectives on termine notre travail

# *Chapitre I*

# *Généralité et mesure de la qualité d'un signal de parole*

# I.1. Introduction :

Le signal de parole étant un signal réel, continu, d'énergie finie et non stationnaire, Sa structure est complexe et variable dans le temps.

Dans la première partie de ce chapitre, nous avons présentons des Généralités au domaine du signal de la parole (les caractéristiques , les paramètre , numérisation…etc ). Dans les deuxième parties de ce chapitre, nous avons présentons les méthodes d'évaluation (subjective et objective) de la qualité parole des systèmes de télécommunications seront décrites, respectivement.

# I.2. Caractéristiques d'un signal de parole:

La connaissance du système vocal et des propriétés du signal de parole est essentielle pour concevoir des codeurs de parole efficaces. Les propriétés du système auditif humain peuvent aussi être exploitées pour améliorer la qualité perceptive du signal codé.

## I.2.1 Production de la parole :

Le signal vocal est engendré par l'appareil phonatoire avant d'être émis puis détecté par un auditeur. À la perception, l'oreille analyse ce signal et transmet au cerveau les informations nécessaires à son interprétation. La figure I-1présente le système phonatoire qui se compose des poumons, du larynx et du conduit vocal formé par le pharynx ainsi que les cavités buccales et nasales.

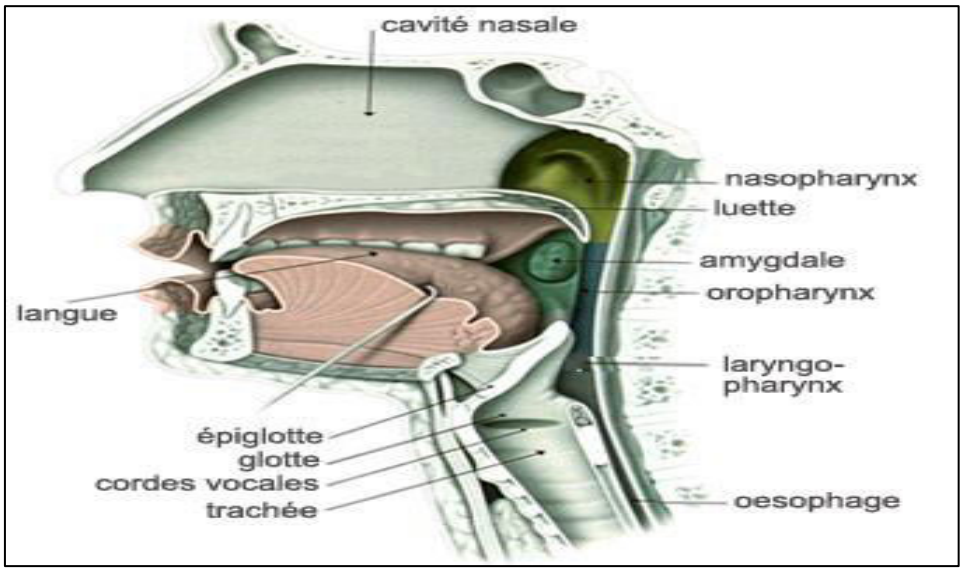


**Figure I-1**: Système phonatoire

La parole est généralement produite par expiration de l'air à travers la glotte et le conduit vocal. L'air venant des poumons est modulé par vibrations des cordes vocales et par déformation (élargissement ou resserrement) du conduit [1].

L'appareil respiratoire fournit l'énergie nécessaire à la production du son, en poussant de l'air à travers la trachée artère. Au sommet de celle-ci se trouve le larynx où l'air est modulé avant d'être appliqué au conduit vocal. Les cordes vocales sont en fait deux lèvres symétriques placées en travers du larynx.

Ces lèvres peuvent fermer complètement le larynx et, en s'écartant progressivement, déterminent une ouverture triangulaire appelée glotte. L'air y passe librement pendant la respiration et la voix chuchotée, ainsi que pendant la phonation des sons non voisés (des consonnes). Les sons voisés (par exemple, les voyelles) résultent au contraire d'une vibration périodique des cordes vocales.

## I.2.2 Modèle de production de la parole :

Dans une grande majorité de codeurs de parole, la production de parole est modélisée par une opération de filtrage, où une source excite un filtre représentant le conduit vocal [2].

La source et le filtre sont considérés comme séparés lors de l'analyse des signaux de parole. Une modélisation précise du conduit vocal et de la source d'excitation est nécessaire pour produire un signal intelligible et naturel. Des études menées dans le but de définir un modèle produisant un signal proche du signal de parole ont permis à Fant [3] de décrire un modèle de production linéaire. Le signal de parole est modélisé par la sortie d'un filtre excité par un train d'impulsions périodiques (à la cadence du pitch) pour les sons voisés et un bruit blanc pour les sons non-voisés (figure I.2).

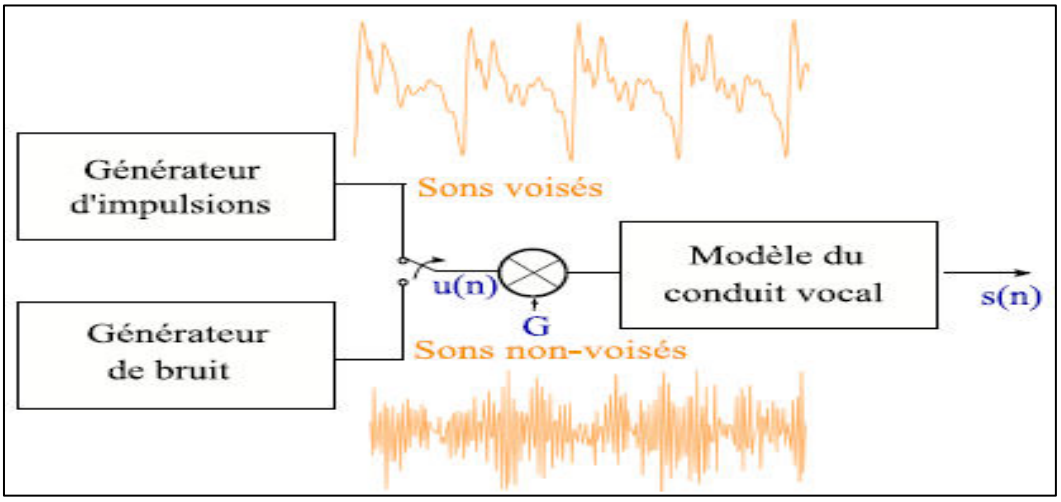


**Figure I-2** : Modèle simple de production de la parole

## I.2.3 Sons voisés et non-voisés :

Une décomposition simplifiée du signal de la parole doit ressortir deux types de sons: voisés et non voisés :

- Les sons voisés, tels que des voyelles, sont produits par le passage de l'air de poumons à travers la trachée qui met en vibration les cordes vocales. Ce mode, qui représente 80% du temps de phonation, est caractérisé en général par une quasipériodicité, une énergie élevée et une fréquence fondamentale (pitch). Typiquement, la période fondamental des différents sons voisés varie entre 2ms et 20ms.

- Les sons non voisés, comme certaines consonnes, dans ce cas les cordes vocales ne vibrent pas, l'air passe à haute vitesse entre les cordes vocales. Le signal produit est équivalent à un bruit blanc.

## I.2.4 Perception de la parole :

La perception de la parole par l'oreille humaine est très complexe. Bien que la quantification d'un signal vocal par une forme d'onde présente des déformations significatives, un auditeur pourra tout de même comprendre la parole synthétisée. Il a été montré que la gamme de fréquences de perception appartenant à l'intervalle [200, 3700] Hz est la plus importante pour l'intelligibilité de la parole. Cette bande de fréquences, où le système auditif est le plus sensible, justifie le taux d'échantillonnage de 8 kHz pour les codeurs de parole .

Les analyses temporelles ou spectrales du signal de parole doivent tenir compte des caractéristiques du système auditif pour améliorer l'efficacité des algorithmes de codage de la parole. Les composantes de phase d'un signal de parole jouent un rôle négligeable dans la perception du langage. L'oreille humaine perçoit principalement la parole grâce aux informations du spectre d'amplitude. Les pics du spectre sont plus importants pour la perception que les vallées du spectre et les basses fréquences plus importantes que les hautes fréquences. Le bruit est moins perceptible dans les zones fréquentielles où le signal de parole a beaucoup d'énergie. On parle alors de masquage de fréquences.

Tout bruit inférieur à un certain seuil pourra ainsi être masqué par le signal et devenir inaudible. On dit que le signal masque le bruit. Le bruit est alors plus toléré dans les zones formantiques que dans celles antifomantiques du signal. Pour cela, un masque fréquentiel dont la forme est semblable à l'enveloppe spectrale du signal est, généralement, utilisé [4].

# I.3. Paramètres du signal de parole :

## I.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 [5] :

- 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.

## I.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 :

$$E = \sum_{n}^{N-1} s^2(n) \qquad \textbf{( I .1 )}$$

N : la taille de la trame.

## I.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.

$$S(k) = S(f = \frac{K}{N}) = \sum_{n=0}^{N-1} s(n).w(n).\exp(-j2\pi nk/N), 0 \le k \le N-1 \qquad \textbf{( I .2)}$$

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:

$$\left|S(K)\right|^2, 0 \le k \le \frac{k}{2} \qquad \textbf{(I .3)}$$

### I.3.4. Intensité d'un signal vocal :

L'intensité d'un son, appelée aussi volume, permet de distinguer un son fort d'un son faible. Elle correspond à l'amplitude de l'onde. L'amplitude est donnée par l'écart maximal de la grandeur qui caractérise l'onde. Pour le son, onde de compression, cette grandeur est la pression. L'amplitude sera donc donnée par l'écart entre la pression la plus forte et la plus faible exercée par l'onde acoustique. Lorsque l'amplitude de l'onde est grande, l'intensité est grande et donc le son est plus fort. L'intensité du son se mesure en décibels (dB). On distingue différentes façons de mesurer l'amplitude d'un son :

- La puissance acoustique : La puissance acoustique est associée à une notion physique. Il s'agit de l'énergie transportée par l'onde sonore par unité de temps et de surface. Elle s'exprime en Watt par mètre carré ($W.\ m^{-2}$) .
- Addition de sons : L'échelle des décibels est une échelle dite logarithmique, ce qui signifie qu'un doublement de la pression sonore implique une augmentation de l'indice d'environ 3 : avec 3 dB de plus, l'intensité est en fait doublée [6].

## I.4. Numérisation du signal de la parole :

La parole est produite par l'articulation des membres phonatoires de l'homme et prend une forme analogique apériodique ; ce qui est impossible pour que la machine puisse l'interpréter ou le prédire car elle ne comprend que du numérique. Pour cela on doit faire un traitement de numérisation sur ce signal. L'une des méthodes les plus utilisés dans la numérisation est la méthode Delta ou MIC qui consiste en trois étapes : l'échantillonnage, la quantification et le codage.

## I.4.1. L'échantillonnage :

L'échantillonnage consiste à transformer une fonction $a(t)$ à valeurs continues en une fonction â($t$) discrète constituée par la suite des valeurs $a(t)$ aux instants d'échantillonnage $t = kT$ avec $k$ un entier naturel (figure - I .3 ). Le choix de la fréquence d'échantillonnage n'est pas aléatoire car une petite fréquence nous donne une présentation pauvre du signal.

