République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Djilali Bounaama Khemis Miliana

Faculté des Sciences et de la Technologie
Département de Technologie

Mémoire du Projet de Fin d'Etudes
Pour l'obtention du diplôme de

Master

En
« Télécommunications »
Option :
« Systèmes de Télécommunications »

Titre :

Compression et codage de la parole par la Transformée KLT

Réalisé par :

BOUASLI Salim
NOUMERI Ahmed

Encadré par :

Mme A.BOUNIF

Année Universitaire: 2015/2016

Dédicace

Je dédie ce mémoire à :

Mon père et Ma mère, qui a œuvré pour ma réussite, de par son amour, son soutien, tous les sacrifices consentis et ses précieux conseils, pour toute son assistance et sa présence dans ma vie, reçois à travers ce travail aussi modeste soit-il, l'expression de mes sentiments et de mon éternelle gratitude.

Mes frères et sœurs qui n'ont cessé d'être pour moi des exemples de persévérance, de courage et de générosité.

Mes professeurs de l'UKM qui doivent voir dans ce travail la fierté d'un savoir bien acquis.

Bouasli Salim

Je dédie ce travail :

A la mémoire de mon père qui a souhaité vivre pour longtemps juste pour nous voir Qu'est-ce que nous allons devenir.

A ma mère et mes frères et sœurs.

A tous mes amis avec lesquels j'ai partagé mes moments de joie et de bonheur

Que toute personne m'ayant aidé de près ou de loin, trouve ici l'expression de ma reconnaissance.

Noumeri Ahmed

Remerciements

Nous remercions tout d'abord le grand Dieu pour l'achèvement de ce mémoire.

Nous exprimons nos gratitudes à Monsieur le président de jury d'avoir accepté examiné ce mémoire.

Nous remercions Messieurs les membres de jury, d'avoir accepté de prendre part à ce jury ainsi que pour l'intérêt qu'ils l'ont portés à ce travail.

Nous remercions Mme BOUNIF, notre encadreur, pour ses conseils et suggestions avisés qui nous ont aidés à mener à bien ce travail, et d'avoir rapporté à ce mémoire ces remarques et conseils.

Résumé

Dans ce mémoire, nous implémentons un algorithme de codage et de compression de la parole basé sur la transformation KLT. Cette transformation permet d'obtenir une bonne qualité de compression. Elle se présente comme la transformation optimale utilisée pour tester les performances d'autres transformations.

Une étude comparative de la KLT avec la transformation DCT a été faite. Les résultats obtenus ont montré que la qualité du codage et de compression donnée par la transformation KLT est supérieure à celle de la transformation DCT.

Mots clés : KLT, DCT, codage, compression, parole, vocal, audio, transformée, algorithme.

Abstract

In this paper we study a speech coding and compression algorithm based on the KLT transform. This transform provide a good compression quality. It can be presented as an optimal transform used for testing others transform.

A comparative study between KLT and DCT transform is done. The obtained results show that the quality of coding based on the KLT transform is superior to that based on the transformation by DCT.

Keywords: KLT, DCT, coding, compression, speech, voice, audio, transformed, algorithm.

Liste des abréviations

AR: Auto **R**égressif.

ARMA : **A**uto **R**égressif à **M**oyenne **A**justée.

ACC : Analyse en **C**omposantes **C**urvilinéaires.

ACP : Analyse en **C**omposantes **P**rincipales

ACELP : Algebraic **C**ode-**E**xcited **L**inear **P**rediction

LPC : **L**inear **P**redictive **C**oding (**C**odage de Prédiction Linéaire).

TFCT : Transformée de **F**ourier à **C**ourt **T**erme.

CCITT : Comité **C**onsultatif **I**nternational pour la **T**éléphonie et la **T**élégraphie.

CER : Critère d'**E**rreur de **R**econstruction**.**

KLT : Karhunen-**L**oeve **T**ransform.

CELP : Code **E**xcited **L**inear **P**rediction (prédiction linéaire avec excitation par code).

DCT : Discrete Cosine Transform(**T**ransformation **C**osinus **D**iscrète).

DSP : **D**igital **S**ignal **P**rocessor

EFR : **E**nhanced **F**ull **R**ate (**P**lein **D**ébit **A**mélioré).

MIC : **M**odulation par **I**mpulsion et **C**odage (PCM).

MICDA : MIC **D**ifférentiel **A**daptatif (ADPCM).

MIPS : Million d'**I**nstructions **P**ar **S**econde.

MPEG : Moving **P**icture **E**xpert **G**roup.

MP3 : Codeur MPEG-1 Couche III.

SB-MICDA : Sous-**B**ande MICDA.

EQM : **E**rreur **Q**uadratique **M**oyenne (**M**ean **S**quare **E**rror).

GSM : Global **S**ystem for **M**obile Communication

LD-CELP : Low-**D**elay **CELP** (CELP à délai réduit).

SNR : Signal to **N**oise **R**atio (Rapport signal sur bruit)

PPM : Prediction by **P**artial **M**atching **(P**rédiction par **P**oursuite **P**artielle).

MOS : Mean **O**pinion **S**core.

Liste des figures

Liste des tableaux

Table de Matière

Chapitre 2 : Notions sur le codage et la compression de la parole

Chapitre 3 : Théorie de la KLT

Chapitre 4 : Résultats expérimentaux

Les progrès technologiques des moyens de communication depuis les années 1990 ont entraîné une augmentation significative de la taille des données envoyées (logiciels, musique, télévision, films, …etc). Pour satisfaire le nombre important d'utilisateurs des diverses services de communication, le son et l'image doivent être compressés tout en gardant une très bonne qualité.

Pour les premières connexions sur le réseau Internet, le temps mis pour échanger des documents volumineux (plus de 100 Koctets) était important. Ainsi, la compression des données à attirer l'attention des chercheurs et a pris plus d'importance. Le temps de récupération d'un document comprimé et envoyé par courrier électronique sera amélioré par rapport au document non comprimé. Deux solutions ont été trouvées pour satisfaire les besoins des utilisateurs. La première solution consiste à utiliser les fibres optiques dans la transmission tels que les liens OC–48 (qui fait passer les données à une vitesse de 2.4 Gbit/s ou Gbps), OC– 192 (9.6 Gbps) ou les liens expérimentaux OC–768 (38.4 Gbps). Deuxièment, des algorithmes de compression adaptatifs aux caractéristiques des données ont été élaborés.

L'un des plus anciens code de compression, c'est le code Morse, inventé par Samuel Finley Breese Morse (1791 – 1872). Les symboles utilisés par ce code sont : le point (·), le trait (–), et une pause servant à délimiter les caractères d'un message.

La compression de données consiste à réduire le débit d'information d'un signal en utilisant des méthodes qui transforment un message long en un message court, sans perdre d'information importante. Le codage des données est un vaste sujet qui a fait l'objet de nombreux ouvrages et articles.

L'étude des techniques de codage et de la compression et leur mise en pratique nécessitent des connaissances théoriques telles que l'algèbre linéaire, la théorie des probabilités et le calcul d'intégral.

Comme le traitement de la parole regroupe plusieurs disciplines de recherche, ce mémoire sera organisé comme suit :

Dans le premier chapitre, nous présentons une description du signal vocal. La section 1 est consacrée à la production et la perception du signal de la parole. La spécificité et les

différentes représentations de la parole seront discutées dans la section 2. Nous terminerons ce chapitre par la présentation des propriétés de la parole.

Le domaine du codage et de la compression de la parole est très diversifié, le deuxième chapitre récapitule les notions de base sur ce sujet. Nous présenterons la compression sans et avec pertes, la notion de la redondance et le codage de base sur un modèle du signal. Nous décrivons ensuite les grandes familles des codeurs de base à savoir : Nous entamerons le codage par transformée d'une façon générale.

Le troisième chapitre sera consacré à la transformation KLT. Les notions théoriques concernant cette dernière seront évoquées en détail. Enfin, l'apport de cette transformation aux techniques de compression sera discuté.

Le quatrième chapitre présentera les résultats expérimentaux de la méthode de compression utilisée. Pour évaluer les performances, une étude comparative avec la transformée cosinus sera faite. Pour terminer ce chapitre, nous présenterons le résultat des tests d'écoutes pour comparer la qualité de compression de méthode utilisée.

Une dernière partie consacrée aux conclusions générales et aux perspectives viendra pour clôturer le travail.

Chapitre 1
GENERALITE SUR LE SIGNAL DE LA PAROLE

1.1. Introduction

La communication parlée est un privilège et moyen principal de dialogue entre les êtres humains. Son apparition est attachée à notre existence sur terre. Grâce à son appareil auditif et en recevant un son seulement, l'être vivant peut localiser et éviter un danger non visible. En parlant, une riche information (geste, émotion, sourire, etc) est directement transmise à l'auditeur en face à nous.

Le traitement de la parole est un domaine multidisciplinaire qui se situe au croisement du traitement du signal numérique et du traitement du langage. Le son peut être produit et perçu instantanément par le cerveau, et pour cela le traitement de la parole tend à remplacer ces fonctions par des systèmes automatiques.

Dans ce chapitre nous essayons de présenter et comprendre la particularité du signal vocal. Dans la section II, nous décrivons le phénomène de production de la parole. Ensuite, nous parlons du fonctionnement de l'appareil phonatoire. La section III est consacrée à la classification des différents sons. Dans la section suivante nous parlerons des systèmes de traitement automatique de la parole. Les classes phonétiques sont décrites dans la cinquième section. Enfin, nous présenterons les paramètres essentiels du signal vocal qui sont nécessaires au traitement de ce dernier.

1.2. PRODUCTION DE LA PAROLE

La parole peut être décrite selon les différents aspects suivants :

1.2.1 Aspect Anatomo-physiologique

Le système vocal humain peut être décomposé en quatre sous-systèmes élémentaires donnés comme suit :

1.2.1.1. Poumons et conduit trachéo-bronchique

La trachée-artère est un conduit cylindrique qui relie le larynx et les bronches qui se ramifient à l'intérieur des poumons.

1.2.1.2. Larynx

Est un ensemble de muscles et de cartilages mobiles qui entourent une cavité située à la partie supérieure de la trachée.

1.2.1.3. Cordes vocales

En sorte de lèvres symétriques placées en travers du larynx en s'écartant, elles déterminent une ouverture triangulaire appelée glotte. Les sons voisés résultent d'une vibration périodique des cordes vocales dans un plan horizontal obéissant à un phénomène d'oscillation et de relaxation [1].

1.2.1.4. Conduit vocal

Est un ensemble de cavités situées entre la glotte et les lèvres qu'on peut distinguer ainsi : La cavité nasale (formée des fosses nasales), la cavité buccale, la cavité pharyngienne. Les deux dernières forment le conduit oral, qui possède un volume et une géométrie extrêmement variable grâce à la grande mobilité de la langue [2].

1.2.2. Fonctionnement de l'appareil phonatoire

La parole résulte d'une série de mouvement des appareils respiratoire et articulatoire.

1.2.2.1. L'appareil vibrateur

Le processus de production du son nécessite trois éléments :

- une soufflerie,
- les sources vocales,
- les cavités supra-glottiques

L'appareil respiratoire est à l'origine de la soufflerie **(**expulsion de l'air pulmonaire à travers la trachée : souffle phonatoire produit, soit par l'abaissement de la cage thoracique, soit dans le cadre de la projection vocale par l'action des muscles abdominaux). Il existe deux sortes de sources vocales.

Les sources vocales sont constituées du larynx et des sources de bruit. Le larynx est un ensemble de muscles et de cartilages mobiles. Il se trouve dans le cou, entre le pharynx et la trachée, et en avant de l'œsophage (Figure1.1). Les cordes vocales sont deux lèvres

symétriques (structures fibreuses) situées dans le larynx. Ces lèvres se rejoignent en avant et en s'écartant l'une de l'autre sur leur partie arrière forment une ouverture triangulaire nommée glotte. Le larynx et les plis vocaux (cordes vocales) forment notre « appareil vibrateur »

Le larynx contrôle le flux d'air lors de la respiration, protège les voies respiratoires et produit une source sonore pour la parole. Lors de la production d'un son voisé (ou sonore), comme c'est le cas, par exemple, pour les phonèmes [z], [v] et pour les voyelles, les plis vocaux s'ouvrent et se ferment périodiquement, obstruant puis libérant par intermittence le passage de l'air dans le larynx. Le flux continu d'air pulmonaire prend ainsi la forme d'un train d'impulsions de pression, nos cordes vocales vibrent.

Lorsque les cordes vocales sont écartées mais ne vibrent pas, un bruit est généré dans le conduit vocal. Lors de la production d'un son non-voisé (ou sourd), comme c'est le cas, par exemple, pour les phonèmes [s] ou [f], les plis vocaux sont écartés et l'air pulmonaire circule librement en direction des structures en aval. Par contre, si les cordes vocales sont rapprochées le bruit est généré au niveau de ces dernières et une voix chuchotée est produite.

En revanche, Le dernier élément principal de notre appareil vibrateur est l'épiglotte. Lors de la déglutition, cette dernière agit comme un clapet qui se rabat sur le larynx, conduisant les aliments vers l'œsophage en empêchant leur passage dans la trachée et les poumons [3].

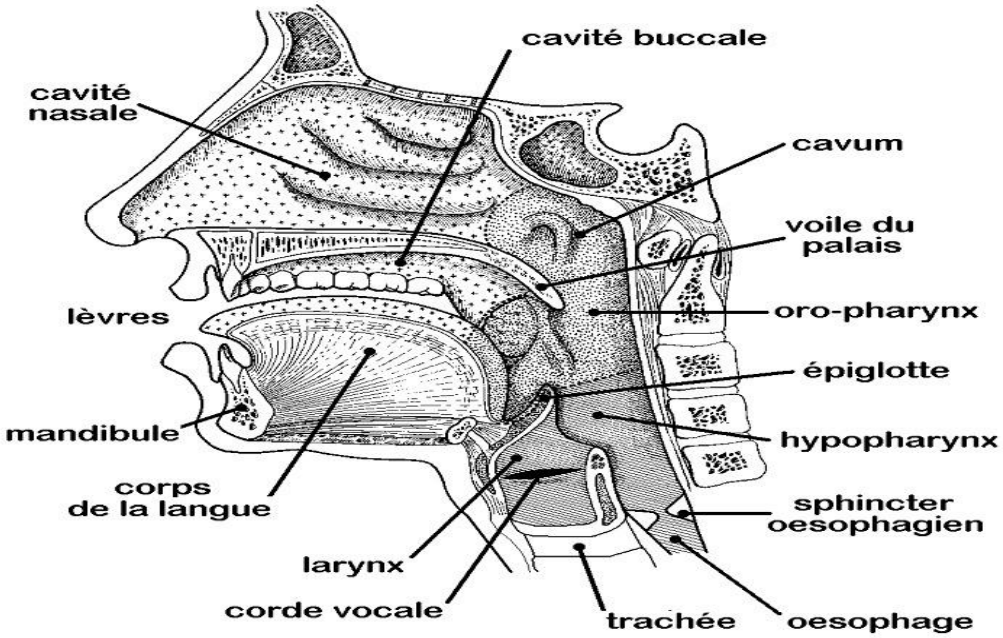


Fig.1.1: Anatomie de l'appareil phonatoire

1.2.2.2. Le résonateur

L'air pulmonaire, modulé par l'appareil vibrateur, est appliqué à l'entrée du conduit vocal. Ce dernier est constitué des cavités pharyngales (laryngé-pharynx et oropharynx situés en arrière-gorge) et de la cavité buccale (espace qui s'étend du larynx jusqu'aux lèvres). Pour la réalisation de certains phonèmes, le voile du palais (le velum) et la luette qui s'y rattache, s'abaissent, permettant ainsi le passage de l'air dans les cavités nasales (fosses nasales et rhinopharynx ou nasaux-pharynx). Ces différentes cavités forment le résonateur. IL constitué des organes mobiles, nommés articulateurs, qui en modifiant sa géométrie et donc ses propriétés acoustiques, mettent en forme le son laryngé (ou son glottique) en une séquence de sons élémentaires. Il est qualifié d'être le lieu de naissance de la parole. Ces derniers peuvent être interprétés comme la réalisation acoustique d'une série de phonèmes, unités linguistiques élémentaires propres à une langue. Les articulateurs principaux sont la langue, les lèvres, le voile du palais et la mâchoire [4].

1.3. Classification des sons de la parole

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

1.3.1. 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 quasi-périodicité, une énergie élevée et une fréquence fondamentale (pitch). Typiquement, la période fondamental des différents sons voisés varie entre 2ms et 20ms.

1.3.2. 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.

1.4. Systèmes automatiques de traitement de la parole

Les techniques modernes de traitement du son essayent de concevoir des systèmes automatiques qui effectuent les fonctions mentionnées sur la figure (1.2).

