République algérienne démocratique et populaire
وزارة التـعليـم العالـي والبحـث العلمـــي
Ministère de l'enseignement supérieur et de la recherche scientifique
بلحاج بوشعيب جامعة عين تموشنت
Université–Ain Temouchent- Belhadj Bouchaib
Faculté des Sciences et de la Technologie
Département : d'Electronique et des Télécommunications

Projet de Fin d'Études
Pour l'obtention du diplôme de Master en :
Domaine : SCIENCES & TECHNOLOGIES
Filière : TELECOMMUNICATION
Spécialité : RESEAU ET TELECOMMUNICATION
Thème

Etude et implémentation d'un schéma de codage de la parole par ondelettes

Présenté Par :
1) M. BENDADA Ahmed
2) M. BECHOUIREF Mohamed El Amine

Devant le jury composé de :

Dr BENGHENIA Hadj Abdelkader	M C B	UAT.B.B (Ain Temouchent)	Président
Dr BENTAIEB Samia	M C B	UAT.B.B (Ain Temouchent)	Examinatrice
Dr BOUKHOBZA Abdelkader	M C A	UAT.B.B (Ain Temouchent)	Encadrant

Année Universitaire 2023/2024


بسم الله الرحمن الرحيم

Je dédie ce travail,

À mes Parents, à mes frères Yasser & Abdelkader,

J’ai le grand honneur à dédie celui qui m'a beaucoup aidé et qui était et est toujours le seul soutien et celui qui a dit « Jusqu'à ce que j'atteigne cet objectif, je n'abandonnerai jamais » HANENE

À mes chères et meilleurs amies ROMAISSA ; SARRA et KOUIDER.

À mon binôme Amine

À mes amis de Sièges d’étude.

Ahmed

Je dédie ce travail

À mon roi Papa B. DJAMEL ma reine Mama SALIHA qui m'ont soutenu durant toutes ces années d'études et qui ont été toujours présent pour me pousser vers le haut et à être toujours plus forte.

À ma chère sœur KARIMA ; et à mon frère BOUHADJAR.

À mes chères et meilleurs amies ROMAISSA, SARRA, BOUCHRA, et HANENE qui sont toujours là pour moi.

À mes professeurs et nos Dr. BOUKHOBZA Abdelkader, pour leur encadrement attentif, leurs conseils éclairés Leur expertise et leur disponibilité ont été d'une grande valeur, et nous avons beaucoup appris de leurs précieux enseignements.

À mon binôme AHMED

À tous ceux qui m'ont assisté dans mes études.

Md AMINE

REMERCIEMENT

On remercie dieu le tout puissant de nous avoir donné la santé et la volonté d'entamer et de terminer ce mémoire. Nous remercions également nos très chers Parents qu'ils ont apporté la force et l'aide tout au long de mont parcours universitaire en fournissant mes besoins afin d'étudier dans les meilleures conditions.

Tout d'abord, ce travail ne serait pas aussi riche et n'aurait pas pu avoir le jour sans l'aide et l'encadrement de ***Mr. BOUKHOBZA****, on le remercie, pour toute l'aide apportée, pour la qualité de son encadrement exceptionnel, pour sa disponibilité, ses remarques fructueuses et ses directives précieuses, qui ont contribué efficacement à l'avancement de ce travail*

.

J'ai remercié vont aussi à ***Dr. FEROUANI*** *pour l'honneur de présider le jury et à* ***Dr. BENGHENIA*** *comme examinateur et Président.*
J'adresse aussi mes remerciements à tous les enseignants qui ma éduquée ***(Mme : BENHMAD.A ; BELOUATI.M ; BENYOUCEF.F ; Dr BADIR ; Mme SAHRAOUI ; M MALIOUI et Dr. MERADI)****.*

Enfin, nous adressons nos sincères sentiments de gratitude à toutes les personnes qui ont participé de près ou de loin à la réalisation de ce modeste travail.

B.Ahmed

B.Mohamed El Amine

Résume

ملخص

أدّى التَّقدم التِّكنولوجي فِي مَجال الإتِّصالات إلى ظُهور تَطبِيقات الإتصالات والوَسائِط المُتعدِدة. يَهدفُ تَمْثِيل إشارة الكَلام فِي تَشْفير الكَلام إلى تَقْليل عددَ ال «bits» المَطلوبَة مع الحِفاظ عَلى جَوْدة إدراكِيّة عالِية. فِي هَذا العَمل سَوْف نَقُوم بدِراسَة وتَنْفيذ نِظام ترْمِيز يَعتَمد على التَّحوِيل المَوْجي. في الوَاقع، تمَّ تَقْديم خَوارِزْمِيتيْن للتَرمِيز. الأولى هي خَوارِزْميَة التَّرمِيز الكْلاسيكيَّة والثَّانية هي تَرْميز الشَجرة الصَّفْرية (SPIHT) .قمنا بدراسة هذه التجارب على عده اشارات صوتيه ، بعد ذلك نقوم بالمقارنة بين الخوارزميتين للوصول الى افضل خوارزميه لتشفير الكلام.

Résume

Les avancées technologiques dans le domaine de la communication ont engendré des applications de communication et multimédia. La représentation du signal vocal dans le codage vocal vise à réduire le nombre de bits nécessaires tout en préservant une qualité de perception élevée. Dans ce travail, nous allons étudier et implémenté un système de codage qui est basé sur la transformation en ondelettes. En effet, deux algorithmes de codage sont présentés. Le premier est un algorithme de codage classique et le deuxième est le codage par arbre de zéro (SPIHT). Nous avons étudié ces expériences sur plusieurs signaux audio, après quoi nous avons comparé les deux algorithmes pour arriver au meilleur algorithme d'encodage de la parole.

Summary

Technological advances in the field of communication have generated communication and multimedia applications. The representation of the speech signal in speech coding aims to reduce the number of bits required while preserving a high perception quality. In this work, we will study and implement a coding system that is based on the wavelet transform. Indeed, two coding algorithms are presented. The first is a classical coding algorithm and the second is the zero-point tree coding (SPIHT). We studied these experiments on several audio signals, after which we compared the two algorithms to arrive at the best speech encoding algorithm.

Table de matière

Introduction Générale 1
CHAPITRE 1 4
1.1 Introduction 5
1.2 Production de la parole 5
1.2.1 Sons voisés 6
1.2.2 Sons non voisés 6
1.3 Principe de perception auditive 7
1.3.1 Concept de perception auditive 9
A. Réception 9
B. Transporte 9
C. Traitement 10
1.4 Méthodes codage de la parole 10
1.4.1 Codeurs en formes d'ondes 10
A. Codage temporel 10
B. Codage par transformée 11
1.4.2 Codage paramétrique 11
A. Extraction de caractéristiques 12
B. Quantification des paramètres 12
C. Codage 12
1.4.3 Codage hybride 12
A.Combinaison des approches 13
B. Segmentation du signal 13
C. Adaptation dynamique 13
1.5 Utilisation de la modélisation psychoacoustique 13
1.6 Conclusion 14
CHAPITRE 2 15
2.1 Introduction 16
2.2 Transformée en ondelettes continue 16
2.3 Transformée en ondelettes discrète 18
2.3 Codage de la parole par ondelettes 20
2.3.1 Décomposition DWT 20
2.3.2 Seuillage 20
2.3.3 Quantification 21

2.3.4 Codage entropique 21
2.4 Codage par arbre de zéros 21
2.4.1 Codage progressif 23
2.4.2 Codage SPIHT 23
a. Principe du SPIHT 24
b. Algorithme de codage 25
A. Initialisation 255
B. Passage de Triage 255
C. Passage de Raffinement 26
C. Mise à jour du Niveau de Quantification 26
D. Décodage 26
2.7 Conclusion 26
CHAPITRE 3 27
3.1 Introduction 28
3.2 Critères d'évaluation de la qualité du signal de la parole 28
3.2.1 Critères objectives 28
A. SNR (rapport signal sur bruit) 28
B. SSNR (Segmental Signal-to-Noise Ratio) 28
3.2.1 Critères subjectives 29
3.2.3 Taux de compression 30
3.3 Base de données et logiciel de calcul 30
3.4 Résultats expérimentaux 30
3.4.1 Algorithme de codage par ondelettes classique 30
3.4.2 Algorithme de codage SPIHT 33
3.4.3 Comparaison en l'algorithme de codage par ondelettes classique et SPIHT…. 36
3.1 Conclusion 37
Conclusion Générale 39
Bibliographie 40

Liste des figures

Figure 1. 1 : Une coupe du conduit vocal.....5
Figure 1.2: Le spectre d'un son voisé.....6
Figure 1.3:Le spectre d'un son non voisé.....7
Figure 1. 4:Schéma de production et rétroaction auditive. 8
Figure 1. 5: Perception et analyse du son par l'être humain.....8
Figure 1.6: Champ auditif humain.....9
Figure 1.7: Schéma de principe du codeur DPCM.....11
Figure 1. 8: Le modèle LPC de production de la parole. 12
Figure 1. 9: Courbe du seuil d'audition absolu de l'oreille humaine..14

Figure 2.1: Fonction d'échelle et ondelette de Haar (à gauche) et leur contenu fréquentiel (à droite). 17
Figure 2.2: Fonction d'échelle et ondelette Daubechies 4 (à gauche) et leur contenu fréquentiel (à droite). 17
Figure 2.3: Banc de filtres implémentant la DWT : (a) analyse, (b) synthèse.....18
Figure 2.4: Décomposition d'un signal de parole à 4 niveaux ondelettes. 19
Figure 2.5: Schéma de codage de la parole en ondelette 21
Figure 2. 6: Modèle de dépendance inter-bandes.....22
Figure 2.7: La relation de descendance dans l'algorithme SPIHT.....24

Figure 3. 1: Système PESQ pour l'évaluation des performances d'un codeur/décodeur de parole.....30
Figure 3.2: Signal «SA1 » synthétisé avec l'algorithme de codage par ondelettes classique pour les différents taux de compression.....32
Figure 3.3: Signal «SI1027 » synthétisé avec l'algorithme de codage par ondelettes classique pour les différents taux de compression.....33
Figure 3.4: Signal « SX37 » synthétisé avec l'algorithme de codage par ondelettes classique pour les différents taux de compression.....33

Figure 3.5: Signal «SA1 » synthétisé avec l'algorithme compression SPIHT pour les différents taux de compression. 35
Figure 3. 6: Signal «SI1027 » synthétisé avec l'algorithme compression SPIHT pour les différents taux de compression. 35
Figure 3.7: Signal « SX37 » synthétisé avec l'algorithme compression SPIHT pour les différents taux de compression. 36
Figure 3.8: Signal « SX37 » synthétisé avec les deux algorithmes de compression par ondelettes pour un taux de compression 14. 37

Liste des tableaux

Tableaux 3.1: Performances de compression en termes de PESQ, SNR et SSNR de l'algorithme de codage par ondelettes classique. 31
Tableaux 3. 2: Performances de compression en termes de PESQ, SNR et SSNR de l'algorithme de codage SPIHT. 34

Abréviations

- **ADM** Adaptive Delta Modulation
- **ADPCM** Adaptive Pulse Code Modulation
- **APC** Adaptive Predictive Coding
- **CELP** Code-Excited Linear Prediction
- **Codec** The encoder and the decoder
- **dB** Decibel
- **DCT** Discrete Cosine Transform
- **DM** Delta Modulation
- **DPCM** Differential Pulse Code Modulation).
- **DSP** Digital Signal Processing
- **DWT** Discrete Wavelet Transform
- **EZW** Embedded zerotree of wavelet transforms
- **ITU-T** Union Internationale des Télécommunications Secteur Télécommunications
- **LIP** (LIC) List of insignificant coefficients
- **LIS** (LIS) List of insignificant sets
- **LSP** (LSC) List of significant coefficients
- **LPAS** Linear Prediction Analysis-by-Synthesis
- **LPC** Linear Predictive Coding
- **MOS** Mean Opinion Score
- **PCM** Pulse Code Modulation
- **PESQ** Perceptual Evaluation of Speech Quality
- **SNR** Signal to Noise Ratio
- **SPIHT** Set partitioning in hierarchical trees
- **SPL** Sound Pressure Level
- **SSNR** Segmental Signal-to-Noise Ratio
- **TOC**(CWT) Continuous wavelet transform
- **TOD** (DWT) Discrete wavelet transform

Introduction Générale

Introduction Générale

L'évolution rapide des technologies de communication et le besoin croissant de transmettre des informations de manière efficace et sécurisée ont conduit au développement de méthodes de codage de plus en plus sophistiquées. Parmi ces méthodes, le codage de la parole. Ce dernier occupe une place prépondérante, en raison de son importance dans des domaines variés tels que les télécommunications, les systèmes de reconnaissance vocale, et les applications multimédia.

Un système de codage de la parole comprend deux parties : le codeur et le décodeur (codec). Le codeur analyse le signal pour en extraire un nombre réduit de paramètres pertinents qui sont représentés par un nombre restreint de bits pour archivage (stockage) ou transmission. Le décodeur utilise ces paramètres pour reconstruire un signal de parole synthétique. L'objectif dans le codage de la parole est de représenter le signal vocal avec un nombre restreint de bits tout en gardant une qualité perceptuelle acceptable.

Dès le début des années 70 apparaît la technique de Prédiction Linéaire LPC (Linear Predictive Coding), supposant que le signal de parole peut être considéré comme la sortie d'un filtre tout-pôle. Un codeur LPC particulièrement performant utilise des techniques d'analyse par synthèse LPAS (Linear Prediction Analysis-by-Synthesis). Dans un système de codage LPAS, le signal de parole à transmettre est codé en recherchant la meilleure excitation possible d'un filtre de synthèse, dont les coefficients sont déterminés par une analyse de Prédiction Linéaire. Le plus connu de ces codeurs LPAS est le codeur de type CELP (Code-Excited Linear Prediction) conçu pour le codage de la parole à des débits allant de 4 à 16 Kbits/s. Pour des débits inférieurs à 4 Kbits/s, les codeurs paramétriques sont plus efficaces. Pour atteindre ces taux de compression, les systèmes de codage paramétriques se basent sur la connaissance du processus de production de la parole. La technique consiste à extraire du signal de parole les paramètres les plus pertinents permettant au décodeur de le synthétiser. Les performances des codeurs paramétriques, appelés aussi vocodeurs, dépendent de la précision des modèles de production de parole. La plupart des codeurs paramétriques sont basés sur le codage prédictif linéaire (LPC), connus sous le nom de vocodeurs prédictifs. Dans ce type de vocodeurs, le conduit vocal est modélisé par un filtre tout-pole. Grâce au développement de processeurs DSP (Digital Signal Processing) performants, des modèles mathématiques très proches du système phonatoire humain ont pu être mis au point et utilisés dans des algorithmes de codage de plus en plus complexes. Bien que les progrès technologiques soient considérables, la complexité algorithmique a encore un impact sur le coût d'un encodeur/décodeur de parole.

