

وزارة التعليم العالي والبحث العلمي

BADJIMOKHTAR-ANNABAUNIVERSITY
UNIVERSITE BADJI MOKHTAR ANNABA



جامعة باجي مختار - عنابة

Année : 2017

Faculté: Sciences de l'Ingéniorat
Département: Electronique

MEMOIRE

Présenté en vue de l'obtention du diplôme de : MASTER

Intitulé
Codage et compression du signal de la parole

Domaine : Sciences et Technologie
Filière : Electronique
Spécialité: Télécommunication avancée

Par : Mr Benhadid Mohamed Brahim

DEVANT Le JURY

Président: Mr .K.Saouchi

MCAU. Annaba Directeur de

mémoire:Mr. Bouchaala Ali

MAAU. Annaba

Examineur : Mme. Zermi.

NMCB.U. Annaba

REMERCIEMENTS

Louange à Allah, Seigneur de l'univers, le tout Puissant, le Miséricordieux, qui me protège, m'inspire et me comble de ses bienfaits. Je lui en rends toujours grâce.

La réalisation de ce mémoire a été possible grâce au plusieurs personnes à qui je voudrais témoigner toute ma reconnaissance.

*Je tiens à remercier chaleureusement mon directeur de mémoire, madame **BOUCHAALA ALI** Maître Assistance classe A à l'université d'Annaba, pour son grand soutien dans la réalisation de ce travail, par son soutien moral et technique ainsi que par sa disponibilité ; tous ces facteurs ont grandement contribué à l'aboutissement de ce projet, qu'il soit assuré de ma reconnaissance et mon respect.*

*Mes remerciements s'adressent à l'ensemble des membres de mon jury d'avoir accepté de juger et d'évaluer ce modeste travail. Je remercie sincèrement monsieur **SAOUCHI KADOUR** Maître de Conférences classe A à l'université d'Annaba, de m'avoir fait l'honneur de présider mon jury ; qu'il trouve ici l'expression de mon profond respect. Je remercie également monsieur **ZERMI NARIMA** Maître Assistant classe A à l'université d'Annaba d'avoir accepté de me donner un peu de son précieux temps pour faire part de mon jury en tant qu'examineur.*

Je tiens également à exprimer mes remerciements et ma reconnaissance ;

- *Mon père et ma mère de son motivation et l'encouragement*
- *à tous les membres du laboratoire **LERICA** (département d'Electronique), que j'ai eus le plaisir de côtoyer durant ces années,*
- *à tous ceux qui ont participé de près ou de loin à l'élaboration de ce mémoire.*

REMERCIEMENTS

Sommaire

Chapitre 1

Description et caractéristique du signal de la parole

Introduction.....	Erreur ! Signet non défini.
1.. Production de la parole	Erreur ! Signet non défini.
1.1. Le processus de production	Erreur ! Signet non défini.
1.2. Les différentes étapes de production de la parole....	Erreur ! Signet non défini.
1.3. Les organes de production de la parole	Erreur ! Signet non défini.
1.3.1. Le larynx.....	Erreur ! Signet non défini.
1.3.2. Les muscles du larynx.....	Erreur ! Signet non défini.
2. Principe de perception auditive	Erreur ! Signet non défini.
3. caractéristique de signal de parole.....	Erreur ! Signet non défini.
3.1. La variabilité du signal de parole.....	Erreur ! Signet non défini.
3.1.1. Variabilité intra-locuteur	Erreur ! Signet non défini.
3.1.2. Variabilité interlocuteur	Erreur ! Signet non défini.
4. Redondance du signal de parole.....	Erreur ! Signet non défini.
4.1. Les sons voisés	Erreur ! Signet non défini.
5. Traits acoustiques :	Erreur ! Signet non défini.
5.1. La fréquence fondamentale	Erreur ! Signet non défini.
6. Formants :	Erreur ! Signet non défini.
6.1. L'énergie	Erreur ! Signet non défini.
6.2. Le timbre	Erreur ! Signet non défini.
6.3 Intensité	Erreur ! Signet non défini.
7. Taux de passage par zéro	Erreur ! Signet non défini.

Chapitre2

Etat de l'art

Compression et Formats

La compression des données : Etat de l'art.	Erreur ! Signet non défini.
Introduction :	Erreur ! Signet non défini.
1-2- Principaux types de compression:	Erreur ! Signet non défini.

Sommaire

1.2.1 La compression sans perte:.....	Erreur ! Signet non défini.
1.2.2 le codage huffman :	Erreur ! Signet non défini.
1.2.3La compression avec perte :	Erreur ! Signet non défini.
2. Les formats de fichiers de son.....	Erreur ! Signet non défini.
2.1. Les Codecs sans Perte :	Erreur ! Signet non défini.
2.2. Les Codecs avec Perte.....	Erreur ! Signet non défini.
3. Plus de détails sur quelques Formats Audio :.....	Erreur ! Signet non défini.
3.1. WAV : waveform audio format.....	Erreur ! Signet non défini.
3.2. Le codeur MP3 (MPEG Layer-3)	Erreur ! Signet non défini.
3.4.OGG	Erreur ! Signet non défini.
3.5. AAC	Erreur ! Signet non défini.
3.6. FLAC.....	Erreur ! Signet non défini.

Chapitre3

Théorie de la transformation en ondelettes

Introduction.....	Erreur ! Signet non défini.
1. Définition	Erreur ! Signet non défini.
1.1 De La Transformée de Fourier à la Transformée en Ondelette.....	Erreur ! Signet non défini.
1.2-Historique des ondelettes	Erreur ! Signet non défini.
1.3- La Transformée d'Ondelette	Erreur ! Signet non défini.
1.4-Propriétés :	Erreur ! Signet non défini.
3 .La fonction fenêtre	Erreur ! Signet non défini.
3 . la transformation en ondelettes TOC:.....	Erreur ! Signet non défini.
4. Transformée en Ondelettes Discrète	Erreur ! Signet non défini.
5. .la comparaison entre la TFCT et TOC:	Erreur ! Signet non défini.
6. Les avantages de la transformé en ondelette :	Erreur ! Signet non défini.
7quelques types d'ondelettes	Erreur ! Signet non défini.
7. 1Ondelette de Haar:	Erreur ! Signet non défini.
7.2. Ondelette de Morlet.....	Erreur ! Signet non défini.
7.3. Ondelette de Daubechies :	Erreur ! Signet non défini.

Chapitre IV

Compression de signal de la parole par la transformée en ondelettes

1- Introduction.....	Erreur ! Signet non défini.
1.2 Méthode de compression par ondelettes :	Erreur ! Signet non défini.
1.2. Choix des ondelettes	Erreur ! Signet non défini.
1.3 Choix du niveau de décomposition.....	Erreur ! Signet non défini.
2 .Calcul des coefficients de la transformation en ondelettes	Erreur ! Signet non défini.
2.1 Troncature des coefficients.....	Erreur ! Signet non défini.
2.1.1-Sélection du seuil :.....	Erreur ! Signet non défini.
2-1.2-Seuillage dépendant du niveau.....	Erreur ! Signet non défini.
2-1.3 Les méthodes de seuillage.....	Erreur ! Signet non défini.
2.1.4 . Seuillage dur	Erreur ! Signet non défini.
2.5 Encodage des coefficients d'ondelettes tronqués :	Erreur ! Signet non défini.
2.6 Critères d'évaluation des performances	Erreur ! Signet non défini.
3. Le rapport signal sur bruit.....	Erreur ! Signet non défini.
3.1 Le rapport signal sur bruit de crête (PSNR).....	Erreur ! Signet non défini.
3.1 L'erreur quadratique moyenne à racine normalisée (NRMSE).....	Erreur ! Signet non défini.
3.2 Taux de compression(CR).....	Erreur ! Signet non défini.