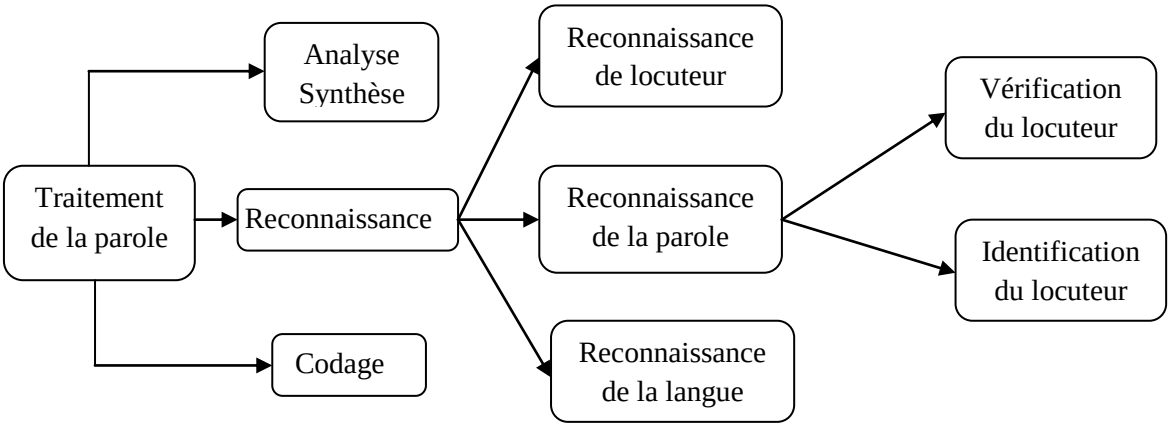


Fig.1.2 : Différents traitements automatiques de la parole

- **Les analyseurs de parole** mettent en évidence les caractéristiques du signal vocal tel qu'il est produit. Ils sont utilisés soit comme composant de base de systèmes de codage, de reconnaissance ou de synthèse.
- **Les systèmes de reconnaissance de la parole** décodent l'information portée par le signal vocal à partir des données fournies par l'analyse. On les classes en fonction de l'information que l'on cherche à extraire du signal vocal:

La reconnaissance du locuteur qui est l'identification (vérifier que la voix analysée correspond bien à la personne qui est sensée la produire) ou la vérification du locuteur (déterminer qui, parmi un nombre fini et préétabli de locuteurs, a produit le signal analysé.).

- **Il y a aussi la reconnaissance du locuteur** dépendante du texte (la phrase à prononcer pour être reconnue est fixée dès la conception du système), la reconnaissance avec texte dicté (la phrase à prononcer est fixée lors du test), reconnaissance indépendante du texte (la phrase à prononcer n'est pas précisée) et reconnaissance indépendante du texte (la phrase à prononcer n'est pas précisée).

La synthèse de parole est une technique informatique qui permet de créer de la parole artificielle à partir de n'importe quel texte. Pour obtenir ce résultat, elle s'appuie à la fois sur des techniques de traitement linguistique, notamment pour transformer le texte orthographique en une version phonétique prononçable sans ambiguïté, et sur des techniques de traitement du signal pour transformer cette version phonétique en son numérisé écoutable

sur un haut-parleur. Il y a deux types de synthétiseurs : les synthétiseurs de parole à partir d'une représentation numérique, inverses des analyseurs, dont la mission est de produire de la parole à partir des caractéristiques numériques d'un signal vocal telles qu'obtenues par analyse, et les synthétiseurs de parole à partir d'une représentation symbolique (texte ou concept), inverse de reconnaissance de parole et capables en principe de prononcer n'importe quelle phrase sans qu'il soit nécessaire de la faire prononcer par un locuteur humain au préalable [3].

- **Enfin,** le rôle des codeurs est de permettre la transmission ou le stockage de parole avec un débit réduit, ce qui passe tout naturellement par une prise en compte judicieuse des propriétés de production et de perception de la parole [5]. C'est ce qu'on va détailler dans les prochains chapitres.

1.6. Classes phonétiques

L'aspect phonétique décrit la parole selon son mode de production. On cherche à la caractériser par des sons élémentaires appelés phonèmes.

1.5.1. Le phonème

La plupart des langues naturelles sont composées à partir de sons distincts, les phonèmes. Un phonème est la plus petite unité présente dans la parole et susceptible de distinguer différents mots. Par exemple le « i » dans : riz , Pari et mari,…

Le nombre de phonèmes est toujours très limité, normalement inférieur à cinquante. Par exemple, la langue française comprend 36 phonèmes. Un système de phonèmes est un ensemble d'images acoustiques emmagasinées dans le cerveau du locuteur dans la mesure où celui-ci maîtrise sa langue.

La production d'un phonème donné laisse toutefois place à une certaine variabilité sur le plan acoustique [6].

1.5.2. Classes phonétiques

La phonétique s'intéresse à regrouper les éléments en classe. Chaque groupe contient des unités ayant des caractéristiques communes.

Les différents sons de la parole sont regroupés en classes phonétiques en fonction de leurs caractéristiques principales. Ces caractéristiques représentent des différences qui sont

suffisamment importantes pour qu'il soit possible de classer les différents sont visibles sur un spectrogramme selon leur classe respective en très peu de temps et sans aucune écoute de la phrase correspondante. Les différentes classes phonétiques présentes en français et en anglais sont :

1.5.2.1. Les voyelles

Cette classe correspond, à quelques nuances supplémentaires près, aux voyelles de l'écrit. Elles se caractérisent principalement par le voisement qui crée des formants. Ces formants, qui sont des zones fréquentielles de forte énergie, correspondent à une résonance dans le conduit vocal de la fréquence fondamentale produite par les cordes vocales.

Ces formants peuvent s'élever jusqu'à des fréquences de 5 kHz mais c'est principalement les formants en basses fréquences qui caractérisent les voyelles. Cette caractéristique permet d'ailleurs de distinguer grossièrement les voyelles en fonction de leur premier et deuxième formant.

1.5.2.2. Les occlusives

Les phonèmes de cette classe se caractérisent oralement par la fermeture du conduit vocal, fermeture précédant un brusque relâchement. Les occlusives sont donc constituées de deux parties successives une première partie de silence, correspondant à l'occlusion effective, et une deuxième partie d'explosion, au moment du relâchement. Les occlusives peuvent être voisées, à la manière des voyelles, ou sourdes, c'est à dire non voisées. Les occlusives voisées peuvent également être appelées occlusives sonores.

1.5.2.3. Les fricatives

Dans cette classe sont regroupés les sons produits par la friction de l'air dans le conduit vocal lorsque celui-ci est rétréci au niveau des lèvres, des dents ou de la langue. Cette friction produit un bruit de hautes fréquences et peut être voisée ou sourde.

1.5.2.4. Les sonantes

Cette classe est en fait constituée, pour simplification, du regroupement des trois sous-classes que sont les semi-consonnes, les liquides et les nasales.

1.5.2.5. Les semi-consonnes (ou semi-voyelles ou glissantes)

Elles ont la structure acoustique des voyelles mais ne peuvent en jouer le rôle car elles ne sont que des transitions vers d'autres voyelles qui sont les véritables noyaux syllabiques. D'un point de vue syntaxique, une règle stricte de la langue française veut que deux voyelles ne puissent jamais se suivre. Cette règle est très largement respectée dans la construction des mots mais présente, comme toute règle, quelques exceptions. La classe des semi-consonnes a été créée pour pallier ces exceptions de manière gracieuse. Les semi-consonnes sont évidements sonores.

1.5.2.6. Les liquides

Les liquides sont très similaires aux voyelles et aux semi-consonnes mais leur durée et leur énergie sont généralement plus faibles. Elles sont sonores.

1.5.2.7. Les nasales

Les phonèmes sont formés par passage de l'air dans le conduit vocal depuis les cordes vocales. Ce passage exclut normalement toute connexion du conduit normal, le conduit buccal, avec le conduit nasal. Ce dernier peut cependant être employé, dans un nombre limité de cas puisque sa physionomie ne permet pas de créer des sons autrement qu'en modifiant le volume de la caisse de résonances qu'il constitue par l'intermédiaire de la langue,

1.5.2.8. Les affriquées

Cette classe est, elle aussi, propre à l'anglo-américain mais les affriquées peuvent également être observées dans le français québécois. Les affriquées sont composées d'une occlusive immédiatement suivie par une fricative de durée cependant plus faible que celle de la véritable fricative [7].

1.6. Paramètres du signal de parole

La parole est un signal continu, d'énergie finie, non stationnaire. Sa structure est complexe et variable dans le temps:

- périodique (pseudo-périodique) pour les sons voisés.
- aléatoire pour les sons fricatifs.

- impulsionnelle dans les phases explosives des sons occlusifs.

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

La fréquence fondamentale représente la fréquence du cycle d'ouverture/fermeture des cordes vocales. Elle caractérise seulement les sons voisés, et 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.

1.6.1. L'énergie

Elle correspond à 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 :

$$E = \sum_{n=0}^{N-1} s^2(n) \qquad \textbf{Eq. 1. 1}$$

Avec N : la taille de la trame.

s(n) : signal de la parole.

1.6.2. Le spectre

La représentation fréquentielle de l'intensité de la voix définit l'enveloppe spectrale ou le spectre, 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 dont l'expression est :

$$\hat{S}(k) = \hat{S}\left(f = \frac{k}{N}\right) = \sum_{n=0}^{N-1} s(n).w(n).exp(-2j\pi nk/N), \;\; 0 \leq k \leq N-1 \qquad \textbf{Eq. 1. 2}$$

Où ; N : Le nombre de points prélevés.

S(k) : Spectre complexe.

s(n) : Segment analysé.

w(n) : Fenêtre d'analyse temporelle.

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

$$\left|\hat{S}(k)\right|^2 \qquad 0 \leq k \leq \frac{N}{2} \qquad \textbf{Eq. 1. 3}$$

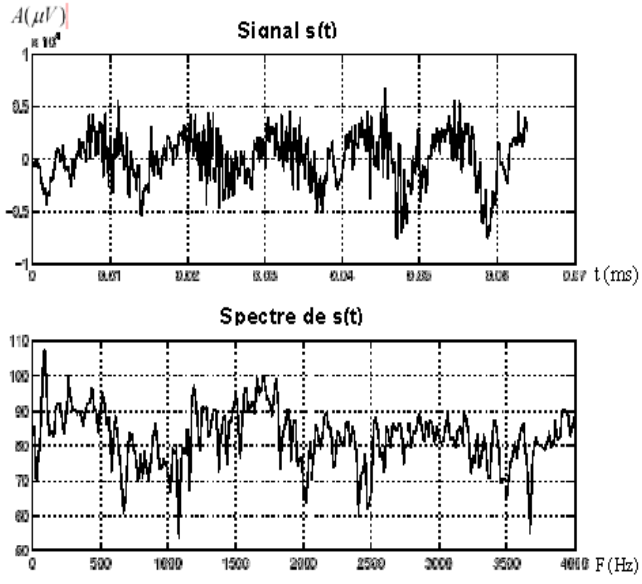


Fig.1.3: Son non voisé et son spectre

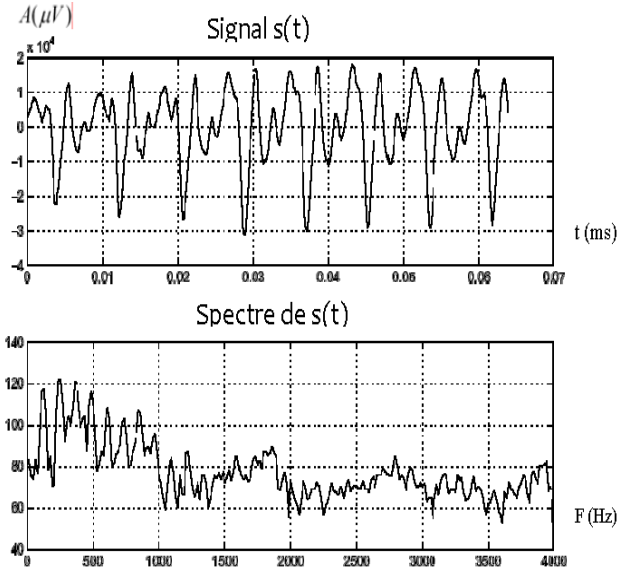


Fig.1.4: Son voisé et son spectre

1.7. Propriétés statistiques du signal vocal

L'audiogramme ou l'évolution temporelle du signal vocal ne fournit pas directement les traits acoustiques du signal, il est nécessaire de mener un ensemble de calculs statiques. Le signal vocal peut être vu comme un processus aléatoire non stationnaire [5].

1.7.1. Densité de probabilité

Si $N_\varepsilon$représente le nombre d'échantillons de *x(n)* dont l'amplitude est comprise entre $[\varepsilon - \Delta\varepsilon/2,\ \varepsilon + \Delta\varepsilon\ /\ 2]$ alors que $n \in [-N,\ N]$, la densité de probabilité du signal x supposé ergodique et stationnaire est donnée par :

$$P_x(\varepsilon) = \lim_{N\to\infty}[N_\varepsilon/(2N+1)] \qquad \textbf{Eq. 1. 4}$$

1.7.2. Valeur moyenne et variance

La valeur moyenne d'un signal stationnaire vaut:

$$\mu_x = \int_{-\infty}^{+\infty} \varepsilon P_x(\varepsilon)\, d\varepsilon = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x(n) \qquad \textbf{Eq. 1. 5}$$

Pour le signal vocal cette moyenne est supposé nulle, elle ne contient aucune information utile.

La variance est donnée par:

$$\sigma_x^2 = \int_{-\infty}^{+\infty} \varepsilon^2 P_x(\varepsilon)\, d\varepsilon = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x^2(n) \qquad \textbf{Eq. 1. 6}$$

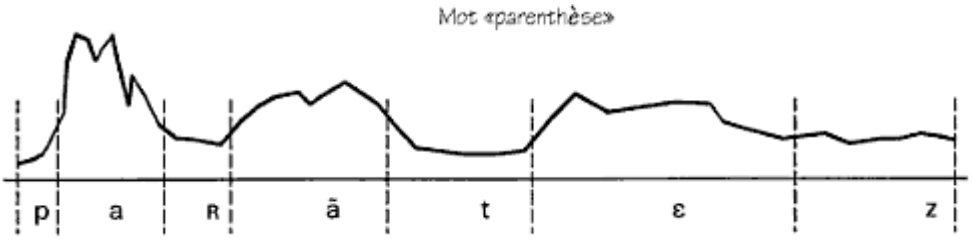


Fig.1.5 : Variance du mot parenthèse.

1.7.3. Fonction d'auto-corrélation

La fonction d'auto-corrélation d'un signal ergodique et stationnaire s'exprime par:

$$\emptyset_{xx}(k) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x(n).x(n+k) \qquad \textbf{Eq. 1. 7.}$$

L'estimation sur un nombre fini de N échantillons peut être calculée par:

$$\emptyset_{xx}(k) = \frac{1}{N-K} \sum_{n=0}^{N} x(n).x(n+k) \qquad \textbf{Eq. I. 8.}$$

La fonction d'auto-covariance est définie et estimée par des formules identiques après avoir soustrait la moyenne μ_x et comme μ_x = 0 alors ces deux notions peuvent être confondues. On a aussi:

$$\sigma_x^2 = \emptyset_{xx}(0) \qquad \boldsymbol{Eq.1.9}$$

Le coefficient d'auto-corrélation est défini par :

$$\rho_x(k) = \frac{\emptyset_{xx}(k)}{\emptyset_{xx}(0)} \qquad \boldsymbol{Eq.1.10}$$

Cette valeur est comprise entre 1 et -1.

1.7.4. Densité spectrale de puissance

La densité spectrale de puissance S_{xx} (θ) est la transformée de fourrier de la fonction d'auto-corrélation:

$$S_{xx}(\theta) = \sum_k \emptyset_{xx}(k) \exp(-jk\theta), \quad \theta = \omega T_e \qquad \textbf{Eq.1.11}$$

$$\hat{S}_{xx}(\theta) = \sum_{-k}^{k} \emptyset_{xx}(k) w(k) \exp(-jk\theta) \qquad \textbf{Eq.1.12}$$

Où ; w(k) est une fonction de fenêtrage.

1.9. Automatisation de la Parole

La parole est produite par l'articulation des organes phonatoires de l'homme et prend une forme analogique apériodique ; ce qui est impossible pour que la machine puisse l'interpréter ou la 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 utilisée dans la numérisation est la méthode Delta ou MIC qui consiste en trois étapes : l'échantillonnage, la quantification et le codage.

1.8.1. 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 (Fig.1.6). 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) [8]. 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. »

$$F_{eh} \geq 2F_{max} \qquad \textbf{Eq. 1. 13}$$

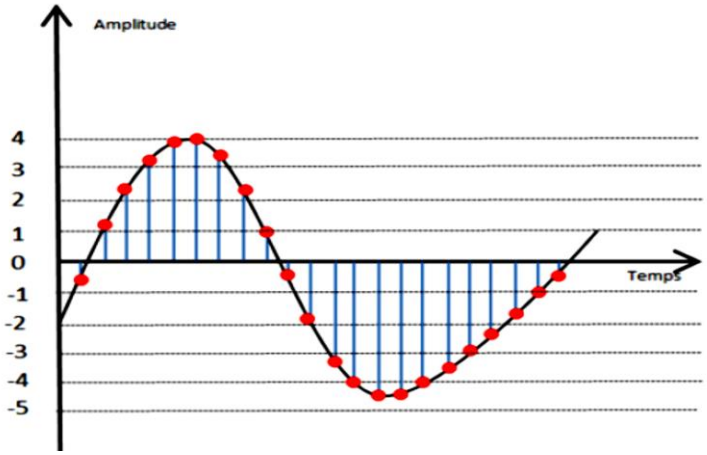


Fig.1.6: Echantillonnage d'un signal

Pour la téléphonie, on estime que le signal garde une qualité suffisante lorsque son spectre est limité à 3400 Hz et l'on choisit f_e= 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 jusqu'à 20 kHz et l'on utilise des fréquences d'échantillonnage de 44.1 ou 48kHz. Pour les applications multimédia, les fréquences sous-multiples de 44.1kHz sont de plus en plus utilisées : 22.5 kHz, 11.25 kHz [9].

1.8.2. 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 (Fig.1.7). 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.