Parmi les techniques de compression, les compressions basées sur la transformée en ondelettes produisent d'excellentes résultats grâce à leur propriété de compacter l'énergie sur différents niveaux hiérarchiques. Ces techniques de compression permettent non seulement d'avoir une efficacité de codage élevé, mais elles ont également la capacité d'être extensible en fonction de la qualité ou de la taille requise.

Ce travail de fin d'étude se concentre sur l'étude et l'implémentation d'un système de codage de la parole basé sur la transformée en ondelettes et le codage par arbre de zéro. Le choix de

ces techniques se justifie par leur capacité à compresser efficacement l'information tout en préservant les caractéristiques essentielles de la parole.

Le document est structuré en trois chapitres principaux :

- **Chapitre 1 :** Ce chapitre introduit les concepts fondamentaux de la perception de la parole et les différentes méthodes de codage existantes. Une attention particulière est portée sur le modèle perceptuel, qui permet d'augmenter l'efficacité de compression en classant les informations comme pertinente et non-pertinente.
- **Chapitre 2 :** Dans ce chapitre, nous explorons en détail la transformée en ondelettes et ses applications dans le codage de la parole. Nous discutons également le codage par arbre de zéro, en mettant en lumière ses avantages par rapport à d'autres méthodes de codage.
- **Chapitre 3 :** Ce dernier chapitre est dédié à l'analyse des performances des différents schémas de codage par ondelettes. Nous effectuerons des expérimentations sur Matlab pour évaluer l'efficacité du système de compression en termes de taux de compression et de qualité perceptuelle.

CHAPITRE01

1.1 Introduction

En raison des caractéristiques du conduit vocal humain, le signal de parole est fortement redondant. Ces redondances permettent aux algorithmes de codage de compresser le signal en enlevant l'information non pertinente contenue dans le signal. Donc la connaissance du système vocal et des propriétés du signal de parole est essentielle pour concevoir des codeurs efficaces. Les propriétés du système auditif humain peuvent également être exploitées, pour améliorer la qualité perceptuelle du signal codé. Avant d'aborder le problème de codage de parole plus précisément, quelques caractéristiques du signal vocal sont présentées, et permettront de mieux apprécier les différentes techniques de codage présentées par la suite. Une partie simple de la théorie acoustique est exposée et les notions de phonème, de formant, de son voisé, non voisé et de pitch sont définis.

1.2 Production de la parole

D'un point de vue physique, la parole se manifeste comme une variation de la pression de l'air produite et émise par les articulations. Lorsqu'un individu parle, sous la supervision du système nerveux central qui reçoit constamment des informations, les poumons sont le générateur du système de production de la parole, ils fournissent l'énergie nécessaire à la production du son, en poussant de l'air à travers la trachée-artère, au sommet de laquelle se trouve le larynx où la pression de l'air est régulée avant d'être appliquée au conduit vocal, constitué pour la plupart des sons des cavités nasales et buccales. L'entrée est fournie par la glotte avec une fréquence du pitch (fondamentale) (F0) spécifique. Le conduit vocal travail comme un instrument de musique produisant un son. En effet, les différentes formes du conduit vocal produisent des sons différents. La cavité buccale joue le rôle majeur pour former les différentes formes du conduit vocal. Pour produire des sons nasaux, la cavité nasale est souvent incluse dans le conduit vocal, elle est connectée en parallèle avec la cavité buccale. Une coupe du conduit vocal est montrée dans la figure 1.1.

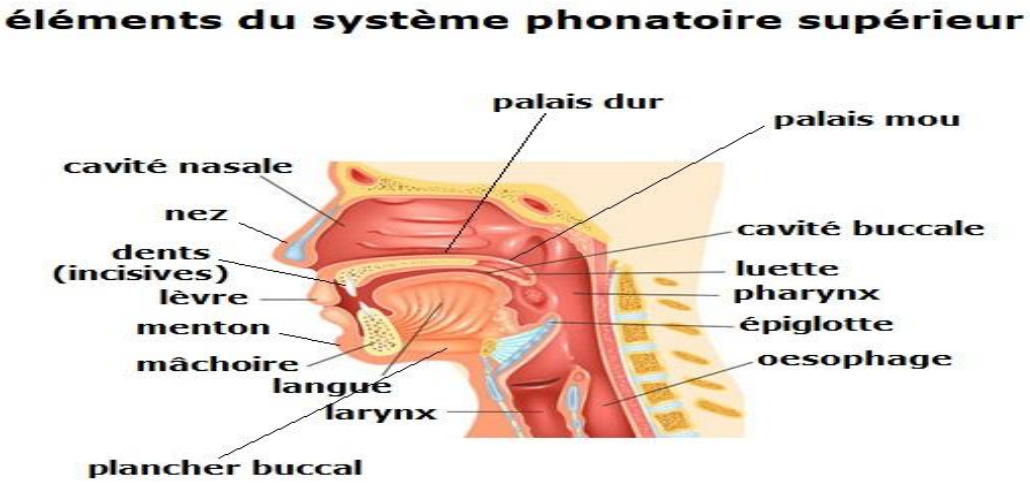

Figure 1. 1 : Une coupe du conduit vocal.

La voix peut être classée en deux catégories principales : les sons voisés et les sons non voisés [1].

1.2.1 Sons voisés

Ces sons sont générés par la vibration des cordes vocales lors de la parole. Les voyelles, les semi-voyelles et les consonnes liquides (/m/, /n/, /l/, /r/, etc.) sont des sons vocaux. Dans ces vibrations, l'air traverse les cordes vocales, qui se déplacent pour générer le son.

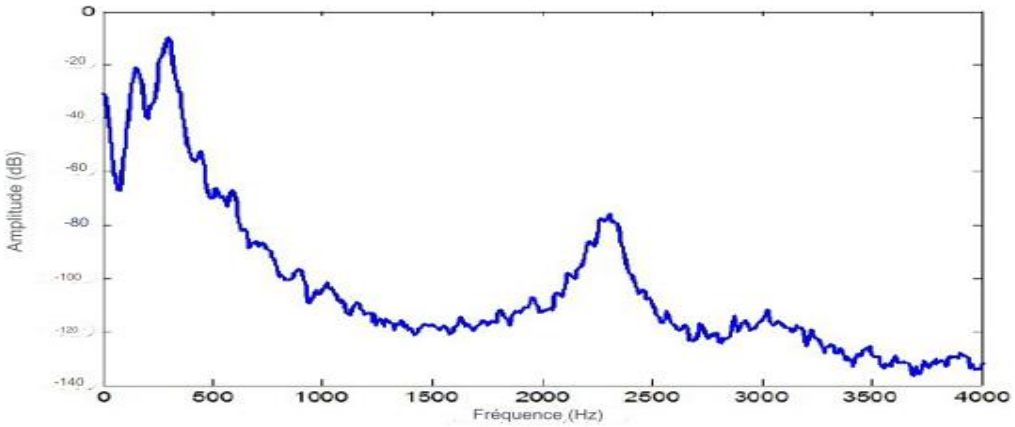

Figure 1.2: Le spectre d'un son voisé.

1.2.2 Sons non voisés

Ces sons sont produits lorsque l'air traverse le conduit vocal sans faire vibrer les cordes vocales. Les non-voyelles incluent certaines consonnes telles que /p/, /t/, /k/, /f/, /s/, /ʃ/, etc. Dans ces sons, les cordes vocales ne vibrent pas, mais l'air est plutôt expulsé par une contraction du conduit vocal, produisant un son sans ton.

Lorsque le conduit vocal est stimulé par des impulsions périodiques de pression, il en résulte vibrations dans les cordes vocales, la pression s'accumule puis se relâche de manière inattendue l'ouverture de la glotte crée des sons appelés voix, qui sont des sons compris entre autres voyelles.

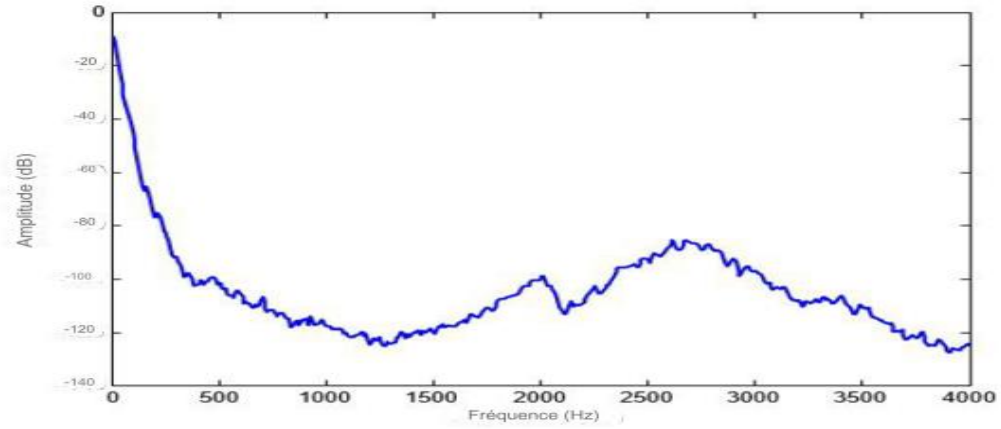

Figure 1.3: Le spectre d'un son non voisé.

Les pics épars sont présents, le premier correspondant à la fréquence fondamentale, tandis que les autres sont des fréquences appelées formants. Il est essentiel d'utiliser les trois premiers formants pour décrire un spectre vocal, tandis que les formants d'ordre supérieur ont différentes applications, comme la reconnaissance vocale.

Le conduit vocal se rétrécit, ce qui provoque des sons similaires à des consonnes. De plus, les cordes vocales ne vibrent pas, elles restent écartées. C'est pourquoi ces sons sont apériodiques. Ils sont généralement assimilés à un bruit blanc à la sortie d'un filtre composé de la partie du conduit vocal située entre la constriction et les lèvres.

Le spectre d'un son non voisé n'a pas de pitch, contrairement au voisé, mais il conserve quelques caractéristiques de ce dernier d'un point de vue informatique. La figure 1.3 présente le spectre d'un tel son non voisé.

1.3 Principe de la perception auditive

La parole peur être décrite comme le résultat de l'action volontaire et coordonnée d'un certain nombre de muscles. Cette action se déroule sous le contrôle du système nerveux central qui reçoit en permanence des informations par rétroaction auditive et par les sensations kinesthésiques, ce principe est présenté sur la figure 1.4 [2].

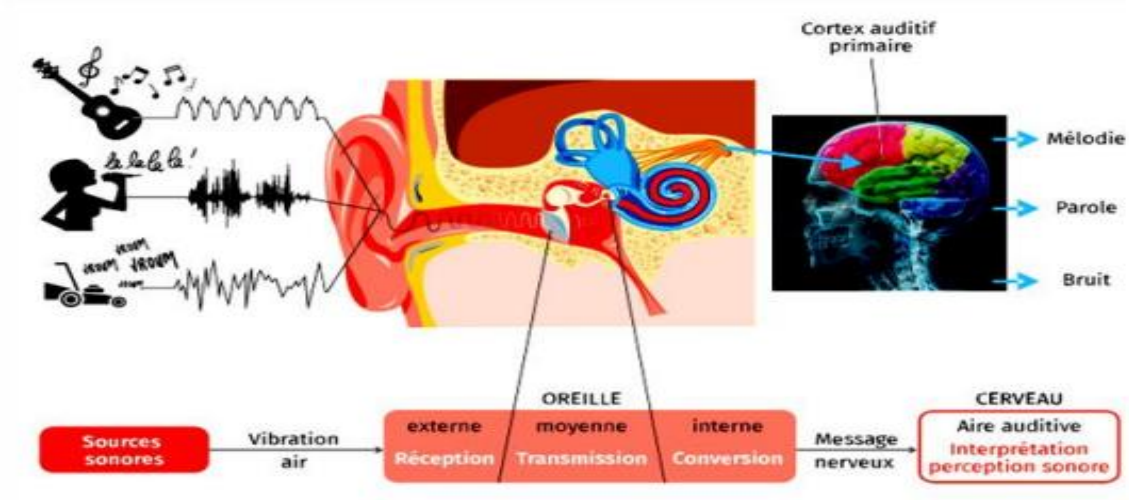

Figure 1. 4: Schéma de production et rétroaction auditive.

Les ondes sonores sont recueillies par l'appareil auditif, ce qui provoque les sensations auditives. Ces ondes de pression sont analysées dans l'oreille interne qui envoie au cerveau l'influx nerveux qui en résulte. Le phénomène physique induit alors un phénomène psychique grâce à un mécanisme physiologique complexe. Le mécanisme de l'oreille interne (marteau, étrier, enclume) permet une adaptation d'impédance entre l'air et le milieu liquide de l'oreille interne. Les vibrations de l'étrier sont transmises au liquide de la cochlée.

Celle-ci contient la membrane basilaire qui transforme les vibrations mécaniques en impulsions nerveuses. La membrane s'élargit et s'épaissit au fur et à mesure que l'on se rapproche de l'apex de la cochlée. Les fibres nerveuses aboutissent à une région de l'écorce cérébrale, appelée aire de projection auditive, et située dans le lobe temporal. En cas de lésion de cette aire, on peut observer des troubles auditifs.

Les fibres nerveuses auditives afférentes « de l'oreille au cerveau » et efférentes « du cerveau vers l'oreille » sont partiellement croisées : chaque moitié du cerveau est mise en relation avec les deux oreilles internes. Entre l'arrivée des signaux vibratoires aux oreilles et la sensation du son dans le cerveau, a lieu le phénomène de traitement des signaux par le système nerveux. Cela signifie que la vibration physique de l'air ne parvient pas de façon brute au cerveau. Elle est transformée, comme décrit sur la figure 1.5 :

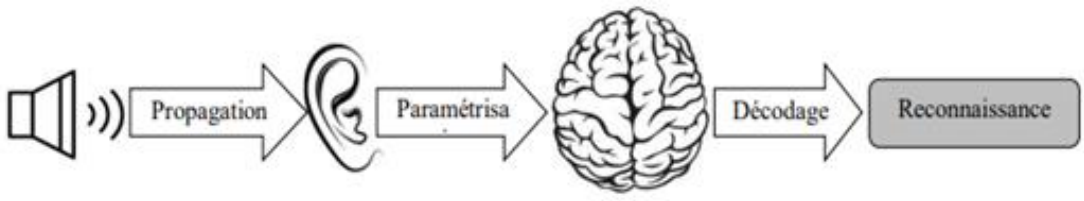

Figure 1. 5: Perception et analyse du son par l'être humain.