Chapitre V

Simulations – résultats et Perspectives

1.premier étape de la simulation :	Erreur ! Signet non défini.
2.deuxièmeétape :	Erreur ! Signet non défini.
3.Troisièmeétape :	Erreur ! Signet non défini.
4 .Quatrième étape :	Erreur ! Signet non défini.
5.Cinquième étape :	Erreur ! Signet non défini.
6. Sixième étape :.....	Erreur ! Signet non défini.
7.Septième étape ;	Erreur ! Signet non défini.
La simulation :	Erreur ! Signet non défini.

Sommaire

Le signal de parole utilisé :	Erreur ! Signet non défini.
L'ensemble des résultats sont illustrés dans le tableau suivant : ..	Erreur ! Signet non défini.
Remarque:	Erreur ! Signet non défini.
2. Conclusion et Perspectives:	Erreur ! Signet non défini.

Listes de figures

Listes des figure

Figure 1:l'appareil phonatoire **Erreur ! Signet non défini.**

Figure 2:schema de larynx..... **Erreur ! Signet non défini.**

Figure 3:les muscles intrinsèques de larynx. **Erreur ! Signet non défini.**

Figure 4:schéma de production et rétroaction auditive **Erreur ! Signet non défini.**

Figure 5:perception et analyse du son par l'être humain. .. **Erreur ! Signet non défini.**

Figure 6:champ auditif humain. **Erreur ! Signet non défini.**

Figure 7:Forme d'onde d'une trame de signal vocal voisé et son spectre d'amplitude.

Les positions du fondamental (F0) et des formants (F1, F2, F3 et F4) sont

indiquées2.4.2.Les sons non voisés..... **Erreur ! Signet non défini.**

Figure 8:Forme d'onde d'une trame de signal vocal non voisé et son spectre d'amplitude. **Erreur ! Signet non défini.**

Figure 9:Densité de probabilité à long terme du signal vocal (trait fin) et densités de probabilité utilisées pour l'approcher : loi de Gauss (trait fort), loi de Laplace (pointillé) et loi Gamma (tirets). **Erreur ! Signet non défini.**

Figure 10 :Densité de probabilité à long terme du signal vocal (trait fin) et densités de probabilité utilisées pour l'approcher : loi de Gauss (trait fort), loi de Laplace (pointillé) et loi Gamma (tirets). **Erreur ! Signet non défini.**

Figure 11:Fonction fenêtre..... **Erreur ! Signet non défini.**

Figure 12:Décomposition d'un signal en approximation et détails **Erreur ! Signet non défini.**

Figure 13:phase de synthèse au niveau 1..... **Erreur ! Signet non défini.**

Figure 14:Réglage du seuil global..... **Erreur ! Signet non défini.**

Figure 15:Exemple de calcul de seuil dépendent du niveau de décomposition

Erreur ! Signet non défini.

Figure 16:signal de parole de 2db **Erreur ! Signet non défini.**

Figure 17:coefficient de détails de décomposition des ondelettes haar et db4 et db8 de 5 eme niveau de décomposition **Erreur ! Signet non défini.**

Figure 18:coefficient de détails de haar et db4 et db8 après l'application de seuillage de valeur 0.3..... **Erreur ! Signet non défini.**

Figure 19:signal de parole original sans compression..... **Erreur ! Signet non défini.**

Figure 20:signal compressé **Erreur ! Signet non défini.**

Listes de figures

Listes des tableaux

Tableau 1:Le spectre de fréquence audible: **Erreur ! Signet non défini.**

Tableau 2:Comparatif entre la taille /qualité des principaux formats audio utilisés.

..... **Erreur ! Signet non défini.**

Introduction générale

Dans de nombreuses applications qui s'occupent de la transmission de données, on utilise des abréviations pour espérer de gagner en rapidité et en flux. En d'autre terme on a recours à l'utilisation d'informations plus compactes pour communiquer.

A l'heure actuelle, si les espaces de stockage ne posent pas de réels problèmes, ce n'est pas le cas des flux de données sur les réseaux ; En effet, les réseaux (Lan, Internet ...) sont de plus en plus utilisés pour échanger des informations. Cela pose cependant un problème important : celui du temps de transmission. Le débit des réseaux étant limité, on cherche à réduire la taille des données à envoyer afin de rendre la transmission plus rapide et par conséquent plus satisfaisante ; Les techniques de compression de données répondent indiscutablement à ce défi.

Pour cela, deux orientations concernant les méthodes de compression sont envisageables :

- La première est dite « **lossless** », c'est à dire **sans perte** et elle est appliquée dans tous les domaines où la perte d'information est inacceptable : Les exécutable, Les textes etc... en sont des exemples.

Elle consiste à analyser des données et à en déduire une écriture de taille inférieure. L'une de ces méthodes est appelée compression de Huffman (inventée par le mathématicien du même nom). Nous allons en faire un aperçu dans la suite du manuscrit.

- La seconde, qui engendre une perte ou une altération de l'information, est utilisée pour tout ce qui concerne l'image et le son. Elle consiste dans un premier temps à enlever toute l'information à laquelle les capteurs humains ne sont pas sensibles (fréquences inaudibles, surfaces insignifiantes ...) puis à écrire une nouvelle donnée qui ne pourra plus reprendre son format (qualité) original. Il y a eu suppression de l'information de manière irréversible. C'est cette forme de compression que nous détaillerons principalement ici. Malgré tout, on est quand même en droit de se poser une question : La compression a-t-elle réellement un avenir, dans ce domaine qu'est l'informatique, lorsque le progrès nous inonde chaque jour de nouvelles technologies de plus en plus performantes? En effet, si l'on revient ne serait-ce que dix ans en arrière, les disques dur atteignaient difficilement 500 ou 600 Mo tandis que les connections internet, aussi rares qu'onéreuses, étaient loin de permettre de

Introduction générale

manipuler de telles données. Dans 10 ans, il ne serait pas étonnant de trouver des disques dur de plusieurs milliers de Go, et des connections permettant de transmettre, via Internet, plusieurs Mo par seconde. La relative perte de qualité occasionnée par la compression pourrait être à l'origine d'un retour à une qualité supérieure quitte à perdre un peu de place.

En résumé, l'objectif de la compression/codage des signaux numériques est la réduction des coûts de transmission et de stockage. Cependant il ne faut perdre de vue qu'un système de compression est toujours le résultat d'un compromis entre quatre critères principaux et qui sont: La qualité du signal restitué, débit, complexité et retard.

La compression des signaux trouve de nombreuses applications telles que la transmission sur téléphonie numérique fixe ou mobile, la transmission sur les réseaux par paquets (internet), la vidéoconférence, la radio et la télévision, ainsi que diverses applications de stockage.

L'objectif de notre sujet de master est l'étude des aspects de la compression du signal de la parole ainsi que l'application et la simulation en Matlab d'algorithmes en se basant sur l'utilisation de la transformation en ondelettes. Des critères d'évaluation ont été adoptés.

Chapitre I : Description et caractéristique du signal de la parole

Introduction

L'information portée par le signal de parole peut être analysée de bien des façons. On distingue, généralement, plusieurs niveaux de description non exclusifs : Acoustique, Phonétique et bien d'autres .

Dans ce chapitre nous allons, dans un premier temps, décrire les processus de production et de perception auditive de la parole puis nous donnerons un aperçu sur les notions de phonétique. Nous terminerons par la conversion de la parole en signal électrique et nous rappellerons quelques outils de base utilisés en traitement de signaux acoustiques.

1.. Production de la parole

La parole peut être décrite comme le résultat de l'action volontaire et coordonnée d'un certain nombre de muscles des appareils respiratoires et articulatoires. 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.