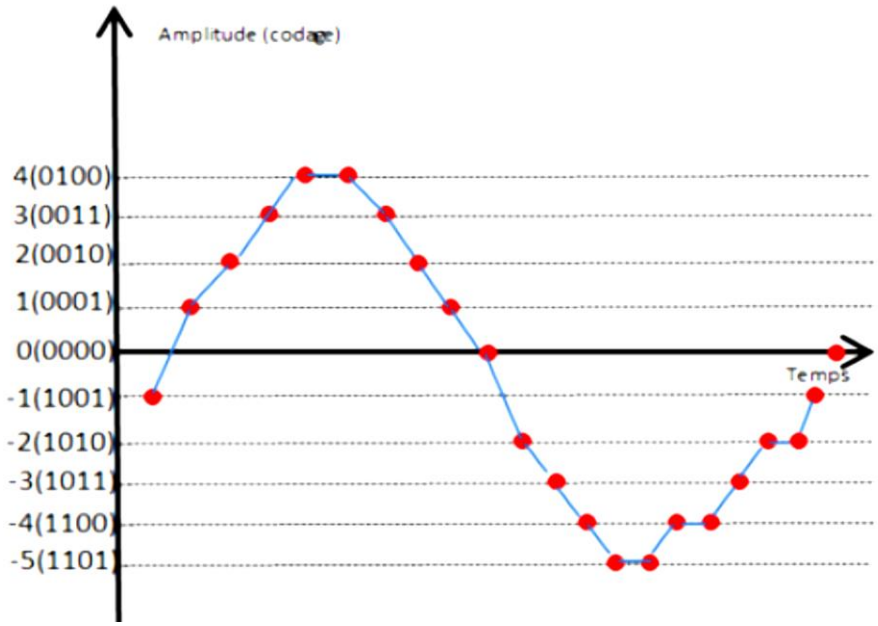


Fig.1.7: Quantification d'un signal échantillonné

1.8.3. Codage des données

C'est la représentation binaire des valeurs quantifiées qui permet le traitement du signal sur machine. Les notions de codage de l'information font l'objet du chapitre suivant.

1.9. Conclusion

Le traitement de la parole est aujourd'hui une composante fondamentale des sciences de l'ingénieur. Dans ce chapitre, nous avons exposés les notions et les caractéristiques fondamentales qui permettent le traitement du signal vocal. Son traitement est situé au croisement du traitement du signal numérique et du traitement du langage (c'est-à-dire du traitement de données symboliques), cette discipline scientifique a connu depuis les années 60 une expansion fulgurante, liée au développement des moyens et des techniques de télécommunications.

L'importance particulière du traitement de la parole dans ce cadre plus général s'explique par la position privilégiée de la parole comme vecteur d'information dans notre société humaine.

Chapitre 2
Notion sur le codage et la compression de la parole

2.1. Introduction

A l'heure actuelle, la puissance des processeurs augmente plus vite que les capacités de stockage, et énormément plus vite que la bande passante des réseaux, car cela demande d'énormes changements dans les infrastructures de télécommunication. Ainsi, pour pallier ce manque, il est recommandé de réduire la taille des données en exploitant la puissance des processeurs plutôt qu'en augmentant les capacités de stockage et de transmission des données.

Le signal de parole est complexe, redondant et possède une grande variabilité. Les traits caractéristiques et invariants doivent être extraits du signal de parole pour permettre le fonctionnement efficace des systèmes de traitement de ce signal. Cette procédure consiste à le paramétrer acoustiquement.

Dans ce chapitre, nous allons présenter brièvement des techniques de codage de la parole et d'audio. Comme, il existe différents types de codeurs, nous n'allons pas récapituler tous les codeurs existants ou de les présenter en détail. Nous allons nous limiter aux techniques les plus répandues.

2.2. Définition de la compression

La compression de données ou codage de source consiste à réduire le volume de données à stocker, à traiter ou à transmettre, tout en préservant l'information nécessaire à la récupération du signal pendant le décodage (*décompression* : l'opération inverse de la compression).

Les différents algorithmes de compression sont basés sur 3 critères :

- **Le taux de compression :** c'est le rapport de la taille du fichier compressé sur la taille du fichier initial.
- **La qualité de compression :** sans ou avec pertes (avec le pourcentage de perte).
- **La vitesse** de compression et de décompression.

2.3. Communication et stockage numérique

Aujourd'hui, la plupart de systèmes de transmission ou de stockage sont numérisés. Cette technique offre des avantages considérables par rapport aux techniques analogiques. Notamment, le signal peut être protégé contre les dégradations lors de la transmission ou du stockage.

La figure 2.1 montre le schéma général de la transmission et du stockage numérique. Puisque le signal de la parole est analogique de nature, pour le traiter, il doit être numérisé. Cette conversion analogique-numérique se fait en deux étapes. Premièrement, le son sera échantillonné en respectant la loi de Shannon (la fréquence d'échantillonnage doit être deux fois plus grande que la largeur de bande du signal analogique). L'étape suivante consiste à représenter chaque échantillon par un mot binaire (la quantification où l'amplitude du signal est discrétisée). Le nombre de bits utilisés par échantillon détermine le taux de quantification. Cette opération introduit une distorsion par rapport au signal analogique. Pour rendre ce phénomène imperceptible, on augmente le nombre de bits par échantillon. L'inconvénient principal de la numérisation est la quantité d'information binaire importante lors de la transmission ou du stockage. Le codage de source(le nombre de bits représentant le signal est réduit) permet de surmonter ce problème. Pendant le codage de canal, des bits redondants sont ajoutés au signal pour assurer la protection contre les erreurs éventuelles de transmission ou de stockage.

En réception, les opérations inverses sont exécutées. Le décodage de canal permet de détester ou corriger des erreurs de transmission ou de stockage. Le décodage de source reconstruit le signal à partir de sa représentation compressée. Enfin, le signal numérique est reconverti dans une forme analogique [10].

Dans la suite de ce chapitre, nous ne nous intéressons pas au codage et décodage de canal. Nous supposons que le signal à l'entrée du décodeur de source est identique à celui à la sortie du codeur de source.

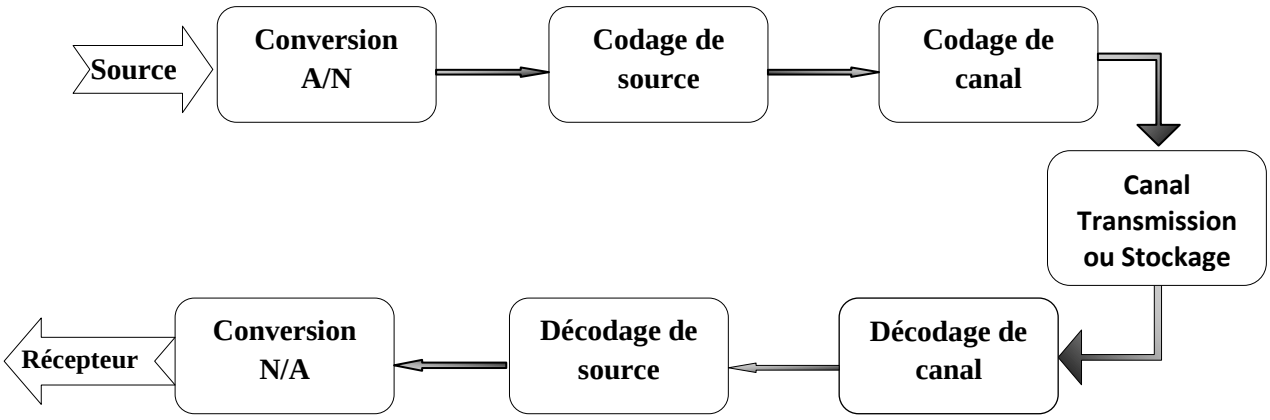


Fig.2.1: Schéma des transmissions et stockage numérique

2.4. Taux de compression

Le taux de compression τ est relié au rapport entre la taille *b* du fichier comprimé *B* et la taille a du fichier initial *A*. Le taux de compression est généralement exprimé en pourcentage. Un taux de 50 % signifie que la taille *b* du fichier comprimé *B* est la moitié de *a*. La formule pour calculer ce taux est :

$$\tau = 1 - \left({}^{b}/_{a}\right) \qquad \textbf{Eq. 2. 1}$$

Exemple :

a=550Mo, b= 250Mo

$$\tau = 1 - \left({}^{250}/_{550}\right) = 54\%$$

L'algorithme utilisé pour transformer *A* en *B* est destiné à obtenir un résultat *B* de taille inférieure à *A*. Il peut paradoxalement produire parfois un résultat de taille supérieure : dans le cas des compressions sans pertes, il existe *toujours* des données incompressibles, pour lesquelles le flux compressé est de taille supérieure ou égale au flux d'origine.

2.5. Classification des algorithmes de compression

2.5.1. Compression symétrique / asymétrique

Dans le cas de la compression symétrique, la même méthode est utilisée pour compresser et décompresser l'information, il faut donc la même quantité de travail pour chacune de ces opérations. C'est ce type de compression qui est généralement utilisée dans les transmissions de données.

La compression asymétrique demande plus de travail pour l'une des deux opérations, la plupart des algorithmes requiert plus de temps de traitement pour la compression que pour la décompression. Des algorithmes plus rapides en compression qu'en décompression peuvent être nécessaires lorsque l'on archive des données auxquelles on accède peu souvent (pour des raisons de sécurité par exemple).

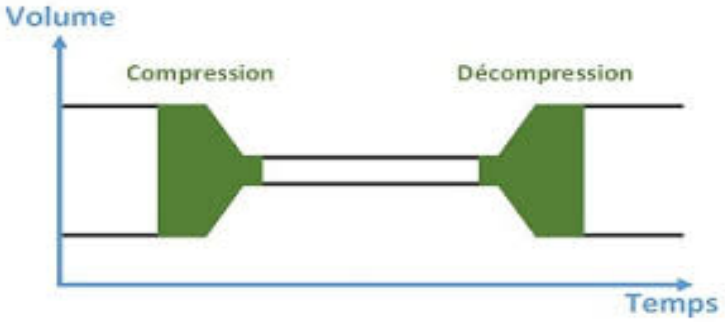


Fig.2.2: Compression de type symétrique.

2.5.2. Compression physique / logique

On considère généralement la compression comme un algorithme capable de comprimer des données dans un minimum de place (compression physique), mais on peut également adopter une autre approche et considérer qu'en premier lieu un algorithme de compression a pour but de recoder les données dans une représentation différente plus compacte contenant la même information (compression logique).

La distinction entre compression physique et logique se base sur la façon dont les données sont compressées ou plus précisément comment elles sont réarrangées ?

La compression physique est exécutée exclusivement sur les informations contenues dans les données. Cette méthode produit typiquement des résultats incompréhensibles qui apparemment n'ont aucun sens. Le résultat d'un bloc de données compressées est plus petit que l'original car l'algorithme de compression physique a retiré la redondance qui existait entre les données elles-mêmes.

La compression logique est accomplie à travers le processus de substitution logique qui consiste à remplacer un symbole alphabétique, numérique ou binaire par un autre. Changer "United State of America" en "USA" est un bon exemple de substitution logique car "USA" est dérivé directement de l'information contenue dans la chaîne "United State of America" et garde la même signification. La substitution logique ne fonctionne qu'au niveau du caractère ou plus haut et est basée exclusivement sur l'information contenue à l'intérieur même des données.

On peut encore distinguer les algorithmes qui travaillent au niveau statistique et ceux qui opèrent au niveau numérique. Pour les premiers, la valeur des motifs ne compte pas. Ce sont les probabilités qui comptent, et le résultat est inchangé par substitution des motifs tandis que pour les seconds, les valeurs des motifs influent sur la compression (par exemple MP3), et les substitutions sont interdites.

Enfin le critère de classification le plus pertinent est basé sur la perte des données. On peut distinguer trois types de compression (codage) qui seront présentés comme suit :

2.5.2.1. Compression sans perte

Les algorithmes de compression sans perte (connu aussi sous le nom de non destructible, réversible, ou conservative) sont des techniques permettant une reconstitution exacte de l'information après le cycle de compression / décompression.

– Codage statistique :

Le but est de :

* Réduire le nombre de bits utilisés pour le codage des caractères fréquents.

* Augmenter ce nombre pour des caractères plus rares.

Exemple

Certaines informations sont plus souvent présentes que d'autres dans les données que l'on veut compresser. Le codage MIC entre dans ce type, nous le présenterons dans la section

Le taux de compression des algorithmes sans perte est en moyenne de l'ordre de 40% pour des données de type texte. Par contre, ce taux est insuffisant pour les données de type multimédia. Il faut donc utiliser un nouveau type de compression pour résoudre ce problème : la compression avec perte.

2.5.2.2. Compression avec perte

Les objectifs de la compression avec pertes sont d'éliminer les données non pertinentes pour ne transmettre que ce qui est perceptible. Les signaux audio et vidéo contiennent une quantité importante de données redondantes.

Ce type de compression, comme pour la compression sans perte, élimine l'information redondante et introduit une dégradation indiscernable à l'œil (ou à l'oreille) avec un taux de compression très élevé. Donc, elle ne s'applique qu'aux données « perceptibles », en général sonores ou visuelles, qui peuvent subir une modification, parfois importante, sans que cela soit perceptible par un humain. Les données originales ne peuvent pas être retrouvées, donc la perte d'information est irréversible c.-à-d. non conservative.

Cette technique est fondée sur une idée simple : seul un sous-ensemble très faible de sons possibles est exploitable par l'oreille. Par exemple, il n'est pas intéressant de coder avec fidélité un bruit qui n'est qu'une redondance. Un codage éliminant cette redondance et la restituant à l'arrivée reste recommandé, même si le signal reconstruit n'est pas identique au son original.

Il existe des algorithmes de compression consacrés à des usages particuliers, dont en voici 3 :

– Compression du son (Audio MPEG, ADCPM ...).

– Compression des images fixes (JPEG,...).

– Compression des images animées (MPEG, ...).

Les formats MPEG sont des formats de compression avec pertes pour les séquences vidéo. Ils incluent à ce titre des codeurs audio, comme les célèbres MP3 ou AAC (voir figure.2.3 , qui peuvent parfaitement être utilisés indépendamment, et bien sûr des codeurs vidéo. 0

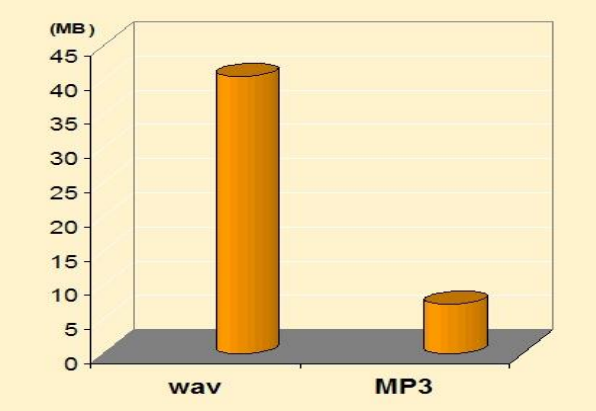


Fig.2.3: Comparaison des tailles d'un fichier audio non compressé (en PCM dans un conteneur WAV) et compressé (en MP3).

2.5.2.3. Compression presque sans perte

On peut classer ce genre compression entre la compression conservative et la compression non conservative, mentionnées précédemment. Elle permet de conserver toute la signification des données d'origine, tout en éliminant une partie de leur information.

Les algorithmes de compression presque sans perte sont spécifiques à un type de données particulier, Par exemple le Monkey's Audio permet de compresser sans perte les données audio du wave PCM : il n'y a pas de perte de qualité, le morceau de musique est exactement celui d'origine.

2.5.3. Méthodes de codage

On a deux types de codage: le codage de la source d'information (qui transforme la source dans une forme alternative, meilleure pour la transmission ou pour la mémorisation) et le codage de canal (qui augmente la robustesse du message contre les erreurs de transmission). Les méthodes de compression sans pertes font partie de la première catégorie. Les codes créés doivent avoir quelques propriétés [12] :

- être uniquement décodables.
- être décodables instantanément.
- être compacts.

Si le code n'est pas compact il doit être le plus efficient possible. Pour apprécier cette efficience il faut calculer la longueur moyenne du code. Cette quantité peut être appréciée à l'aide de l'entropie de la source en utilisant le théorème de Shannon sur le codage dans l'absence du bruit. Parmi les codes pour la compression de données, le plus ancien et peut être le plus utilisé encore, est le code de Huffmann. Celui-ci a élaboré son fameux algorithme de codage en 1952, comme réponse à une question posée par l'un de ses professeurs quand il était étudiant à MIT [13].

2.5.3.1. Codage par répétition

2.5.3.1.1. Codage RLE

La compression RLE est utilisée par de nombreux formats d'images. Elle est basée sur la répétition d'éléments consécutifs. Une première valeur (codée sur un octet) donne le nombre de répétitions, une seconde valeur donne la valeur à répéter (codée sur un octet).

La phrase suivante 'oooooohhhhhhhhhh' donnerait '6o1lh', elle est très utile dans ce cas-là. Par contre dans 'onde' cela donne '1o1n1d1e', elle s'avère ici très coûteuse.

2.5.3.1.2. Codage par modélisation de contexte

2.5.3.1.2.1. Prédiction par poursuite partielle (PPM)

La prédiction par poursuite partielle est basée sur une modélisation de contexte pour évaluer la probabilité des différents symboles. En connaissant le contenu d'une partie d'une source de données (fichier, flux…), un PPM est capable de deviner la suite, avec plus ou moins de précision. Un PPM peut être utilisé en entrée d'un codeur arithmétique par exemple.

La prédiction par reconnaissance partielle donne en général de meilleurs taux de compression que des algorithmes à base de Lempel-Ziv, mais est sensiblement plus lente.

2.5.3.1.2.2. Pondération de contextes

La pondération de contextes consiste à utiliser plusieurs prédicteurs (par exemple des PPM) pour obtenir l'estimation la plus fiable possible du symbole à venir dans une source de données (fichier, flux…). Elle peut être principalement réalisée par une moyenne pondérée, mais les

meilleurs résultats sont obtenus par des méthodes d'apprentissage automatique comme les réseaux de neurones.

La pondération de contextes est très performante en termes de taux de compression, mais est d'autant plus lente que le nombre de contextes est important.

Actuellement, les meilleurs taux de compression sont obtenus par des algorithmes liant pondération de contextes et codage arithmétique, comme PAQ [15].