Il reste très difficile de nos jours de dire comment l'information auditive est traitée par le cerveau. On a pu par contre étudier comment elle était finalement perçue, dans le cadre d'une science spécifique appelée psycho acoustique. Sans vouloir entrer dans trop de détails sur la contribution majeure de la psycho acousticiens dans l'étude de la parole, il est intéressant d'en connaître les résultats les plus marquants. Ainsi, l'oreille ne répond pas également à toutes les fréquences. La figure 1.6 présente le champ auditif humain, délimité par la courbe de seuil l'audition et celle du seuil de la douleur. La fréquence d'échantillonnage maximale utile pour un signal auditif (~ 32000 Hz) est fixée par sa limite supérieure en fréquence (~16000 Hz, variable selon les individus).

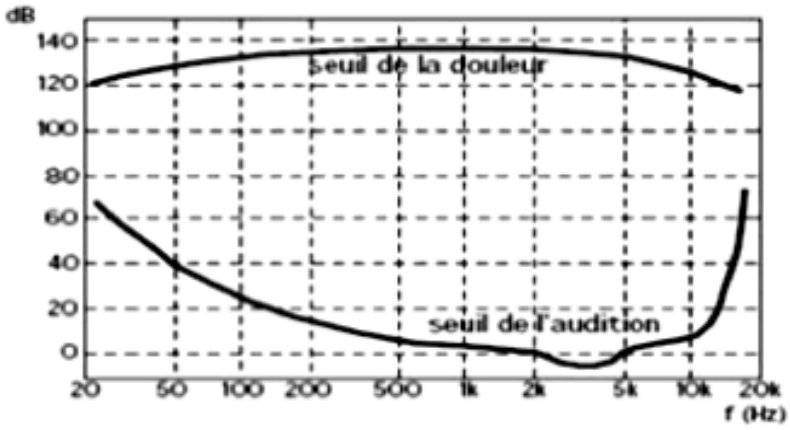

Figure 1.6: Champ auditif humain.

A l'intérieur de son domaine d'audition, l'oreille ne présente pas une sensibilité identique à toutes les fréquences. La figure 1.6 fait apparaître les courbes d'égale impression de puissance auditive - physiologie auditive (aussi appelée sonie, exprimée en sones) en fonction de la fréquence. Elles révèlent un maximum de sensibilité dans la plage [500 Hz, 10 kHz], en dehors de laquelle les sons doivent être plus intenses pour être perçus.

1.3.1 Concept de perception auditive

La perception auditive est la capacité d'identifier et d'interpréter des informations et des significations et de les relier à des sons qui sont reçus par l'ouïe et l'organe sensoriel. Oreilles, sous forme d'ondes de fréquence pouvant être entendues, transmises dans l'air ou par d'autres moyens.

La perception auditive est l'un des facteurs les plus importants qui aident les humains à s'adapter et à s'adapter à l'environnement, à travers :

- Comprendre le discours des autres et apporter des réponses appropriées.
- Réaliser les processus d'apprentissage, de développement et d'éducation.
- Déterminer le temps et le lieu en donnant un sens aux choses en termes de proximité, de distance ou de destination, et du moment de leur apparition.
- Se protéger en identifiant et en évitant les sources de menaces sonores, comme le bruit des animaux ou autres.

Le processus de perception auditive passe par trois étapes fondamentales :

A. Réception

A cette étape, l'oreille reçoit des informations sous forme d'ondes émises par l'homme ou par divers autres moyens, ce qui constitue la première étape de l'étape de perception auditive. Il convient de noter que l'air joue un rôle majeur dans cette étape. Il aide à transmettre et à recevoir des ondes à travers l'oreille jusqu'à l'oreille interne.

B. Transporte

Les neurones transmettent à ce stade un signal spécifique qui constitue l'information entendue, et le délivrent au corps géniculé médial, dans le thalamus.

B. Traitement

À ce stade, l'information est envoyée au cortex auditif du lobe temporal, pour être traitée et interagir avec elle, en émettant les réponses correctes en conséquence.

1.4 Méthodes codage de la parole

Le codage de la parole fait référence au processus de conversion du signal audio de la parole humaine en une forme numérique pour le stockage, la transmission ou le traitement ultérieur. Ce processus est essentiel dans de nombreux domaines, y compris les télécommunications, la reconnaissance vocale, la compression audio et la synthèse vocale. Le codage de la parole implique généralement la sélection d'un ensemble de caractéristiques pertinentes dans le signal audio et leur représentation sous forme numérique à l'aide d'algorithmes de traitement du signal.

L'objectif est de réduire la quantité de données nécessaire pour représenter de manière efficace et fidèle la parole, tout en minimisant la perte de qualité perceptible.

Traditionnellement les codeurs de parole sont divisés en trois grandes classes, en fonction de la manière dont ils représentent et traitent les caractéristiques acoustiques du signal vocal [3]. Voici une brève description de ces trois :

1.4.1 Codeurs en formes d'ondes

Ce type de codeurs essaye de reproduire la forme d'onde du signal d'entrée à coder. Ils sont conçus pour être indépendants du signal, ainsi, ils peuvent être employés pour coder une large variété de signaux. Généralement, ils sont de faible complexité, ils fournissent des signaux de parole de bonne qualité à des débits au-dessus de 16 kbps. Le codage en formes d'ondes peut être effectué aussi bien dans le domaine temporel que dans le domaine fréquentiel.

A. Codage temporel

Dans ce type de codage, l'accent est mis sur la représentation temporelle des signaux vocaux. Il peut s'agir de techniques telles que l'échantillonnage du signal à des intervalles réguliers, la quantification de l'amplitude du signal à chaque échantillon, et la transmission de ces valeurs d'échantillonnage. Les codecs basés sur le codage temporel peuvent être simples et

efficaces en termes de consommation de bande passante, mais peuvent ne pas capturer toutes les nuances et les détails du signal vocal.

Les méthodes de codage les plus connues dans le domaine temporel sont : le codage PCM (Pulse Code Modulation), le codage APCM (Adaptive Pulse Code Modulation), le codage DPCM (Differential Pulse Code Modulation), le codage ADPCM (Adaptive Differential Pulse Code Modulation), le codage DM (Delta Modulation), le codage ADM (Adaptive Delta Modulation) et le codage APC (Adaptive Predictive Coding).

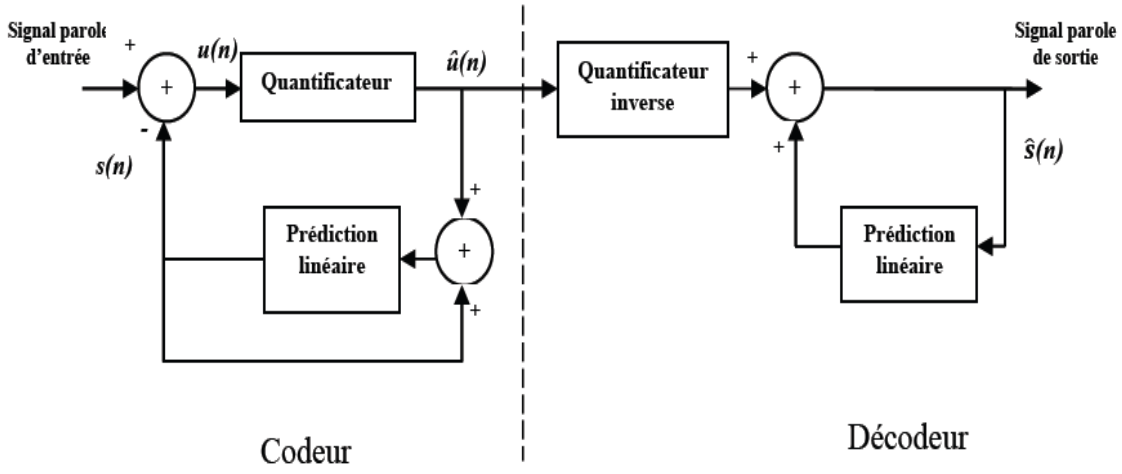

Figure 1.7: Schéma de principe du codeur DPCM

B. Codage par transformée

Ce type de codage transforme le signal de la parole en utilisant des techniques mathématiques pour le rendre plus facile à coder. Les coefficients obtenus par la transformation sont quantifiés et codés. Le nombre de bits utilisé pour coder chaque composante transformée peut varier de manière dynamique.
Les codeurs de la parole utilisant le codage par transformée 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 codeurs dans le domaine fréquentiel sont divisés en deux groupes : Les codeurs en sous bande (subband coders) et les codeurs par transformée (transform coders).

Les principales transformations utilisées pour la compression sont la transformée en ondelettes, la transformation DCT et la DWT. La transformée en ondelettes discrètes (DWT), qui fait l'objet de ce travail, fournit une décomposition multi-échelle du signal vocal, permettant ainsi une meilleure gestion des détails à différentes résolutions.

1.4.2 Codage paramétrique

Les techniques de codage paramétrique se concentrent sur la modélisation des caractéristiques acoustiques du signal vocal à l'aide de paramètres. Au lieu de coder chaque échantillon de manière brute, ces systèmes extraient des caractéristiques importantes du signal

vocal, telles que les formants, les coefficients cepstraux, etc., et les transmettent à la place. Ces modèles paramétriques peuvent être plus efficaces en termes de compression de données tout en préservant la qualité perçue du signal. Voici la description :

A. Extraction de caractéristiques

Le signal vocal est analysé pour extraire des paramètres représentant ses caractéristiques importantes. Cela peut inclure les formants, les coefficients cepstraux, les coefficients de prédiction linéaire (LPC), etc.

B. Quantification des paramètres

Les paramètres extraits sont quantifiés et représentés à l'aide d'une quantification adaptée à chaque type de paramètre.

C. Codage

Les paramètres quantifiés sont ensuite transmis ou stockés. Certains codecs utilisent des techniques de prédiction pour réduire davantage la quantité d'informations à transmettre.

La figure 1.8 montre le modèle LPC de production de la parole.

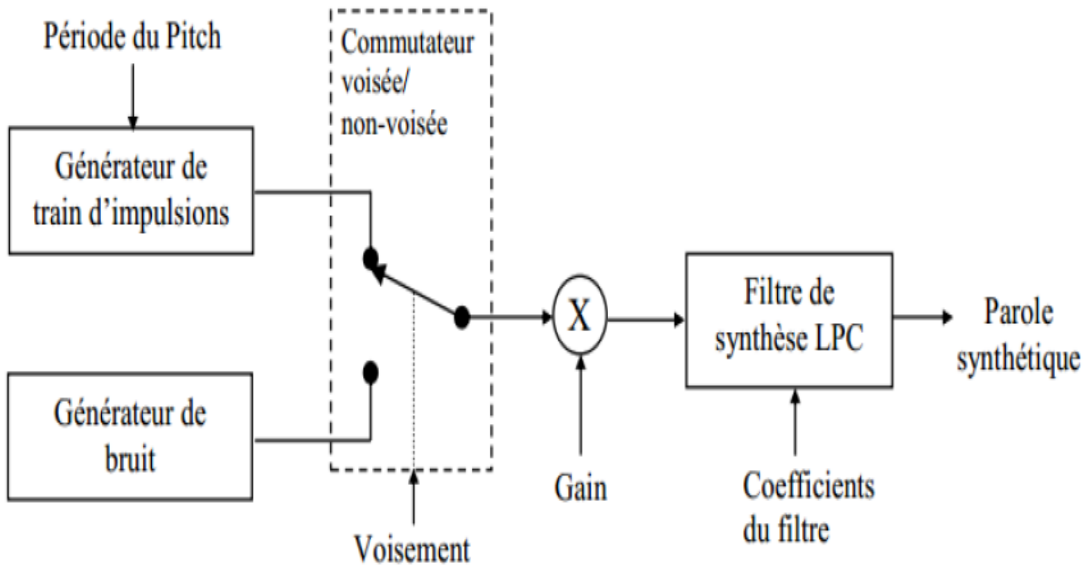

Figure 1. 8: Le modèle LPC de production de la parole.

1.4.3 Codage hybride

Les systèmes de codage hybride combinent généralement des éléments des deux approches précédentes. Ils peuvent utiliser à la fois des techniques de codage temporel et paramétrique pour tirer parti des avantages de chaque méthode. Par exemple, certains codecs hybrides peuvent utiliser le codage temporel pour les parties du signal où la précision est essentielle, tandis que le codage paramétrique est utilisé pour les parties où une plus grande compression est souhaitée. Voici la description :

A. Combinaison des approches

Les systèmes de codage hybride utilisent à la fois des techniques de codage temporel et paramétrique.

B. Segmentation du signal

Le signal vocal peut être segmenté en parties, où chaque partie est traitée différemment en fonction de son contenu. Par exemple, les parties du signal contenant des transitions rapides peuvent être traitées avec un codage temporel, tandis que les parties plus stables peuvent être représentées par des paramètres.

C. Adaptation dynamique

Les codecs hybrides peuvent ajuster dynamiquement les techniques de codage utilisées en fonction des caractéristiques du signal vocal en temps réel, ce qui permet une efficacité accrue de la compression tout en préservant la qualité perçue.

1.5 Utilisation de la modélisation psychoacoustique

Les modèles perceptuels par transformée utilisent la modélisation psychoacoustique afin de ne pas transmettre de l'information superflue que l'oreille humaine ne peut entendre. L'encodeur effectue une analyse du signal dans le domaine de la transformée afin de créer un modèle psychoacoustique qui décrit le comportement perceptuel de l'oreille humaine en fonction : de la fréquence, de l'intensité et du temps. Cette modélisation psychoacoustique tient compte des limites et des faiblesses de l'oreille humaine afin de transmettre uniquement les composantes essentielles du signal [4-7].

La figure 1.8 montre la courbe du seuil d'audition absolu. Cette courbe représente le niveau de pression acoustique (Sound Pressure Level, SPL) moyennen dB (decibel) pour qu'un son sinusoïdal pur soit perçu par l'oreille humaine. Puisque chaque individu possède une courbe unique qui varie selon l'âge, la figure 1.9 représente le seuil d'audition absolu moyen.