Par contre une très grande fréquence nous donne des mêmes valeurs, redondance, de certains échantillons voisins donc il faut prélever suffisamment de valeurs pour ne pas perdre l'information contenue dans a(t). Le théorème suivant traite cette problématique :

**Théorème** (de Shannon)**.** La fréquence d'échantillonnage assurant un non repliement du spectre doit être supérieure à 2 fois la fréquence haute du spectre du signal analogique.

$$f_{ech} = 2 \times f_{max} \qquad \textbf{(I .4)}$$

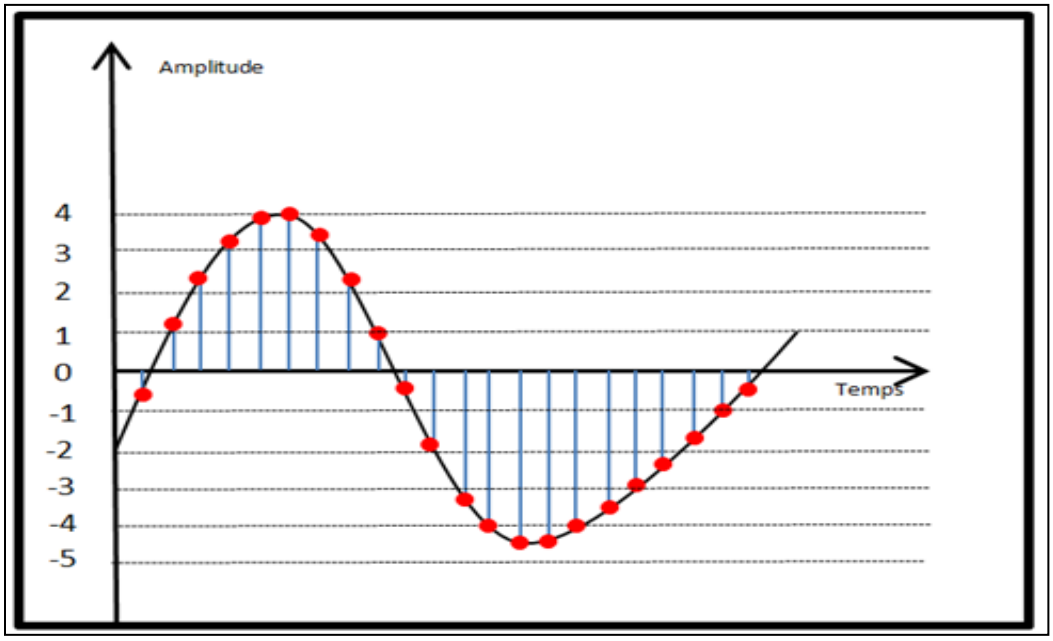


**Figure I-3** :un signal échantillonné

Pour la téléphonie, on estime que le signal garde une qualité suffisante lorsque son spectre est limité à 3400 Hz et l'on choisit fe = 8000 Hz. Pour les techniques d'analyse,de synthèse ou de reconnaissance de la parole, la fréquence peut varier de 6000 à 16000 Hz

Par contre pour le signal audio (parole et musique), on exige une bonne représentation du signal jusque 20 kHz et l'on utilise des fréquences d'échantillonnage de 44.1 ou 48 kHz . Pour les applications multimédia, les fréquences sous-multiples de 44.1 kHz sont de plus en plus utilisées : 22.5 kHz, 11.25 kHz [7].

### I.4.2. Quantification :

Cette étape consiste à approximer les valeurs réelles des échantillons selon une échelle de *n* niveaux appelée échelle de quantification. Il y a donc 2n valeurs possibles comprises entre $-2n-1$ et $2n-1$ pour les échantillons quantifiés (figure. I .4 ). L'erreur systématique que l'on commet en assimilant les valeurs réelles de l'écart au niveau du quantifiant le plus proche est appelé bruit de quantification.

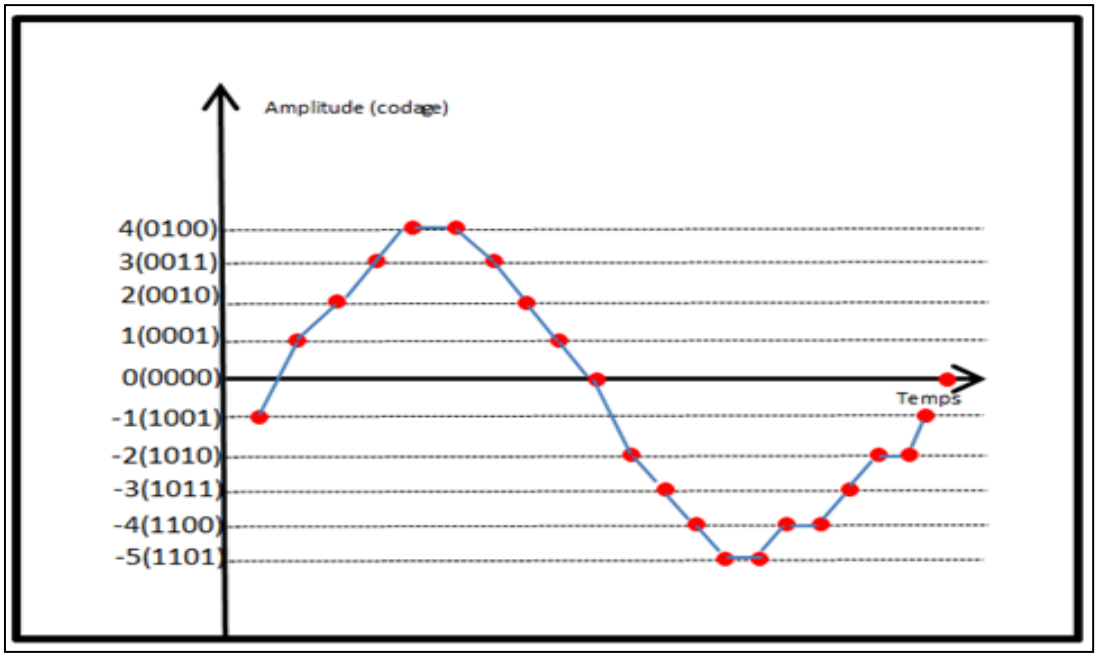


**Figure I-4** : un signal quantifié

### I.4.3. Codage :

C'est la représentation binaire des valeurs quantifiées qui permet le traitement du signal sur machine (fig. I .4 ).

## I.5. Mesurée la qualité de la parole :

Ce n'est pas facile d'évaluer la qualité perçue de la voix. La meilleure manière de l'évaluer est d'avoir de vraies personnes qui font l'évaluation parce que la qualité est un concept très subjectif. UIT-T a défini plusieurs standards qui permettent une évaluation de la qualité de la communication vocale.

Les réseaux de télécommunication modernes fournissent un large ensemble de services vocaux utilisant beaucoup de systèmes de transmission. Le déploiement rapide des technologies numériques en particulier a conduit a un besoin accru d'évaluation des caractéristiques de transmission des nouveaux équipements de communication. UIT-T a

défini plusieurs standards qui permettent une évaluation de la qualité de la communication vocale.

## I.5.1. Méthodes subjectives :

### I.5.1.1 Score moyen d'opinion (Mean Opinion Score, MOS) :

En 1996, UIT-T a défini la méthodologie de déterminer le degré de satisfaction concernant une certaine ligne téléphonique, sous le nom de score moyen d'opinion (Mean Opinion Score, MOS) (la norme P-800). Cette méthode peut s'appliquer a toute forme possible de dégradation : perte de paquets, distorsions de circuit, erreurs de transmission, distorsions environnementales, écho, distorsions de codage etc.

La procédure d'évaluation est basée sur des tests subjectifs dans lesquels la qualité est notée par un ≪panel ≫d'expérimentateurs. Les valeurs suivantes sont assignées en fonction de la qualité perçue de la connexion [8]:

| | |
|---|---|
| ***Excellent*** | 5 |
| ***Bien*** | 4 |
| ***Acceptable*** | 3 |
| ***Pauvre*** | 2 |
| ***Mauvais*** | 1 |

**Tableau I .1**: Les valeurs assignées pour la qualité perçue du MOS

## I.5.2. Méthodes objectives :

Les mesures objectives de qualité des signaux vocaux les plus communément utilisées sont citées et classées dans le tableau( I .2)

| Mesures dans le domaine temporel | Mesures dans le domaine fréquentiel | Mesures dans le domaine perceptuel |
|---|---|---|
| SNR<br>segSNR | IS<br>CD<br>WSS<br>LLR | BSD, MBSD<br>PESQ<br>PSQM |

**Tableau (I-2)** : Classification des critères d'´evaluation objective

Les critères temporels et fréquentiels se basent essentiellement sur l'évaluation de la qualité en termes de comparaison de distorsion de formes entre signal de référence et signal débrutie, sans tenir compte de l'aspect perceptif. Certes, c'est une condition nécessaire mais non suffisante dans la mesure ou deux signaux pratiquement de même forme peuvent être perçus différemment [9], d'ou l'intérêt d'introduire le facteur psycho-acoustique pour tout système ayant pour objectif de conserver la qualité de la parole.

Diverses mesures objectives perceptuelles sont élaborées conduisant a de bonnes corrélations avec la perception humaine. Elles sont essentiellement dédiées au codage de la parole, mais trouvent leur application en de bruitage de la parole.

### I.5.2.1. SNR segmental (segSNR) :

Le SNR (Signal to Noise Ratio) segmental segSNR est la mesure de qualité objective la plus utilisée dans le domaine temporal. Il définit la moyenne des segSNR issus de plusieurs segments de court durée (15 `a 20 ms) :

$$SNRseg = \frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\frac{\sum_{i=mN}^{mN+N-1}s^2(i)}{\sum_{i=mN}^{mN+N-1}(s(i)-\hat{s}(i))^2} \qquad \textbf{(I .5)}$$

o`u s(i), ˆs(i), N et M sont respectivement le signal de référence, le signal débrutie, la longueur d'un segment et le nombre total de segments.

Le SNR segmental souffre de deux limitations : d'abord si le signal de parole contient des segments de silence, ce qui est tr`es probable, le s(i) sera nul et n'importe quelle quantité de bruit entrainera un SNR en dB négatif pour ce segment ; du coup le SNR total sera biaise par cette quantité. Ce problème peut être résolu partiellement en choisi-ssant un seuil d'´energie au delà duquel le SNR segmental sera calcul´e. Ensuite, il faut nécessairement que les deux signaux compares soient alignes temporellement car ce critère est tr`es sensible aux déphasages.

### I.5.2.2. PESQ (Perceptual Evaluation of Speech Quality)

PESQ est une méthode objective de prédiction de la qualité subjective pour la téléphonie avec bande passante réduite et codeurs vocaux. PESQ combine les meilleurs aspects de PSQM et PAMS. Compare au PSQM, PESQ prend en compte, en plus, le filtrage, le délai variable, les distorsions de codage et les erreurs de canal.

Le processus clé de PESQ, comme pour les méthodes apparentées, est la transformation des signaux original et dégrade en représentations psychophysiques proches de celles des signaux auditifs du système auditif humain.
Suite a des tests mènes de façon extensive, PESQ s'avère fournir des précisions acceptables dans les types d'applications suivants : l'évaluation et la sélection des codeurs, tests dans des réseaux actifs (avec connexions numériques ou analogiques pour la communication VoIP), tests des réseaux émules et prototype.

Le score PESQ est produit sur une échelle similaire au MOS, avec des valeurs situées entre -0,5 et 4,5. Les valeurs habituelles sont entre 1,0 et 4,5, les scores MOS obtenus dans des expériences subjectives sur la qualité de l'écoute.

### I.5.2.3. PSQM (Perceptual Speech Quality Measure) :

Le PSQM standardise par UIT-T en 1998 concernant la mesure de la qualité des codeurs vocaux dans la bande téléphonique. Cette méthode a été initialement développée par la société KPN aux Pays-Bas.

PSQM utilise des représentations psychophysiques qui sont aussi prés que possible des représentations internes humaines pour les signaux vocaux. Une valeur zéro de PSQM signifie qu'aucune dégradation 'est présente, et une valeur de 6,5 signifie un canal totalement inutilisable.

Les valeurs PSQM peuvent être transposées sur une échelle de type MOS afin de pouvoir obtenir une estimation de la qualité subjective [10].

### I.5.2.3. BSD et MBSD :

La mesure BSD (Bark Spectral Distortion) est parmi les premiers critères `a avoir incorpore des notions en relation avec notre système d'audition dans l'´evaluation de la qualité de la parole Le BSD a pour objectif de mesurer la distorsion entre le signal de référence et celui code, dans le domaine de Bark. La sensation de force sonore connue sous le nom de sonie est mise en jeu pour calculer cette distorsion.

En effet, la distorsion totale est la moyenne de la distance euclidienne entre la sonie du signal de référence et celle du signal débrutie.

Le MBSD (Modified Bark Spectral Distortion) introduit le seuil de masquage du bruit pour calculer la distorsion dans le BSD; l'idée est de ne tenir compte que de la distorsion audible. Effectivement, tout ce qui est au-dessous du seuil de masquage du bruit est imperceptible `a l'oreille humaine. Par conséquent, la distorsion totale est la moyenne de la différence entre les sonies du signal de référence et du signal débrutie pondérée par un paramètres annulant lorsque la distorsion est inaudible [11] .

### I.5.2.4. Mesure d'Itakura Saito :

La mesure d'Itakura Saito repose sur l'analyse LPC. Son expression fait intervenir le modèle tout pôle du signal de référence s et celui du signal test´e y. Soient P(w), Pˆ (w) les densités spectrales de puissance du modèle AR du signal de référence et du signal de test. La distance d'Itakura Saito est donnée par :

$$d_{IS}(\mathrm{P}(w), \mathrm{P\hat{}}(w)) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{\mathrm{P}(w)}{\mathrm{P\hat{}}(\mathrm{w})} - \log\frac{P(w)}{\mathrm{P\hat{}}(\mathrm{w})} - 1\right]dw \qquad \textbf{(I .6)}$$

### I.5.2.5. Distance cepstrale :

La distance cepstrale est principalement utile pour représenter la distribution de l'erreur au cours du temps. Les coefficients cepstraux c(i) peuvent être calcules `a partir des coefficients de prédiction linéaire a(i) `a l'aide de la relation suivante :

$$c(i) = -a(1) \qquad \textbf{(I .7)}$$

$$c(i) = -a(i) - \sum_{k=1}^{i-1}(1-\frac{k}{i})c(i-k)a(k), 1 \le i \le p \qquad \textbf{(I .8)}$$

Considérons les coefficients cepstraux Ct(i) et Cr(i) calcules respectivement sur les trames d'indice i du signal-test `a ´évaluer et de la référence. La distance cepstrale d'ordre 2 entre ces deux signaux est donnée par [12] :

$$d_{\text{cep}} = \sum_{i=1}^{p}(Ct(i) - \text{C}r(i))^2 \qquad \textbf{(I .9)}$$

Où est l'ordre des coefficients LPC. Suite `a cette ´écriture, la distance cepstrale est tout simplement la distance euclidienne entre les coefficients cepstraux génères récursivement `a partir de l'analyse LPC.