2.6. Techniques du codage de la parole

Pratiquement tous les codeurs de la parole sont des codeurs avec pertes. Dans la suite nous présenterons les techniques les plus répandues et les techniques utilisent la décomposition en ondelettes ou des modèles sinusoïdaux. Le lecteur intéressé peut consulter des ouvrages spécialisés pour une description plus détaillée des techniques différentes du codage de la parole.

2.6.1. Codage temporel

Dans ce codage, la forme temporelle (d'onde) du signal est conservée. L'Erreur Quadratique Moyenne (EQM) est utilisée comme critère d'erreur d'évaluation de qualité. Ces techniques fournissent des débits élevés (16-64 kbps).

2.6.2. Le codeur MIC

Le MIC est la plus simple méthode de codage (Modulation par Impulsion et Codage) (recommandation G711 du CCITT5). Elle code le signal dans la bande téléphonique à 64 kbps en utilisant une loi logarithmique de quantification (loi A en Europe, loi « μ » Amérique du Nord, au Japon et en Australie).

2.6.3. Le codeur MICDA

Le débit peut être d'avantage réduit en utilisant la quantification adaptative et le codage prédictif. L'algorithme MICDA (MIC Différentiel Adaptatif) utilisant de telles techniques a été normalisé par le CCITT sous la recommandation G721.

La prédiction est réalisée par un modèle ARMA dont les paramètres sont remis à jour à chaque instant d'échantillonnage par un algorithme adaptatif du gradient. La qualité du signal codé reste à peu près la même que celle de l'algorithme MIC, cependant le débit est réduit à 32 kbps.

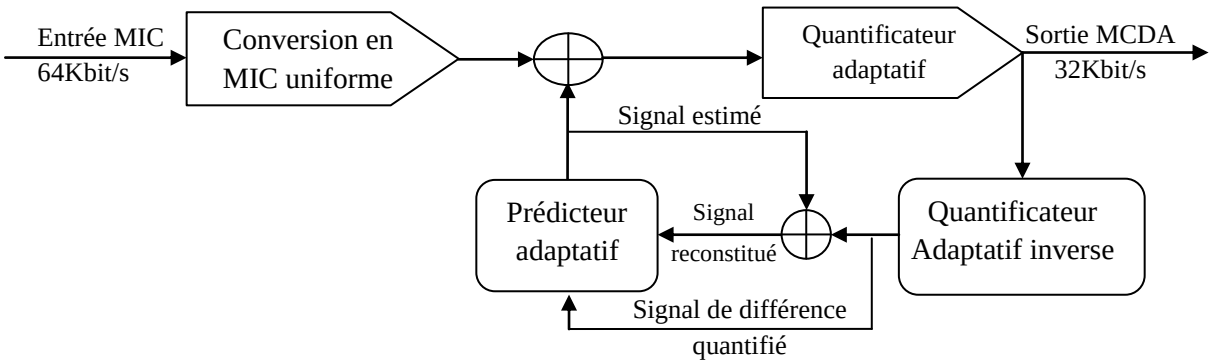


Fig.2.4: Schéma de principe d'un codeur MICDA.

2.6.4. Le codeur SB-MICDA

La version modifie du codeur MICDA est également utilisée pour le codage de la parole dans la bande élargie. Le signal de parole est limité à 7 kHz et échantillonné à 16 kHz. L'algorithme SB-MICDA (Sous-Bande MICDA) (recommandation G722 du CCITT) sépare l'entrée en deux sous-bandes par des filtres miroirs en quadrature.

Après un sous-échantillonnage à 8 kHz, chaque sous-bande est codée par un codeur MICDA. En réception, les signaux de chaque sous-bande sont décodés, traités par les filtres de synthèse et additionnés. Il fonctionne également aux débits de 56 et 48 kbps ayant une qualité inférieure [12].

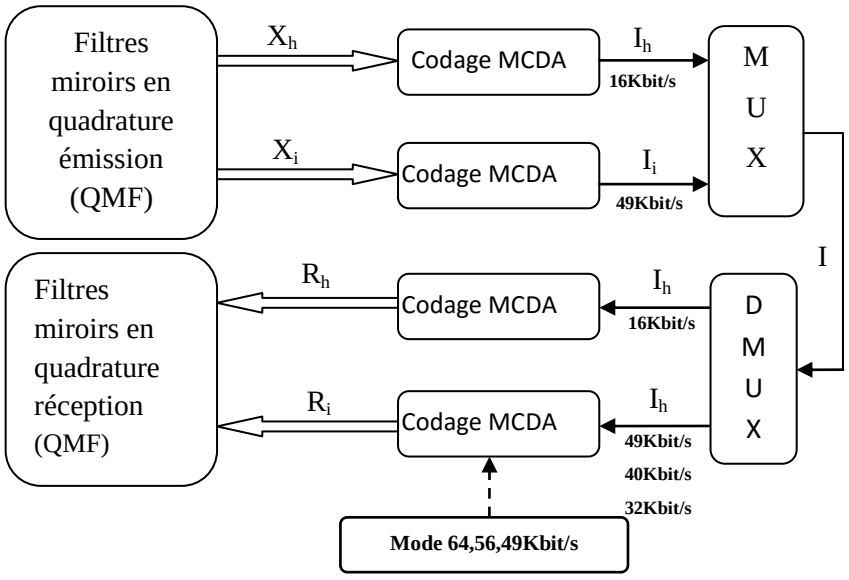


Fig.2.5: Schéma de principe d'un codeur/décodeur SB/MICDA.

2.6.5. Le codage en sous-bandes

Les codeurs en sous-bandes décomposent l'entrée dans un faible nombre (typiquement 2 -- 8) de bandes de fréquence. Les signaux dans les sous-bandes sont ensuite sous-échantillonnés et codés. En réception, les signaux de chaque sous-bande sont décodés, traités par les filtres de synthèse et additionnés. Le codage des signaux dans les sous-bandes est fait typiquement par des algorithmes de codage temporel, comme dans le cas de l'algorithme SB-MICDA. Cependant, il est bien possible d'utiliser d'autres types de codeurs.

2.6.6. Le codage transformé

Ce type de codage utilise une transformation linéaire pour projeter les données dans un espace transformé. Les coefficients obtenues par la transformation sont quantifiés et codés.

Les codeurs de la parole utilisant le codage transformé décomposent chaque trame de l'entrée en des composantes principales (Typiquement en 64 - 512 composantes fréquentielles) à l'aide d'une transformation unitaire. A la réception, la transformation inverse est appliquée sur les coefficients décodés.

Les principales transformations utilisées pour la compression sont la transformée en ondelettes, la transformation DCT et la KLT.

2.6.6.1. Codage utilisant la décomposition en ondelettes

Il permet de réaliser des décompositions temps-fréquence non-uniformes. , Ainsi il est possible d'adapter la décomposition soit aux caractéristiques du signal, soit aux propriétés de l'ouïe humaine. La décomposition en ondelettes (ou en paquets d'ondelettes) permet l'analyse et la synthèse du codage en sous-bandes ou du codage transformé.

2.6.6.2. Codage basé sur la prédiction linéaire

Les codeurs de prédiction linéaire sont fondés sur le modèle de la production de la parole, présenté sur la figure 2.6. L'excitation est composée d'une source quasi-périodique et du bruit blanc stationnaire ou du bruit transitoire. Le type du bruit (stationnaire ou transitoire) et le mélange de deux types d'excitation dépendent du type du son. L'effet de la glotte peut être modélisé par un filtre AR non causal [Gardner et Rao, 1997]. En ignorant la correspondance en phase, ce filtre est souvent remplacé par un filtre causal contenant un pôle de deuxième ordre [Atal et Hanauer, 1971]. Le conduit vocal peut être bien modélisé par un filtre AR causal, ne contenant que des pôles. Le conduit nasal est modélisé par un filtre ARMA possédant des pôles

et des zéros. Le rayonnement aux lèvres a un eut d'accentuation des fréquences hautes modélisé par un filtre contenant un zéro à 1 [13].

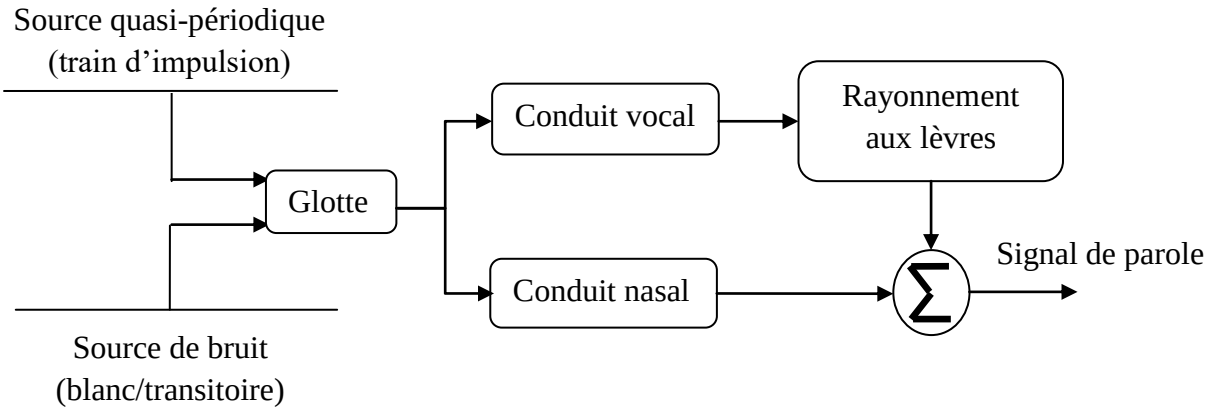


Fig.2.6: Modèle de la production de la parole.

Dans le cas de la modélisation autorégressif de la parole [Atal et Hanauer, 1971] (indiqué sur la figure 2.7), tous les effets précédemment considérés sont modélisés par un filtre AR, de façon plus-au-moins justifié. Cette structure sera détaillée dans la section suivante. L'effet d'accentuation aux lèvres est compensé par l'atténuation à la glotte, d'où il résulte une atténuation des fréquences hautes, modélisée par un ou deux pôles. Les zéros du couplage nasal peuvent être approximés par des pôles supplémentaires, de plus l'oreille est moins sensible à la localisation des zéros que celle de pôles. La qualité de la modélisation dépend de la représentation de l'excitation. L'avantage de la modélisation AR est qu'elle possède une méthode de calcul efficace (basé sur la prédiction linéaire) pour déterminer les coefficients du filtre comme on le verra dans la section 5.7.

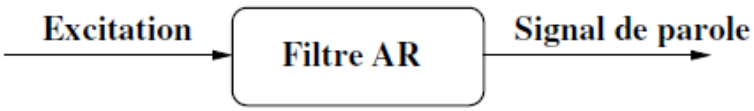


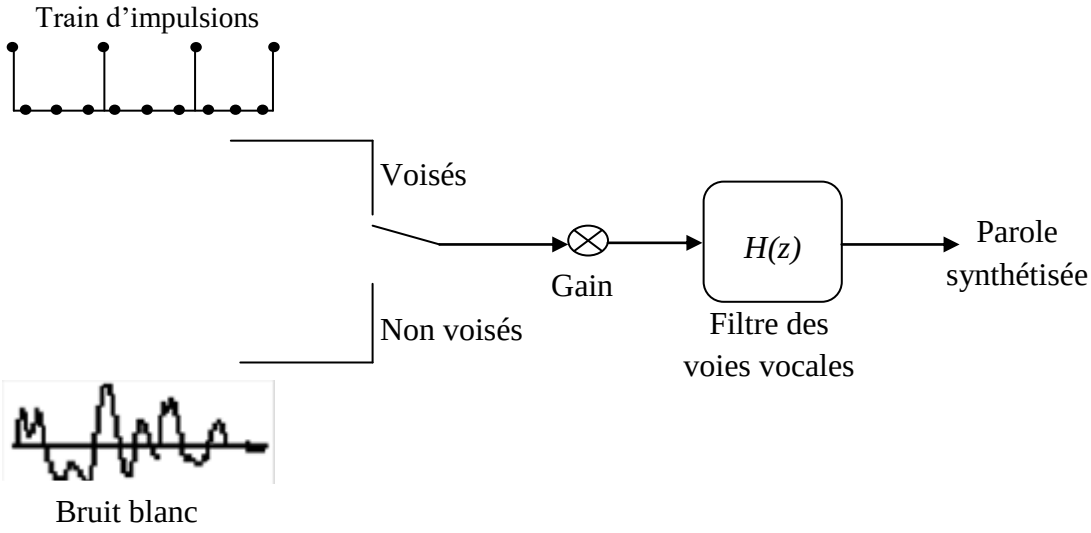
Fig.2.7: Modèle autorégressif de la production de la parole.

Les codeurs de prédiction linéaire transmettent au décodeur les paramètres du filtre AR et la description de l'excitation. Le décodeur est capable de synthétiser le signal de parole à partir des paramètres transmis.

2.6.7. Codage LPC paramétrique

Les codeurs paramétriques fonctionnant à bas débit (1-2 kbps) ne transmettent l'excitation que de façon grossière. L'exemple classique est le codeur LPC8-10 (norme FS9-1015) à 2 kbps qui construit l'excitation soit comme du bruit blanc stationnaire, soit comme un train d'impulsions dont la périodicité correspond à la fréquence fondamentale estimée. Ce modèle trop simple n'est pas capable de bien représenter ni les sons transitoires, ni les sons mixtes contenant du bruit et des composantes quasi-parodiques à la fois. De plus, le filtre AR ne modélise pas la structure de phase de la glotte. Dû à ces problèmes, la qualité du signal synthétisé manque de clarté, il est perçu comme un bruit et présente des artefacts de tonalité. De plus des erreurs de la classification bruit/périodique et de l'estimation de la fréquence fondamentale dégradent d'avantage la qualité. Le codeur a surtout des applications dans la communication militaire ou il est suffisant de garder l'intelligibilité de la parole [15].

Le codeur MELP10 [McCree et Barnwell III, 1995], [Supplee et al., 1997] ,la nouvelle norme du DOD à 2 kbps, remédie aux problèmes de LPC-10 en utilisant une représentation plus souple de l'excitation. Le mélange de la composante impulsionnelle et la composante de bruit dans l'excitation dépend de la bande de fréquence (5 bandes sont définies entre 0- 4000 Hz). La composante impulsionnelle peut être périodique ou apériodique. La dernière joue un rôle dans la modélisation des consonnes plosives non-voisées et des zones de transition entre sons. Le codeur extrait et transmet les amplitudes des 10 premières harmoniques du signal résiduel. Cela peut compenser les défauts de la modélisation tout pôle et améliorer la représentation du signal résiduel. Pour corriger les défauts de la modélisation encore d'avantage, le décodeur utilise un filtre adaptatif de renforcement des formants et un autre filtre dont le but est d'étaler l'énergie des impulsions sur une période de pitch. La qualité du codeur est sensiblement meilleure à celle du codeur LPC-10 [14].



Principe du codage LPC

2.6.8. Codage LPC hybride

Les codeurs LPC hybrides transmettent la forme d'onde de l'excitation. Pour réaliser une réduction importante du débit, l'excitation est fortement quantifiée. Un prédicteur du pitch est souvent appliqué sur le résidu d'analyse LPC, et l'on quantifie le second résidu ainsi obtenu. La qualité du signal codé peut être améliorée en utilisant un critère d'erreur perceptuel et une boucle fermée d'analyse-synthèse.

La technique d'analyse par synthèse quantifie le résidu en minimisant l'erreur sur le signal synthétique au lieu de minimiser l'erreur entre le résidu original et sa version quantifiée.

La norme GSM 06.10 définit un codeur à 13 kbps utilisant de telles techniques. La qualité du signal codé est loin d'être transparente, mais acceptable pour la radiotéléphonie.

Les codeurs CELP et ses dérivations permettent de réduire le débit jusqu'à environ 4 kbps en utilisant la quantification vectorielle pour coder le résidu. Les normes les plus importantes à mentionner sont la FS1016 4.8 kbps, la LD-CELP (norme G728) à 16 kbps, les codeurs ACELP (normes GSM EFR à 12.2 kbps, G729 à 8 kbps, G723.1 à 6.3 et 5.3 kbps, et G722.2 dans la bande élargie à 13-24 kbps), et les codeurs CELP de la norme MPEG-4 (3.85-12.2 kbps dans la bande téléphonique et 10.9-23.8 kbps dans la bande élargie) [13].

2.7. Conclusion

Alors, que choisir ? Tout dépend de l'utilisation du signal compressé. Lorsque qu'on veut travailler avec un format numérique tel qu'il est utilisé dans le milieu de l'audio professionnel, il est rigoureusement conseiller d'utiliser une technique de compression sans pertes.

Mais si l'on veut faire du stockage massif de donnée ou de la transmission de signaux via les canaux dont nous disposons actuellement (internet), il est préférable d'utiliser une méthode de compression avec pertes.

Dans ce chapitre, nous avons présenté les techniques les plus utilisées pour le codage et la compression de la parole. Comme le nombre et le type de codeurs-décodeurs est élevé, nous n'avons pas pu les présenter tous.

Notre travail repose sur la compression par la KLT, donc cette transformation sera exposée en détail dans le chapitre suivant.

Chapitre 3
Théorie de la KLT

3.1. Introduction

La KLT (Karhunen-Loève Transform) est la transformation linéaire réversible optimale qui permet d'enlever la redondance dans un signal par décorrélation des échantillons. Elle est connue aussi sous le nom de l'analyse en composantes principales (ACP ou PCA en anglais). Elle consiste à transformer des variables liées (dites "corrélées" en statistique) en nouvelles variables décorrélées les unes des autres et projetées sur des axes principaux "composantes principales", constituant une base orthonormée.

Il s'agit d'une approche à la fois géométrique (les variables étant représentées dans un nouvel espace, selon des directions d'énergie maximale) et statistique (la recherche portant sur des axes indépendants expliquant au mieux la variabilité — la variance — des données). Lorsqu'on veut compresser un ensemble de N variables aléatoires, les n premiers axes de l'analyse en composantes principales sont un meilleur choix, du point de vue de l'énergie ou de la variance.