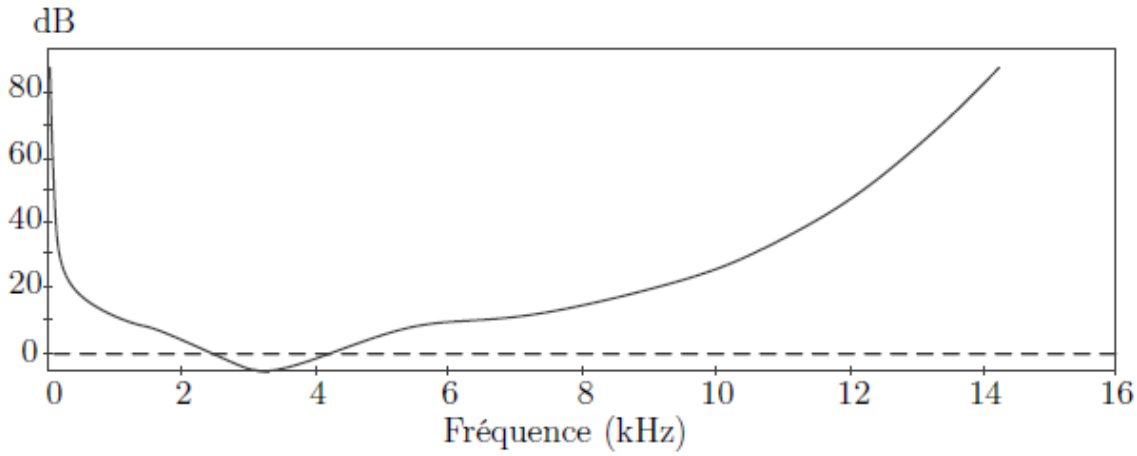

Figure 1. 9: Courbe du seuil d'audition absolu de l'oreille humaine.

La figure 1.9 montre que l'oreille humaine possède une plage de sensibilité à des fréquences allant de 20 Hz à 16 kHz et que le niveau de sensibilité diffère selon la fréquence. Le niveau de sensibilité maximale se situe autour de 1kHz à 5 kHz.

La courbe de la figure 1.9 correspond à une écoute dans un environnement calme. En présence de sons multiples, la courbe se modifie et le phénomène de masquage survient. Le phénomène se produit lorsqu'un son empêche la perception d'un autre son qui autrement serait audible. Les modèles perceptuels par transformée exploitent ces phénomènes de masquage afin de réduire l'information à transmettre.

1.6 Conclusion

La compression de la parole est nécessaire et se réalise grâce à différentes méthodes qui permettent l'extraction de la redondance du signal dans le domaine temporel, et la modélisation du signal dans le domaine fréquentiel.

Dans ce chapitre nous avons présenté les concepts fondamentaux de la perception de la parole et les différentes méthodes de codage existantes. Nous avons également défini le model perceptuel qui permet de distinguer les parties pertinente et non pertinente au sens du système auditif humain. Une attention particulière est portée dans le chapitre suivant au codage par ondelettes, une technique prometteuse pour la compression des données.

CHAPITRE 02

2.1 Introduction

Dans un schéma de compression de données, la transformation est une étape clé qui a pour objectif de réduire les corrélations entre les échantillons. Elle concentre les variations (énergies) du signal sur quelques échantillons et répartie presque uniformément les échantillons corrélés. La réduction de la redondance est en général obtenue en transformant la donnée originale d'une forme sous une autre représentation.

La transformée en ondelettes est parmi les méthodes courantes de transformation. Elle consiste à grouper l'énergie du signal en plusieurs sous bandes. C'est la transformée requise par les algorithmes de compression d'ondelettes SPIHT et EZW, objet de ce chapitre.

2.2 Transformée en ondelettes continue

La transformée en ondelettes est similaire à la transformée de Fourier (et encore plus à la transformée de Fourier locale) avec une fonction de mérite complètement différente. La différence principale est la suivante : la transformée de Fourier décompose le signal en sinus et en cosinus, c'est-à-dire en fonctions localisées dans l'espace de Fourier ; contrairement à la transformée en ondelettes qui utilise des fonctions localisées à la fois dans l'espace réel et dans l'espace de Fourier [8]. De manière générale, la transformée en ondelettes peut être exprimée avec l'équation suivante :

$$W(s,\tau) = \int_{-\infty}^{\infty} f(x)\psi_{s,\tau}(t)\, dt \qquad (2.1)$$

Où le symbole * désigne le conjugué complexe et ψ est une fonction donnée. L'ondelette est définie par :

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{|s|}}\psi\left(\frac{t-\tau}{s}\right) \qquad (2.2)$$

Cette équation représente un ensemble de fonctions (appelé atomes d'ondelette), pouvant être obtenues par dilatation et déplacement d'une seule fonction $\psi(t)$, appelée *ondelette mère*. La variable τ représente le *déplacement temporel*, et la variable s représente la *dilatation temporelle* : Si s>1, on a un élargissement de s, et si 0<s<1, on a un rétrécissement de s. Si s est négatif, on a une dilatation combinée avec une inversion de temps.

On obtient de cette manière une analyse dont la résolution fréquentielle et temporelle est variable. C'est grâce à cette propriété que la théorie des ondelettes a connu un tel succès : faire varier τ permet d'analyser localement le signal, et ce, à différentes échelles grâce à s. Cette échelle s permet de faire évoluer la résolution temps-fréquence de l'analyse en faisant varier la taille du support de $\psi_{s\tau}$ d'un rapport de $\frac{1}{s}$.

On peut montrer que si la fonction analysante (l'ondelette) est convenablement choisie, la transformation en ondelettes est inversible et la fonction peut être reconstruite après analyse suivant l'équation :

$$f = C_{\psi}^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{+\infty} W(s,\tau)\psi_{s,\tau} da\, db \qquad (2.3)$$

Avec $< f, \psi_{s,\tau} >$ est le produit scalaire entre la fonction et $\psi_{s,\tau}$.

Le coefficient C_ψ est donné par l'équation :

$$C_\psi = \int_{-\infty}^{+\infty} \frac{|\widehat{\psi}(\omega)|^2}{\omega} d\omega \qquad (2.4)$$

De plus $\psi(\omega)$ est la transformée de Fourier de $\psi(t)$.

Quelques fonctions d'échelle et d'ondelettes sont présentées dans les figures figure 2.1 , *figure 2.2* ci-dessous. La famille Daubechies est la famille la plus connue des ondelettes orthonormales [9].

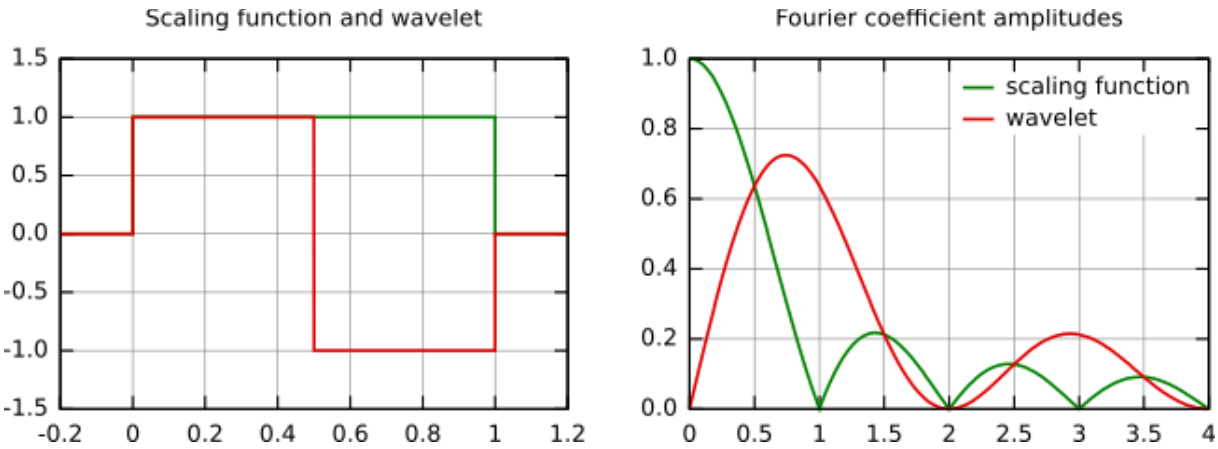

Figure 2.3: Fonction d'échelle et ondelette de Haar (à gauche) et leur contenu fréquentiel (à droite).

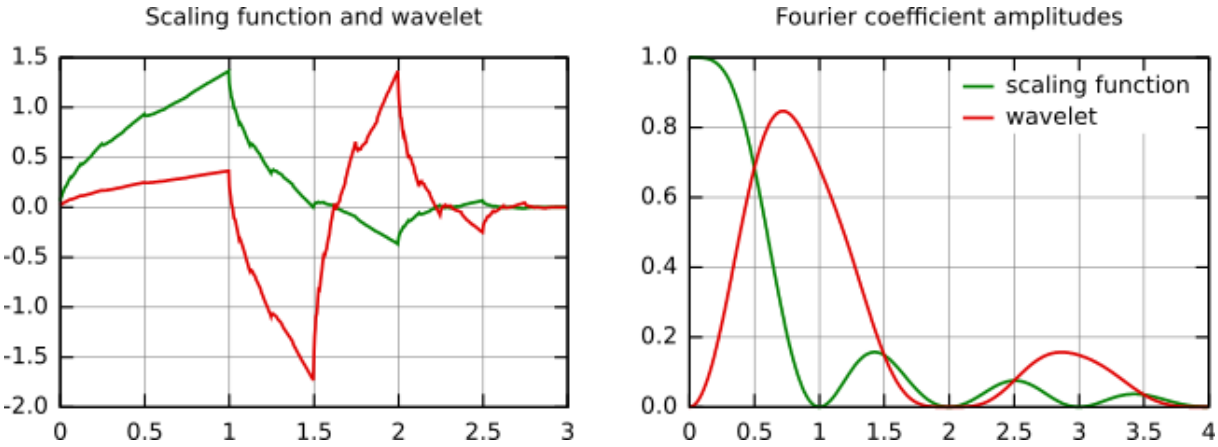

Figure 2.4: Fonction d'échelle et ondelette Daubechies 4 (à gauche) et leur contenu fréquentiel (à droite).

2.3 Transformée en ondelettes discrète

La transformée en ondelettes discrète (TOD, en anglais : Discrete Wavelet Transform, ou DWT) était introduite pour surmonter le problème de redondance de la TOC. Cette redondance mobilise une grande quantité de ressources de calculs. La TOD, au contraire, fournit suffisamment d'information, tant pour l'analyse que pour la reconstruction du signal original, en un temps de calcul notablement réduit. La TOD est considérablement plus simple à implémenter que la TOC. Dans la transformation en ondelette, on parle souvent d'approximation et de détails. L'approximation est à haute échelle, les composantes de basse fréquence du signal. Les détails sont à basses échelles, les composantes de hautes fréquences.

La décomposition en ondelettes discrète d'un signal $x(t)$est exprimée par :

$$x(t) = \sum_{n=-\infty}^{\infty} a(n)\varphi(t-n) + \sum_{=0}^{\infty}\sum_{n=-\infty}^{\infty} d(k,n)2^{k/2}\,\psi(2^k t - n) \qquad (2.5)$$

Les coefficients d'approximation $a(n)$et les coefficients d'ondelettes $d(k,n)$ à l'échelle *k* sont calculés à l'aide des produits scalaires :

$$a(n) = \int_{-\infty}^{\infty} x(t)\varphi(t-n)dt \qquad (2.6)$$

$$d(k,n) = 2^{k/2}\int_{-\infty}^{\infty} x(t)\psi(2^k t - n)dt \qquad (2.7)$$

Avec $\varphi(t)$est appelée la fonction d'échelle (passe-bas) et $\psi(t)$ est la fonction ondelettes (passe-bande). Dans la pratique, l'implémentation de la décomposition en ondelettes discrètes est basée sur un banc de filtres à deux canaux (généralement RIF) appliqués en cascade.

La figure 2.3 illustre une décomposition en ondelettes discrète à trois niveaux du signal discret $x(n)$.

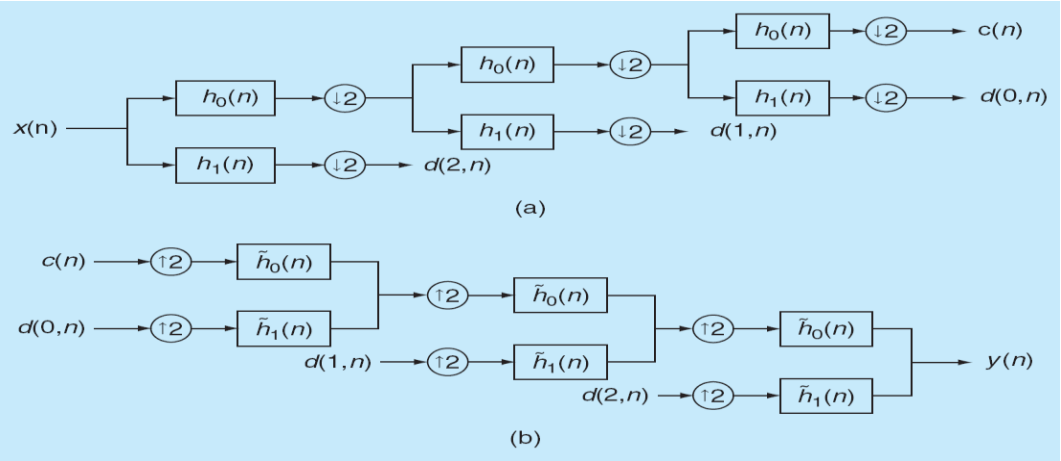

Figure 2.5: Banc de filtres implémentant la DWT : (a) analyse, (b) synthèse.

Le principe est d'appliquer d'abord un banc de filtres $[H_0; H_1]$ sur le signal. Le résultat du filtrage par H_0 (qui sera supposé ici être un filtre passe-bas) représente les coefficients d'approximation et le résultat par H_1 (supposé passe haut) représente les coefficients de détail

du signal. On applique de nouveau ce banc de filtres sur le résultat du filtrage par H_0 et cette opération est répétée un nombre fini de fois. A chaque étage de décomposition en ondelettes on change en fait d'échelle dans la représentation du signal analysé.

Le schéma de synthèse se construit simplement de proche en proche, en utilisant une structure en cascade symétrique à celle utilisée pour l'analyse. La figure présente également la partie de reconstruction. Dans cette partie, les filtres appliqués $[\tilde{H}_0 ; \tilde{H}_1]$ doivent assurer la reconstruction parfaite (c.à.d. $\hat{x}(n) = x(n-k)$). La fonction d'ondelette d'analyse $\psi(t)$ associée à ces filtres est définie par :

$$\psi(t) = \sqrt{2} \sum_n h_1(n) \varphi(2t - n) \qquad (2.8)$$

Et la fonction d'échelle $\varphi(t)$qui est donnée par :

$$\varphi(t) = \sqrt{2} \sum_n h_0(n) \varphi(2t - n) \qquad (2.9)$$

L'ondelette et la fonction d'échelle de synthèse, $\tilde{\psi}(t)$ et $\tilde{\varphi}(t)$, sont définies par les mêmes équations, mais en utilisant les filtres $\tilde{H}_i$au lieu des H_i.

La figure ci-dessous représente un signal de parole décomposé à 4 niveaux d'ondelettes.