1.1. Le processus de production

De façon simple, on peut résumer le processus de production de la parole comme un système dans lequel une ou plusieurs sources excitent un ensemble de cavités. La source sera soit générée au niveau des cordes vocales soit au niveau d'une constriction du conduit vocal. Dans le premier cas, la source résulte d'une vibration quasi-périodique des cordes vocales et produit ainsi une onde de débit quasi-périodique. Dans le second cas, la source sonore est soit un bruit de friction soit un bruit d'explosion qui peut apparaître s'il y a un fort rétrécissement dans le conduit vocal ou si un brusque relâchement d'une occlusion du conduit vocal s'est produit.

L'ensemble de cavités situées après la glotte, dites les cavités supra glottiques, vont ainsi être excitées par la ou les sources et "filtrer" le son produit au niveau de ces sources. Ainsi, en changeant la forme de ces cavités, l'homme peut produire des sons différents. Les acteurs de cette mobilité du conduit vocal sont communément appelés les articulateurs.

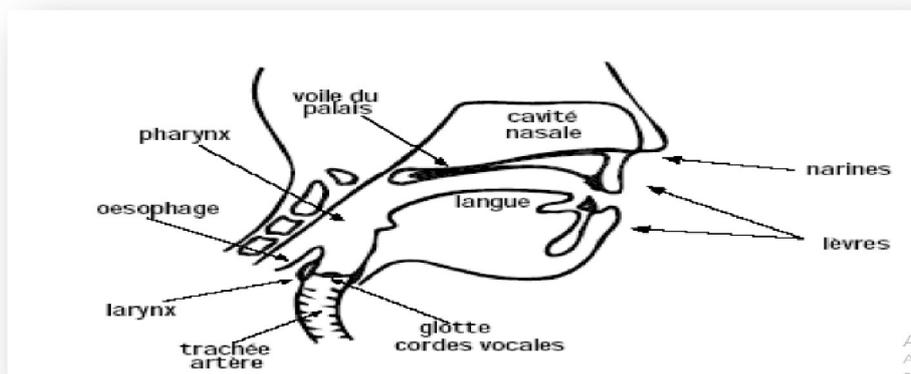


Figure 1:l'appareil phonatoire

On peut donc résumer le processus de production de la parole en trois étapes essentielles. La génération d'un flux d'air qui va être utilisé pour faire naître une source sonore (au niveau des cordes vocales ou au niveau d'une constriction du conduit vocal). C'est le rôle de la soufflerie.

La génération d'une source sonore sous la forme d'une onde quasi-périodique résultant de la vibration des cordes vocales ou/et sous la forme d'un bruit résultant d'une constriction ou d'un brusque relâchement ou occlusion du conduit vocal : C'est le rôle de la source vocale.

La mise en place des cavités supra glottiques (conduits nasal et vocal) pour obtenir le son désiré (c'est principalement le rôle des différents articulateurs du conduit vocal).

1.2. Les différentes étapes de production de la parole

L'appareil respiratoire fournit l'énergie nécessaire à la production de sons, en poussant de l'air à travers la trachée-artère. Au sommet de celle-ci se trouve le larynx où la pression de l'air est modulée avant d'être appliquée au conduit vocal. Le larynx est un ensemble de muscles et de cartilages mobiles qui entourent une cavité située à la partie supérieure de la trachée. Les cordes vocales sont en fait deux lèvres symétriques placées en travers du larynx. Ces lèvres peuvent fermer complètement le larynx et, en s'écartant progressivement, déterminer une ouverture triangulaire appelée glotte. L'air y passe librement pendant la respiration et la voix chuchotée,

ainsi que pendant la phonation des sons non voisés. Les sons voisés résultent au contraire d'une vibration périodique des cordes vocales.

Le larynx est d'abord complètement fermé, ce qui accroît la pression en amont des cordes vocales, et les force à s'ouvrir, ce qui fait tomber la pression, et permet aux cordes vocales de se refermer ; des impulsions périodiques de pression sont ainsi appliquées, au conduit vocal, composé des cavités pharyngienne et buccale pour la plupart des sons. Lorsque la luette est en position basse, la cavité nasale vient s'y ajouter en dérivation.

Dans la suite de cette section, nous allons définir au mieux les organes intervenants dans ce processus.

1.3. Les organes de production de la parole

La parole est essentiellement produite par deux types de sources vocales. La première, plus sonore, est celle qui prend naissance au niveau du larynx suite à la vibration des cordes vocales. La seconde, moins sonore, prend naissance au niveau d'une constriction du conduit vocal ou lors d'un relâchement brusque d'une occlusion du conduit vocal. On parlera dans ces cas de sources de bruit.

1.3.1. Le larynx

Le larynx est un organe situé dans le cou qui joue un rôle crucial dans la respiration et dans la production de parole. Le larynx est plus spécifiquement situé au niveau de la séparation entre la trachée artère et le tube digestif, juste sous la racine de la langue. Sa position varie avec le sexe et l'âge : il s'abaisse progressivement jusqu'à la puberté et il est sensiblement plus élevé chez la femme.

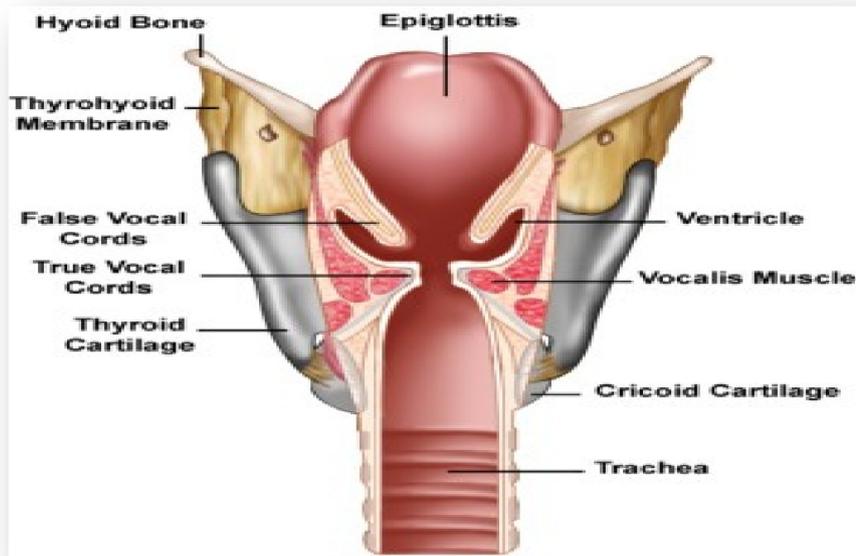


Figure 2:schéma de larynx.

Il est constitué d'un ensemble de cartilages, il est constitué d'un ensemble de cartilages entourés de tissus mous. La partie la plus proéminente du larynx est formée de la thyroïde. La partie antérieure de cartilage est communément appelée la "pomme d'Adam". On trouve, juste au-dessus du larynx, un os en forme de 'U' appelé l'os hyoïde. Cet os relie le larynx à la mandibule par l'intermédiaire de muscles et de tendons qui joueront un rôle important pour élever le larynx pour la déglutition ou la production de parole.

La partie inférieure du larynx est constituée d'un ensemble de pièces circulaires, le

cricoïde, sous lequel on trouve les anneaux de la trachée artère.

Le larynx assure ainsi trois fonctions essentielles :

- _ Le contrôle du flux d'air lors de la respiration
- _ La protection des voies respiratoires
- _ La production d'une source sonore pour la parole

I.4. la modélisation globale passe par une modélisation élémentaire de chaque étage de l'organe phonatoire à savoir :

- La source d'excitation (la glotte).

- Le conduit vocal.
- Le rayonnement des lèvres.
- La conduite nasale.