Nous consacrons ce chapitre à l'étude théorique de la KLT car notre travail repose sur le concept fondamental de cette transformation. Dans la deuxième section, nous donnerons un petit historique sur la KLT.

La troisième section, nous présenterons la formulation de la KLT. Ainsi, nous montrerons que la rotation de l'espace contenant les échantillons corrélés d'un signal de parole peut enlever la corrélation. La base des vecteurs du nouvel espace définit la transformation linéaire de la donnée.

Les vecteurs de base de la KLT sont les vecteurs propres de la matrice de covariance des données analysées. En diagonalisant cette matrice, la KLT enlève la corrélation entre les échantillons voisins. La section 4, montre que la KLT minimise le taux de quantification (donnée par l'entropie du signal). L'entropie d'une variable aléatoire discrète et d'un processus aléatoire continu seront définis. La Notion du gain de codage par KLT est aussi abordée. En outre, l'effet de troncature et de l'intercorrélation entre trames sur le codage et la compression

est aussi mentionné. Nous terminerons ce chapitre en citant les critères d'évaluation de la qualité de compression.

3.2. Historique

L'ACP prend sa source dans un article de Karl Pearson publié en 1901(intéressé par la régression et les corrélations entre plusieurs variables). Il utilisa les corrélations pour décrire et résumer l'information contenue dans ces variables.

La transformée de Karhunen-Loève ou de transformée de Hotelling, a été de nouveau développée et formalisée dans les années 1930 par Harold Hotelling. La puissance mathématique de l'économiste et statisticien américain le conduira aussi à développer l'analyse canonique, généralisation des analyses factorielles dont fait partie l'ACP [16].

Dans les diverses applications de la KLT, elle agit essentiellement sur les données pour:

- Les décrire et visualiser.
- Les décorréler : obtenir une nouvelle base constituée d'axes non corrélés entre eux.
- Les débruiter : en considérant que les axes que l'on décide d'ignorer représentent un bruit.
- Les compresser : en supprimant la redondance.

3.3. Décorrélation des données

3.3.1. Décorrélation inter-canaux audio

La décorrélation entre canaux audio est parfaitement réalisée à l'aide de la KLT dans la phase de prétraitement [17,18]. Pour un signal corrélé, la transformation KLT est adoptée, car elle est théoriquement la méthode optimale pour dé-corréler les signaux de différents canaux. La figure 3.1 illustre la manière dont la KLT est appliquée sur des signaux audio multicanaux. Les colonnes de la matrice de transformation KL sont composées de vecteurs propres calculés à partir de la matrice de l'inter-covariance associée aux signaux audio multicanaux d'origine.

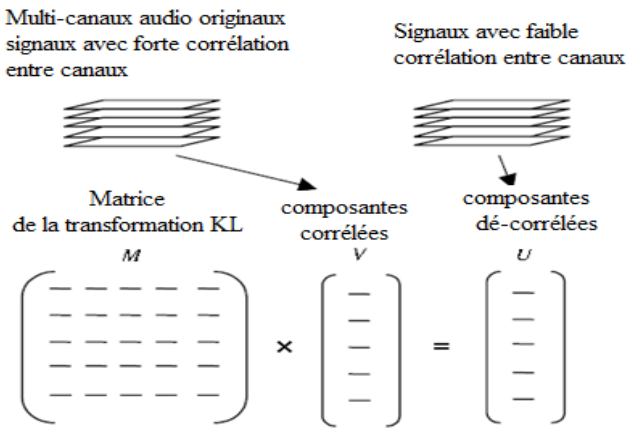


Fig.3.1: Décorrélation des signaux multicanaux par la KLT

Si le signal audio d'entrée possède *n* canaux, on forme une matrice M de la transformation KL de dimension $n \times n$ composée de *n* vecteurs propres de la matrice d'inter-covariance associée à ces *n* voies. Le vecteur dont les *n* éléments représentent la valeur du i^{eme} échantillon des canaux 1, 2,. . . , *n*, respectivement est *V(i)* tel que :

$$V(i) = [x_1, x_2, \dots, x_n]^T \qquad \textbf{Eq.3.1}$$

3.3.2. Décorrélation des échantillons d'un signal

Le codage par transformée est une des méthodes les plus importantes pour la compression du signal parole avec perte. La transformation linéaire de Karhunen-Loeve (KLT) est optimale au sens de la minimisation de l'erreur quadratique moyenne (erreur de reconstruction).

Pour tout vecteur de données X ={x_1, x_2,..., x_N }, c.-à-d. de dimension N, la KLT est définie par :

$$Y = V^T(X - m_X) \qquad \boldsymbol{Eq.3.2}$$

Où ;

m_X : est la moyenne du vecteur de la donnée originale X.

V^T : matrice des vecteurs propres de la matrice d'autocorrélation.

Y : projection de x sur l'espace transformée.

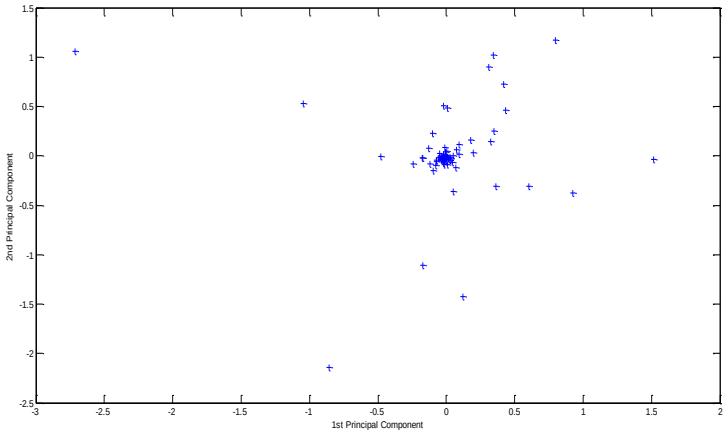


Fig.3.2. : Dispersion des échantillons d'un signal de parole dans l'espace transformé par KLT

La figure ci-dessus montre la projection des échantillons d'un signal de parole sur les deux composantes principales (sur deux axes principaux) obtenues par la KLT). Nous remarquons clairement la décorrélation de la majorité des échantillons.

3.4. La transformée KL

3.4.1. Formulation de la KLT

Cette procédure s'effectue en deux étapes :

1) La première étape consiste au réarrangement du signal de la parole en une matrice contenant les trames considérées non stationnaires (transformer un vecteur en une matrice de dimension $m \times p$). Par exemple on décompose un signal vocal de 30000 échantillons, en 58 trames de 512 échantillons. Dans ce cas, on a une matrice X de dimension 512 x 58.
2) La deuxième étape correspond à l'analyse de la matrice obtenue précédemment. Avant d'entamer le traitement, il faut supposer que X est de type aléatoire.

Supposons qu'on a une matrice $X = [x_{ij}]$ de type variable aléatoire de dimension m × p, et on veut trouver une transformation linéaire V_d pour compresser X à Z définie par :

$$Z = V_d^T (X - \bar{X}) \qquad \textbf{Eq.3.3}$$

Où ; $\bar{X} = E(X)$ est la moyenne de X, l'operateur $E(.)$ désigne l'espérance mathématique, et $V_d = (v_1, v_2, \ldots, v_d) \in R_{m \times d}$ est une transformation à trouver.

L'équation (3.3) met en œuvre une réduction de dimension de X à Z. L'inversons du processus permet de trouver Z. Et nous pouvons reconstruire X par : $\hat{X} = V_d Z + \bar{X}$ au moyen de l'équation (3.3).

Maintenant, la clé de la mise en œuvre de la compression est de trouver la transformation V_d. Nous adoptons toujours le Critère d'Erreur de Reconstruction (CER) tel qu'il est utilisé dans la KLT et de le minimiser afin d'aboutir à une transformation optimale V_d. Le CER de V_d est donné comme suit:

$$\begin{aligned} RCE(V_d) &= E\left\{\left\|X - \hat{X}\right\|^2\right\} = E\{\|Z - V_d Z - \bar{X}\|^2\} = E\{\|X - \bar{X} - V_d V_d^T (X - \bar{X})\|^2\} \\ &= E\left\{tr\left(\left(X - \bar{X} - V_d V_d^T (X - \bar{X})\right)\left(X - \bar{X} - V_d V_d^T (X - \bar{X})\right)^T\right)\right\} \\ &= E(\|X - \bar{X}\|^2) - tr\left(V_d^T E\left((X - \bar{X})(X - \bar{X})^T\right) V_d\right) \end{aligned} \qquad \textbf{Eq. 3. 4}$$

Où ; *tr*(.) représente la trace d'une matrice, et I la matrice identité de dimension d × d, et $\|X\| = \left(\sum_{i=0}^{m} \sum_{j=1}^{p} |x_{ij}|^2\right)$, est la norme Frobenius. Et les deux dernières lignes de l'équation (3.4) sont issues des deux caractères qui suivent, de la trace et de la norme Frobenius:

1) $\|X\| = tr(XX^T)$

2) $tr(XZ) = tr(ZX)$

3) $I = V_d^T V_d$

Mettons $R = E\left((X - \bar{X})(X - \bar{X})^T\right)$, c.-à-d., une matrice de covariance générale pour un type de matrice variable aléatoire X.

De l'équation (3.4), la minimisation du (CER) par rapport à V_d est équivalente à la maximisation de $J(V_d)$ défini par :

$$J(V_d) = tr(V_d^T R V_d) \qquad \textbf{Eq.3.5}$$

Sous la contrainte que $V_d^T V_d = I$. Les équations (3.4) et (3.5) donnent une optimisation identique à la formulation de la KLT sauf que la matrice de covariance de la KLT construite à l'aide d'un vecteur de type variable aléatoire est remplacée par une nouvelle matrice de covariance générale (MCG : matrice identité). En effet, la dérivation suivante de la transformée est identique à celle de la KLT. Toutefois, pour l'exhaustivité de la description, nous donnons toujours la reformulation ci-dessous.

Pour maximiser $J(V_d)$ sous des contraintes, nous définissons une fonction Lagrangien comme suit:

$$L(V_d) = J(V_d) - \sum_{j=1}^{d} \lambda_j \left(V_j^T V_j - 1\right) \qquad \textbf{Eq. 3.6}$$

Où ; λ_j sont des multiplicateurs de Lagrange. Par différenciation de (3.6) par rapport à V_j et en annulant les dérivées correspondantes, nous obtenons :

$$RV_j = \lambda_j V_j \qquad j = 1,2,\ldots..,d \qquad \textbf{Eq. 3.7}$$

Ceci représente un système en fonction des valeurs et vecteurs propres par rapport à (λ_j, v_j). En tenant compte que R (matrice d'auto-corrélation) est symétrique et semi-définie positive, ses valeurs propres sont non négatives. Donc, l'équation (3.6) admet comme valeur maximale :

$$J(V_d) = \sum_{j=1}^{d} \lambda_j \qquad \textbf{Eq. 3.8}$$

Ce qui correspond à un minimum de $RCE(V_d) = \sum_{j=d+1}^{m} \lambda_j$.

Ici, nous prenons les d vecteurs propres de R correspondant aux premières plus grandes valeurs propres pour construire la transformée nécessaire $V_d = (V_1, V_2, \ldots, V_d)^T$ qui donne une erreur de reconstruction minimale (REC) et en même temps conserve au maximum les informations originales des données. Par conséquent, les données initiales sont efficacement comprimées.

3.4.2. Procédure de décorrélation

Comme nous l'avons dit, la KLT peut enlever la corrélation entre les échantillons d'un signal original de données pour des fins de compression. En fait, notre matrice KLT possède également une telle propriété comme analysé ci-dessous.

Réécrivons Z en lignes de $[(z_1)^T, (z_2)^T, \ldots, (z_d)^T]^T$,

Où ;

$$Z_j = V_j^T (X - \bar{X}) \quad j = 1,2,\dots.,d \qquad \textbf{Eq. 3.9}$$

Ensuite, pour chaque Z_i et Z_j , on calcule la matrice d'auto-corrélation définie par :

$$E\left(Z_i Z_j^T\right) = E\left[V_i^T (X - \bar{X})(X - \bar{X})^T V_j\right] = V_i^T R V_j = \lambda_j V_i^T V_j = \begin{cases} \lambda_j & i = j \\ 0 & i \neq j \end{cases} \qquad \textbf{Eq. 3.10}$$

Dans le développement ci-dessus, on a utilisé l'équation (3.5). L'équation (3.7), nous informe que chaque deux vecteurs lignes différents Z_i et Z_j de Z sont décorrélés, ce qui est utile à la compression ultérieure. En particulier, lorsque la matrice X est concaténée dans un vecteur, notre matrice KLT est réduite.

Si les vecteurs propres qui forment la matrice de transformation sont ordonnés dans l'ordre décroissant des valeurs propres correspondantes, les vecteurs transformés sont classés par rapport à leur importance pour la synthèse du signal original avec l'erreur minimale. La compression consiste à conserver uniquement les composantes qui donnent une restauration désirable du signal évaluée par un certain critère de qualité.

3.4.3. Reconstruction réelle

Généralement, dans une mise en œuvre réelle, nous n'avons qu'un ensemble limité d'échantillons de la matrice de données$\{X_i \ , i = 1,2,\dots.,N\}$. Donc, la matrice de covariance (R) et la moyenne $\bar{X}$ seront respectivement estimées par l'ensemble des échantillons de données comme suit:

$$R = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})(X_i - \bar{X})^T \ \ et \ \ \bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \qquad \textbf{Eq. 3.11}$$

Où ; X_i représente les vecteurs des échantillons du signal à compresser et $\bar{X}$ leur moyenne.

En les substituant dans l'équation (3.7), et en résolvant cette dernière, nous pouvons obtenir une transformation V_d pour l'ensemble de données.

Pour implémenter la décompression ou la reconstruction, nous utilisons l'équation suivante :

$$\hat{X} = V_d Z + \bar{Z} \qquad \textbf{Eq. 3.12}$$

Elle permet la restauration du signal de la parole originale avec un minimum d'erreurs.

3.4.4. Calcul de la KLT

3.4.4.1. Estimation de la covariance

Le calcul de la KLT est généralement effectué en trouvant les vecteurs propres de la matrice de covariance, ceci nécessite une estimation de la matrice de covariance. Si le signal complet est disponible, comme cela est le cas pour le codage d'un signal 1D, la matrice de covariance peut être estimée à partir de n échantillons de données en tant que :

$$\left[\hat{C}\right]_x = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T \qquad \textbf{Eq. 3. 13}$$

Où ; X est un vecteur des échantillons de données. Si seulement des parties du signal sont disponibles, il faut s'assurer que l'estimation est représentative du signal entier. Dans l'extrême, si un seul vecteur de données est utilisé, alors une seule valeur propre non nulle existe, et son vecteur propre est tout simplement la version réduite du vecteur de données.

3.4.4.2. Calcul des vecteurs propres

Bien qu'il soit au-delà de la portée de ce chapitre pour fournir une étude détaillée des algorithmes pour extraire les valeurs et vecteurs propres, nous allons présenter un bref aperçu des méthodes générales couramment utilisées. Pour des explications plus détaillées, le lecteur est renvoyé vers des références plus spécialisées.

Une approche simple est la méthode de Jacobi. Il développe une séquence de matrices de rotation, $[P]_i$, qui diagonalise [C] comme :

$$[D] = [V]^T[C][V] \qquad \textbf{Eq. 3. 14}$$

Où ;

[D] : est la matrice diagonale désirée

et $[V]=[P]_1[P]_2[P]_3 \cdot \cdot$ tel que chaque $[P]_i$ tourne dans un plan pour éliminer l'un des éléments hors diagonale.

C'est une technique itérative qui s'arrête lorsque les valeurs hors diagonale sont proches de zéro avec une certaine tolérance. En fin, la matrice [D] contient les valeurs propres sur les diagonales et les colonnes de [V] sont les vecteurs de la base KLT.

Bien que cette technique soit assez simple, pour les grandes matrices, la complexité de calcul de l'algorithme est importante pour qu'il converge.

Une approche plus efficace pour les grandes matrices symétriques divise le problème en deux étapes. L'algorithme de Householder peut être appliqué pour réduire une matrice symétrique en une forme tri-diagonale dans un nombre fini d'étapes. Une fois que la matrice est sous cette forme plus simple, une méthode itérative telle que la factorisation **QL** peut être utilisée pour générer les valeurs et vecteurs propres. L'avantage de cette approche est que la factorisation de la matrice tri-diagonale simplifiée nécessite généralement moins d'itérations que la méthode de Jacobi [19].

3.4.5 Performances des transformées

En soi, une transformation orthonormale n'affecte pas la compression des données. Les blocs du signal sont simplement transformés d'un ensemble de valeurs à un autre et, pour des transformations réversibles, le retour permet la reconstruction. Afin de réduire le nombre de bits pour représenter un signal, les coefficients sont quantifiés, subissant une perte irréversible, puis codés pour une représentation plus efficace. Avant la dé-corrélation des données par la KLT, un prétraitement de données peut être nécessaire. Pour examiner les effets de cette efficacité supplémentaire, nous pouvons utiliser des mesures d'information de Shannon [20].

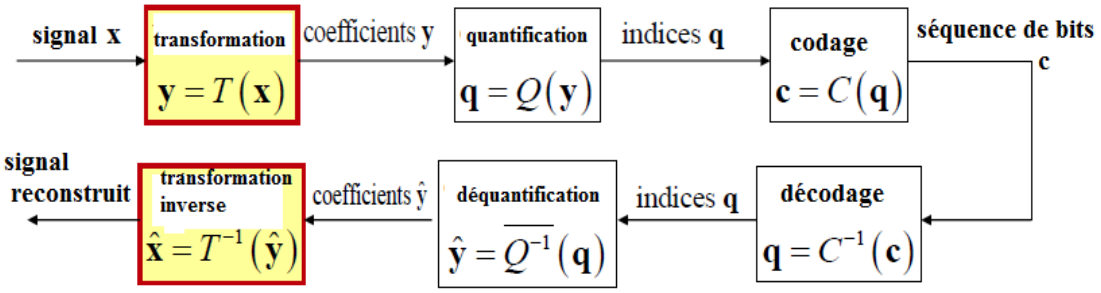


Fig.3.3: Structure d'un codeur /décodeur par transformé

Le schéma ci-dessus montre les constituants d'un système de codage et décodage d'un signal par transformation.