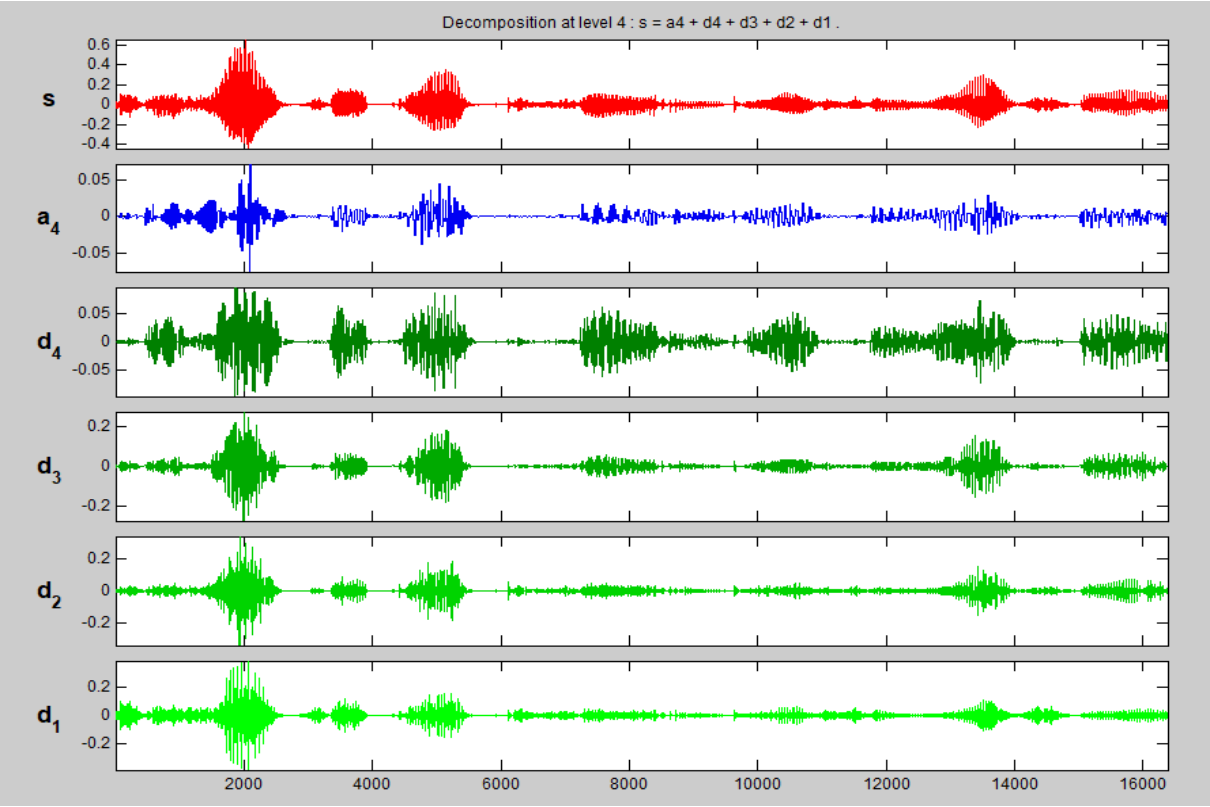

Figure 2.6: Décomposition d'un signal de parole à 4 niveaux ondelettes.

2.4 Codage de la parole par ondelettes

Récemment, de nombreuses méthodes basées sur les ondelettes ont été conçues afin de compresser les signaux vocaux [10]. La caractéristique la plus importante de la transformée en ondelettes, par rapport à la compression des données, est qu'elle a tendance à concentrer l'énergie du signal de l'entrée en un nombre relativement petit de coefficients d'ondelettes. Par rapport au signal complet, le codage de ces coefficients nécessite moins de ressources binaires, tout en conservant une qualité satisfaisante du signal reconstruit. Dans un schéma de codage de la parole en ondelette, trois étapes sont couramment utilisées :

2.4.1 Décomposition DWT

Dans cette étape, le banc de filtres est conçu puis la décomposition DWT est effectuée sur le signal de la parole. En plus des coefficients du banc de filtres, les performances de codage de la parole sont étroitement liées aux filtres employés et aux nombres de niveaux de décomposition effectués [11] [12]. Il a été recommandé dans [11] que le nombre adéquat de niveaux décomposition dans la compression audio doit être inférieurs ou égal à cinq.

2.4.2 Seuillage

Le seuillage est une technique non linéaire simple dans laquelle chaque coefficient est comparé à un seuil ; si le coefficient est inférieur au seuil, alors il est mis à zéro ; sinon il est maintenu ou modifié. En général, les méthodes de seuillage peuvent être classées en deux catégories, à savoir, méthodes de seuillage globales et par niveau. En seuillage global, le signal décomposé est classé en fonction d'une valeur seuil unique estimée à partir de ses coefficients d'ondelettes. En seuillage de niveau, cette valeur est évaluée pour chaque niveau de l'arbre de décomposition en ondelettes.

Dans cette étude, un seuillage global est appliqué. La valeur du seuil est donnée par :

$$\lambda = \sigma\sqrt{2log(N)} \qquad (2.10)$$

Où N est le nombre d'échantillons dans le signal et σ est la ***variance*** du signal.

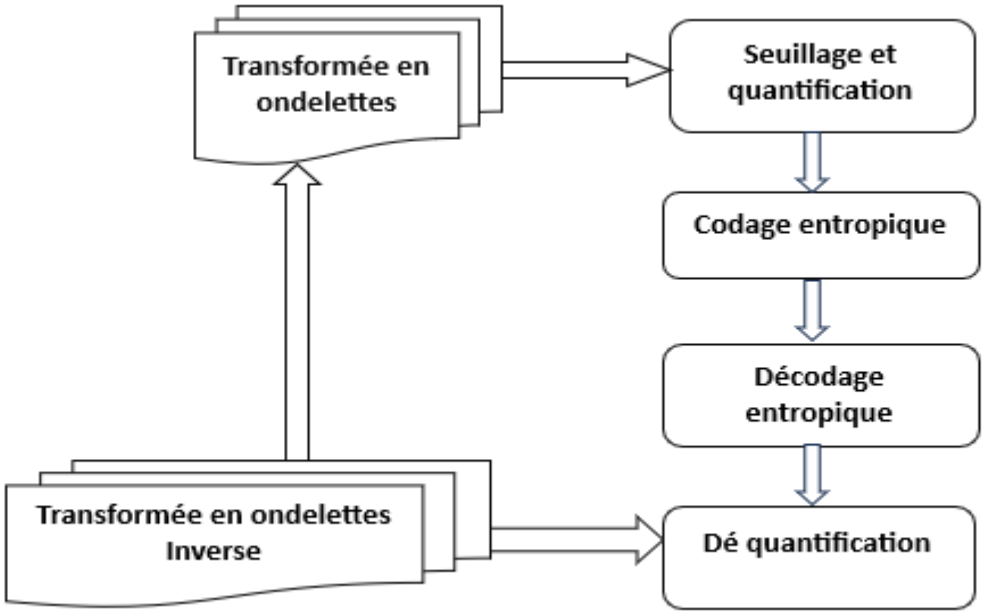

Figure 2.7: Schéma de codage de la parole en ondelette

2.4.3 Quantification

Le but de cette technique est de masquer les informations non pertinentes du signal. Cela minimise le nombre de bits nécessaires pour stocker les coefficients transformés en réduisant la précision de ces valeurs. Dans une quantification uniforme, les coefficients d'ondelettes sont quantifiés en utilisant le pas calculé par :

$$\triangle = \frac{Xmax - Xmin}{L} \qquad (2.11)$$

Où $Xmax$ et $Xmax$ sont respectivement le maximum et le minimum des valeurs dans le signal et L est le nombre de niveaux de quantification.

2.4.4 Codage entropique

Cette technique réduit la redondance statistique dans les coefficients quantifiés en utilisant les méthodes de codage à longueur variable telles que le codage de Huffman, le codage arithmétique, …, etc.

2.5 Codage par arbre de zéros

Le premier algorithme interbandes pour les images 2D se nomme EZW et a été proposé en 1993 par Jerry Shapiro [13]. Le codage EZW exploite complètement la notion de multi-résolution associée aux ondelettes. Leur schéma de codage utilise un modèle simple pour caractériser les dépendances interbandes parmi les coefficients d'ondelettes localisés dans les sous-bandes ayant la même orientation. Le modèle est basé sur l'hypothèse des arbres de zéros, la création d'un arbre de zéros part du principe que si un coefficient d'ondelette à une échelle plus grossière n'est pas significatif par rapport à un seuil T, alors il est fortement

probable que tous les coefficients d'ondelettes aux échelles plus fines qui ont la même orientation (horizontale, verticale ou diagonale) et la même localisation spatiale soient aussi non significatifs par rapport à ce même seuil T. En pratique, la probabilité que ce phénomène survienne est très élevée.

Plus spécifiquement, dans un système à sous-bandes hiérarchique, à l'exception de la sous-bande de plus basse fréquence (A3), chaque coefficient dans une échelle donnée peut être relié avec un ensemble de coefficients dans l'échelle plus fine suivant qui a la même orientation spatiale (cf., figure 2.6). Le coefficient dans l'échelle grossière est appelé 'parent' et tous les coefficients ayant la même orientation spatiale dans l'échelle plus fine suivant sont appelés 'fils'.

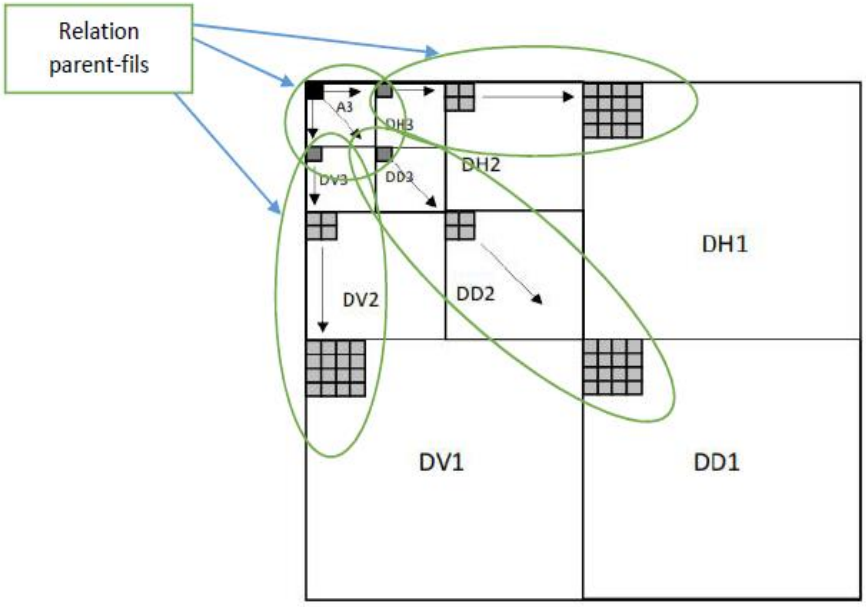

Figure 2. 8: Modèle de dépendance inter-bandes.

La figue montre un arbre de quatre obtenu après trois niveaux de décomposition en ondelettes. La flèche pointe du parent vers l'enfant puis vers le petit-enfant. Une relation similaire existe également pour les sous-bandes LH3 et HH3. Cependant cette relation n'existe pas dans la sous-bande LL3.

Pour un parent donné, l'ensemble des coefficients aux échelles plus fines qui ont la même orientation spatiale sont appelés « descendants ». De même, pour chaque fils à une échelle donnée, l'ensemble des coefficients aux échelles plus grossières qui ont la même orientation spatiale sont appelés « ancêtres ».

Les relations parent-fils dans une structure pyramidale sont représentées en figure 2.6. A l'exception de la sous-bande de plus basse fréquence, chaque parent admet quatre fils dans la sous-bande de même orientation et de résolution juste supérieure. En ce qui concerne la sous-bande de plus basse fréquence (A3), chaque coefficient admet trois fils dans chacune des trois sous-bandes correspondant à la même résolution comme indiqué dans la figure 2.6.

2.5.1 Codage progressif

Les algorithmes EZW et SPIHT exploitent la structure hiérarchique pyramidale entre les différents niveaux des transformées en ondelettes. Pour cela, ils utilisent l'encodage imbriqué (appelé également encodage progressif) pour envoyer en premier lieu les informations concernant les niveaux supérieurs de la transformée, puis ils envoient progressivement les informations concernant les niveaux plus basses, et affinent les informations déjà envoyées. Cela permet ainsi la transmission progressive de l'image.

En effet, un encodeur utilisant un encodage imbriqué envoie progressivement des détails améliorant la qualité de l'image au fur et à mesure où il fait sortir *des données binaires.*

Les exemples suivants montrent l'avantage de cette technique d'encodage.

Supposons que nous avons trois utilisateurs qui attendent d'un serveur une certaine image comprimée, mais ils ont besoin des qualités d'images différentes. Le premier a besoin de la qualité contenue dans un fichier de 10Ko. Les qualités de l'image voulue par le second et troisième utilisateur sont contenus respectivement dans des fichiers de 20Ko et 50Ko.

2.5.2 Codage SPIHT

L'algorithme SPIHT a été proposé par Saïd et Pearlman en 1996 pour la compression d'image 2D avec et sans perte [14]. Cet algorithme repose sur la même idée que celle de Shapiro (EZW) pour caractériser les dépendances entre les coefficients d'ondelettes. Cependant, il utilise les trois principes de base suivant :

- Un rangement partiel par amplitude des coefficients d'ondelettes de la TO 2D (Résultant de la quantification par approximations successives),
- Un partitionnement dans des arbres hiérarchiques (à chaque seuil appliqué les arbres sont triés sur la base de leur signification en deux catégories d'arbre)
- Un ordonnancement de la transmission des bits de raffinement (l'amplitude de chaque coefficient significatif est progressivement raffinée).

Pour caractériser les relations parent enfant dans les sous bandes. Les ensembles suivants de coordonnées sont utilisés :

- $O(i,j)$: Ensemble des coordonnées de tous les enfants du nœud (i, j). Il s'exprime de la même façon que celui de l'EZW.
- $D(i,j)$: Ensemble des cordonnées de tous les descendants du nœud (i, j) (type A d'arbres de zéros).

- $L(i,j) = D(i,j) - O\ (i,j)$: L'ensemble des descendants à l'exception des enfants (type B d'arbre de zéros).