# I.6.Conclusion :

Ce chapitre a été consacré à l'étude de signal parole : production de la parole par l'appareil phonatoire, modélisation et les paramètres essentielle caractérisant le signal parole . L'étapes de numériseation d'un signal parole , revient à prendre des valeurs de ce signal (echantillons) à des instants donnés (instants d'échantillonnage) et souvent de façon régulière (période d'échantillonnage) .

La qualité vocale peut être évaluée par des méthodes subjectives et objectives. Bien que les méthodes objectives ( SegSNR ,PESQ ,PSQM ,...etc) permettent un gain de temps, les méthodes subjectives(MOS) restent les plus fiables, la perception sonore et vocale étant un phénomène subjectif. De plus, les modèles objectifs les plus aboutis cherchent à corréler les résultats fournis par les tests subjectifs. Le chapitre suivant aborde les types des codage de la parole les plus répandus qui conforme à l'IUT (Internation union Telecommunucation).

# Chapitre II
# Types de codages de la parole

# II.1. Introduction :

Le codage de parole permet la réduction de débit de transmission du signal dans de 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 .

Les codeurs de parole exploitent les redondances du signal de parole pour améliorer leur efficacité. Différentes techniques peuvent être utilisées. Certains codeurs modélisent le système de production de parole par extraction de paramètres décrivant le signal vocal .

Dans pce chapitre on explore les différents éléments utilisés dans une chaîne de transmission numérique, en s'appuyant sur les différentes techniques de codages de la parole notamment les codeurs qui conforment à l'IUT (International Union of Télécommunications) comme PCM, DPCM, ADPCM ..

# II.2. Chaîne Transmission numérique :

La transmission numérique a fait pour l'objet de la communication d'un point (émetteur) vers un autre (récepteur), et d'une information discrète provenant d'une source qui ne dispose que d'un nombre fini n de caractères alphabet , en assurant généralement un débit élevé, une bon-ne robustesse, une mobilité et un coût faible.

Afin de garantir une bonne qualité de service le système de communication doit faire face aux différentes sources de distorsion rencontrées lors de l'émission, la propagation et la réception [13].

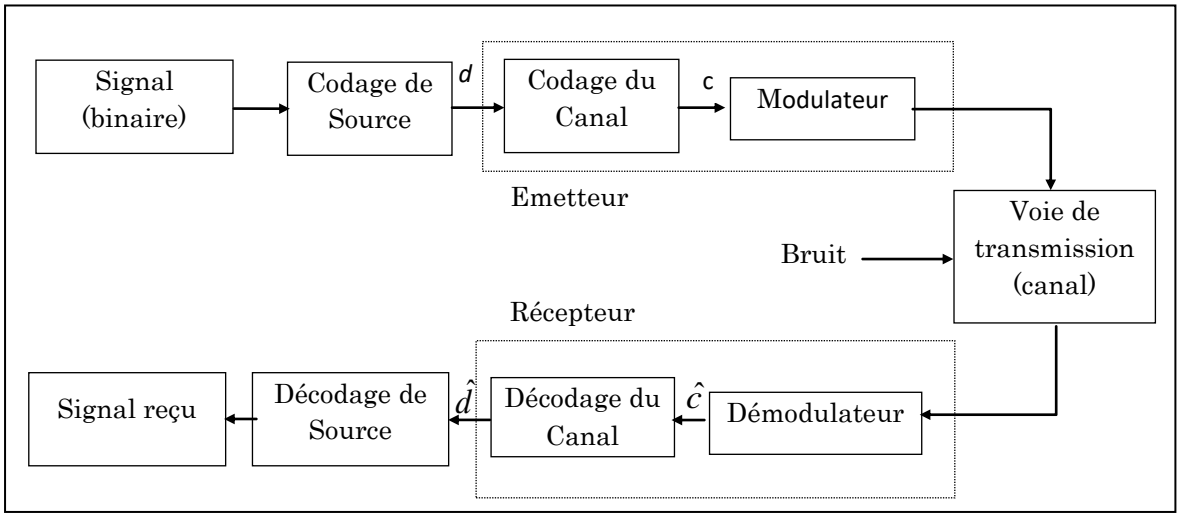


**Figure II.1** : Schéma général de tout système de transmission numérique

## II.2.1. La Source :

Le signal de parole étant par nature analogique, il doit tout d'abord être converti dans l'objectif d'une transmission numérique. La Figure II.2 rappelle les étapes de cette conversion analogique-numérique où $F_e = 1/T_e$ désigne la fréquence d'échantillonnage.

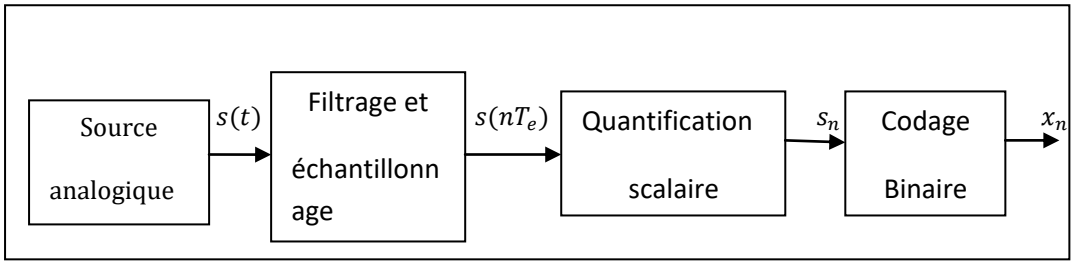


**Figure II.2** : Principe de la numérisation d'une source analogique

## II.2.2. Codage de source :

Sert à comprimer le signal numérique en exploitent les redondances fréquentes dans le signal numérique [14] Il existe deux types de codage de source :

- **Codage de source sans perte** : Les données sont compressées et ensuite décompressées, l'information originale contenue dans les données a été préservée. Aucune donnée n'a été perdue ou oubliée. Les données n'ont pas été modifiées Parmi les plus connues de ces méthodes figurent le codage RLE, le codage de HUFFMAN ou bien encore le codage LZW .
- **Codage de source avec perte** : la utilisation peut être envisage si la distorsion qui en résulte mainmise un critère de codage appelé codage avec perte (lossy coding) ou code intelligente est utilisé par fois pour transmettre des signaux de parole ou séquence vidéo avec débit très faible. On notera L la longueur moyenne des mots codes c'est-à-dire le nombre moyen d'éléments de C utilisées pour transmettre un symbole de S .

$$L = E[I_K] = \sum_{K=1}^{N} P_K I_K \qquad \textbf{(II.1)}$$

Où $I_K$ est la longueur de mot code de symbole $S_K$ émis avec une probabilité $P_K$ .

- **redondance :** Dans la littérature, la notion intuitive de la redondance est souvent utiliséeElle se rapport a la corrélation statistique ou dépendance entre les échantillons de la source. D'après la croyance intuitive la compression de données est équivalente a l'élimination de la redondance présente dans le signal et les données sans redondance ne peuvent pas être compressées.

  Cette intuition n'est pas entièrement vraie puisqu'il est possible de compresser mêmes des sources sans mémoire (c'est-a-dire des sources dont les échantillons ne sont pas

corrèles) a l'aide de la quanti cation vectorielle [15]. Egalement, on utilise la notion de redondance perceptuelle 1 laquelle l'on peut éliminer dans le signal sans introduire une dégradation de la qualité perceptuelle.

**a. Codage NRZ (no return to zéro):**

Son principe consiste à coder le ‘’1’’ par +V et ‘’-1’’ par –V.

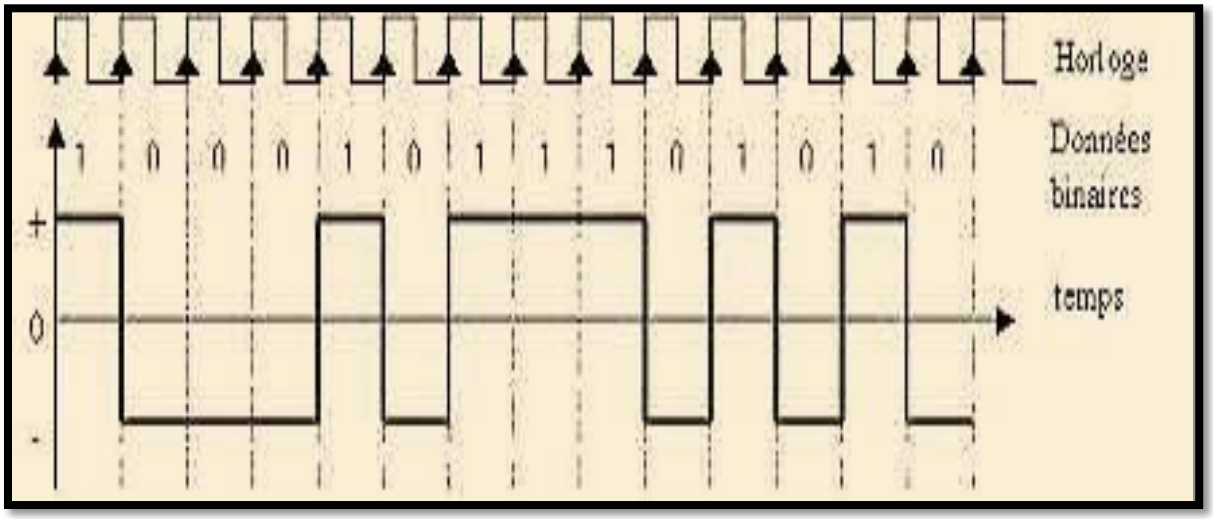


**Figure II.3:** Exemple de codage NRZ.

Le codage NRZ améliore légèrement le codage binaire de base en augmentant la différence d’amplitude du signal entre les 0 et les 1. Toutefois les longues séries de bit identiques (0 ou 1) provoquent un signal sans transition pendant une longue période de temps, ce qui peut engendrer une perte de synchronisation.

Le débit maximum théorique est le double de la fréquence utilisée pour le signal : on transmet deux bits pour un hertz. [16]

**b. Codage Manchester :**

Dans le codage Manchester, l’idée de base est de provoquer une transition du signal pour chaque bit transmis. Un 1 est représenté par le passage de -V à +V, un 0 est représenté par le passage de +V à -V.

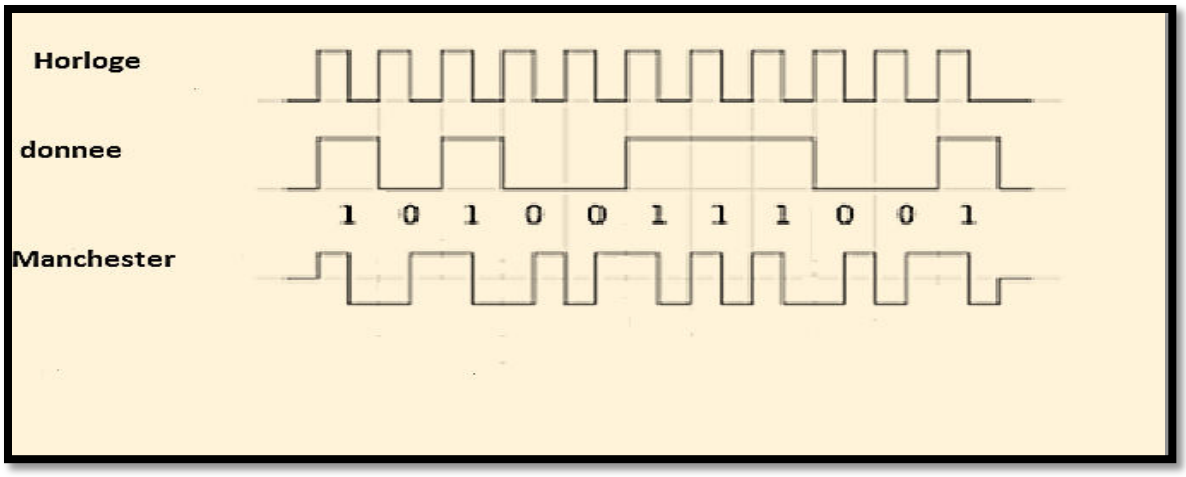


**Figure II.4:** Exemple de codage Manchester.

La synchronisation des échanges entre émetteur et récepteur est toujours assurée, même lors de l'envoi de longues séries de 0 ou de 1. Par ailleurs, un bit 0 ou 1 étant caractérisé par une transition du signal et non par un état comme dans les autres codages, il est très peu sensible aux erreurs de transmission. La présence de parasites peut endommager le signal et le rendre incompréhensible par le récepteur, mais ne peut pas transformer accidentellement un 0 en 1 ou inversement.