Où;

$T(\boldsymbol{x})$: est une transformation généralement inversible.

$Q(\mathbf{y})$: Quantification non inversible, introduit une distorsion.

T^{-1} : Transformée inverse.

3.4.5.1. Théorie de l'information

L'information véhiculée par une observation de certains processus aléatoires est liée à sa probabilité d'occurrence. Si une observation était presque certaine de se produire, ce qui veut dire, la probabilité était proche de 1, ce ne serait pas très instructif. Toutefois, s'il était tout à fait inattendu, l'observation transmettrait beaucoup plus d'informations. Shannon formalise cette relation entre la probabilité d'un événement, P(x), et le contenu de l'information I(x), tel que :

$$I(x) = -logP(x) \qquad \boldsymbol{Eq.3.15}$$

Si le logarithme est pris par rapport à la base 2, l'information, I (x), est mesurée en unités de bits.

Une variable aléatoire, X, est une collection de tous les événements possibles et leurs probabilités associées. L'information moyenne d'une variable aléatoire peut être calculée comme :

$$H(x) = \sum_i P(x_i)\, I(x_i) = -\sum_i P(x_i)\, logP(x_i) \qquad \boldsymbol{Eq.3.16}$$

Où ; la somme est prise à travers tous les événements possibles. L'information moyenne est appelée l'entropie du processus.

L'entropie est utile pour déterminer les mesures de rendement théoriques des méthodes de compression. Shannon a montré que l'entropie donne une borne inférieure sur le nombre moyen de bits requis pour coder les événements d'un processus aléatoire sans introduire d'erreur. En d'autres termes, il faut au moins autant de bits par événement, en moyenne, que l'entropie pour représenter un ensemble d'observations.

Toutefois, ces mesures ne sont pas directement applicables aux coefficients d'une transformation arbitraire. Ils sont définis pour des événements discrets, alors que les coefficients, étant donné que ce sont des valeurs à virgule flottante, doivent être considérés comme des échantillons à valeurs réelles des distributions continues. Comme la probabilité d'un tel échantillon de valeur réelle est égale à zéro, l'entropie discrète est indéfinie. Au lieu de cela, nous définissons l'entropie différentielle [21].

$$h(x) = \int_{-\infty}^{+\infty} p(s) log\ p(s)\, ds \qquad \textbf{Eq. 3.17}$$

Pour les distributions d'échantillons telles que la loi gaussienne, uniforme ou des distributions de Laplace, l'entropie différentielle est de la forme :

$$h(x) = \frac{1}{2} log \sigma_x^2 + k \qquad \textbf{Eq. 3.18}$$

Où ;

σ_x^2 : est la variance de la variable aléatoire et k est une constante de dépendance de la distribution (par exemple, par une gaussienne, $k = \frac{1}{2} log_2 2\pi e$) [1].

Une bonne transformation devrait donc minimiser la somme des entropies différentielles pour les coefficients résultants. En raison de l'expression logarithmique, ce qui équivaut à réduire au minimum le produit des variances des coefficients. Cependant, rappelons que pour toute transformation orthonormale, l'énergie totale est conservée, de sorte que la somme des variances des coefficients est constante. Une mesure de l'efficacité de la transformation est le gain de codage [22]. Il est défini comme le rapport entre la moyenne algébrique des écarts (indépendant de la transformation) et la moyenne géométrique des écarts (dépendante de transformation):

$$G_w = \frac{\frac{1}{N}\sum_{i=1}^{N} \sigma_{y_i}^2}{\left(\prod_{i=1}^{N} \sigma_{y_i}^2\right)^{1/N}} \qquad \textbf{Eq. 3.19}$$

Pour le signal original, avant toute transformation, toutes les variances sont approximativement égales donnant un gain d'unité de codage. Toute augmentation dans l'un des écarts de coefficient doit être compensée par une baisse équivalente de l'un ou plusieurs des autres écarts pour une transformation orthonormale. La moyenne arithmétique est donc la même, mais la moyenne géométrique diminue résultant en un gain de codage supérieur à un.

Pour une énergie donnée du signal, la réduction du produit de variances maximise le gain de codage. A l'inverse, la maximisation du gain de codage réduit la limite inférieure du nombre de bits requis pour coder la parole. Ainsi, afin de minimiser le produit des écarts pour donner une somme constante, il faut maximiser la variance du premier coefficient. Ensuite, sous réserve de la contrainte d'orthonormalité, maximiser la variance du second coefficient, et ainsi de suite.

Cette procédure est rien de plus que d'extraire les composantes principales ou, de façon équivalente, la génération du KLT. Par conséquent, la KLT, par décorrélation des données, produit un ensemble de coefficients qui minimise l'entropie différentielle des données.

3.4.5.2. Quantification

Dans le codage par transformation, les coefficients de transformation sont quantifiés pour effectuer la réduction des données. Lorsque la transformation est réversible, la quantification ne l'est pas, et donc introduit une erreur. Soit **y** l'ensemble des valeurs des coefficients quantifiés pour un bloc. La reconstruction du bloc est calculé par :

$$\hat{x} = [W]\hat{y} \qquad \textbf{Eq. 3.20}$$

L'expression de l'Erreur Quadratique du bloc est donnée par :

$$\varepsilon^2 = \|\hat{x} - x\|^2 = (\hat{x} - x)^T(\hat{x} - x) = ([W]\hat{y} - [W]y)^T([W]\hat{y} - [W]y)$$
$$= (\hat{y} - y)^T(\hat{y} - y)[W]^T[W] = (\hat{y} - y)^T(\hat{y} - y) = \|\hat{y} - y\|^2 \qquad \textbf{Eq. 3.21}$$

Ainsi, l'erreur quadratique de reconstruction est la même que l'erreur quadratique des coefficients de la transformation orthonormale.

Les coefficients quantifiés sont généralement codés en utilisant un procédé sans perte, tel que le codage arithmétique ou un codage de Huffman (déjà mentionné au chapitre 2). Ces procédés peuvent, au mieux, réduire le nombre moyen de bits de l'entropie des coefficients quantifiés.

Afin d'illustrer l'avantage de réaliser la KLT avant quantification, on calcule l'entropie totale pour un certain nombre d'intervalles de quantification sur les données originales et les données transformées. Pour cet exemple, une étape intermédiaire, la quantification uniforme est utilisée lorsque la valeur de quantification est calculée comme suit :

$$\hat{y} = q\ round\left(\frac{y}{q}\right) \qquad \textbf{Eq. 3.22}$$

En fonction de la largeur de l'intervalle de quantification q, où la fonction d'arrondissement round(x) renvoie l'entier le plus proche à la valeur réelle x. Pour une erreur quadratique de quantification donnée, l'entropie de bits par échantillon est inférieure pour les données transformées que pour les données d'origine.

3.4.5.3. L'erreur de troncature décodage de quantification

Une autre approche pour réduire les données et donc l'introduction d'erreur, est la suppression complète d'un certain nombre de coefficients avant la quantification. Cela veut dire que M coefficients seulement de N devront être retenus. L'erreur quadratique résultante (estimée) est calculée comme suit :

$$E[\varepsilon^2] = E\left[\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2\right] = \frac{1}{N}E\left[\sum_{i=1}^{M}(y_i - \hat{y}_i)^2 + \sum_{i=M+1}^{N}(y_i - 0)^2\right]$$

$$= E\frac{1}{N}\left[\sum_{i=M+1}^{M}(y_i)^2\right] = \frac{1}{N}\sum_{i=M+1}^{N}\sigma_i^2 \qquad \boldsymbol{Eq.3.23}$$

Rappelons que pour la KLT les variances des coefficients, σ_i^2, sont les valeurs propres, λ_i, de la matrice de covariance. Pour minimiser l'erreur quadratique estimée, les M coefficients correspondant aux M plus grandes valeurs propres doivent être conservés.

Notons que la minimisation ci-dessus est valable pour toute transformation dont M bases des vecteurs balayent l'espace de dimensions M définies par les M plus grandes composantes principales (vecteurs propres pour les M plus grandes valeurs propres). Cependant, seule la KLT assure que les coefficients restants peuvent être codés avec un nombre minimum de bits, car elle minimise l'entropie différentielle des coefficients [19].

3.4.5.4. Taille de la trame

La question reste de quelle taille utiliser pour les trames de la parole. Plus la longueur de trame est grande, plus la décorrélation est bonne, par conséquent le gain de codage est important. Cependant, le nombre des opérations arithmétiques pour les transformations inverse et directe augmente de façon linéaire avec le nombre d'échantillons dans la trame. En outre, la taille de la matrice de covariance est le carré du nombre des échantillons. Non seulement le calcul des vecteurs propres nécessite plus de ressources, mais le nombre d'échantillons pour obtenir une estimation raisonnable de la matrice de covariance augmente de façon significative.

De plus, si l'ensemble des vecteurs de la base KLT doit être gardé pour la reconstruction du signal, la taille de l'ensemble de base est également nécessaire. Par conséquent, il existe un compromis entre les exigences de calcul et le degré de décorrélation pour déterminer la taille de trame.

3.5. Evaluation de la qualité de la parole

L'évaluation des performances d'un algorithme de codage d'un signal vocal dépend des paramètres suivants :

- taux de bits
- qualité de parole reconstruite
- complexité de l'algorithme
- retard introduit

Pour les codeurs de parole haute qualité à bas débit, les algorithmes sont très complexes et s'implantent en temps réel bas débit (par exemples sur DSP d'au moins 12 MIPS). Le retard introduit (codage + décodage) est entre 50 et 60 ms

3.5.1. Classification de la qualité de parole

La parole peut être divisée en classes de qualité comme suit :

- Haute qualité (radiodiffusion) > 64 kbits / s
- qualité réseau : comparable à la parole analogique > 16 kbits / s
- qualité communications : parole très naturelle, hautement intelligible, pour télécoms > 4.8 kbits/s.
- qualité synthétique : intelligible mais non naturelle, perte de reconnaissance du locuteur.

On peut évaluer la qualité de la parole selon deux types de critères : Objectifs et subjectifs

3.5.2. Critères Objectifs

Dans ce type on trouve :

- le Rapport Signal sur Bruit (SNR : utilisé en parole non stationnaire), SNR segmental (SNR moyenné sur des tranches temporelles).

3.5.3. Critères Subjectifs

Pour tenir compte des propriétés perceptuelles de l'oreille, des tests d'écoute en double aveugle avec référence cachée sont utilisées pour évaluer la qualité du son. Le critère MOS (Mean Opinion Score) est largement utilisé pour décrire la qualité d'un signal vocal. Il est calculé par la moyenne des notes que donnent plusieurs auditeurs du son reconstruit. Il nécessite 12 à 24 auditeurs entraînés voire 32 à 64 pour normalisation, avec un intervalle de

confiance allant jusqu' à 95%. La note que peut prendre une parole écoutée dans l'échelle MOS est donnée dans le tableau 2.1.

Note	Qualité de la parole	Niveau de dégradation
5	Excellente	imperceptible
4	bonne	perceptible mais non gênant
3	Moyenne	légèrement gênant
2	Mauvaise	gênant
1	Très mauvaise	très gênant

Tableau.2.1: l'échelle MOS

3.6. Conclusion

Dans ce chapitre nous avons présenté la théorie de la transformation KLT. Comme introduction, on a entamé le problème de la décorrélation des données.

La transformée KL, est une des méthodes les plus importantes pour la compression du signal de parole avec pertes. Malgré qu'elle est optimale en sens de l'erreur quadratique moyenne de reconstruction minimale, la KLT (Karhunen-Loeve) ou analyse en composante principale (PCA) peut être difficilement utilisée dans la compression de la parole dû à sa vitesse lente dans la recherche de la matrice de covariance construite par des données d'apprentissage. Nous avons terminé ce chapitre par l'exposition des différents critères d'évaluation de la qualité de la parole tels que : MSE, SNR et la cotation MOS.

Chapitre 4
Résultats expérimentaux

4.1. Introduction

Après avoir présenté la théorie de la transformée KLT dans le chapitre précèdent, cette partie sera consacrée à la mise en œuvre de cette dernière à la compression de la parole. Pour cela nous avons utilisé un signal son en langue française choisi d'une base de données de test qui est « la bise et le soleil se …… », dont le locuteur est une femme.

Notre première expérience consiste à lire un fichier .wav, à l'écouter et à le représenter temporellement.

Ensuite, nous entamons la simulation de différentes expériences de compression.

4.2. Outils de travail

4.2.1. Logiciel ''MATLAB''

Entre 1985 et 1990, plusieurs logiciels interactifs de calcul scientifique sont apparus sur le marché dont MATLAB est l'un d'entre eux.

La syntaxe et la structure de son langage de programmation offrent les mêmes possibilités que les langages polyvalents de programmation structurée comme le Pascal, le C ou le Basic.

Sa simplicité fait de lui un outil de choix pour la mise au point de logiciels scientifiques. Ses nombreuses bibliothèques "toolboxes" de fonctions préexistantes, simplifient et rendent plus fiable la résolution des problèmes par l'utilisateur. De plus, ses fonctions graphiques puissantes et simples d'utilisation, permettent une visualisation immédiate des résultats, sous forme de graphiques en deux ou trois dimensions. De plus, sa disponibilité à un prix raisonnable sur la plupart des ordinateurs existants, et sa portabilité totale, qui permet au même programme MATLAB d'être exécuté sur n'importe quel ordinateur [23].

4.2.2. MATLAB et les fichiers audio

Matlab (Matrix Laboratory) est le logiciel utilisé pour l'implémentation de notre algorithme de compression, il présente plusieurs avantages pour le traitement du signal, il a été choisi pour sa simplicité et sa puissance.

Vu le nombre de fonctions offertes et facilitant la manipulation d'audio, nous avons décidé de se servir de Matlab. En fait, il est très aisé de lire, afficher, filtrer un son sous Matlab.

A cela s'ajoutent d'autres opérations qui sont résumées comme suit :

Fonction	Fonctionnalité
[y,Fe,B] = wavread ('filname')	Charger un fichier audio dans le vecteur y. (Fe fréquence d'échantillonnage, B nombre de bits).
[N, p]=size(y)	Donner le nombre des pistes p, et le nombre d'échantillons N.
plot(y)	Représenter graphiquement un signal y en fonction des indices.
t= (0 : N-1)*Fe ; plot (t, y)	Représenter graphiquement un signal y en fonction du temps.
Y=wavrecord(n*Fe,Fe,'filname')	Pour enregistrer un son (n: durée(s), Fe: Fréquence d'échantillonnage, dans un fichier nommé filename).
Soundsc(filename,Fs)	Pour lire le signal audio 'filename'

Tableau.4.1: fonctions Matlab pour le traitement d'audio.

4.3. Résultats

Pour donner plus de précisions à nos résultats, nous avons appliqué l'algorithme de compression par KLT à deux types de sons : le premier concerne une voix féminine et le deuxième une voix masculine.

4.3.1. Locutrice femme

L'audiogramme du signal analysé est donné par la figure 4.1

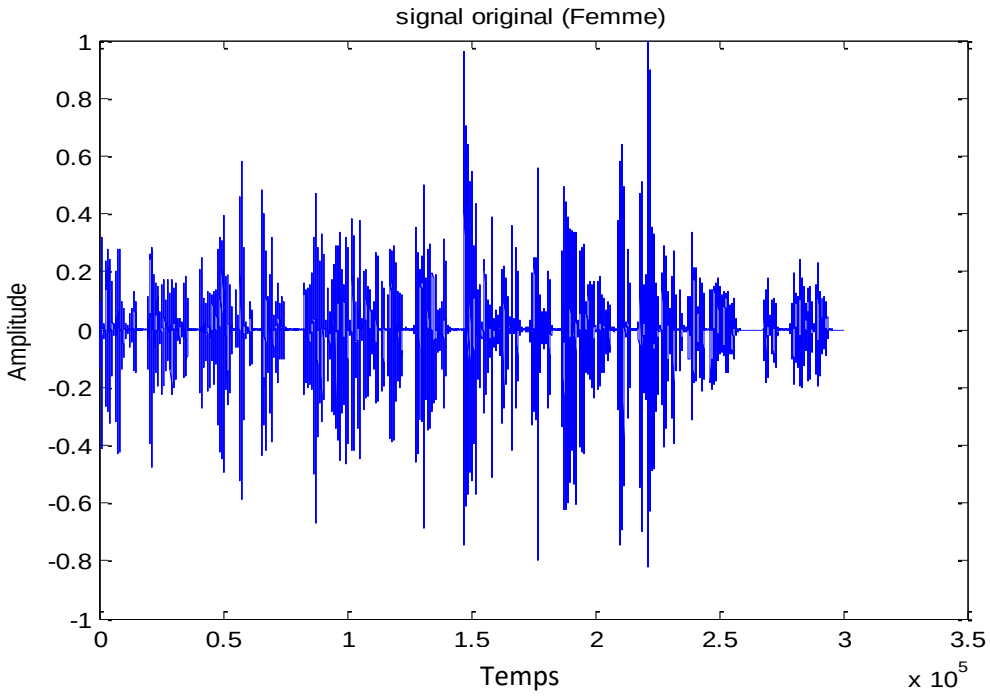


Fig.4.1. L'audiogramme du signal analysé

Ce signal sera décomposé en trames de durées permettant d'avoir la quasi-stationnarité.

Puisque la qualité de compression est fonction du nombre de coefficients retenus pour chaque trame du signal, nous allons faire trois essais pour voir l'effet de ce phénomène :

Pour pouvoir commenter les graphes, nous avons tronqué le signal original à 5000 échantillons lors des tracés des différentes simulations.