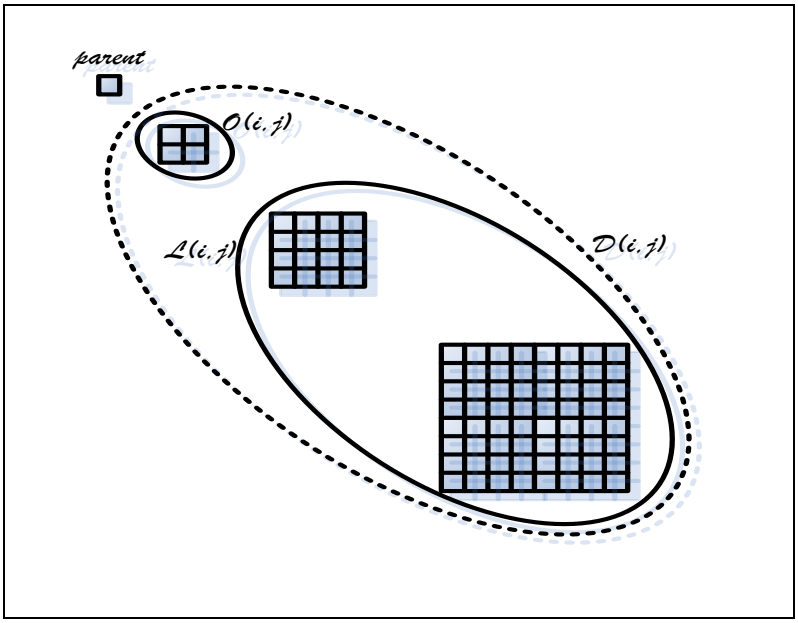

Figure 2.9: La relation de descendance dans l'algorithme SPIHT.

a. Principe du SPIHT

L'algorithme SPIHT maintient trois listes de coefficients : la liste des coefficients significatifs LSP, la liste des coefficients non significatifs LIP et la liste des ensembles non significatifs LSP

O (i, j) (Offspring) : ensemble de coordonnées de toute les descendants direct du nœud (i, j); (i.e., childrenonly) Excepté à la plus haute et plus basse niveau de la pyramide, nous avons :

$$O(i,j) = \{(2i,2j),(2i,2j+1),(2i+1,2j),(2i+1,2j+1)\}. \qquad (2.12)$$

D(i, j) (descendants) : ensemble de coordonnées de tous les descendants du nœud (i, j); (i.e., children, and the following générations).

H : l'ensemble de coordonnées de toute racine d'arbre d'orientation spatial (nœuds au niveau le plus élevé de pyramide); parents L (i, j) (the leaves) = D (i, j) - O(i, j) .

Dans LIP et LSP il y a des pixels individuels, et LIS représente soit l'ensemble D(i, j) ou L(i,j), pour les différencier, nous disons que d'une entrée de LIS est de type A si elle représente D(i, j), et de type B si elle représente L(i, j).

Durant le passage de triage les pixels dans LIP sont testés-lesquels étaient signifiants dans le passage prédécesseur, et ceux décidés significatifs, à l'aide de la fonction de signifiance :

$$S_n(T) = \begin{cases} 1, & max\{|C_{i,j}|\} \geq 2^n \\ 0; & otherwise \end{cases} \qquad (2.13)$$

Sont déplacés à LSP, similairement pour les ensembles LIS, sont évalués successivement et ceux décidés significatifs, au moins l'un des pixels rencontrés significatif lors de passage du triage, sont décomposés en autres sous-ensembles. Les nouveaux sous-ensembles sont réintégrés dans LIS, et si l'un a un seul pixel, il réintégré dans LIP ou LSP.

b. Algorithme de codage

A. Initialisation

$$n = \left\lfloor \log_2(\max_{i,j}\{|c_{i,j}|\}) \right\rfloor \qquad (2.14)$$

– LSP = ∅, une liste vide

– LIP : tous les coefficients de la sous-bande la basse fréquence

– LIS : tous les coefficients de la LIP qui ont des descendants (marqués comme type A, par défaut)

Dans toute les liste chaque entrée est identifié par leur coordonné (i,j).

B. Passage de Triage

B.1) Pour chaque coefficient (i, j) de la LIP :

B.1.1) Ecrire S_n(i, j) ;

B.1.2) Si S_n(i, j) = 1, déplacer (i, j) dans la LSP et écrire le signe de $c_{i,j}$

B.2) Pour chaque coefficient (i, j) de la LIS faire :

B.2.1) Si le coefficient est de type A, alors

- Ecrire S_n(D(i, j));
- Si $S_n = (D(i,j)) = 1$ alors
  - Pour tout $(k,l) \in O(i,j)$faire:
    - Ecrire S_n(k, l) ;
    - Si S_n(k, l) = 1, ajouter (k, l) à la LSP et écrire le signe de $c_{i,j}$;
    - Si S_n(k, l) = 0, ajouter (k, l) à la fin de la LIP ;
  - Si L(i, j) ≠ ∅, déplacer (i, j) à la fin de la LIS comme une entrée de type B, et passer à l'étape (B.2.2) **;** sinon, retirer (i, j) de la LIS ;

B.2.2) Si l'entrée est de type B, alors

- Ecrire S_n(L(i, j))
- Si S_n(L(i,j)) = 1, alors
  - Ajouter tous les $(k,l) \in O(i,j)$ à la fin de la LIS comme entrée de type B Retirer (i, j) de la LIS.

C. Passage de Raffinement

Pour tous les coefficients (i, j) de la LSP à l'exception de ceux ajoutés autours du dernier storting pas (e. g., avec la même n), écrire le n^{ième} bit le plus significatif de $\left|c_{i,j}\right|$;

D. Mise à jour du Niveau de Quantification

Décrémenter n par 1 et retourner à l'étape 2.

E. Décodage

Le décodeur utilise un algorithme similaire à l'encodeur mais en remplaçant les sorties par des entrées. Il a également pour tâche de mettre à jour le signal reconstruit à chaque arrivée d'une nouvelle information à l'entrée.

1.6 Conclusion

Dans ce chapitre, nous avons commencé par la description de la transformée en ondelettes. La transformée en ondelettes continues et discrète sont brièvement exposées. Nous avons présenté le schéma de codage basé sur la transformation en ondelettes. Ensuite, nous avons présenté le principe du codage par arbre zéro et l'algorithme SPIHT. Cet algorithme sera employé dans le chapitre suivant pour le codage de la parole.

CHAPITRE 03

3.1 Introduction

Le codage de la parole est un aspect crucial du traitement du signal et des communications numériques. Il permet de représenter efficacement les signaux vocaux pour leur transmission ou stockage, tout en conservant la qualité de la parole.

Après la présentation du principe de codage de la parole par ondelettes dans le chapitre précédent, cette partie sera consacrée à l'évaluation des performances de cette technique codage.

D'abord, nous allons présenter les résultats de compression par ondelettes classique.

Ensuite, nous allons détailler les performances de compression de l'algorithme SPIHT.

Une comparaison des performances de codage de ces deux algorithmes est également effectuée à la fin de ce chapitre.

3.2 Critères d'évaluation de la qualité du signal de la parole

L'évaluation de la qualité du signal de la parole est cruciale pour assurer la clarté et l'intelligibilité dans les systèmes de communication vocale. Plusieurs critères sont utilisés pour évaluer la qualité du signal de la parole, chacun ayant des approches objectives et subjectives. Les mesures subjectives, telles que le MOS, restent essentielles pour capturer les perceptions humaines, tandis que les méthodes objectives comme le SNR, et le SSNR. Le choix des critères dépend souvent de l'application spécifique et des ressources disponibles pour les évaluations. Les critères employés dans ce travail sont succinctement présentés ci-dessous :

3.2.1 Critères objectives

A. SNR (rapport signal sur bruit)

Le critère objectif le plus couramment utilisé est le rapport signal sur bruit SNR (Signal to **N**oise Ratio). Est un indicateur de la qualité de la transmission d'une information. Il est défini comme le rapport entre la puissance du signal original et la puissance du signal d'erreur. Il est exprimé par :

$$SNR = 10log\left(\frac{x^2(n)}{x^2(n)-\hat{x}^2(n)}\right) \quad (3.1)$$

Où (n) est le signal original, $\hat{x}(n)$ est le signal reconstruit après compression.

B. SSNR (Segmental Signal-to-Noise Ratio)

Quand on détermine le SNR global, on calcul la puissance globale des deux signaux, original et synthétisé, sans tenir compte de la répartition mutuelle des énergies dans le domaine du

temps. De ce fait, on peut rencontrer des zones de faibles énergies dans lesquelles la différence entre le signal original et le signal synthétisé est importante par rapport au signal original. La faible valeur des énergies respectives des deux signaux n'affecte pas le rapport signal/bruit, cependant la dégradation du signal est audible. Pour remédier à cet inconvénient, on définit le *rapport signal/bruit segmental* (SSNR). Il prend davantage en compte les zones de faible énergie du signal que le rapport signal à bruit global. Plutôt que de calculer un seul rapport SNR sur l'ensemble du signal, le SSNR est calculé pour chaque segment de longueur fixe et ensuite moyenné. La formule pour le SSNR est :

$$SSNR = \frac{1}{N}\sum_{i=1}^{N} 10log\ 10\frac{x_i^2(n)}{x_i^2(n)-\hat{x}_i^2(n)} \qquad (3.2)$$

Où N est le nombre de segments, $x_i(n)$est le signal original dans la *i*-ème segment, et $\hat{x}_i(n)$ est le signal reconstruit dans le *i*-ème segment.

Il ne faut surtout pas considérer ces mesures comme des critères suffisamment représentatifs de la qualité d'un système de codage de la parole. Elles utilisent des relations mathématiques qui ne tiennent pas compte des propriétés de l'audition humaine. Néanmoins des tests d'écoute ont ainsi montré qu'une amélioration du rapport signal/bruit entraînait certainement un accroissement de la qualité du signal synthétisé sur le plan de perception auditive. De toute façon, ces mesures de SNR doivent être accompagnées par des méthodes de mesure subjective de la qualité.

3.2.2 Critères subjectives

Plusieurs méthodes de mesure de la qualité existent, la plus principalement utilisée est la note moyenne d'opinion (En anglais MOS : **M**ean **O**pinion **S**core). Le test MOS fournit à l'auditeur cinq niveaux d'appréciation possibles {1 : Mauvais ; 2 : Médiocre ; 3 : Passable ; 4 : Bon ; 5 : Passable). Le moyennage du score sur un nombre important d'auditeurs, donne une note entre 1 et 5 de l'agrément d'écoute.

Les valeurs de MOS sont fiables car elles sont basées sur la perception humaine. Un grand nombre d'auditeurs est requis, de sorte qu'une évaluation raisonnable puisse être faite. Ceci peut être long (demande beaucoup de temps) et cher. Par conséquent, diverses mesures objectives ont été développées et ont comme but de renvoyer la même valeur que celle du test MOS. Parmi eux, on trouve le PESQ normalisé par ITU-T (Union Internationale des Télécommunications – Secteur Télécommunications) en Février 2001. Il est adopté comme la recommandation ITU-T P.862 [15]. Il a été montré que le PESQ peut fournir des résultats fortement corrélés avec les évaluations subjectives du test MOS.

Pour évaluer la qualité d'un signal reconstruit par un codeur de parole en utilisant le PESQ, deux entrées sont exigées : le signal reconstruit ou signal à tester, et un signal original. La méthode de test est de prendre le signal de parole synthétisé et on le transmettre à travers le système PESQ et on le compare avec le signal de parole original, comme illustre la figure 3.1.

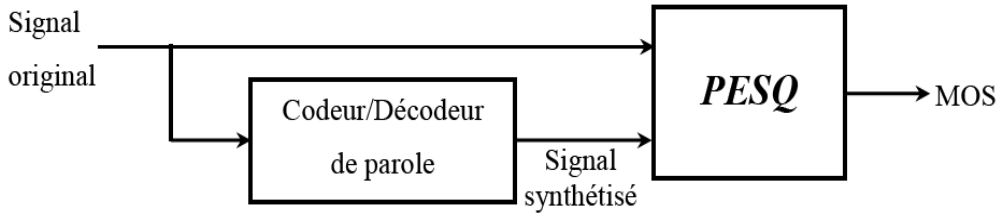

Figure 3. 1: Système PESQ pour l'évaluation des performances d'un codeur/décodeur de parole.

3.2.3 Taux de compression

Le terme taux de compression a différentes significations selon le domaine dans lequel il s'applique. Selon le type de données (caractères, image, son, vidéo, etc.), ce taux peut être très différent. En informatique, le taux de compression est le rapport du volume de données avant et après compression :

$$TC = \frac{nombre\ de\ bits\ du\ signal\ original}{nombre\ de\ bits\ du\ signal\ reconstruit} \qquad (3.3)$$

3.3 Base de données et logiciel de calcul

Dans ce travail, les tests ont été réalisés sur huit (03) signaux de parole prononcés en langue anglaise, à savoir : « SA1 », « SI1027 », et « SX37 ». Ces fichiers sont sélectionnés à partir de la base de données TIMIT [16].

Matlab est logiciel interactif de calcul scientifique employé dans ce travail. Sa simplicité fait de lui un outil de choix pour la mise au point de algorithmes 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.

3.4 Résultats expérimentaux

Dans cette section nous allons exposer les résultats de compression des algorithmes de compression de la parole étudiées dans ce travail.

3.4.1 Algorithme de codage par ondelettes classique

Dans cette section, les performances de l'algorithme de compression par ondelettes classique de la figure 2.5 sont détaillées. Elles sont exprimées en termes de PESQ, SNR et SSNR dans les tableaux 3.1 pour les trois signaux test.

	TC	SNR	SSNR	PESQ
« SA1 »	10	15.50	9.83	2.21
	12	11.24	6.11	1.86
	14	7.26	3.06	0.97
« SI1027 »	10	15.76	8.47	2.41
	12	11.30	5.16	1.92
	14	7.12	2.48	1.41
« SX37 »	10	18.40	7.92	2.81
	12	14.15	4.89	2.11
	14	10.20	2.58	1.14

Tableaux 3.1: Performances de compression en termes de PESQ, SNR et SSNR de l'algorithme de codage par ondelettes classique.

Il tout à fait naturel que les performances de compression se dégrade lorsque le taux de compression augmente et cela est confirmée par inspection visuelle des signaux reconstruits des figures 3.2, 3.3, et 3.4 et vice versa.