Le modèle global de la production de la parole est schématisé par la figure suivante :

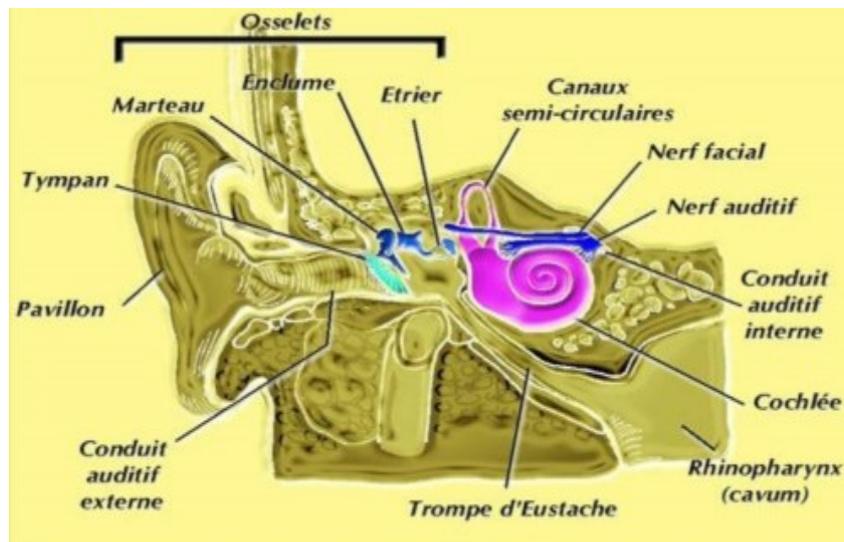
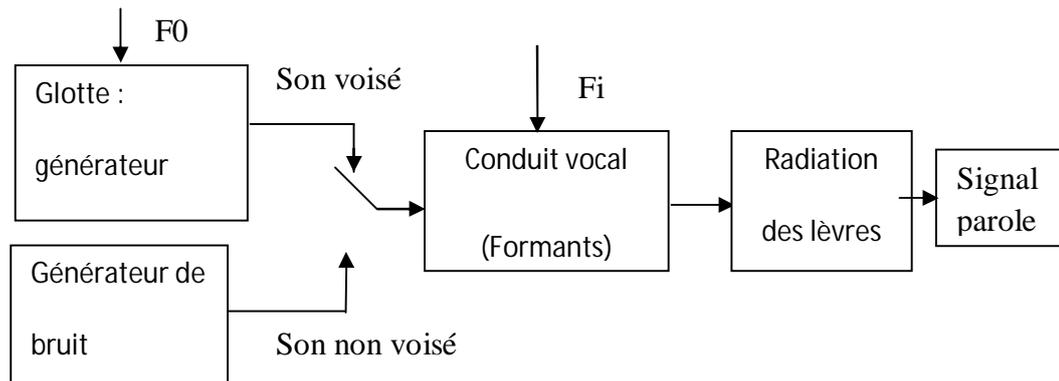


Figure 3: Représentation fonctionnelle du modèle de production de la parole

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

envoi 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

1.3.2· Les muscles du larynx

Les mouvements du larynx sont contrôlés par deux groupes de muscles. On distingue Ainsi les muscles intrinsèques, qui contrôlent le mouvement des cordes vocales et des muscles à l'intérieur du larynx, et les muscles extrinsèques, qui contrôlent la position du larynx dans le cou.

La figure 3 nous représente les muscles intrinsèques. Les cordes vocales sont ouvertes par une paires de muscles (les muscles cricoaryténoïde postérieur) qui sont situés entre la partie arrière du cricoïde et le cricoaryténoïde.

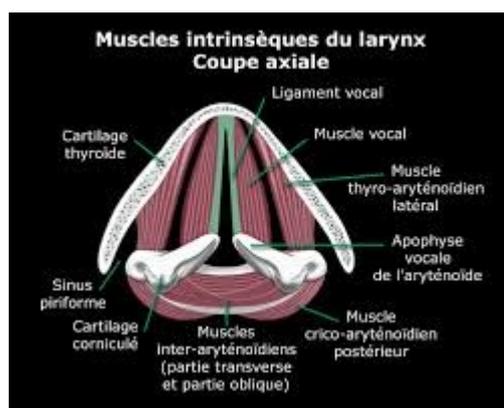


Figure 4:les muscles intrinsèques de larynx.

2. Principe de perception auditive

La parole peut 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 4.

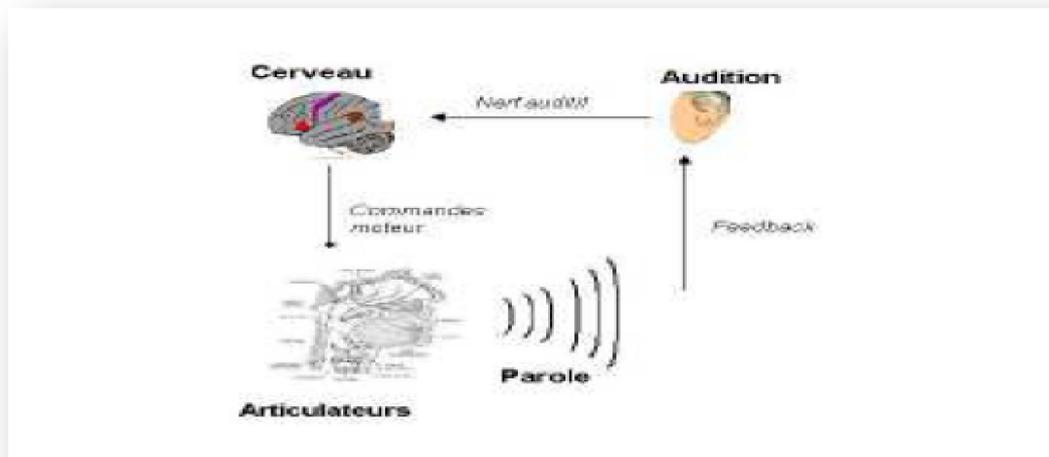


Figure 5: schéma de production et rétroaction auditive

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 5.

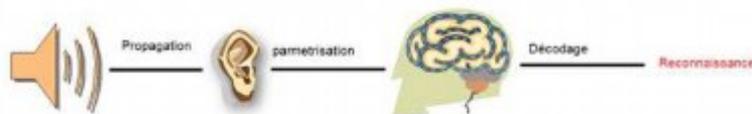


Figure 6: perception et analyse du son par l'être humain.

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 1.9 présente le champ auditif humain, délimité par la courbe de seuil de l'audition et celle du seuil de la douleur. Sa limite supérieure en fréquence (~16000 Hz, variable selon les individus) fixe la fréquence d'échantillonnage maximale utile pour un signal auditif (~32000 Hz).

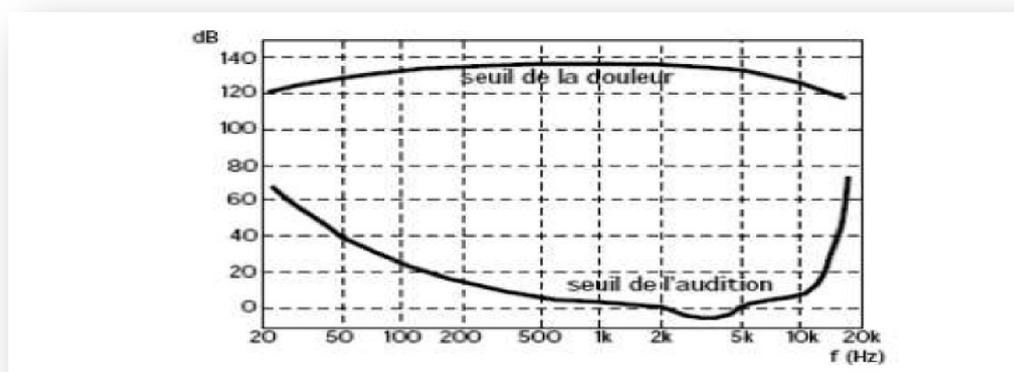


Figure 7: champ auditif humain.