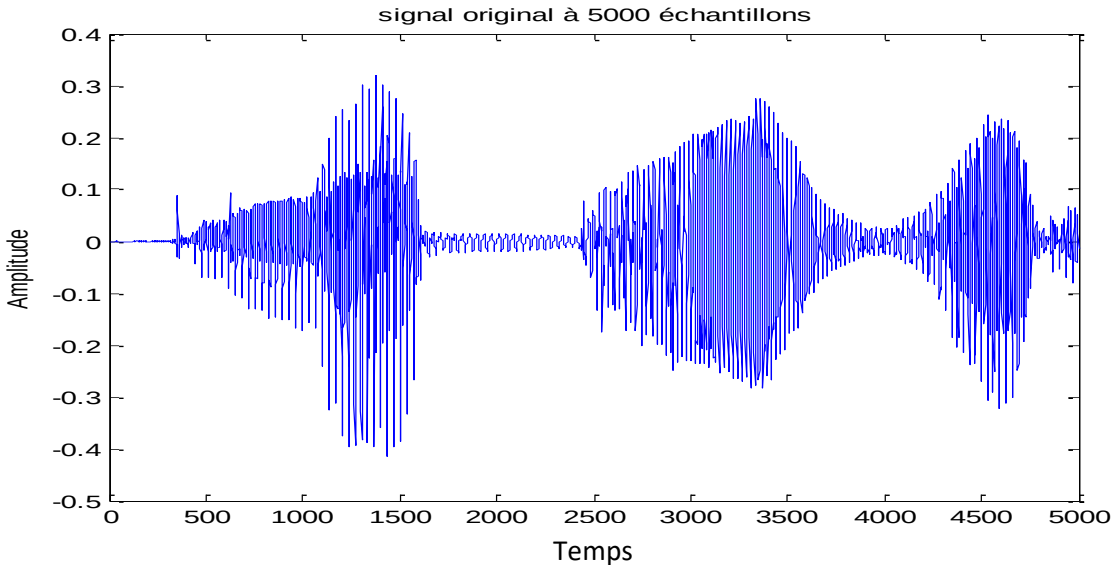


Fig.4.2: L'audiogramme du signal analysé tronqué à 5000 échantillons

Nous allons comparer, au travers des expériences qui suivent les performances de la KLT par rapport à la DCT en taux de compression. Pour cela, nous réaliserons la compression par les deux méthodes en conservant uniquement les coefficients les plus importants et représenter les résultats comme suit :

4.3.1.1. Compression avec 20 coefficients

Soit N=20, représente le nombre de coefficients retenus par trame de 512 échantillons. Chaque trame est analysée individuellement par DCT (et KLT), les coefficients transformés sont tronqués à 20 (les autres sont mis à zéros). La reconstruction est faite par transformation inverse après compression. Pour retrouver le signal complet, nous avons concaténé les différentes trames. Les résultats de cette expérience sont illustrés dans la figure 4.3.

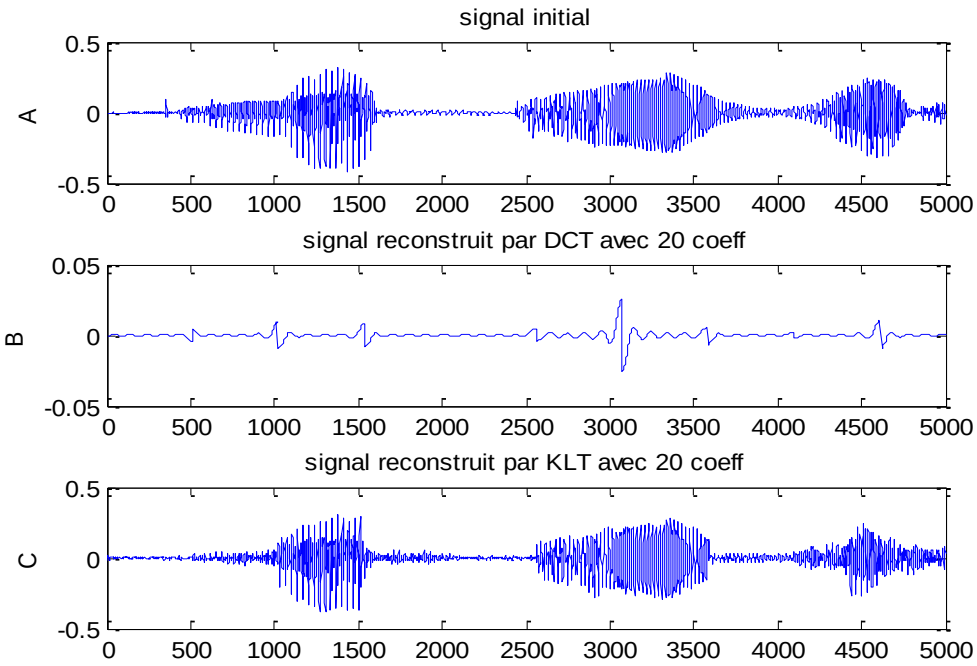


Fig.4.3: Reconstruction d'un signal de parole avec 20 coefficients

(A) Signal original, (B) reconstruit avec DCT, (C) reconstruit avec KLT

Nous constatons sur la figure 4.3 que le signal reconstruit avec KLT s'approche le mieux au signal original.

Du tableau 4.2, les tests d'écoute montre que la qualité du signal est meilleure pour la KLT. Nous notons que les tests MOS ont été effectués par nous-mêmes et en s'aidant de nos collègues.

Puisque le signal obtenu dans cette expérience est de mauvaise qualité, nous allons augmenter le nombre de coefficients et nous prenons dans l'essai suivant N=80 coefficients

4.3.1.2. Compression avec N=80

Pour ce cas, nous avons retenu 80 coefficients (N=80) pour chaque trame afin de reconstruire le signal initial. Nous remarquerons à travers la figure 4.4, que le signal reconstruit par KLT est très similaire au signal original, à un point où on ne peut pas remarquer la différence entre les deux audiogrammes, par contre celui reconstruit par DCT commence à s'améliorer et se rapprocher du signal initial.

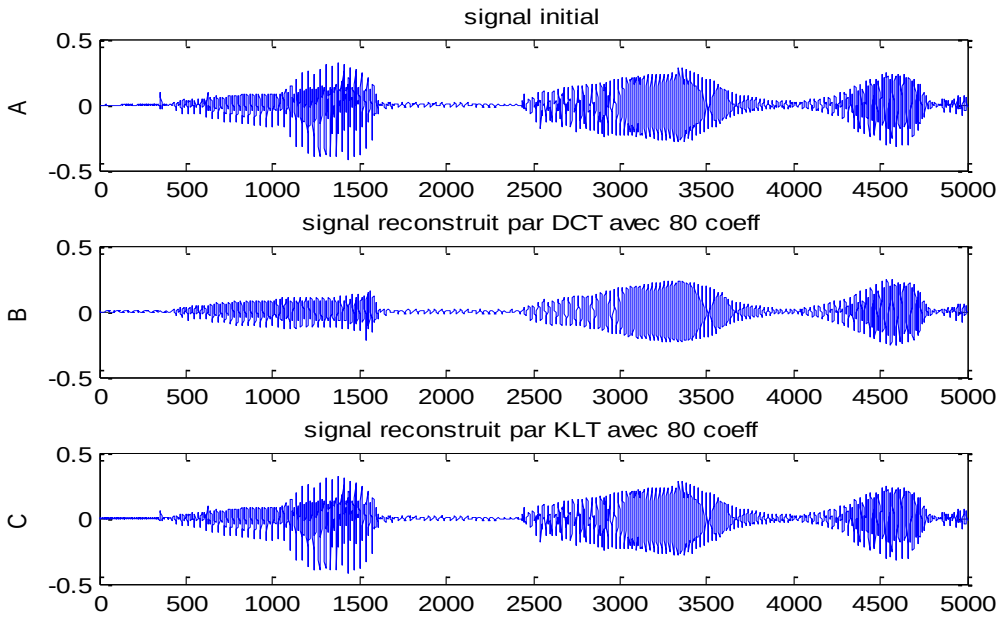


Fig.4.4: Reconstruction d'un signal de parole par DCT et KLT
avec 80 coefficients retenus
(A) Signal original, (B) reconstruit avec DCT, (C) reconstruit avec KLT

4.3.1.3. Compression avec N=100

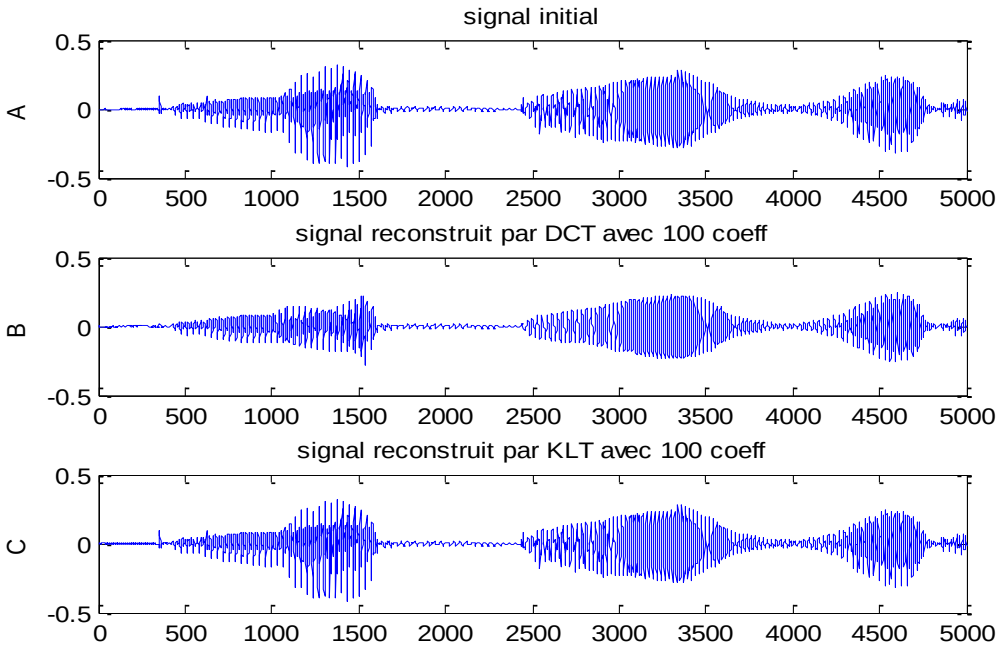
Pour cet essai, nous avons représenté chaque segment de parole par 100 coefficients (N=100). Nous sommes arrivés à une situation optimale, comme le montre la figure 4.5, tels que les deux audiogrammes (signal original et signal reconstruit par KLT) ont la même forme et sont indiscernables. Et nous remarquons que la transformée DCT donne un signal reconstruit très proche à l'original.

Pour la figure 4.5 on a :

(A) : Signal original,

(B) : signal reconstruit avec DCT

(C) : signal reconstruit avec KLT



Figu.4.5: Reconstruction d'un signal de parole avec 100 coefficients

4.3.2. Locuteur homme

L'audiogramme du signal analysé est donné par la figure 4.6

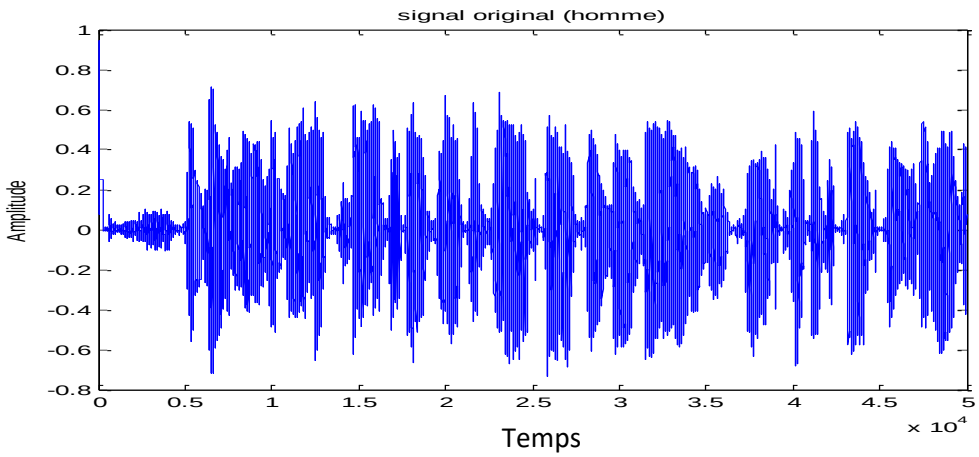


Fig.4.6: L'audiogramme du signal analysé pour la voix d'un homme

Ce signal est une phrase en Anglais prononcé par un homme, il a été choisi d'un corpus de données test.

Nous allons refaire les trois essais précédents pour ce signal. Les résultats de simulation sont donnés comme suit :

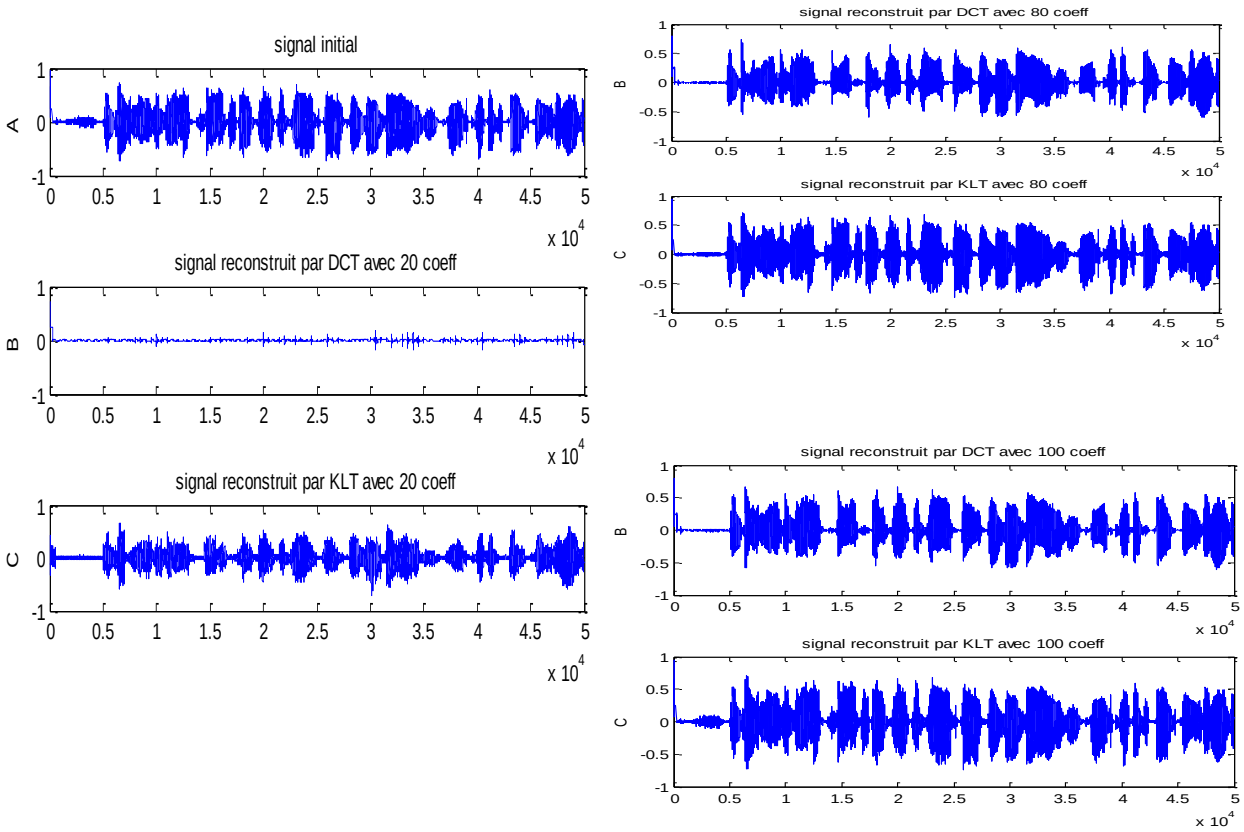


Fig.4.7: Reconstruction d'un signal de parole homme avec 20, 80 et 100 coefficients

En comparant, les graphes de cette expérience à ceux obtenus pour la voix d'une femme, nous constatons que les résultats sont identiques, pour les trois cas de simulation(les coefficients retenus : N=20, N=80 et N=100). Cela veut dire que la KLT donne toujours les meilleurs résultats par rapport à la DCT.

4.3.3. Simulation de la compression du point de vue énergie du signal

La figure 4.8 illustre la variation de l'énergie en fonction du nombre de coefficients retenus pour les voix masculine et féminine. Les courbes montrent une croissance de celle-ci avec l'augmentation du nombre de coefficients. Ceci est évident, car tous les coefficients contribuent à donner l'énergie du signal. Si le signal est tronqué, l'énergie est influencée.