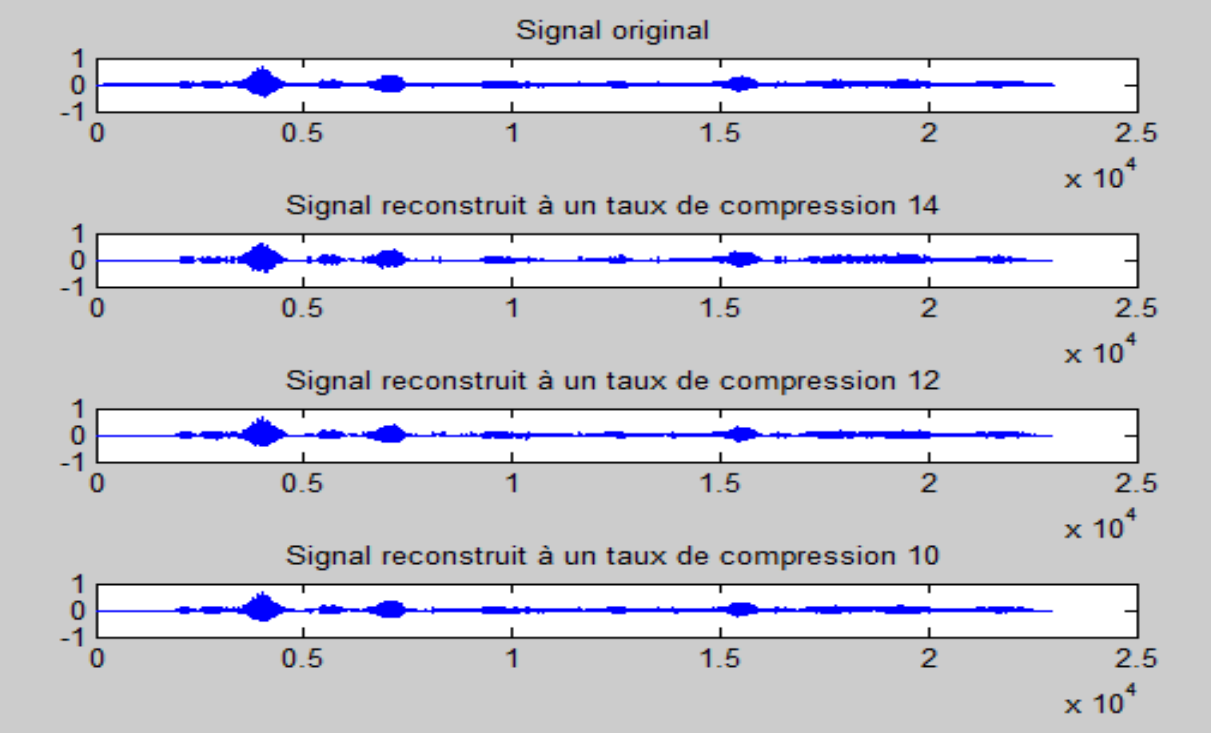

Figure 3.2: Signal «SA1 » synthétisé avec l'algorithme de codage par ondelettes classique pour les différents taux de compression.

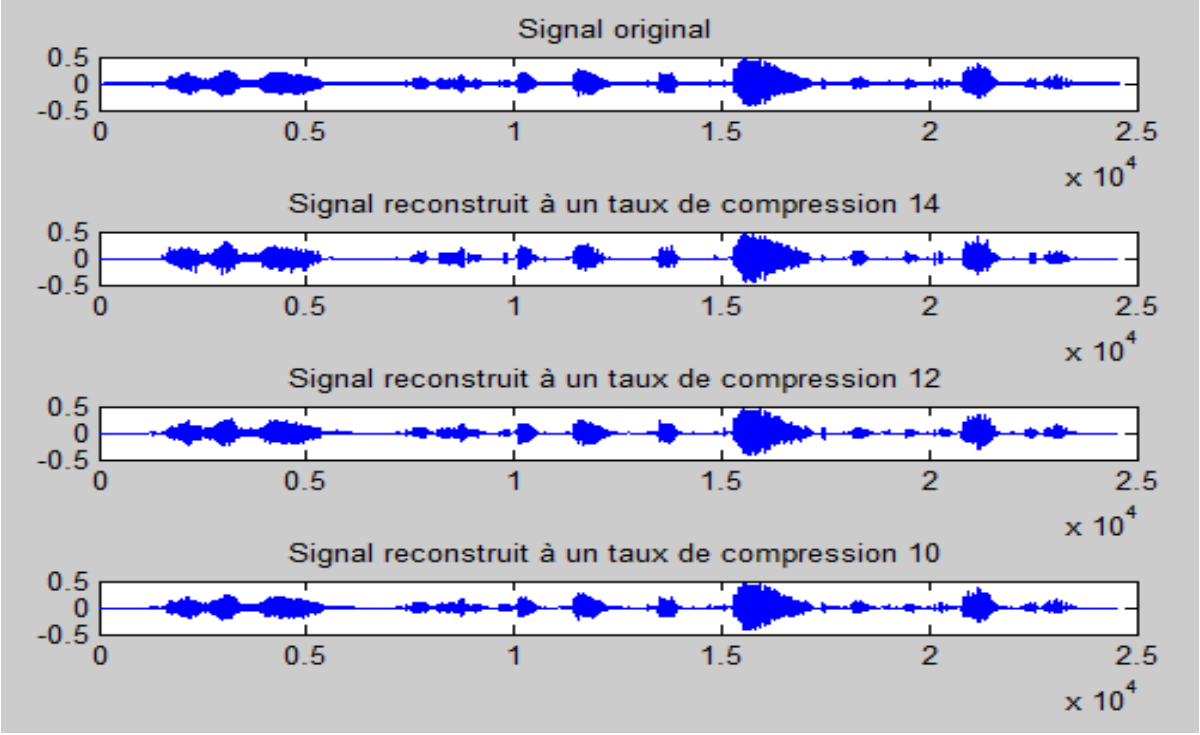

Figure 3.3: Signal «SI1027 » synthétisé avec l'algorithme de codage par ondelettes classique pour les différents taux de compression.

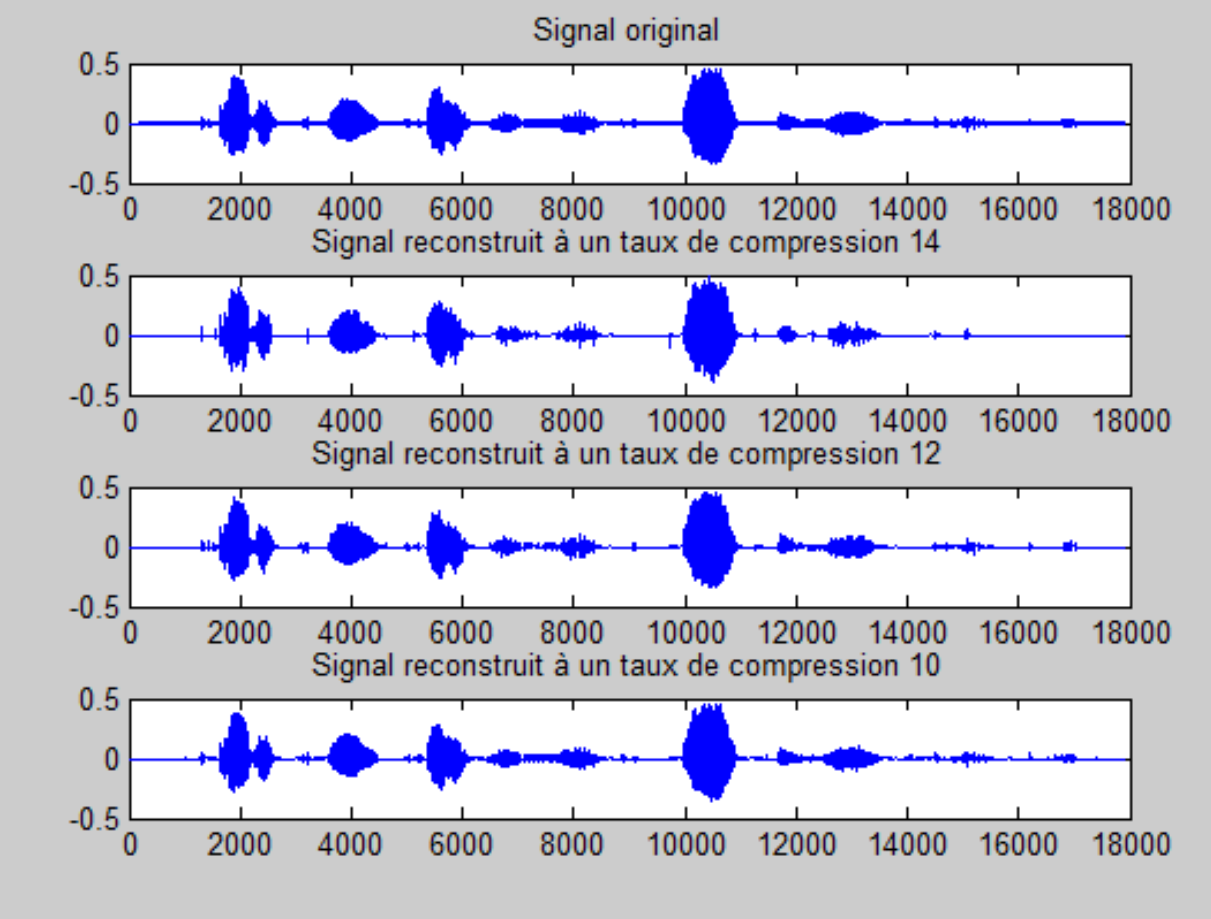

Figure 3.4: Signal « SX37 » synthétisé avec l'algorithme de codage par ondelettes classique pour les différents taux de compression.

3.4.2 Algorithme de codage SPIHT

Dans cette section, les performances de l'algorithme de compression par ondelettes SPIHT présenté dans la section 2.5.1 sont détaillées. Elles sont exprimées en termes de PESQ, SNR et SSNR dans les tableaux 3.2 pour les trois signaux test.

Il faut souligner que, dans un système à sous-bandes hiérarchique unidimensionnel, à l'exception de la sous-bande de plus basse fréquence, chaque coefficient dans une échelle donnée est relié à 2 coefficients (au lieu de 4 dans le cas bidimensionnel) dans l'échelle plus fine suivant qui a la même orientation spatiale.

	TAUX	SNR	SSNR	PESQ
« SA1 »	10	14.28	8.82	2.20
	12	12.48	7.28	1.98
	14	11.02	6.13	1.60
« SI1027 »	10	15.27	8.18	2.36
	12	13.19	6.77	2.18
	14	11.29	5.46	1.90
« SX37 »	10	19.84	9.26	2.77
	12	17.48	7.48	2.50
	14	15.82	6.26	2.11

Tableaux 3. 2: Performances de compression en termes de PESQ, SNR et SSNR de l'algorithme de codage SPIHT.

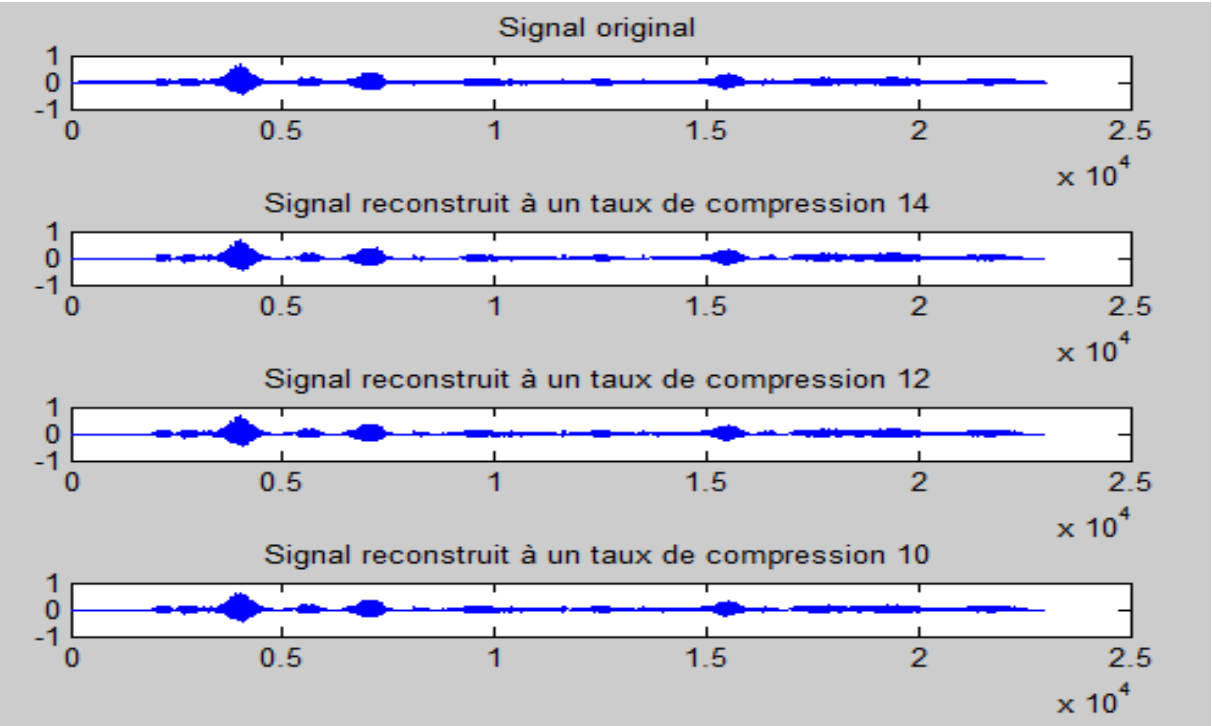

Figure 3.5: Signal «SA1 » synthétisé avec l'algorithme compression SPIHT pour les différents taux de compression.

Les figures 3.5, 3.6, et 3.7 illustrent les changements survenus sur les signaux synthétisés lorsque le taux de compression change. On peut facilement remarquer que les signaux reconstruits à un taux de compression 10 présentent moins de dégradations et sont plus proche au signal original.