A l'intérieur de son domaine d'audition, l'oreille ne présente pas une sensibilité identique à toutes les fréquences. La figure 7, 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.

Après avoir décrit les différentes étapes dans la genèse des phonèmes de la parole, Nous allons exposer les caractéristiques spectrales, en intensité ainsi que le débit de parole à long terme.

Tableau 1:Le spectre de fréquence audible:

Nom	Bande passante	Commentaires
Fréquences subsoniques	1Hz - 20Hz	Ne sont pas perçues par l'oreille humaine. Sont produites par exemple par les tremblements de terre et par les gros orgues à canne des églises.
Très basses fréquences	20Hz-40Hz	C'est l'octave la plus basse que l'oreille est en mesure d'entendre. Entrent dans cette zone les harmoniques plus basses de la grosse caisse de la batterie et les notes basses du piano, ainsi que le bruit du tonnerre et de la climatisation.
Basses fréquences	40Hz - 160Hz	Presque toutes les fréquences basses de la musique appartiennent à cette zone.
Fréquences moyennes basses	160Hz - 315Hz	Le Do central du piano (261 Hz) entre dans cette zone. Cette zone contient plusieurs informations du signal sonore qui peut être gravement altéré par un usage erroné de l'égaliseur.
Fréquences moyennes	315Hz - 2.5KHz	L'oreille est sensible à cette zone. Cette bande, prise à part, restitue un son similaire en qualité à celui du téléphone.
Fréquences moyennes hautes	2.5KHz - 5KHz	La courbe iso phonique atteint dans cette zone son point culminant, c'est donc là où l'oreille est le plus sensible. Egaliser dans cette zone un instrument en

		augmente la présence dans le mix, le mettant au premier plan par rapport aux autres.
Hautes fréquences	5KHz - 10KHz	C'est la zone où on perçoit la brillance parce qu'elle contient plusieurs des harmoniques des notes produites dans les bandes précédentes. L'énergie acoustique contenue dans cette zone est très limitée. On trouve dans cette zone les consonnes "s", "t" et "c".
Très hautes fréquences	10KHz - 20KHz	Encore moins d'énergie acoustique dans cette zone. Sont seulement présentes les harmoniques plus hautes de certains instruments. Cependant si on élimine cette bande (par exemple au moyen d'un filtre) un mix deviendrait opaque.

3. caractéristique de signal de parole

3.1. La variabilité du signal de parole

À contenu phonétique égal, le signal vocal est très variable pour un même locuteur (variabilité intra locuteur) ou pour des locuteurs différents (variabilité interlocuteur).

3.1.1. Variabilité intra-locuteur

Une même personne ne prononce jamais un mot deux fois de façon identique. La vitesse d'élocution en détermine la durée. Toute affection de l'appareil phonatoire peut altérer la qualité de la production. Un rhume teinte les voyelles nasales; une simple fatigue et l'intensité de l'onde sonore fléchit, l'articulation perd de sa clarté. La diction évolue dans le temps: l'enfance, l'adolescence, l'âge mûr, puis la vieillesse, autant d'âges qui marquent la voix de leurs sceaux.

3.1.2. Variabilité interlocuteur

Est encore plus flagrante. Les différences physiologiques entre locuteurs, qu'il s'agisse de la longueur du conduit vocal ou du volume des cavités résonnantes, modifient la production acoustique. En plus, il y a la hauteur de la voix, l'intonation et

l'accent différent selon le sexe, l'origine sociale, régionale ou nationale. Enfin toute parole s'inscrit dans un processus de communication où entrent en jeu de nombreux éléments comme le lieu, l'émotion, l'intention, la relation qui s'établit entre les interlocuteurs. Chacun de ces facteurs détermine la situation de communication, et influe à sa manière sur la forme et le contenu du message.

La variabilité des sons de parole est en partie attribuée au phénomène de coarticulation que nous aborderons plus longuement par la suite. La coarticulation rend compte d'un chevauchement des gestes articulatoires sur l'axe temporel. Les indices acoustiques vont donc être distribués sur le signal. Au niveau des traits phonétiques, on parlera d'assimilation.

Ces trois caractéristiques, directionnalité, continuité et variabilité, compliquent la reconnaissance des mots parlés. Pour que les mots soient correctement identifiés, deux problèmes majeurs doivent être résolus : la segmentation et la catégorisation.

3.1.3 La continuité

La production d'un son est fortement influencée par les sons qui le précèdent et le suivent en raison de l'anticipation du geste articulatoire. L'identification correcte d'un segment de parole isolé de son contexte est parfois impossible. Évidemment il est plus simple de reconnaître des mots isolés bien séparés par des périodes de silence que de reconnaître la séquence de mots constituant une phrase. En effet, dans ce dernier cas, non seulement la frontière entre mots n'est plus connue mais, de plus, les mots deviennent fortement articulés.

4. Redondance du signal de parole

Quiconque a vu une représentation graphique de l'onde sonore a certainement été frappée par le caractère répétitif du signal de parole. Un grossissement à la loupe d'une brève émission de parole donne à voir une succession de figures sonores semblant se répéter à l'excès. Un peu de recul laisse apparaître des zones moins stables qu'il convient, de qualifier de transitoires. Ce qui semblerait de prime abord superflu, s'avère en réalité fort utile. Les répétitions confèrent à ce signal une

robustesse. La redondance le rend résistant au bruit. Dans une certaine mesure, elle fonctionne comme un code correcteur d'erreur, puisqu'un interlocuteur humain sait décrypter un message même s'il est entaché de bruits dus à de possibles interférences.

4.1. Les sons voisés

La hauteur d'un son voisé est liée à la fréquence de vibration des cordes vocales. Cette fréquence est appelée fréquence fondamentale ou pitch. Un son voisé est par définition un signal quasi périodique qui possède un spectre fréquentiel très caractéristique, comme on peut le voir sur l'exemple de la figure 7. La première raie de ce spectre correspond au fondamental (F_0) et les raies suivantes à ses harmoniques (multiples de F_0). L'enveloppe de ces raies possède des maxima locaux appelés formants (localisés en F_i avec $i = 1, \dots, 4$). Si les trois premiers formants sont indispensables pour assurer l'intelligibilité du signal vocal, les formants d'ordres supérieurs jouent quant à eux un rôle important pour la qualité du signal vocal. Deux sons de même intensité et de même hauteur se distinguent par le timbre qui est déterminé par les amplitudes relatives des harmoniques du fondamental.

La fréquence fondamentale (F_0) peut varier :

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

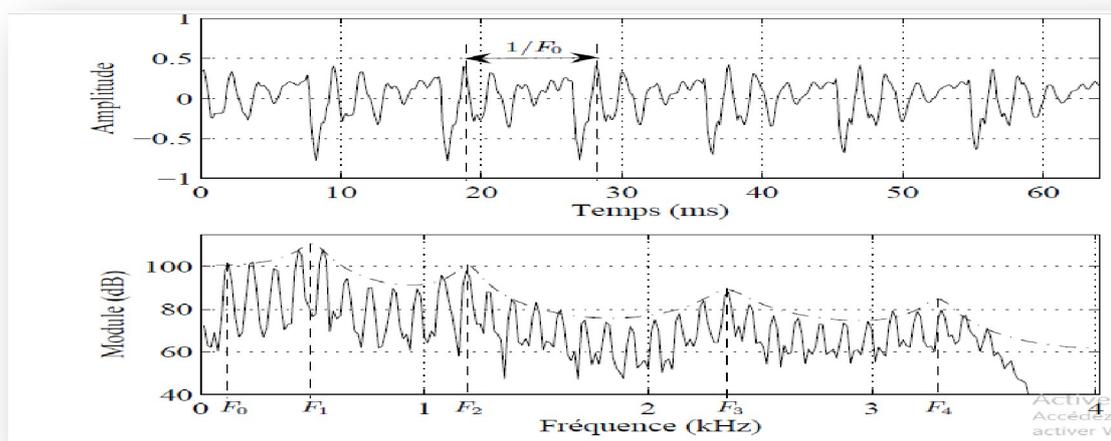


Figure 8 :Forme d'onde d'une trame de signal vocal voisé et son spectre d'amplitude. Les positions du fondamental (F0) et des formants (F1, F2, F3 et F4) sont indiquées.