Toutefois, le codage Manchester présente un inconvénient : il nécessite un débit sur le canal de transmission deux fois plus élevé que le codage binaire. Pour 10 Mbit/s transmis, on a besoin d'une fréquence à 10 Mhz. Ceci le rend difficilement utilisable pour des débits plus élevés [17] .

L'utilisation de ce codage pour une transmission à 1 Gbit/s nécessiterait une fréquence maximale du signal de 1 GHz, ce qui est incompatible avec les possibilités des câblages actuels ainsi qu'avec les normes sur les compatibilités électromagnétiques.

Plus la fréquence du signal est élevée, plus les phénomènes de paradiaphonie pouvant perturber les installations avoisinantes du câble sont sensibles.

## II.2.3. Codage de canal :

Le codage canal a pour rôle de protéger l'information émise contre les perturbations du canal de transmission susceptibles de modifier son contenu. Il s'agit donc de rajouter de la redondance de manière à détecter et éventuellement corriger les erreurs lors de la réception si la stratégie adoptée le permet. Les conditions d'un codage correct sont déterminées par le second théorème de Shannon [18]. L'information $I(a_i)$ issue du codage source est transformée en séquence codée $C(a_i)$.

Deux types de codes peuvent être utilisés: les codes en bloc et les codes convolutifs .
Les codes en blocs sont mis en œuvre par des circuits de logique combinatoire. Des exemples de codes en blocs sont les codes (BCH), Reed-Muller (RM), les codes cycliques, et les codes de Hamming, et Reed Solomon (RS).

. Les codes convolutifs sont mis en œuvre par des circuits de logique séquentielle et sont aussi appelés codes en d'arbres ou en treillis. En général, les codes en bloc et les codes convolutifs peuvent être binaires ou non binaire, linéaire ou non linéaire, et systématique ou non systématiques .

### II.2.4. Modulation numérique :

La modulation numérique est caractérisée par : sa rapidité qui est exprimé en '' bauds'' d'où $R = \frac{1}{T}$ , par son débit il est exprimé en ''bits par seconds '' $D = \frac{1}{T_b}$ où le débit est supérieur ou égal a la rapidité, et la qualité d'une liaison est liée au 'taux d'erreur par bit' d'où T.E.B= $\frac{\text{nombre de bits faux}}{\text{nombre de bits transmis}}$ et par l'efficacité spectrale $\eta = \frac{D}{B}$ ou D est le débit binaire et B est la largeur de la bande occupé par le signal modulé .

- Il ya plusieurs type de modulation numérique :

#### a. Modulation par déplacement de phase (MDP) :

Les Modulations par Déplacement de Phase (MDP) sont aussi souvent appelées par leur abréviation anglaise: PSK pour "Phase Shift Keying" Cette modulation est très utilisée pour transporter des signaux numériques.
Le signal m(t) modulant s'écrit sous la forme suivante :

$$m(t) = A\cos(\omega_0 t + \varphi_0)\cos(\varphi_k) - A\sin(\omega_0 t + \varphi_0)\sin(\varphi_k) \qquad \textbf{(II.2)}$$

Cette expression montre que la phase de la porteuse est modulée par l'argument $\varphi_k$de chaque symbole ce qui explique le nom donné à la MDP. Remarquons aussi que la porteuse en phase $\cos(\omega_0 t + \varphi_0)$ est modulée en amplitude par le signal $A\cos(\varphi_k)$ et que la porteuse en quadrature $\sin(\omega_0 t + \varphi_0)$ modulée en amplitude par le signal $A\,sin(\varphi_k)$ [19] .

On appelle "MDP-M" une modulation par déplacement de phase (MDP) correspondant à des symboles M-aires. La figure II.5 montre différentes constellations de MDP pour M= 2, 4 et 8.

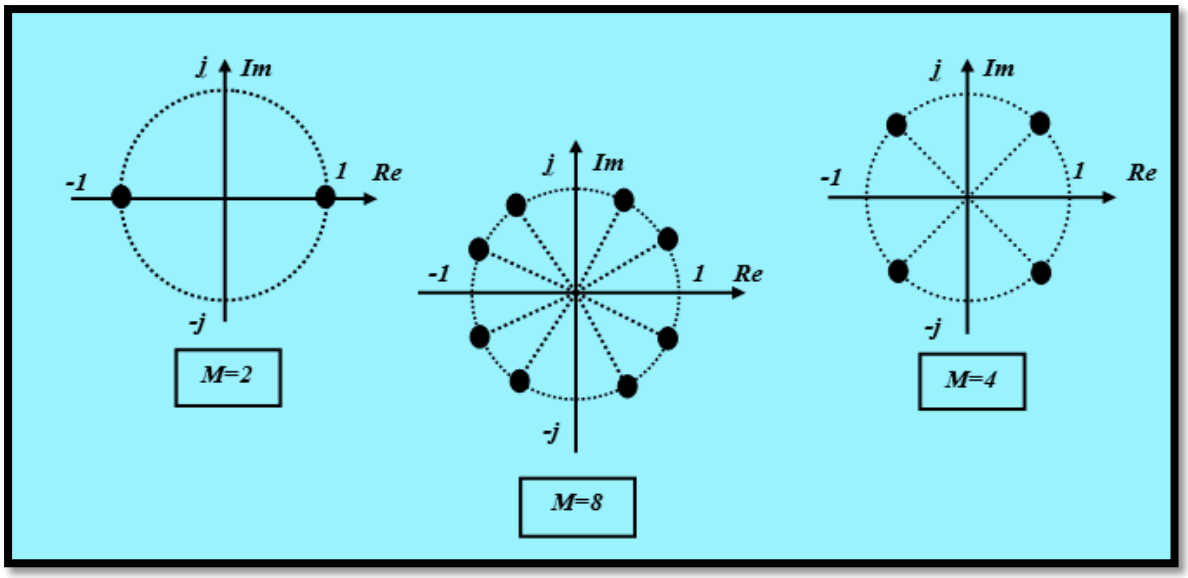


**Figure II.5:** Représentation des différents types de constellation de MDP pour M=2,4 et 8.

## b. Modulation par déplacement de fréquence (MDF) :

Les Modulations par Déplacement de fréquence (MDF) sont aussi souvent appelées par leur abréviation anglaise : FSK pour "Frequency Shift Keying".

L'expression du signal modulé est :

$$m(t) = \cos(2\pi \left(f_0 + \frac{\Delta f}{2} a_k\right) t) \qquad \textbf{(II.3)}$$

On peut aussi définir l'indice de modulation $\eta = \Delta f$ , T qui conditionne la forme de la densité spectrale du signal modulé [20] .

- **La modulation MDF a phase discontinue :**

Dans les Modulations par Déplacement de fréquence, on trouve les MDF à phase discontinue pour lesquelles la phase aux instants de transition kT peut sauter brusquement Exemple d'une modulation MDF discontinue :

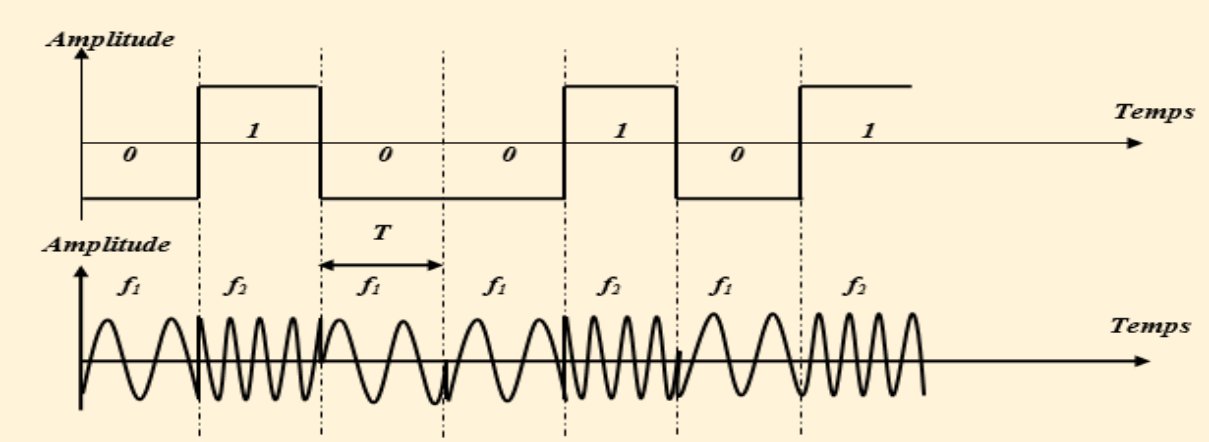


**Figure II.6:** Exemple de modulation MDF a phase discontinu.

## II.2.5. Canaux de transmission :

### II.2.5.1. Canal binaire symétrique :

Le canal binaire symétrique (CBS) est un canal discret dont les alphabets d'entrée et de sortie sont finis et égaux à $\{0,1\}$. On considère dans ce cas que le canal comprend tous les éléments de la chaîne comprise entre le codeur de canal et le décodeur correspondant Figure II.7

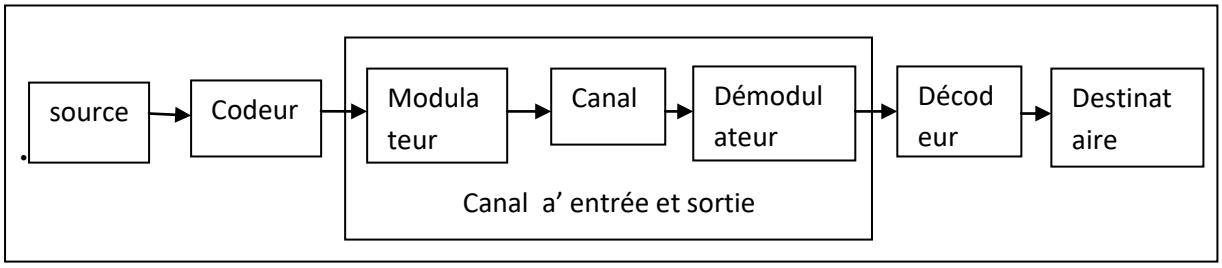


**Figure. II.7:** Description d'un canal binaire symétrique

On note respectivement $a_k$ et $y_k$ les éléments à l'entrée et à la sortie du CBS. Si le bruit et autres perturbations causent des erreurs statiquement indépendantes dans la séquence binaire transmise avec une probabilité p, alors :

$$P_r\,(y_k 0|a_k = 1) = P_r\,(y_k = 1|a_k = 0) = p \qquad (\textbf{II.4})$$

$$P_r\,(y_k = 1|a_k = 1) = p_r\,(y_k = 0|a_k = 0) = 1 - p \qquad (\textbf{II.5})$$

Le fonctionnement du CBS est résumé sous forme de diagramme sur la figure II.8 Chaque élément binaire à la sortie du canal ne dépond que de l'élément binaire entrant correspondant, le canal est appelé sans mémoire[21] .

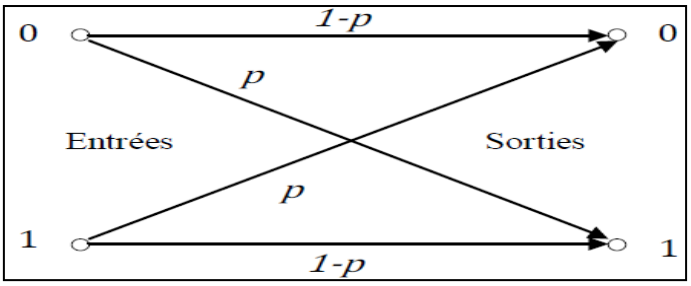


**Figure. II.8:** Diagramme du canal binaire symétrique

### II.2.5.2. Canal à bruit additif blanc gaussien :

Le modèle de canal le plus fréquemment utilisé pour la simulation de transmission numérique, qui est aussi un des plus faciles à générer et à analyser, est le canal à bruit blanc additif gaussien (BBAG). Ce bruit modélise à la fois les bruits d'origine interne (bruit thermique dû aux imperfections des équipements…) et le bruit d'origine externe (bruit d'antenne…). Ce modèle est toutefois plutôt associé à une transmission filaire, puisqu'il représente une transmission quasi-parfaite de l'émetteur au récepteur [22] .