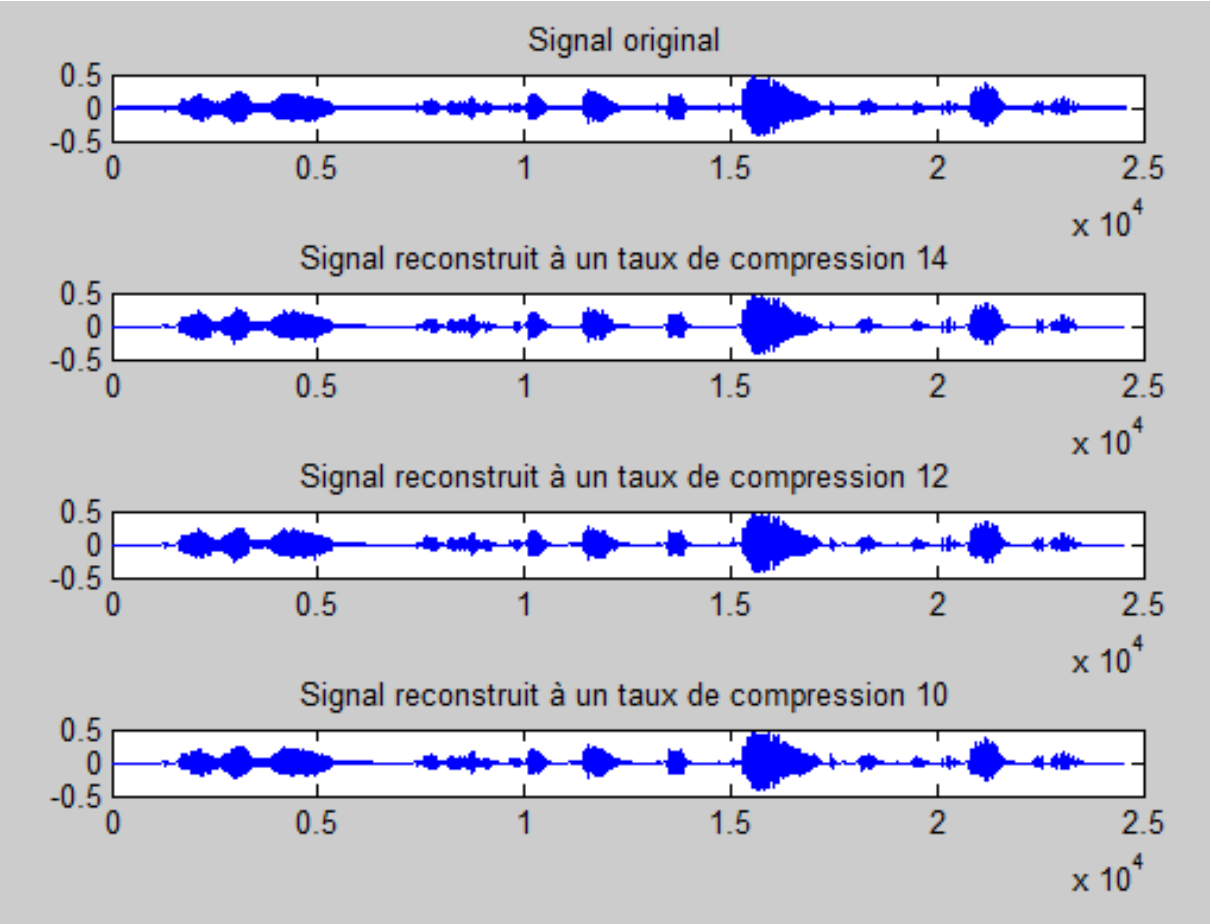

Figure 3. 6: Signal «SI1027 » synthétisé avec l'algorithme compression SPIHT pour les différents taux de compression.

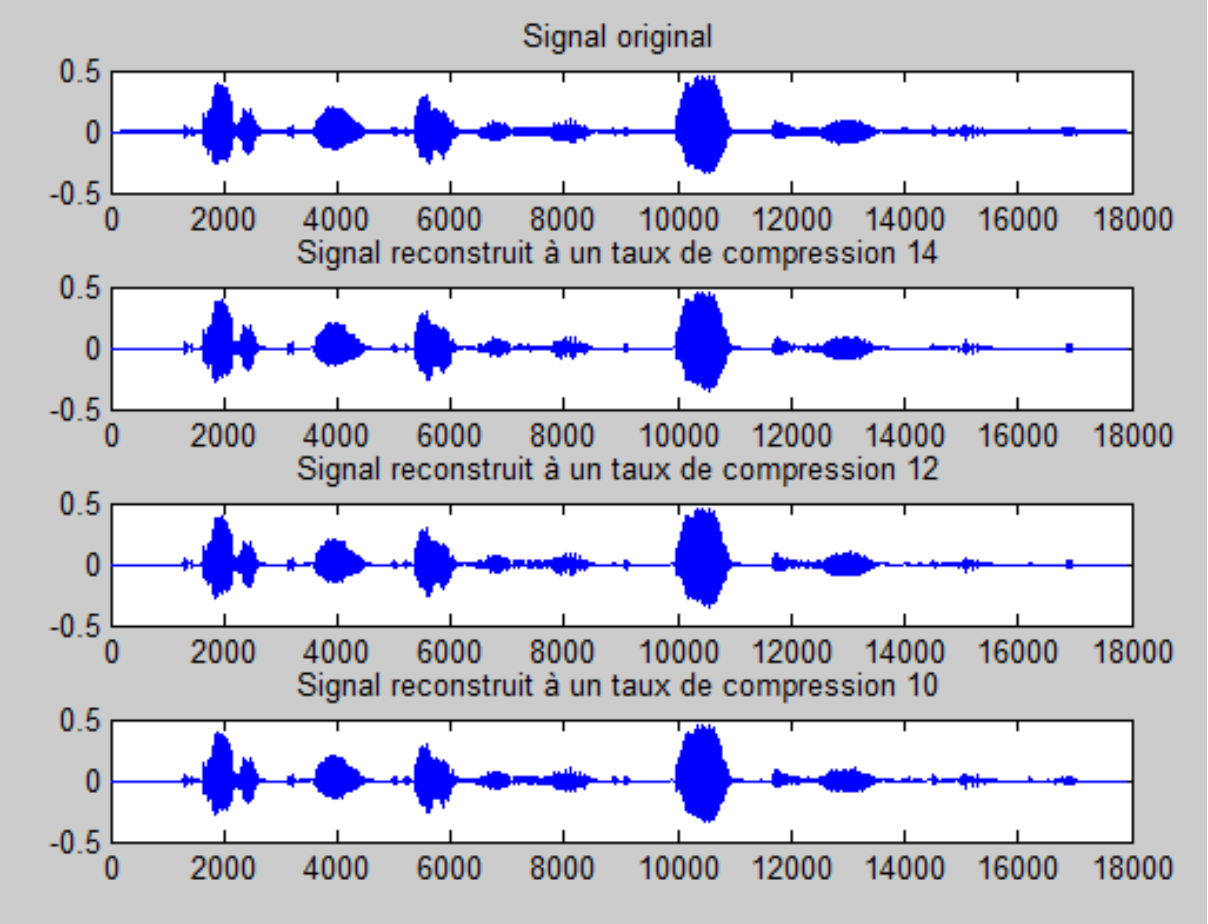

Figure 3.7: Signal « SX37 » synthétisé avec l'algorithme compression SPIHT pour les différents taux de compression.

3.4.3 Comparaison en l'algorithme de codage par ondelettes classique et SPIHT

Dans cette section, nous allons comparer brièvement les performances des deux algorithmes de compression par ondelettes présentés ci-dessus.

A partir des tableaux 3.1 et 3.2, nous avons constaté que l'algorithme SPIHT offre des performances nettement supérieures à l'algorithme de codage par ondelettes classique surtout lorsque le taux de compression augmente. Par exemple, pour le signal « SX37 » et pour un taux de compression 14, nous avons relevé un gain de SNR de 5.60 dB et de SSNR de 3.7 dB.

En termes de PESQ nous avons enregistré une amélioration de 1dB environ ce qui confirme l'amélioration de la qualité subjective du signal reconstruit. Cela est bien illustré lorsqu'on observe les signaux de la figure 3.8.

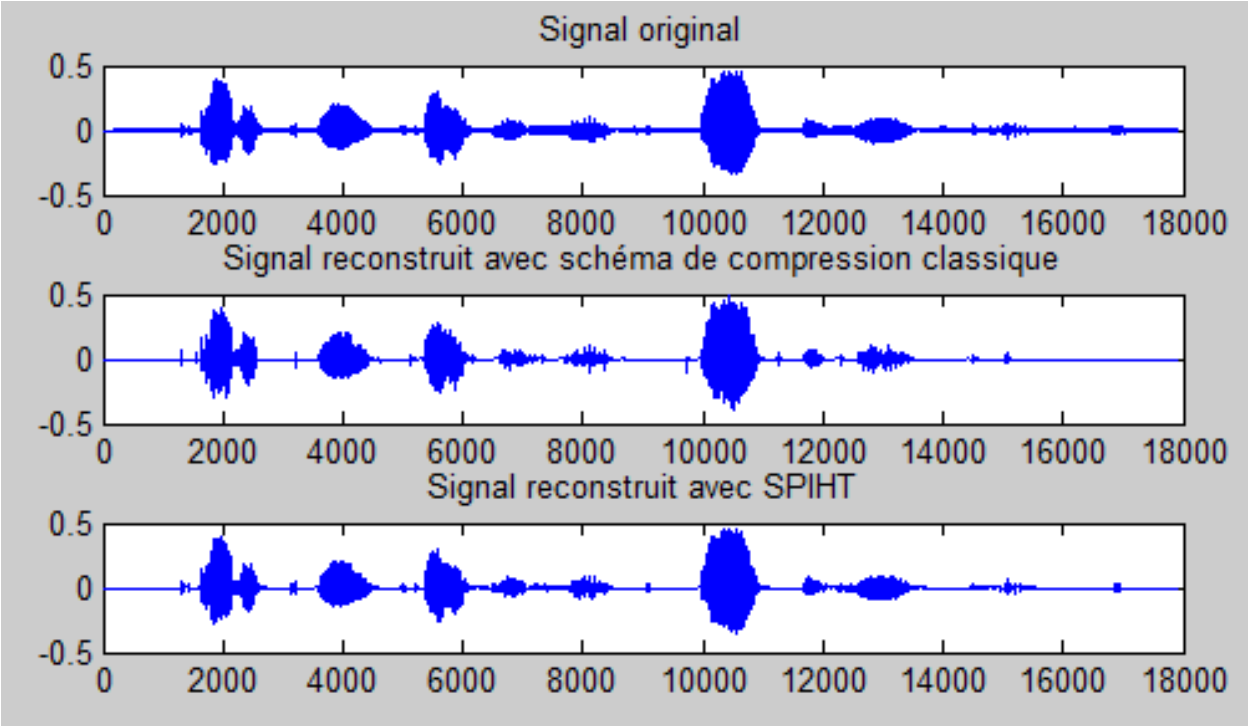

Figure 3.8: Signal « SX37 » synthétisé avec les deux algorithmes de compression par ondelettes pour un taux de compression 14.

3.5 Conclusion

Deux algorithmes de compression de la parole par ondelettes ont été implantés. Les tests montrent l'efficacité de l'algorithme SPIHT. En plus de flexibilité et rapidité, cet algorithme offre des performances excellentes non seulement pour les critères objectives mais également pour les critères subjectives.

Conclusion Générale

Conclusion Générale

Les problèmes critiques qui constituent des contraintes dans les communications sans fil, sont la bande passante, la mémoire de stockage et l'alimentation. L'objectif dans le codage de la parole est associé à la réduction des informations supplémentaires présentes dans le signal afin de représenter le signal vocal avec un nombre restreint de bits tout en mettant à jour sa qualité perceptuelle. Pour cette raison, le codage de la parole sera la question de recherche la plus importante.

Afin de réaliser un codage performant permettant de préserver la qualité, il est nécessaire de prendre en considération certains points essentiels tels que : la complexité des algorithmes décodage, le débit binaire et l'intelligibilité de la parole.

Dans ce travail nous avons focalisé sur la compression par ondelettes. Pour cela nous avons implémenté deux algorithmes de codage de la parole. Le premier est un algorithme de codage classique basé sur les trois étapes clé : transformation, quantification, et codage entropique. Le deuxième est le codage par arbre de zéro (SPIHT).

Le but et de d'évaluer les performances en compression ces deux types codeurs. Des critères objectifs et subjectifs sont employés pour l'évaluation de la qualité de la parole synthétisée.

Les résultats obtenus montrent que les deux codeurs donnent des bonnes performances de compression, avec des scores « SNR, SSNR et le PESQ » désignant une qualité perçue satisfaisante. Nous avons constaté également que le codeur de SPIHT réalise des résultats excellents de compression et offre plus de flexibilité dans le choix de taux de compression.

Avant de terminer cette conclusion, on peut dire que ce travail nous a apporté d'immenses intérêts tant sur le plan théorique qu'expérimental. En effet, de plus qu'il nous a permis de nous approfondir dans la simulation et la programmation, nous avons été amenés, par ce travail, à nous instruire dans un domaine clé dans les télécommunications et les applications multimédias, c'est le codage et le traitement de la parole.

Il y a plusieurs vois pour continuer ce travail, par exemple l'utilisation du model perceptuel qui permet le masquage de l'information non pertinente pourra optimiser les performances et améliorer la qualité subjective du signal synthétisé. La comparaison des résultats de compression obtenus avec les codeurs de l'état de l'art est très importante et permet de mettre en valeur les résultats obtenus.

BIBLIOGRAPHIE

[1] R. de Boite. Traitement de la parole. *Edition de Presses polytechniques et universitaires romandes*, 2000.

[2] M. Denis et al. La psychologie cognitive. *Éditions de la Maison des sciences de l'homme*, p.143-64.

[3] S. BOUASLI, A. NOUMERI. Compression et codage de la parole par la Transformée KLT. *Mémoire de Master, Université de Khemis Miliana*, 2016.

[4] V. VILAYSOUK. Codage de parole par Transformée pour le développement de Codeurs Parole-audio Unifiés. *Thèse de doctorat. Université de Sherbrooke*. 2015.

[5] M. Rossi. Audio. Edition de *Presses Polytechinques et Universitaires romandes*. 2007

[6] D. Pan. A tutorial on MPEG/audio compression. *Multimedia, IEEE, vol. 2, no 2*, p.p.60–74.1995.

[7] T. PAINTER, A. SPANIAS. Perceptual Coding of Digital Audio. *Proceedings of the IEEE, vol. 88, no. 4*, april, 2000.

[8] Charles, C. (2021). De la transformée de Fourier à la transformée par ondelettes : le succès de l'analyse temps-fréquence. inria.hal.science, 2021.

[9] J. R. Winkler. Orthogonal wavelets via filter banks theory and applications. *The University of Sheffield Department of Computer Science*, 2000.

[10] J.D. Gibson. Speech coding methods, standards, and applications. *IEEE Circuits Syst. Mag.* 5,4, 30–49, 2004.

[11] J. I. Agbinya, Discrete wavelet transform techniques in speech processing, *Proceedings of Digital Processing Applications (TENCON'96)*, 1996, *pp*. 514–519.

[12] S. Joseph, P. Anto, The optimal wavelet for speech compression. *Communications in Computer and Information Science, ACC 2011, Part III, CCIS 192*, 406-414, 2011.

[13] BOUCHEMEL. Contribution à la transmission des images compressées : Application aux systèmes de télécommunications. *Thèse de doctorat*, 2018.

[14] W. Pearlman, A. Said. A new, fast, and efficient image codec based on set partitioning in hierarchical trees. *IEEE Trans. on Circuits and Systems for Video Technology*, 1996

[15] ITU-T, Recommendation P.862. Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. Feb. 2001.

[16] NIST, "The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus". Oct. 1990