Un son non voisé ne présente pas de structure périodique, il peut être considéré comme un bruit blanc filtré par le conduit vocal. Son spectre ne possède donc pas de structure particulière et c'est souvent dans les hautes fréquences que le spectre est le plus énergétique comme l'illustre la figure 8

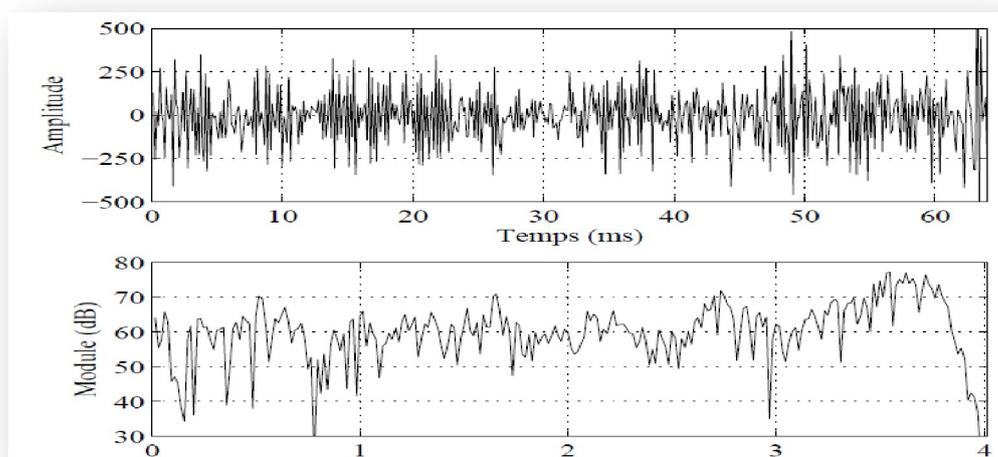


Figure 9 : Forme d'onde d'une trame de signal vocal non voisé et son spectre d'amplitude.

À court terme, le signal vocal peut être considéré comme la réalisation particulière d'un processus aléatoire non-stationnaire. Ses statistiques moyennes doivent être estimées à long terme (au moins plusieurs secondes) et moyennées pour plusieurs locuteurs afin d'être fiables. Le signal de parole étant quasi-stationnaire on peut aussi définir des statistiques à court terme sur la durée d'une trame.

Ces deux types de statistiques sont couramment utilisés dans les techniques de réduction de bruit, les statistiques à court et long terme portant des informations complémentaires.

Pour approcher les statistiques à long terme du signal vocal temporel, les fonctions de densité de probabilité suivantes peuvent être utilisées (la moyenne μ_x est supposée nulle) :

– loi de Gauss :

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{x^2}{2\sigma_x^2}\right), \quad (1.1)$$

– loi de Laplace :

$$p(x) = \frac{1}{\sqrt{2}\sigma_x} \exp\left(-\frac{\sqrt{2}|x|}{\sigma_x}\right), \quad (1.2)$$

– loi Gamma :

$$p(x) = \left(\frac{\sqrt{3}}{8\pi\sigma_x|x|}\right)^{\frac{1}{2}} \exp\left(-\frac{\sqrt{3}|x|}{2\sigma_x}\right), \quad (1.3)$$

Où $2\sigma_x^2$ est la variance de x . **La figure 9** représente ces trois densités de probabilité, dans l'hypothèse d'une moyenne nulle et d'une variance unité, ainsi que la densité de probabilité expérimentale obtenue pour environ 50s de parole (prononcée par 4 locuteurs, 2 femmes et 2 hommes). **La loi Gamma**

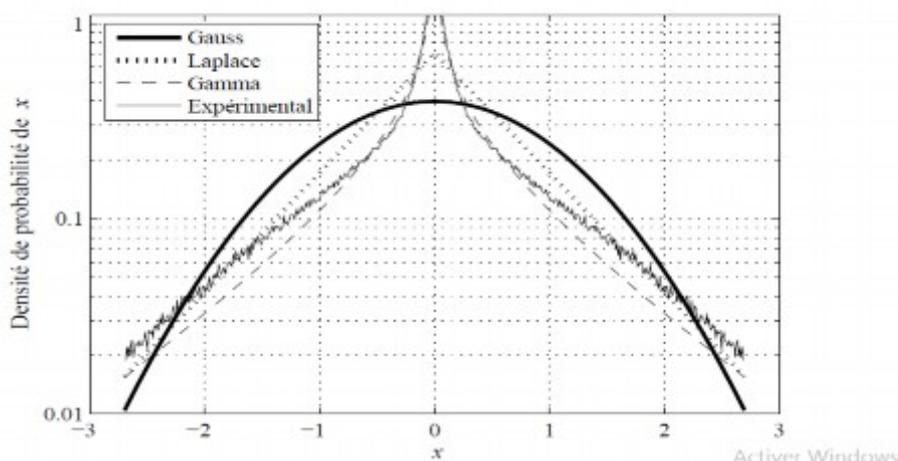


Figure 10 : Densité de probabilité à long terme du signal vocal (trait fin) et densités de probabilité utilisées pour l'approcher : loi de Gauss (trait fort), loi de Laplace (pointillé) et loi Gamma (tirets).

Approche fidèlement la loi expérimentale et la loi de Laplace en est aussi relativement proche.

Cependant, dans la pratique, la loi de Gauss est très souvent choisie à cause des nombreuses simplifications que son utilisation apporte, et ce bien que cette loi soit relativement éloignée de la réalité. On peut noter que la valeur importante en $x = 0$ de

la densité de probabilité à long terme du signal vocal s'explique par la présence de silence dans la parole continue

5. Traits acoustiques :

Le signal de parole est un vecteur acoustique porteur d'informations d'une grande Complexité, variabilité et redondance..... Les caractéristiques de ce signal sont appelées « Traits acoustiques ». Chaque trait acoustique a une signification sur le plan perceptuel.

Parmi ces traits on cite la fréquence fondamentale, fréquence de vibration des cordes Vocales. Ses variations définissent le pitch qui constitue la perception de la hauteur (Ou les sons s'ordonnent de grave à aigu). Seuls les sons quasi-périodiques (voisés) engendrent une sensation de hauteur tonale bien définie.

Autre trait, le spectre fréquentiel dont dépend principalement le timbre de la voix. Le timbre est une caractéristique permettant d'identifier une personne à la simple écoute de sa voix. Le timbre dépend de la corrélation entre la fréquence fondamentale et les harmoniques qui sont les multiples de cette fréquence.

Le dernier trait acoustique est l'énergie correspondant à l'intensité sonore. Elle est Habituellement plus forte pour les segments voisés de la parole que pour les segments non voisés.

Les traits acoustiques du signal de parole sont liés à sa production.

5.1. La fréquence fondamentale

Une analyse d'un signal de parole n'est pas complète tant qu'on n'a pas mesuré l'évolution temporelle de la fréquence fondamentale ou pitch. Physiologiquement, elle correspond à la fréquence laryngienne produite par les vibrations des cordes vocales. C'est la fréquence de vibration des cordes vocales.

Pour les sons voisés, la fréquence fondamentale correspond à la fréquence du cycle D'ouverture/fermeture des cordes vocales.

La fréquence fondamentale (F_0) peut varier :

- de 80 à 200 Hz pour une voix masculine,
- de 150 à 450 Hz pour une voix féminine,

– de 200 à 600 Hz pour une voix d'enfant.