Le signale reçu s'écrit alors :

$$r(t) = s(t) + v(t) \qquad \textbf{( II.6)}$$

où v(t) représente la BBAG, caractérisé par un processus aléatoire gaussien de moyenne nulle, de variance $\sigma_v^2$ et de densité spectrale de puissance bilatérale $\Phi vv = \frac{N0}{2}$ ou $N_0 = 2\sigma_v^2$
La densité de probabilité conditionnelle de r est donnée par l'expression :

$$p(r|s) = \frac{1}{\sqrt{2\pi}\sigma v} e^{-\frac{(r-s)^2}{2\sigma v^2}} \qquad \textbf{(II.7)}$$

### II.2.5.3. Canal de Rayleigh :

Dans les liaisons radio mobiles, les canaux de transmission évoluent en fonction du temps à cause des déplacements aléatoires des entités communicantes et l'existence d'obstacles entre l'émetteur et le récepteur. Il peut en résulter que le signal émis suit plusieurs trajets avant d'arriver au récepteur, conduisant à une variabilité importante du signal reçu due à l'addition de plusieurs signaux déphasés. Lorsque le débit de transmission est suffisamment faible, chaque symbole ne se superpose qu'avec lui-même, au moins sur une portion significative de sa durée. Un canal de Rayleigh permet de prendre en compte ces effets : réflexions multiples, évanouissements, fluctuations à grande et petite échelle et effet Doppler. L'amplitude et la phase du signal reçu apparaissent comme des variables aléatoires qui suivent une loi

de Rayleigh (équation II.8). Ce modèle est particulièrement adapté à une représentation statistique d'un canal radio mobile [23] .

$$p(x) = \frac{x^2}{\sigma^2} exp\left(\frac{-x^2}{2\sigma^2}\right) \quad \textbf{(II.8)}$$

### II.2.6. Démodulation :

La démodulation permet de récupérer chaque symbole émis à partir de chaque signal modulé reçu de durée . Le démodulateur fournit une tension continue ou un symbole binaire dans le cas où sa sortie est quantifiée. Le démodulateur fournit au bloc décodeur une séquence binaire $\widetilde{K}(a_i)$ qui représente l'information émise à laquelle est superposée une séquence d'erreur :

$$E(a_i)\ , \widetilde{K}(a_i) = K(a_i) + E(a_i) \quad \textbf{(II.9)}$$

## II.2.7. Décodage canal :

Comme le décrit le théorème fondamental du codage canal, pour se rapprocher de la capacité du canal de transmission, il est nécessaire de coder l'information avant de la transmettre. Au niveau du récepteur, le décodage canal consiste dans un premier temps à détecter la présence d'erreurs dans l'information et puis dans un deuxième temps de les corriger.

A partir de ces deux actions découlent trois principales stratégies : les stratégies ARQ (Automatic Repeat Request) qui se limitent à détecter la présence d'éventuelles erreurs ,la correction s'effectuant par retransmission des blocs erronés ; les stratégies FEC (Forward Error Correction) mettant en œuvre les codes permettant la détection et la correction des erreurs sans aucune retransmission ; enfin les systèmes hybrides combinent entre les deux techniques [24] .

## II.2.8. Décodage source :

Le décodage source consiste à reconstituer, par l'application de l'algorithme de décodage source (décompression par exemple), l'information originelle à partir de la séquence de substitution «$K(a_i)$».

# II.3. Techniques de codage de signal parole :

L'objectif d'un codec est d'obtenir une bonne qualité de voix avec un débit et un délai de compression les plus faibles possibles. Traditionnellement on distingue 3 classes de techniques de codage de la voix, le codeur forme d'onde, codeur paramétrique et codeur hybride **.**

## II.3.1. Les codeurs d'onde :

Essayent de coder la forme exacte de la forme d'onde de son articulé, sans considérer la nature de la production de la parole et de la perception du langage humain . .
Ces codeurs appropriés mieux au codage de débit binaire élevé depuis des baisses d'exécution rapidement avec la diminution de débit binaire [25] .

## II.3.2. Les codeurs paramétriques :

Modélise le processus de fabrication du son et extrait les paramètres convenables, qui sont transmis au décodeur. Les voici qui sont habitués pour reconstruire une forme d'onde qui est souvent très différente de celle du signal original mais qui produit un son qui est semblable ou près de l'original. Cette technique de codage fonctionne bien pour de bas débits binaires .Augmentant le débit binaire normalement ne traduit pas une meilleure qualité, puisqu'elle est limitée par le modèle choisi.

## II.3.3. Codeurs Hybrides :

Combine la force d'un codeur de forme d'onde avec celle d'un codeur paramétrique comme un codeur paramétrique, il se fonde sur un modèle de production de la parole
comme dans des codeurs de forme d'onde, une tentative est faite d'assortir le signal original avec le signal décodé dans le domaine de temps , Cette technique domine les codeurs de débit binaire moyen.

Les codeurs de la parole opèrent habituellement en simples blocs, qui doivent être accumulés avant que le traitement commence. La taille des blocs d'entrée varie avec le codeur de voix utilisé. Par exemple, le codeur G.729 segmente le discours d'entrée dans 80 échantillons (10ms) et assigne 80 bits par bloc codé, menant à un débit binaire de 8 Kbps [26] .

- le tableau II -1 présentation quelque Les standards de codage de la voix :

| Classe de codage | Technique de codage | Standard | Débit binaire |
|---|---|---|---|
| Forme d'onde | PCM<br>ADPCM | G.711<br>G.726 | 64 kbps<br>16-40 kbps |
| Paramétrique | LPC | FS 1015 | 2.4 kbps |
| Hybride | CELP<br>LD-CELP<br>CS-ACELP | FS 1016<br>G.728<br>G.729 | 4.8 kbps<br>16 kbps<br>8 kbps |

**Tableau II -1** : Les standards de codage de la voix

## II.3.4 Les Codeurs d'un signal de parole :

- Il ya plusieurs types des codeurs de la parole :

### II.3.4.1 Codeur PCM (ITU-T G.711) :

Les recommandations de la méthode PCM (Pulse Code Modulation) ont et regroupes en 1972 dans l'ITU-T G.711 (International Télécommunication Union - Télécommunication standardization) [27]. Le codeur PCM traite des signaux audio sur la bande téléphonique, considérée comme la bande étroite (Narrow band en anglais), allant de 300 a 3400 Hz. A une fréquence d'échantillonnage de 8000 Hz, le standard permet un débit de 64 kbps (avec 8 bits de quantification pour chaque échantillon). La méthodes de codage est dite sans compression. A ce débit, la qualité de parole obtenue constitue la référence par rapport a laquelle sont juges la majorité des autres codeurs.

. Le fonctionnement du codeur PCM est illustre par la figure II.9. Le codeur se base sur une méthodes de compression suivant une loi non linéaire, la loi mu (loi μ en Amerique du Nord et au Japon, similaire a la loi A utilisée en Europe) décrite par l'équation suivant :

$$y[n] = x_{max} \frac{\ln(1+\mu|x[n]|x_{max})}{\ln(1+\mu)} \text{signe}(x[n]) \qquad \textbf{(II.10)}$$

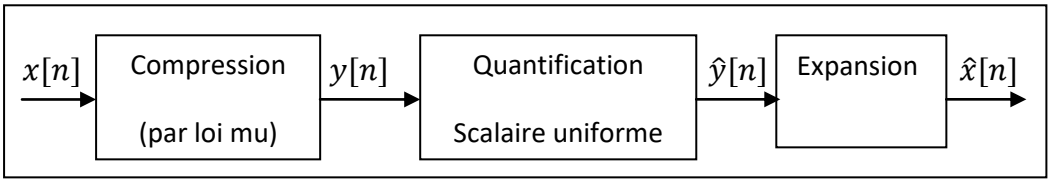


**Figure. II.9**- Schéma bloc du PCM.

### II.3.4.2 Codeur DPCM (Différentia Pulse Code Modulation) :

L'un des challenges du codage de la parole est de pouvoir transférer de la parole avec le moins de débit possible. Pour ce faire, on peut exploiter la corrélation existant entre les échantillons consécutifs du signal de parole. On transmettre donc la différence entre deux échantillons consécutifs permettant ainsi le codage de l'information sur un minimum de bits. Soit $s(n)$ le signal à transmettre. L'émetteur calcule la différence $e(n) = s(n) - s(n-1)$ qui sera ensuite quantifiée puis transmise au décodeur. La transmission de la différence entre deux échantillons consécutifs est une prédiction d'ordre 1.

Plus généralement, on peut estimer $s(n)$ par :

$$\tilde{s}(n) = \sum_{k=1}^{p} a_k \, s(n-1) \qquad \textbf{(II.11)}$$

On parle alors de prédiction linéaire d'ordre p . Les coefficients $a_k$ sont appelés coefficients de prédiction du filtre de prédiction dont la fonction de transfert est :

$$A(z) = \sum_{k=1}^{p} a_k z^{n-k} \qquad \textbf{(II.12)}$$

La Figure II.10 illustre le fonctionnement du codage DPCM , Sur cette figure, on voit que l'erreur de prédiction en entrée du codeur vaut $e(n) = s(n) - \tilde{s}(n)$ .

L'erreur de prédiction est ensuite quantifiée en $e_q(n)$ (avec un débit faible car sa variance est beaucoup plus faible que celle du signal) puis tramise au décodeur [28]. L'erreur de prédiction quantifiée est également ajoutée à un signal prédit $\tilde{s}(n)$ pour reconstruire le signal $s'(n)$ Lorsqu'il n'y a aucune erreur de transmission (canal de transmission parfait), alors le signal reconstruit $s(n) = \hat{s}(n)$, où $\hat{s}(n)$ est le signal donné par le décodeur .

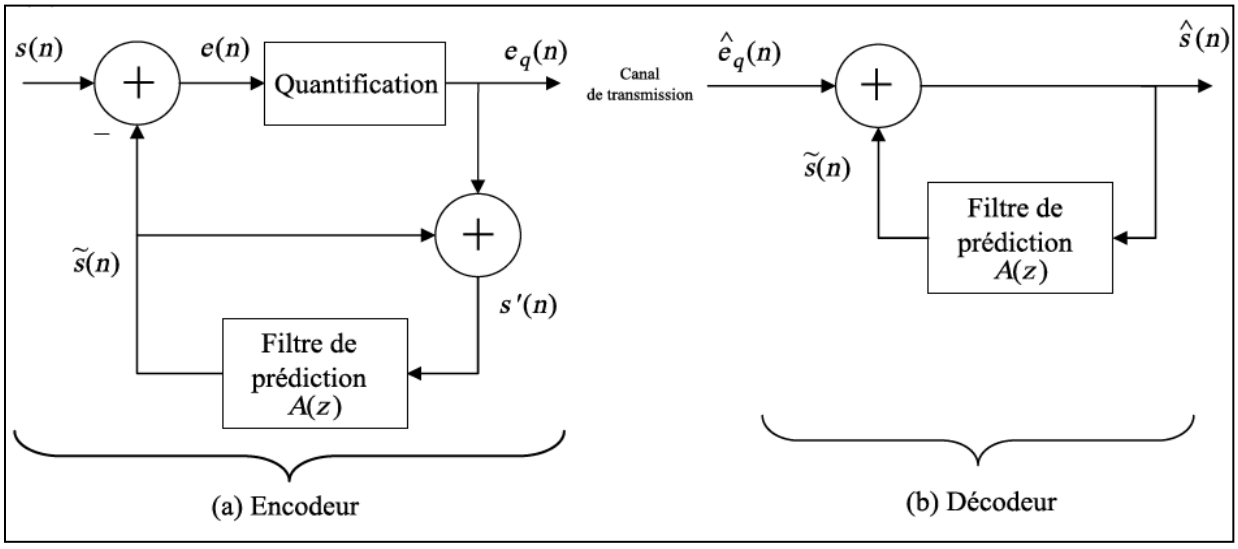


**Figure II.10 :** Schéma d'un DPCM

### II.3.4.3. Codeur ADPCM (Adaptive DPCM) :

Le codage DPCM peut être amélioré en le rendant adaptatif. Le processus adaptatif peut être réalisé selon deux méthodes pouvant être simultanément utilisées : la quantification adaptative et la prédiction adaptative. On distingue deux types de quantification adaptative :
la quantification adaptative progressive qui adapte le pas de quantification en fonction du signal à coder et la quantification adaptative rétrograde basée sur le signal déjà quantifié.
La quantification adaptative progressive est plus précise, puisque basée sur le signal original. Cependant, elle requiert la transmission d'informations du quantificateur au décodeur.

. La quantification adaptative rétrograde est exempte de ce problème de transmission d'informations, mais est moins précise. Cela s'explique par le fait que le signal utilisé pour l'adaptation est déjà entaché par le bruit de quantification [29] .