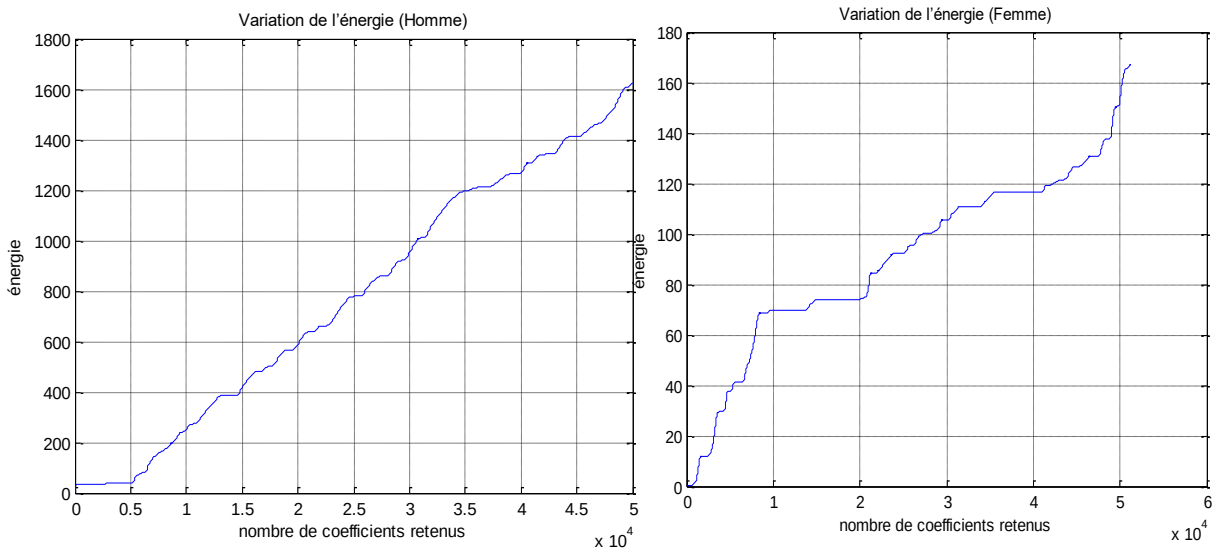


Fig.4.8: La variation de l'énergie en fonction du nombre de coefficients retenus

4.4. Evaluation de la qualité du signal reconstruit

4.4.1 Critères Objectifs

Les tableaux suivants, expriment l'évaluation de la qualité du signal par des critères objectifs pour les voix d'une Femme et d'un Homme. Dans cette expérience, nous avons simulé, trois paramètres : SNR, MSE(EQM) et le taux de compression; pour les trois cas concernant les coefficients retenus : 20, 80 et 100.

	KLT		**DCT**		
Nombre de coefficients retenus **N**	*SNR (dB)*	*MSE*	*SNR (dB)*	*MSE*	*Taux de compression (%)*
20	**0.4452**	**0.0029**	**0.0002**	**0.0033**	**96.0938**
80	**30.3498**	**0.0000**	**4.8500**	**0.0011**	**84.3750**
100	**64.7950**	**0.0000**	**5.8561**	**0.0008**	**80.4688**

Tableau.4.2: l'évaluation de la qualité du signal pour la voix féminine

	KLT		**DCT**		
Nombre de coefficients retenus **N**	*SNR (dB)*	*MSE*	*SNR (dB)*	*MSE*	*Taux de compression(%)*
20	**0.5374**	**0.0287**	**0.0384**	**0.0322**	**96.0938**
80	**21.3341**	**0.0002**	**6.4166**	**0.0074**	**84.3750**
100	**34.1411**	**0.0000**	**7.4903**	**0.0058**	**80.4688**

Tableau.4.3: l'évaluation de la qualité du signal pour la voix masculine

Nous remarquons que, pour les deux tableaux (Tableau 4.2 et Tableau 4.3) qui représentent les résultats d'évaluation de la KLT comparée à la DCT pour deux voix Homme et Femme, une augmentation du SNR proportionnelle au nombre des coefficients, mais avec des degrés différents. Le SNR associé à la KLT augmente de façon assez grande par rapport à celui de la DCT.

Et nous constatons que c'est l'inverse pour le MSE, telle que la diminution de cette dernière est inversement proportionnelle au nombre des coefficients retenus. Comparée à la DCT, la KLT présente des MSE plus faibles. Par exemple, nous remarquons que la valeur du MSE commence à s'annuler pour N= 80 avec la technique KLT, par contre pour la technique DCT ne s'annule même pas avec N=100.

L'évolution de l'erreur quadratique moyenne en fonction du nombre de coefficients retenus est représentée dans la figure 4.9. Le MSE de KLT est représenté en pointillés et celui de la DCT en point continus. Nous observons la décroissance rapide du MSE avec le nombre des coefficients dans le cas de la KLT par rapport à la DCT. Cela traduit la forte redondance du signal de parole. On voit que le MSE s'annule pour 10 coefficients dans le cas de la KLT, alors qu'il ne s'annule que pour 45 coefficients dans le cas de la DCT.

On constate que la KLT donne le même MSE pour les deux types de voix (féminine et masculine). Par contre, le MSE correspondant à la DCT converge mieux pour la parole femme que pour la parole homme. L'EQM_F s'annule pour N=45 alors que L'EQM_H s'annule pour N=50.

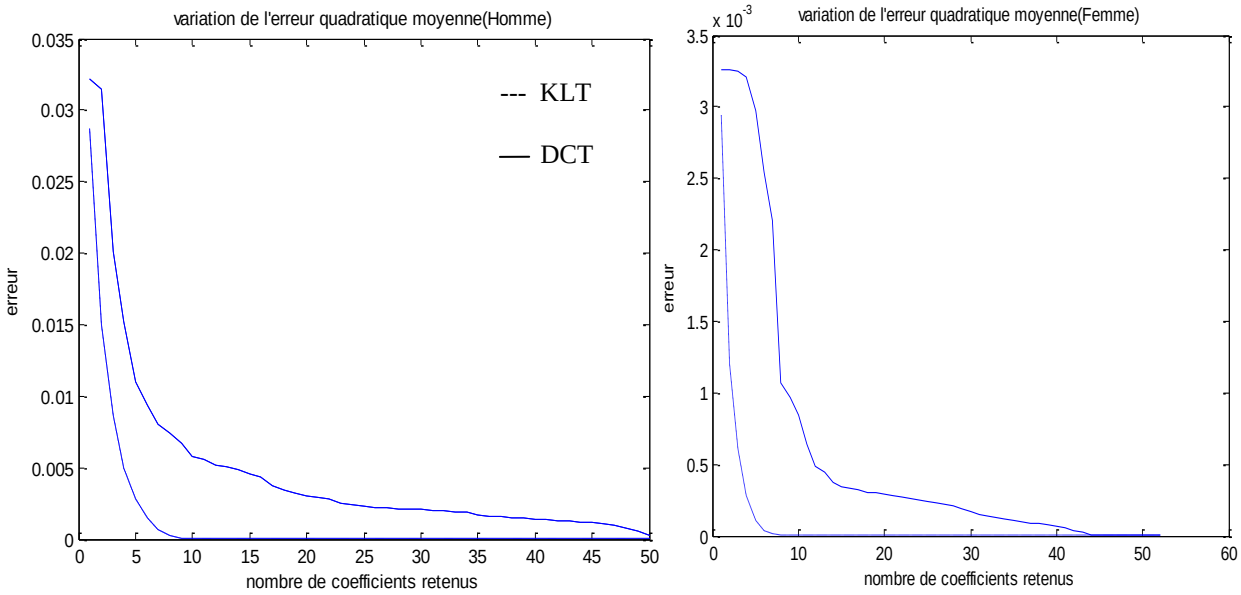


Fig.4.9: La variation de l'erreur quadratique moyenne

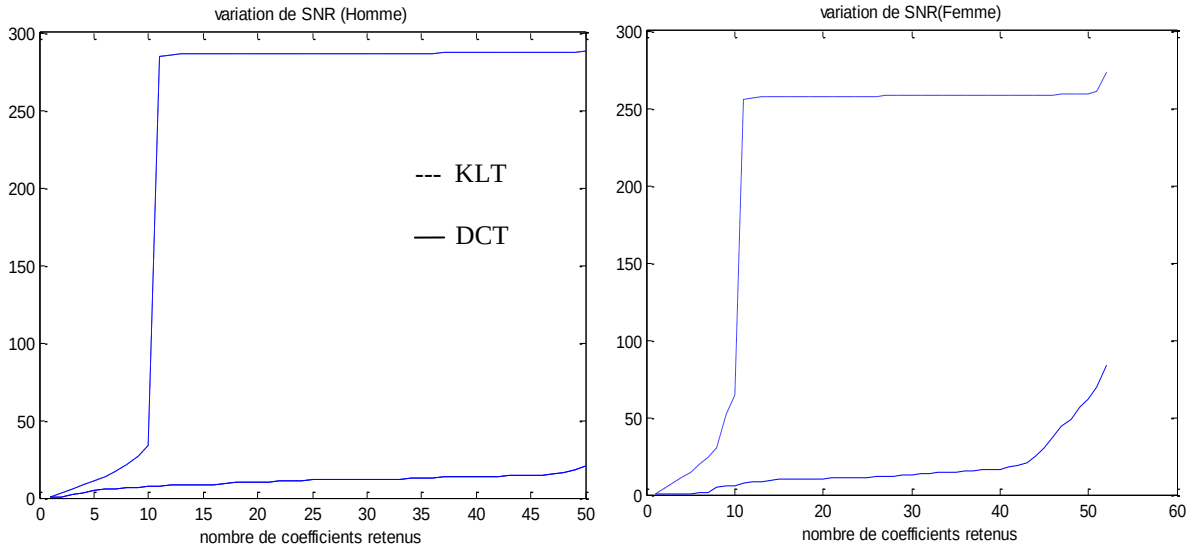


Fig.4.10: La variation du SNR pour une voix d'homme à la gauche et voix d'une femme à la droite

Le SNR de la KLT est représenté en pointillés et celui de la DCT en points continus sur la figure précédente (fig4.10). Le SNR de la KLT est meilleur par rapport à celui de la DCT et marque une stabilité après le 10^{eme} coefficient, avec une augmentation rapide dans le début. Le SNR du DCT croit très lentement. Les mêmes remarques sont valables pour la deuxième voix. En plus, le SNR est plus important dans le cas de la parole femme qu'homme.

4.4.2 L'évaluation MOS (Mean Opinion Score)

Les tests MOS basés sur l'écoute du signal reconstruit, ont été effectués par nous mêmes et en s'aidant de nos collègues.

Pour faciliter les calculs et la préservation du temps, nous allons donner les résultats de test des deux voix sur le même tableau, et calculer la moyenne des notes des différents auditeurs de chaque signal. Les résultats trouvés sont affichés dans le tableau suivant :

	N=20		**N=80**		**N=100**	
Notes	Nombre des auditeurs		Nombre des auditeurs		Nombre des auditeurs	
	KLT	DCT	KLT	DCT	KLT	DCT
1	6	15	0	2	0	0
2	9	0	3	7	0	0
3	0	0	6	6	0	8
4	0	0	5	0	3	4
5	0	0	1	0	12	3
Moyenne	1.6	1	3.2	1.8	4.8	3.6
Qualité	mauvaise	Très mauvaise	moyenne	mauvaise	excellente	bonne

Tableau.4.4: Comparaison entre KLT et DCT par l'évaluation MOS pour deux voix (F et M) avec N=20,80 ,100

Nous remarquons dans le tableau (Tableau 4.4) qui représente les résultats d'évaluation de KLT et DCT pour les deux voix Homme et Femme avec tests MOS. La compression du son par KLT est acceptable pour la majorité des auditeurs, par contre la compression par la DCT présente une dégradation perceptible du son, surtout pour les cas où le nombre des coefficients vaut 80 et 100.

4.5. Conclusion

Nous avons présenté dans ce chapitre une méthode pour l'analyse et la synthèse des signaux de parole en utilisant la technique de compression KLT. Nous avons comparé cette technique à la transformation DCT qui est la plus utilisée. D'après les résultats trouvés, nous concluons que la transformée KLT est une méthode très importante pour la compression avec pertes du signal de parole. Elle optimale par rapport à la DCT. Cependant la KLT, comme une

transformation linéaire optimale dans le sens de l'erreur quadratique moyenne de reconstruction, la KLT peut être difficilement utilisé dans la compression de la parole à cause de sa vitesse lente dans la recherche de la transformation de la matrice de covariance construite par des données d'apprentissage.

Toutes les expériences faites dans ce chapitre ont montré la robustesse de la KLT par rapport à la DCT qui n'est que la transformation sous-optimale de la KLT. L'erreur quadratique moyenne et le SNR étaient meilleurs pour cette dernière.

CONCLUSION GENERALE

Dans ce travail, nous avons abordé un domaine en cours d'expansion ces dernières décennies : compression et codage de la parole. Pour comprendre ce sujet, nous avons présenté les caractéristiques générales du signal de la parole et la notion de codage et la compression d'un signal monodimensionnel dans les chapitres 1 et 2 respectivement. Nous avons travaillé avec la Transformée KLT (Analyse en Composante Principale) qui représente un outil très robuste en s'appuyant sur des fondements mathématiques très solides et qui se caractérise par la notion de décomposition de la matrice d'autocorrélation en valeurs et vecteurs propres. Ceci permet de traiter les phénomènes temporels dont la parole fait partie.

Les résultats obtenus dans le chapitre 4 montrent que la transformée KLT, est une méthode plus efficace pour la compression avec perte du signal de parole. Elle a été montrée optimale sous plusieurs aspects et elle est prise comme base de comparaison, comme limite de performance pour d'autres transformations dites sous optimale. Une comparaison entre la KLT et la DCT a été faite, et a montré la robustesse de la KLT. Malgré cela, la technique de compression par KLT n'est utilisée que dans les laboratoires pour tester les performances des autres transformations. Le processus d'analyse des fonctions mathématiques de la KLT est compliqué et prend du temps.

Tous les efforts de recherches s'orientent vers l'amélioration de l'algorithme de la KLT en réduisant son temps de calcul. Les travaux de recherche sur la KLT doivent trouver des moyens pour réduire le temps d'analyse des équations mathématiques relatives à la KLT. Pour cette raison, nous attendons à ce que dans le futur, il y aura un algorithme correspondant à l'algorithme de la KLT en termes de qualité, mais plus rapide et être utilisable en pratique.

Bibliographie

[1] Mr : MERIANE Brahim, ‘’ Analyse du Signal de Parole par Les Ondelettes « Application Aux Mots Isolés » ‘’ thèse Magister Université de Batna 2009.

[2] BENYOUCEF M, ‘’ Reconnaissance Automatique de Parole pour la Commande Des Systèmes ‘’ thèse Magister université de Batna 1995.

[3] AZIZA Yassamine, ‘’ Modélisation AR et ARMA de la Parole pour une Vérification Robuste du Locuteur dans un Milieu Bruité en Mode Dépendant du Texte ‘’. Thèse Magister Université FERHAT ABBAS –Setif 1- UFAS (ALGERIE) 2013.

[4] Thomas Hueber. Reconstitution de la parole par imagerie ultrasonore et vidéo de l’appareil vocal : vers une communication parlée silencieuse. Thèse de doctorat de l’université Pierre Marie Curie 2009.

[5] R. Boite. Traitement de la parole. Collection Electricité. Presses Polytechniques et Universitaires Romandes, 2000.

[6] BOITE. R & KUNT. M, ‘’ Complément au traité d’électricité, Traitement de la Parole’’.

[7] BUNIET Laurent, ‘’ Traitement automatique de la parole en milieu bruité : Étude De modèles connexionnistes statiques et dynamiques ’’ THÈSE Doctorat de l’Université Henri Poincaré - Nancy 1 1997.

[8] BENAMMAR Ryadh, ‘’ Traitement Automatique De La Parole Arabe
Par Les HMMs: Calculatrice Vocale ‘’. Université Abou Bekr Belkaid Tlemcen 2012.

[9] Bernard Gosselin. Représentation de l’information et quantification des signaux.
Faculté Polytechniques de Mons, 2000. Belgique.

[10] [Aggarwal et al., 1999]. ‘’ Perceptual zerotrees for scalable wavelet coding of wideband audio ‘’. In IEEE Workshop on Speech Coding..

[11] S. Deligne. ‘’Modèles de séquences de longueurs variables: Application au traitement du langage écrit et de la parole ‘’. PhD thesis, École nationale supérieure des télécommunications (ENST), Paris, 1996.

[12] D. Salomon. Data compression. ‘’The Complete Reference, Springer Verlag.‘’ New-York, 1998.

[13] Jeffrey Scott Vitter. ‘’Design and analysis of dynamic Huffman codes.‘’ ACM Transactions on Mathematical Software, Volume 15 , Issue 2 (1989)

[14] Specifications for the Analog to Digital Conversion of Voice by 2,400 Bit /Second Mixed Excitation Linear Prediction. Federal Information Processing Standards Publication (FOPS PUB) Draft-Mai 1998.

[15] A. McCree, K. Truong, E.B. George, T.P. Barnwell, ‘’V. Viswanathan. A 2,4 Kbits/s MELP Coder Candidate for the New U.S ‘’. Federal Standard. Proc. ICASSP-96,1996.

[16] R.D. Dony “The Transform and Data Compression Handbook” Ed. K. R. Rao and P.C. Yip. Boca Raton, CRC Press LLC, 2001.

[17] D. Yang, H. Ai, C. Kyriakakis, and C.-C. Kuo, “An inter-channel redundancy removal approach for high-quality multichannel audio compression,” in AES 109th convention, AES preprint 5238, (Los Angeles, CA), September 2000.

[18] D. Yang, H. Ai, C. Kyriakakis, and C.-C. Kuo, “”An Explorition of Karhunen-Loeve Transform for Multichannel Audio Coding”,” in Conference of Electronic Cinema, SPIE’s International Symposium on Information Technologies, (Boston, MA), November 2000.

[2] Ed. K. R. Rao and P.C. Yip. ‘’The Transform and Data Compression Handbook ‘’Boca Raton, CRC Press LLC, 2001.

[20] Shannon, C.E. ‘’A mathematical theory of communication ‘’. The Bell System Technical J., 27(3):379–423, 623–656, 1948.

[21] Gray, R.M., Source Coding Theory, Kluwer Academic Publishers, Norwell, MA, 1990.

[22] Gersho, A. and Gray, R.M., ‘’Vector Quantization and Signal Compression ‘’. Kluwer Academic Publishers, Norwell, MA, 1992.

[23] Mme Mihoubi Fadila (Née Maouche). ‘’*La reconnaissance automatique de la parole Approche évolutionniste Cas de* l'arabe ‘’. Mémoire de magistère, Université Oum El Bouaghi, 2007.