6. Formants :

Ce sont les fréquences de résonance au niveau du conduit vocal supposé comme un banc de filtres. Ils sont observés par les pics de l'enveloppe spectrale donnée par le spectre du signal.

6.1. L'énergie

Le dernier trait acoustique est l'énergie correspondant à l'intensité sonore. L'énergie de la parole est liée à la pression de l'air en amont du larynx. Elle est habituellement plus forte pour les segments voisés de la parole que pour les segments non voisés. Chaque trait acoustique est intimement lié à une caractéristique perceptuelle.

6.2. Le timbre

Le timbre est une caractéristique permettant d'identifier une personne à la simple écoute de sa voix. Il provient en particulier de la résonance dans la poitrine, la gorge, la cavité buccale et le nez. Le timbre dépend fortement de la corrélation entre la fréquence fondamentale et les harmoniques qui sont les multiples de cette fréquence.

6.3. Intensité

L'intensité d'un son, appelée aussi volume, permet de distinguer un son fort d'un son faible. Elle correspond à l'amplitude de l'onde acoustique. Pour le son, onde de compression, cette grandeur est la pression.

7. Taux de passage par zéro

Le taux de passage par zéro (*zero crossing rate* en anglais) représente le nombre de fois que le signal, dans sa représentation amplitude/temps, passe par la valeur centrale de l'amplitude (généralement zéro). Il est fréquemment employé pour des algorithmes de détection de section voisée/non voisée dans un signal. En effet, du fait de sa nature aléatoire, le bruit possède généralement un taux de passage par zéro supérieur à celui des parties voisées.

Le comptage du nombre de passages par zéro est très simple à effectuer. Dans un premier temps, il faut enlever le décalage d'amplitude (*offset* en anglais), produit

par la majorité des matériels d'acquisition, pour centrer le signal autour de zéro. Ensuite, pour chaque trame, il suffit de dénombrer tous les changements de signe du signal. Pour éliminer certains phénomènes parasites, ont proposé une méthode nommée le *band-crossing*. Un seuil d'amplitude S permet de définir une zone autour du zéro de largeur $2 \times S$ au sein de laquelle les oscillations ne sont pas prises en compte. La formule du *band-crossing* pour chaque fenêtre analysée est donc :

$$\text{BCR}(\text{fenêtre}) = \sum_{n \in \text{fenêtre}} |f(n) - f(n-1)| \text{ avec } f(n) = \begin{cases} 1 & \text{si } n > S \\ f(n-1) & \text{si } -S \leq n \leq S \\ -1 & \text{si } n < -S \end{cases}$$

Équation 1: formule de calcul du Band-Crossing Rate

Cette mesure se montre très intéressante, dans le cadre d'une détection de parole en amont d'un système de reconnaissance, pour la détection de fricative en fin de signal à reconnaître ou d'attaque de plosive.

Chapitre II

Etat de l'art

Compression et Formats

La compression des données : Etat de l'art.

Introduction :

Nous sommes de plus en plus utilisateurs permanents de systèmes qui véhiculent, traitent et stockent des informations de différentes natures sur des supports qui présentent des limites de capacité.

Les besoins actuel dû au développement, nous impose de communiquer en temps réel n'importe où dans le monde entier. La vitesse des communications ne peut pas être augmentée sans l'aide des techniques de compression. Tous les types des signaux intéressants pour les transmissions : les logiciels, les textes, la parole, la musique, les images, doivent être comprimés. La compression ne doit pas conduire à des distorsions saisissables et doit être réalisée avec des taux de compression les plus élevés possibles. Le nombre d'opérations et le volume de mémoire requis par les algorithmes de compression ne doivent pas être très grands.

Quant aux applications elles sont nombreuses et concernent en particulier les réseaux téléphoniques fixes public, la communication avec les mobiles, la radiodiffusion numérique, la transmission sur les réseaux par paquets (internet), la télévision numérique, la compression de données dans les modems haut débit etc.

L'intérêt de la compression de donnée consiste donc à obtenir des fichiers plus légers, afin d'améliorer la vitesse de transfert sur un canal de transmission ou de limiter l'espace de stockage sur un support.

Les domaines de connaissances informatique et mathématiques entrant dans le champ des méthodes de la compression sont nombreux : théorie de l'information, traitement du signal, algèbre linéaire, géométrie fractale, théorie des ondelettes, probabilités, etc.

1-2- Principaux types de compression:

Toutes les méthodes de compression sont dépendantes du type de données à comprimer. Pour différents types de données il faut utiliser différentes méthodes de compression. On a déjà vu que pour les textes ou pour les logiciels il faut utiliser des méthodes de compression sans pertes. Du fait que la musique, la parole ou les images. sont des signaux redondants, On peut donc utiliser des méthodes de compression à pertes.

Le premier système de compression d'un signal musical a été le système NICAM (Nearly Instantaneously Companded Audio Multiplex) à 384 kbit/s par voie promu par la BBC.

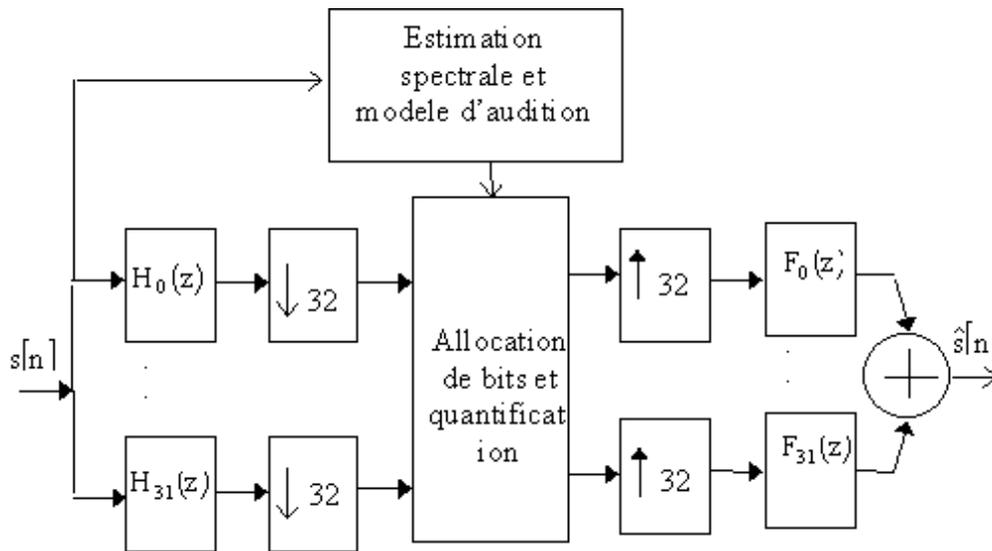


Figure 11 : Schéma de principe du codeur Musicam.

1.2.1 La compression sans perte:

Appelée aussi « compactage », cette solution consiste simplement à coder les données binaires de manière plus concise dans un fichier. Elle permet ainsi de retrouver la totalité des informations après une procédure de décompactage.

Les méthodes ont été établies dans le cadre de la théorie de l'information. Elles se basent sur des méthodes de codage qui se classe en deux types :

- le codage de la source d'information (qui transforme la source dans une forme alternative, meilleure pour la transmission ou pour l'enregistrassions)
- le codage de la voie (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 :

- 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 Huffman. 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.

Le codage de Huffman réalise la compression de données en utilisant un nombre inférieur de bits pour la représentation des caractères avec une apparition plus fréquente.

1.2.2 le codage huffman :

L'idée qui préside au **codage de Huffman** est voisine de celle utilisée dans le [code Morse](#) : coder ce qui est fréquent sur peu de place, et coder en revanche sur des séquences plus longues ce qui revient rarement (entropie). En morse le « e », lettre très fréquente, est codé par un simple point, le plus bref de tous les signes.