. La prédiction adaptative met à jour les paramètres du prédicateur de manière périodique afin de minimiser la variance du signal à coder. De la même manière que dans le cas de la quantification adaptative, on peut réaliser une prédiction adaptative progressive ou rétrograde.
Le codage ADPCM implémenté dans le codeur G.726 (IUT-T 1990) est utilisé dans la téléphonie fixe sans fil (DECT ou Digital Enhanced Cordless Téléphone). Comme autre exemple implémentant la technique DPCM , nous pouvons citer le codeur G.722 (ITU-T 1988) très utilisé en téléphonie sur IP.

- Il y a deux façons d'introduire l'adaptation :

• **Avant adaptation** : l'adaptation est réalisée à la fin de l'émetteur. Ce régime exige la transmission des paramètres d'adaptation (coefficients de prédiction linéaire ou de niveau de signal) comme information adjacente. Il s'agit de l'introduction d'un retard.

• **Adaptation en arrière** : les paramètres d'adaptation sont prélevés sur le signal décodé lui-même, où il n'est pas nécessaire d'envoyer les informations du côté, La figure II.11 montre les codeurs ADPCM, correspondant à une adaptation en avance et en arrière des coefficients de prédiction linéaire.

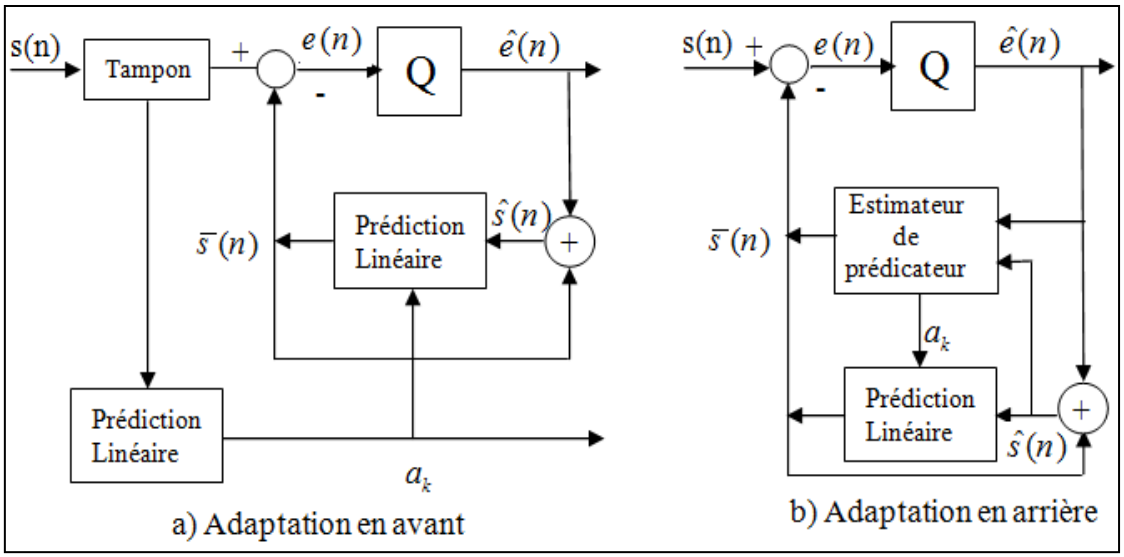


**Figure II.11 :**Avant et en arrière des codeurs ADPCM

### II.3.4.4. Codeur CELP :

Le codage CELP *(* code excited linear prediction ) inventé par Atal et Schroeder dans les années 1980, permet d'améliorer la qualité proposée par les codeurs MPE et RPE pour de faibles débits, les débits des codeurs CELP s'étalant entre 5 et 12 kbit/s. Pour ce faire, comme on peut le constater sur la Figure II.12 (a), le codeur CELP utilise deux dictionnaires d'excitations : un adaptatif et l'autre fixe. Soit $u(n)$ le signal d'excitations global, il s'écrit :

$$u(n) = \beta c_k(n) + g c_a(n) \qquad \textbf{(II.13)}$$

où $\beta$ et $g$ sont les gains respectifs des dictionnaires adaptatif et fixe correspondant aux vecteurs d'excitations $c_k$ et $ca$ respectivement. Le dictionnaire adaptatif sert à modéliser la périodicité du signal (signal voisé) tandis que le dictionnaire fixe ou stochastique modélise les segments non voisés du signal. Il faut toutefois préciser que dans le codage CELP d'origine, la modélisation du pitch est réalisée via un filtre de synthèse long terme [30] .

Le modèle que nous présentons dans cette section est celui implémenté dans le FS1016 CELP (Federal Standard 1016 CELP) (Campbell, Tremain et al. 1991).

Dans le FS1016 CELP, le signal en entrée est échantillonné à 8 kHz puis divisé en trames de 30 ms qui sont elles-mêmes subdivisées en 4 sous-trames. Le codage CELP est un codage AbS où l'analyse LPC est faite sur les 4 sous-trames. L'erreur de prédiction issue de cette analyse alimente ensuite le filtre LTP (long-terme) afin d'extraire le pitch du signal. L'analyse LTP est réalisée sur les sous-trames, car les coefficients du filtre LTP doivent être mis à jour plus fréquemment en raison de la nature très dynamique du pitch. La recherche de l'ordre filtre du LTP correspond à la recherche de l'indice du vecteur d'excitations du dictionnaire

adaptatif. En effet le signal d'excitation adaptatif $c_a(n)$ correspond au signal d'excitation global retardé

$$c_{\mathrm{a}}(\mathrm{n}) = u(n-i) \qquad \textbf{(II.14)}$$

La variable $a$ est l'indice du vecteur d'excitations du dictionnaire adaptatif et k est l'indice du vecteur d'excitations du dictionnaire fixe.

Soient $h_w(n)$ la réponse impulsionnelle du filtre de pondération du signal synthétisé et $h_p(n)$ celle du filtre de pondération perceptuelle [31]. Leurs transformée en Z sont respectivement $1/A(z/\gamma)$ et $A(z)\ /\ A(z/\gamma)$ Si on note $\hat{s}_0(n)$ la valeur initiale du signal synthétisé alors le signal synthétisé pondéré $\hat{s}_w(n)$ s'écrit :

$$\hat{s}_w(n) = h_w(n) * u(n) + \hat{s}_0(n) \qquad \textbf{(II.15)}$$

$$\hat{s}_w(n) = \beta c_{\mathrm{k}}(\mathrm{n}) * h_w(n) + g c_{\mathrm{a}}(\mathrm{n}) * h_w(n) + \hat{s}_0(n) \qquad \textbf{(II.16)}$$

Soit $s(n)$ le signal original et $\hat{s}_w(n)$ le signal après l'avoir pondéré par le filtre perceptuel. On a :

$$\hat{s}_w(n) = s(n) * h_p(n) \qquad \textbf{(II.17)}$$

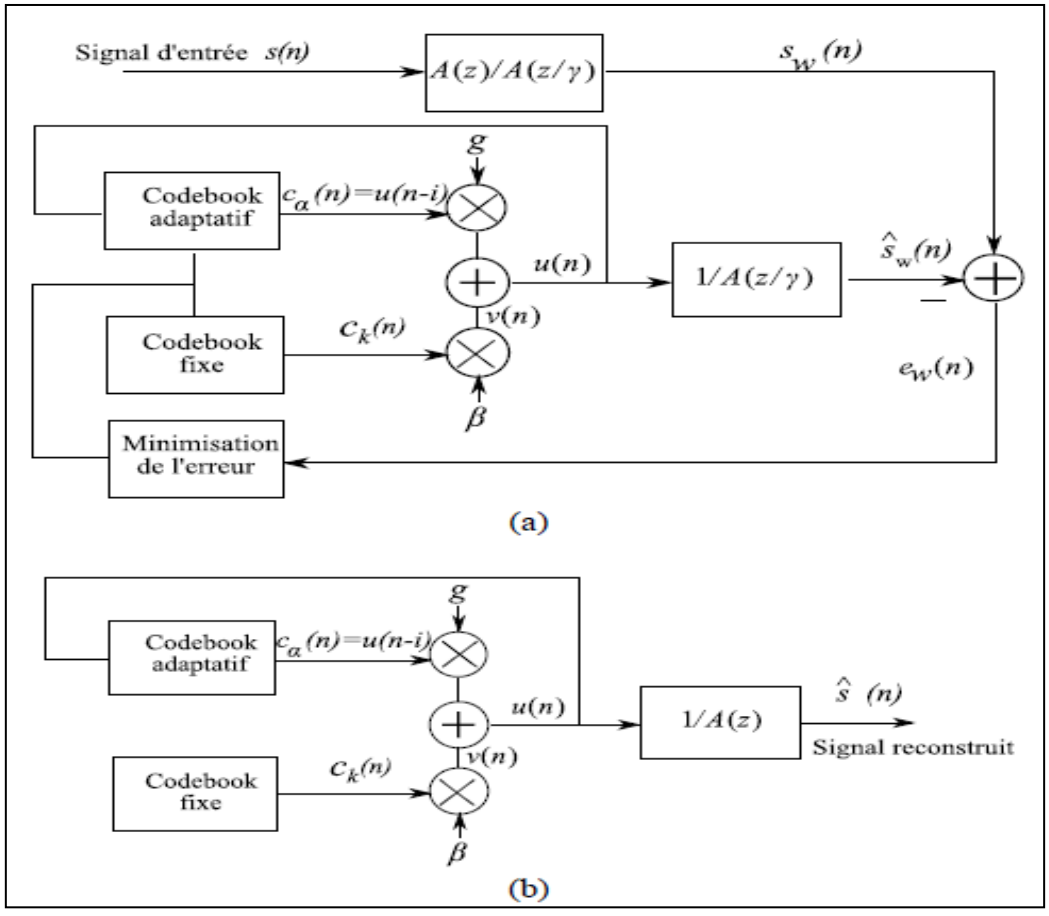


**Figure II.12** : Principe du codage CELP, (a) Encodeur CELP et (b) Décodeur CELP

### II.3.4.5 Codeur AMR-NB :

L'AMR (Adaptive Multi Rate) est un codeur de source reconnu sous la spécification technique 3GPP TS 26.090 [32]. II consiste en un codeur bande étroite a débit variable (AMR-Narrow Band) avec un dispositif de contrôle du débit de la source, un détecteur d'activité vocale (VAD, Voice Activity Detector), un générateur de bruit de confort et un mécanisme de récupération des paquets perdus durant la transmission. Le codec traite des signaux audio sur une bande allant de 300 a 3400 Hz. L'adaptation fonctionne a huit débits allant de 4,75 kbps a 12,2 kbps pour une longueur de trame de 20 ms a 8 kHz. Le codage du signal s'effectue grâce a un codeur prédictif et un dictionnaire d'excitation algébrique

## II.4. Conclusion :

Dans ce chapitre, nous avons présenté les mécanismes de codage de la parole qui peuvent être sépares en trois catégories: le codage de forme d'onde, le codage de source ou le codage hybride . Le codage est un processus permettant de réduire le flux numérique a des débits inferieurs pouvant aller jusqu'a quelques kbps (kilo bits par seconde) avec des niveaux de qualité différents selon le modèle utilisé dans l'algorithme de codage .

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. Le décodeur utilise ces paramètres pour reconstruire un signal de parole synthétique, le type de codage est un paramètre important dans la qualité de signal.

# *Chapitre III*
# *Simulation des différents codeurs*

# III.1. introduction :

Ce chapitre est consacré aux résultats obtenus lors de notre simulation, une évaluation des performances des différents codeurs (PCM , DPCM, ADPCM ) sont présentés. Dont, Pour cela, nous avons utilisé le logiciel MATLAB pour simuler notre chaîne de transmission (figure III.1) . Dans ce chapitre le modèle et les paramètres de simulation et les différents résultats de simulation . Une étude comparative, des codeurs PCM, DPCM et ADPCM en termes de PESQ, BER (Taux d'errer par bit) et l'évaluations de temps d'exécution .

# III.2. Simulation d'une chaine de transmission numérique :

## III.2.1. Modèle et les paramètres de simulation :

Nous avons utilisé des codecs PCM, ADPCM et DPCM, dont leurs coefficients sont convertis en une séquence binaire. Avant d'utiliser la modulation PSK nous introduisons un code correcteur d'erreur (FEC) pour mettre le système plus robuste aux erreurs de canal de transmission, le signal codé est transmis à travers le canal AWGN.

Après démodulation (PSK), décodage de convolution, et décodage de source (PCM, DPCM, ou ADPCM), les données binaires sont reconverties en un signal parole synthétisée. Notre chaine de transmission détaillé et utilisée lors de notre simulation est représenté par la figure (III.1) :

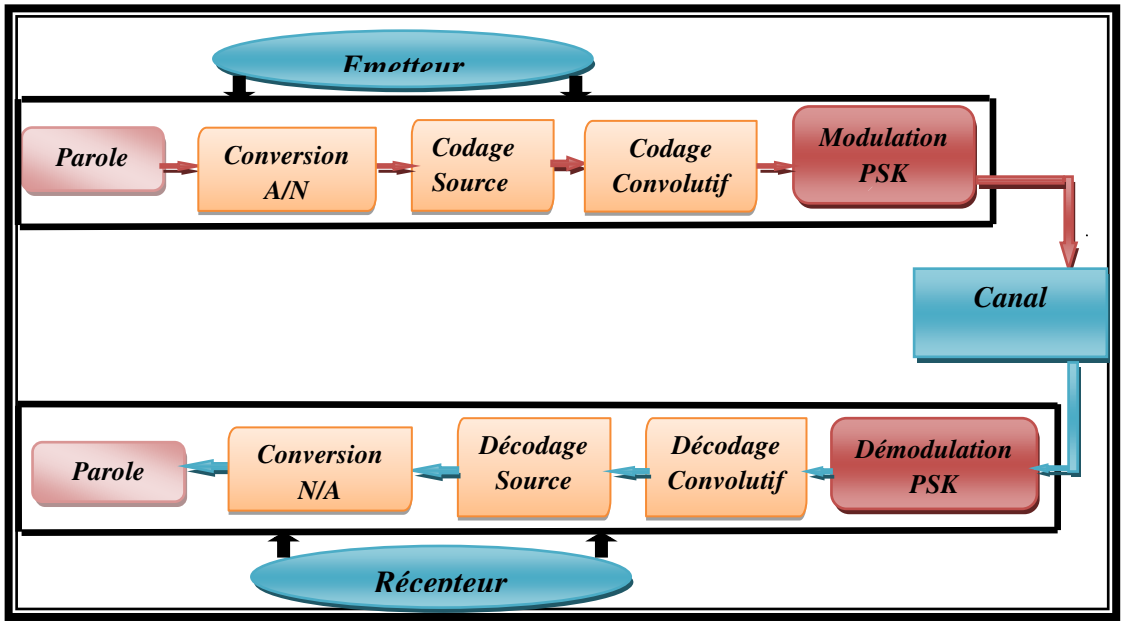


**Figure III.1 -** Modèle de simulation.

Les principaux paramètres de simulation sont résumés dans le Tableau (III.1):

| **Paramètre** | **Caractéristique** |
|---|---|
| Programme | Matlabe 2013 |
| Types des Codes de la source | PCM / DPCM / ADPCM |
| Type de codeur du canal | Convolutif ½ |
| Type du canal transmission | Canal Gaussien à bruit blanc additif (AWGN) |
| Type de modulation | PSK (Phase Shift Keying) |
| Le critère d'évaluation des performances | PESQ / BER ( en fonction de SNR) |

**Tableau III.1**- Les paramètres de simulation

### III.2.2. But de simulation :

Dans ce travail, Nous avons essayé de faire une étude comparative des codages PCM ,DPCM ,ADPCM pour savoir quelle est la techniques de codage soit meilleur point de vue PESQ et BER vis-à-vis SNR?

### III.2.3. Rapport signal sur bruit :

Toute dégradation du rapport signal sur bruit dans une bande de fréquence donnée, perturbe la qualité de l'ensemble du canal donc réduit la capacité globale de l'accès.

Cette diminution de la capacité revient à diminuer d'un bit de la taille du symbole de la constellation, c'est-à-dire a réduire par deux les performances.

Rapport signal bruit => $S/B = 10\, log_{10}(P_S\,(Watt) / P_B(Watt))$.

Où $P_S$ et $P_B$ désignent respectivement les puissances du signal et du bruit.

## III.2.4. Signal de la parole

Le signal de parole qui est" *She had your dark suit in greasy wash water all year*" extrait de la base de données TIMIT [33] sont échantillonnés avec une fréquence d'échantillonnage « Fs» de 16000 échantillons/s ,et sous échantillonné à 8000 Hz (le codeur PCM travil avec le signaux échantillonnées à8000Hz). La figure (III.2) représente le signal de la parole original :

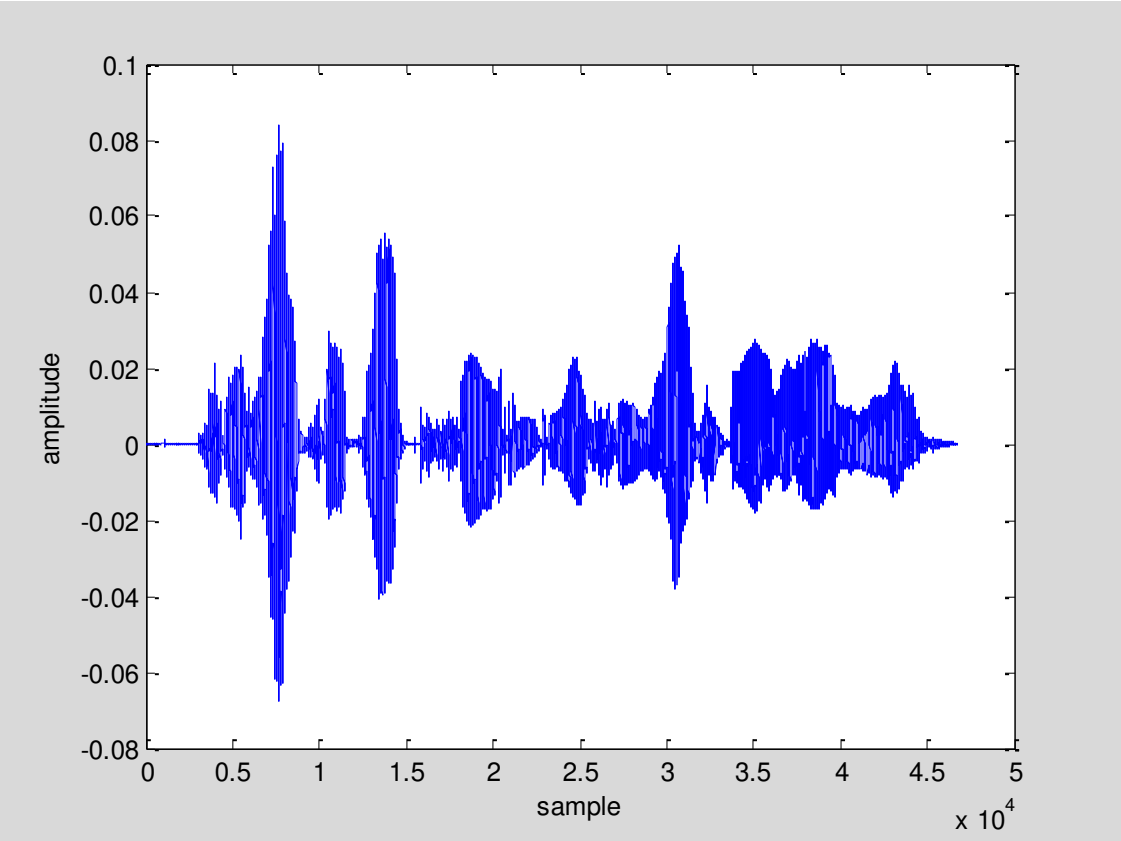


**Figure III.2-** représente le signal parole original

# III.3. Les résultats de simulation et discussions :

## III.3.1 Evaluation de la qualité point de vue PESQ

La figure III.4 illustre une comparaison des performances des techniques des codeurs de la parole.

- Cette simulation consiste à évaluer le PESQ en fonction du rapport signal bruit (SNR) pour différents niveaux d' SNR, de -5 à 30 dB par un pas de 5 dB

- On remarque d'après cette figure (II.4) que l'encodeur ADPCM a donné de meilleur résultat de PESQ mais pour des valeurs d'SNR allant de -5 à 10 et même encodeur a donné de faible résultat pour des valeurs d'SNR allant de 10 à 30 cela en comparaison avec les autres l'encodeur PCM et DPCM..

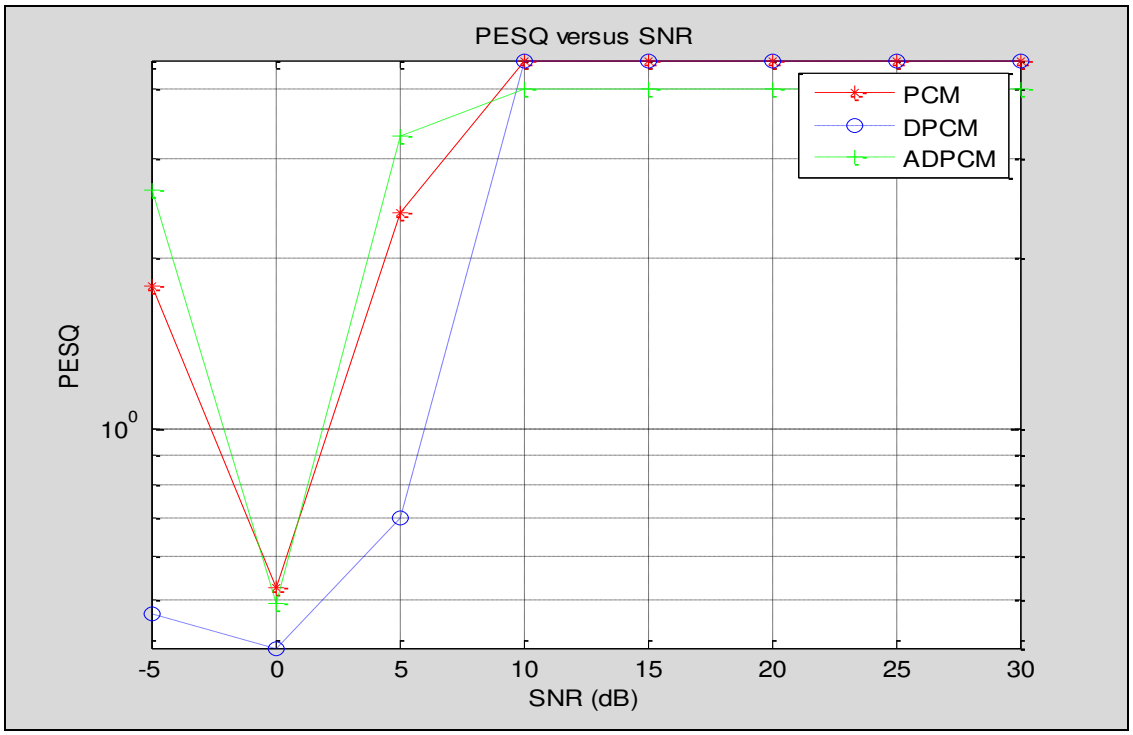


**Figure III.3-** Étude comparative des techniques de codage: PCM ,DPCM ,ADPCM en terme de PESQ versus SNR

- Le tableau III.2 représente les mesures moyennes de PESQ en fonction de SNR pour les technique de codage de la parole mentionnées précédemment ( PCM ,DPCM ,ADPCM ).
- On remarque dans ce tableau que l'encodeur PCM donnée meilleur résultat de moyenne de PESQ cela en comparaison avec les autres l'encodeur DPCM et ADPCM.

| SNR | - 5 | 0 | 5 | 10 | 15 | 20 | 25 | 30 | moyenne |
|---|---|---|---|---|---|---|---|---|---|
| PESQ-PCM | 1.7929 | 0.5260 | 2.4044 | 4.4506 | 4.4506 | 4.4506 | 4.4506 | 4.4506 | 3.3720 |
| PESQ-DPCM | 0.4742 | 0.4129 | 0.8530 | 4.4798 | 4.4798 | 4.4798 | 4.4798 | 4.4798 | 3.0173 |
| PESQ-ADPCM | 2.6416 | 0.4943 | 3.3023 | 3.9800 | 3.9800 | 3.9800 | 3.9800 | 3.9800 | 3.2922 |

**Tableau III.2** - résultats d'évaluation par technique PESQ

## III.3.2 Evaluation du BER en fonction du SNR :

le taux d'erreur par bite BER (Bit Error Rate ), qu'on définit par la relation suivante :

$$BER = \frac{\text{nombre de bits érronés}}{\text{nombre de bits transmis}}$$

Cette simulation consiste à évaluer le taux d'erreur par bit en fonction du rapport signal bruit (SNR) pour différents niveaux d' SNR, de -5 à 10 dB par un pas de 2 dB.

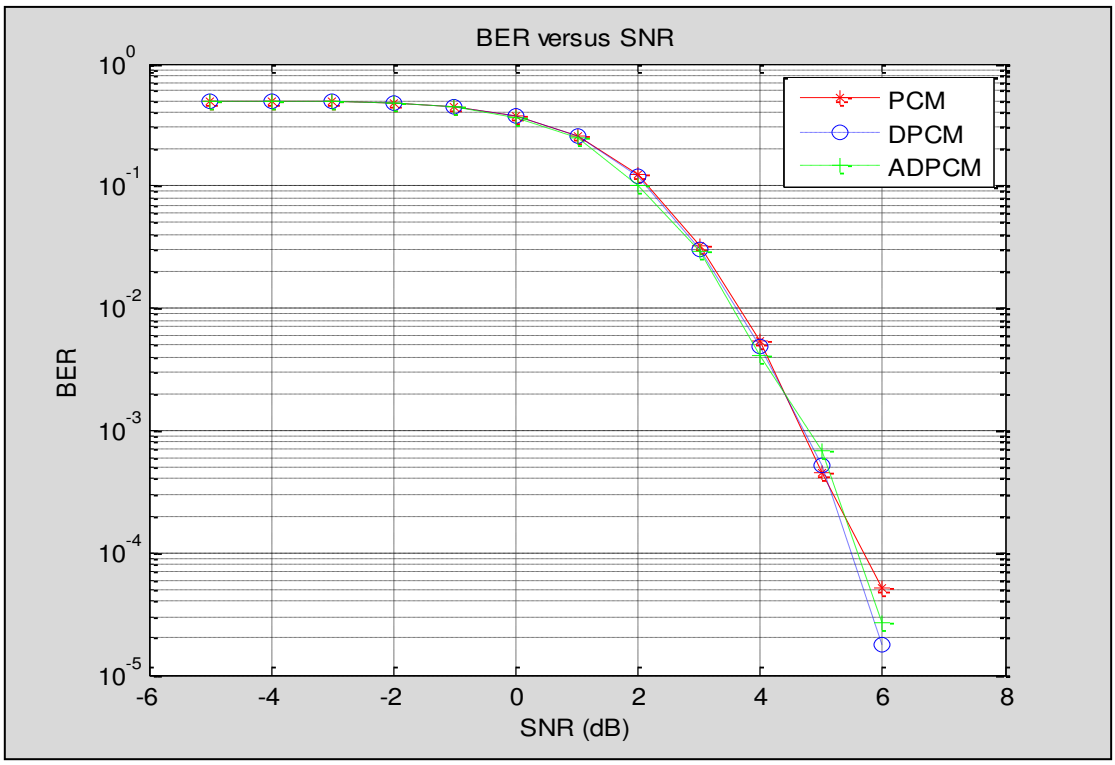


**Figure III.4-** Étude comparative des techniques de codage: PCM , DPCM , ADPCM en terme de BER versus SNR

- On remarque dans cette figure de Convergence les courbes de BER pour des valeurs d'SNR allant de -5 à 5 mais des différent chez d'SNR égale 6

| SNR | -5 | -3 | -1 | 1 | 3 | 5 | 7 | 9 |
|---|---|---|---|---|---|---|---|---|
| BER-PCM | 0.4996 | 0.4907 | 0.4506 | 0.2550 | 0.0330 | 0.0006 | 0 | 0 |
| BER-DPCM | 0.4996 | 0.4890 | 0.4467 | 0.2531 | 0.0318 | 0.0005 | 0 | 0 |
| BER-ADPCM | 0.4966 | 0.4869 | 0.4381 | 0.2436 | 0.0264 | 0.0003 | 0 | 0 |

**Tableau III.3-** résultats d'évaluation par technique BER

D'après le tableau III.3 qui montre les mesures de BER en fonction de SNR pour les techniques codage de la parole mentionnée précédemment (PCM, DPCM, ADPCM), Nous pouvons observer bien que l'encodeur ADPCM donne-nous de faible BER pour toutes les valeurs de SNR par rapport les autres codecs PCM et DPCM .

## III.3.3 Temps d'exécution des codecs :

- Le tableau III.4 représente des valeurs de simulation les temps d'exécution des codecs PCM , DPCM , ADPCM par le technique de MATLAB (tic / toc ) :

| Codeurs | PCM | DPCM | ADPCM |
|---|---|---|---|
| Temps d'exécution [sec] | 2.891296 | 3.977003 | 1.666993 |

**Tableau III.4:** Temps d'exécution des codecs : PCM, DPCM et ADPCM.

- On remarque dans ce tableau que l'encodeur DPCM prenait du long temps d'exécution et l'encodeur ADPCM prendre moins du temps.

# III.4 CONCLUSION :

Ce chapitre a fait l'objet d'une étude comparative des différentes les techniques des codeurs du signal parole ( PCM, DPCM, ADPCM) . On peut résumer les résultats de ce travail :

D'après les mesures moyennes de PESQ on peut conclure que l'encodeur PCM est le meilleur, par rapport au codec DPCM et ADPCM.

D'après les mesures de BER on peut conclure que l'encodeur ADPCM est le meilleur codeur puisqu'il nous donne de faible BER avec toutes les valeurs d' SNR.

D'après l'évaluations en termes de temps d'exécution on peut conclure que l'encodeur ADPCM soit le meilleur et cela en comparaison avec les d'autres codec PCM et DPCM.

# *conclusion générale*

Le but de codage du signal parole est de représenter le signal par le moins de bits possible, mais maintenir un certain niveau de qualité du signal parole décodé. En général, les techniques de codage sont basées sur la réduction de la redondance du signal, afin que le signal résiduel (après réduction de redondance) puisse être encodé avec moins de bits que l'originale. Afin d'évaluer un codeur de parole, nous devons tenir compte de: débit binaire, qualité de la parole décodée, coût de calcul, retard introduit et robustesse contre l'acoustique et la dégradation de canal.

Les mesures objectives de la qualité de signal parole telle que le SNR peuvent être utiles, mais ils ne peuvent pas refléter la façon dont un signal décodé est perçue.

Les codeurs de la parole peuvent être classés en forme d'ondes, paramétriques et codeurs hybrides.

Notre travail est divisé en trois chapitres, le premier chapitre a illustré en général que est ce qu'un signal parole, comment fonctionner l'appareil vocal et la numérisation de signal parole qui parcourent en trois étapes : l'échantillonnage, la quantification et le codage.

Dans le deuxième chapitre nous avons présenté les différentes techniques de codages de la parole notamment les codeurs qui conforment à l'IUT (International Union of Télécommunications) comme PCM, DPCM, ADPCM.

Enfin le dernier chapitre contient les résultats de simulation de la comparaison des différents codeurs PCM, DPCM, ADPCM en mesurant les critères objectifs tels que PESQ, BER et l'évaluation en termes de temps d'exécution et cela à l'aide du logiciel de MATLAB et la base de donnée TIMIT. Les résultats de simulation ont démontré que le meilleur codec est l'encodeur PCM point de vue les mesures moyennes de PESQ. Point de vue BER et temps d'exécution, l'encodeur ADPCM est le meilleur.

Enfin ce mémoire nous avons conclu que l'encodeur ADPCM soit la meilleure technique qui code et élimine la redondance d'un signal de parole à travers un canal bruité (AWGN).

Comme perspective une étude sur d'autres codecs est intéressante.

# *Références bibliographiques*

## Références bibliographiques

[1] CHERADID, Abdellatif. Etude des approches adaptatives liées àla QoE dans le cadre des applications deTéléphonie sur IP. Thèse de doctorat.2013.

[2] ZANGO, Tiraogo Abdoulaye Yves. Évaluation subjective de la qualité: proposition d'un système de référence pour les codecs en bande élargie. 2013. Thèse de doctorat. Rennes 1.

[3] CHEMSA, A., LABBI, Y., HETTIRI, M., et al. New semi-blind approach to optimize turbo decoding for a cauchy α-stable impulsive noise channel. Journal of Fundamental and Applied Sciences, 2018, vol. 10, no 2.

[4] Y.Aziza. Modélisation AR et ARMA de la parole pour une vérification robuste du locuteur dans un milieu bruité en mode dépendant du texte. Mémoire de Magister. Université FERHAT ABBAS Setif.2013.

[5] AMEHRAYE, Asmaa. Débruitage perceptuel de la parole. 2009. Thèse de doctorat. Télécom Bretagne.

[6] KWONG, Mylène. Transmission efficace en temps réel de la voix sur réseaux ad hoc sans fil/Mylène D. Kwong. Library and Archives Canada= Bibliothèque et Archives Canada, Ottawa, 2009.

[7] BRAHIMI, Zouheira et ZELLOUMA, Hadjra. Effet d'HPA sur le système SC-FDMA. 2015.

[8] AZIZA, Yassamine. Modélisation AR et ARMA de la Parole pour une Vérification Robuste du Locuteur dans un Milieu Bruité en Mode Dépendant du Texte. 2018. Thèse de doctorat.

[9] LADHEM Bouchra, (mémoire de Master). Etude l'association des antennes MIMO à Maximisation du rapport signal sur bruit avec la technique Multi-porteuses OFDM. Universite de Telemcen. 2015

[10] VERNEUIL, Jean-Louis. Simulation de systèmes de télécommunications par fibre optique à 40 Gbits/s. Université de Limoges, Limoges, 2003, p. 297.

[11] BOUAOUADJA, N., MADJOUBI, M., KOLLI, M., et al. Etude des possibilités d'amélioration de la transmission optique d'un verre sodocalcique érodé par sablage. 2011..

[12] CHARBIT, Maurice. Communications Numériques et Théorie de l'Information. 2005.

[13] KÖVESI, Balazs, SAOUDI, Samir, BOUCHER, Jean-Marc M., et al. Algorithme rapide de codage conjoint source-canal pour un canal de transmission non-stationnaire. In : GRETSI'97: seizième colloque du groupe d'études du traitement du signal et des images. 1997. p. 1065-1068.

[14] RIMA, Boudjaja. Transmission par paquets de la video en temps réel sur les réseaux sans fils. 2008. Thèse de doctorat. Université de Béjaia-Abderrahmane Mira.

[15] Deng, Chenwei, et al. Visual signal quality assessment. Springer International Pu, 2016..

[16] DIOUF, Madiagne. Conception Avancée des codes LDPC binaires pour des applications pratiques. 2015. Thèse de doctorat. Université Cergy Pontoise; Université Cheikh Anta DIOP.

[17] Marie, O. G. E. R. Docteur en Sciences. Diss. INRIA Rennes, 2007.

[18] BOITE, René. Traitement de la parole. PPUR presses polytechniques, 2000.

[19] BAUDOIN, Geneviève, CERNOCKÝ, Jan, GOURNAY, Philippe, et al. Codage de la parole a bas et tres bas debits. Ann. des Télécommunications, 2000, vol. 55, no 9-10, p. 462-482.

[20] AJGOU, Riadh. Reconnaissance Automatique du Locuteur à Travers les Canaux Digitaux. 2016. Thèse de doctorat. Université Mohamed Khider–Biskra..

[21] AJGOU, Riadh, SBAA, Salim, GHENDIR, Said, et al. Effects of speech codecs on a remote speaker recognition system using a new SAD. In : Proceedings of the 2014 International Conference on Systems, Control, Signal Processing and Informatics II (SCSI'14). Prague, Czech Republic April. 2014. p. 2-4.

[22] Ajgou, Riadh, et al. "Robust speaker identification system over AWGN channel using improved features extraction and efficient SAD algorithm with prior SNR estimation." International Journal of Circuits, Systems and Signal Processing 10 (2016): 108-118..

[23] AJGOU, Riadh, SBAA, Salim, GHENDIR, Said, et al. New Speech Enhancement Method based on Wavelet Transform and Tracking of Non Stationary Noise Algorithm. RECENT ADVANCES on ELECTROSCIENCE and COMPUTERS, 2015, p. 45.

[24] Ajgou, Riadh, et al. "Speaker Recognition System Based on AR-MFCC and SAD Algorithm with Prior SNR Estimation and Adaptive Threshold over AWGN channel." International Conference on Recent Advances in Electrical Engineering and Educational Technologies. 2014.

[25] AFFONSO, Emmanuel T., ROSA, Renata L., et RODRIGUEZ, Demostenes Z. Speech quality assessment over lossy transmission channels using deep belief networks. IEEE Signal Processing Letters, 2017, vol. 25, no 1, p. 70-74.

[26] AJGOU, Riadh, SBAA, Salim, GHENDIR, Said, et al. Novel detection algorithm of speech activity and the impact of speech codecs on remote speaker recognition system. WSEAS TRANSACTIONS on SIGNAL PROCESSING. 2014.

[27] PFAU, Thilo, ELLIS, Daniel PW, et STOLCKE, Andreas. Multispeaker speech activity detection for the ICSI meeting recorder. In : IEEE Workshop on Automatic Speech Recognition and Understanding, 2001. ASRU'01. IEEE, 2001. p. 107-110.

[28] KATES, James M. et AREHART, Kathryn H. The hearing-aid speech perception index (HASPI). Speech Communication, 2014, vol. 65, p. 75-93.

[29] PEINADO, Antonio et SEGURA, Jose. Speech recognition over digital channels: Robustness and Standards. John Wiley & Sons, 2006.

[30] HUANG, D.-Y. et ONG, E. P. Lombard speech model for automatic enhancement of speech intelligibility over telephone channel. In : 2010 International Conference on Audio, Language and Image Processing. IEEE, 2010. p. 429-434.