L'originalité de David A. Huffman est qu'il fournit un *procédé d'agrégation objectif* permettant de constituer son code dès lors qu'on possède les statistiques d'utilisation de chaque caractère.

Le Macintosh d'Apple codait les textes dans un système inspiré de Huffman : les 15 lettres les plus fréquentes (dans la langue utilisée) étaient codées sur 4 bits, et la 16^e combinaison était un code d'échappement indiquant que la lettre était codée en ASCII sur les 8 bits suivants. Ce système permettait une compression des textes voisine en moyenne de 30 % à une époque où la mémoire était extrêmement chère par rapport aux prix actuels (compter un facteur 1 000).

Le défaut du codage Huffman est qu'il doit connaître la fréquence des caractères utilisés dans un fichier avant de choisir les codes optimaux. Et il doit donc **lire tout le fichier** avant de comprimer. Une autre conséquence est que pour décompresser il faut connaître les codes et donc la table, qui est ajoutée devant le fichier, aussi bien pour transmettre que stocker, ce qui diminue la compression, surtout pour les petits fichiers. Ce problème est éliminé par le **codage Huffman adaptatif**, qui modifie sa table au fil des choses. Et peut donc démarrer avec une table de base. En principe il commence avec les caractères à même probabilité.

1.2.3 La compression avec perte :

Elle concerne essentiellement le domaine du multimédia (image, son, vidéo) ; elle consiste en une « réduction » de l'information basée sur notre propre limite de perception.

Dans le domaine audible, puisque l'oreille n'est pas sensible à toutes les fréquences audio (seulement entre 20Hz et 20KHz), et que sa sensibilité varie selon certaines d'entre elle, il est possible de réduire la quantité de données de telle sorte que le résultat soit très ressemblant à l'original, voire identique, pour l'oreille humaine. **La réduction de la redondance serait la loi générale de ce type de compression de données.**

Il y a plusieurs méthodes pour la compression **avec pertes** de données parmi lesquelles on cite les méthodes:

- par codage à pas variable,
- par modulation différentielle en impulsions et codage,
- par codage en sous-bandes,
- par transformée orthogonale (utilisation de l'ondelette.....).

2. Les formats de fichiers de son

Une multitude de formats audio existent à l'heure actuelle. Les principaux formats sont les formats MP3 (MPEG-2 Audio Layer III) et WAV. Le type de format correspond habituellement à l'extension du fichier (c.-à-d. les lettres qui se trouvent après le point dans le nom du fichier, par exemple .mp3, .wav, .ogg, .wma).

Un **codec** est un algorithme de codage et de compression de données vers un format audio. Certains types de fichiers ont un codec spécifique qui leur est associé. Par exemple, le format MP3 utilise toujours le codec MPEG Layer-3, tandis que le format MP4 peut utiliser toute une série de codecs.

Les notions de codec et de format sont le plus souvent interchangeables, plus particulièrement lorsqu'un format utilise toujours un seul et même codec. Cependant, il est nécessaire de bien comprendre la différence entre un format et un codec. En termes simplifiés, un format peut être comparé à un conteneur dans lequel un signal audio ou vidéo peut être stocké à l'aide d'un certain codec.

Certains formats tels que MP4 ou FLV peuvent stocker à la fois des flux audio et vidéo. Si vous ignorez quel programme utiliser pour ouvrir tel ou tel format audio, nous vous recommandons

d'utiliser notre convertisseur audio. Celui-ci supporte quasiment l'ensemble des formats existant actuellement.

On distingue deux types de codecs en fonction du type de compression:

2.1. Les Codecs sans Perte :

Ce groupe de formats enregistre et compresse un son de façon à conserver la qualité originale intacte lors du décodage.

Les formats de codage sans perte les plus courants sont:

- FLAC (Free Lossless Audio Codec),
- APE (Monkey's Audio),
- ALAC (Apple Lossless Audio Codec).

2.2. Les Codecs avec Perte

Lorsqu'un son est compressé avec perte, il subit une certaine modification. La compression peut par exemple couper les fréquences de son non perceptibles par l'oreille humaine. Lors du décodage, le fichier sera différent de l'original par rapport aux informations qui y sont stockées, mais sera quasiment semblable à l'écoute.

Les formats avec perte les plus courants sont:

- MP3
- WMA
- OGG
- AAC

3. Plus de détails sur quelques Formats Audio :

3.1. WAV : waveform audio format

C'est l'un des tout premiers formats audio. Il est principalement utilisé pour stocker des pistes audio non compressées (PCM) comparables aux CD audio en termes de qualité. En moyenne, une minute de son WAV formaté nécessite environ 10 mégaoctets de mémoire. Les CD sont généralement numérisés en format WAV et peuvent être ensuite convertis en MP3 à l'aide d'un convertisseur audio.

3.2. Le codeur MP3 (MPEG Layer-3)

C'est le format audio le plus répandu au monde ; c'est pourquoi nous allons un peu s'attarder sur ce paragraphe.

Le MP3, comme de nombreux autres formats avec perte, compresse la taille du fichier en coupant les sons non perceptibles par l'oreille humaine. À l'heure actuelle, le MP3 ne constitue pas le meilleur format existant en ce qui concerne le rapport taille du fichier et qualité du son. Cependant, étant donné qu'il s'agit du format le plus répandu et le plus compatible avec la majorité des appareils, de nombreuses personnes préfèrent sauvegarder leurs enregistrements sous ce format.

Cette méthode utilise une transformée en cosinus discrète modifiée qui implémente un banc de filtres.

Le résultat est la transparence de la compression des signaux de musique stéréo d'un CD à un facteur de compression de 12 (le débit à l'entrée est de 1411,2 kbit/s et le débit à la sortie est de 112-128 kbit/s).

En 1987, IIS a commencé à travailler au codage perceptuel dans le cadre du programme Eureka, le projet EU 147, Digital Audio Broadcasting (DAB).

L'algorithme créé a été normalisé comme ISO-MPEG Audio Layer-3 (IS 11172-3 et IS 13818-3). Sans compression, les signaux numériques audio sont des représentations des signaux audio analogiques à 16 bit/échantillon. La fréquence d'échantillonnage est de 44,1 kHz pour les CD. Donc on a plus de 1.400 Mbit pour représenter une seule seconde de musique stéréo à la qualité d'un CD. En utilisant le codage audio MPEG on peut obtenir un facteur de compression de 12 sans perdre la qualité du signal. Des facteurs de compression de 24 ou plus conduisent aux reconstructions (après la compression) qui sont meilleur que les reconstructions réalisées dans les cas où seulement la fréquence d'échantillonnage ou la résolution ont été réduites. On peut réaliser ça par codage perceptuel. Celui-ci est basé sur un modèle de perception de l'oreille. En utilisant le codage MPEG audio, on peut obtenir un facteur de compression de: 4 par la couche 1 (384 kbits/s pour un signal stéréo), 6 à 8 par la couche 2 (256 à 192 kbits/s), 10 à 12 par la couche 3 (128 à 112 kbits/s).

3.3. WMA (Windows Media Audio)

Est un format détenu par Microsoft Corporation. Il fut à l'origine présenté pour remplacer le MP3 grâce à ses fonctionnalités de compression plus élevées. Ce fait a toutefois été discrédité par certains tests indépendants. Le format WMA supporte également la protection des données via DRM.

3.4. OGG

Est un format ouvert supportant le codage audio par divers codecs. Le Codec Vorbis est le codec le plus utilisé avec OGG. La qualité de la compression peut être comparée à celle du MP3, mais il n'est pas supporté par un aussi grand nombre de lecteurs et appareils audio.

3.5. AAC

C'est un format audio breveté possédant des capacités plus élevées (nombre de canaux, fréquences non audibles) que le MP3. Il offre en général une meilleure qualité audio tout en ayant une taille de fichier identique. AAC est actuellement l'un des algorithmes de codage avec perte ayant la meilleure qualité. Un fichier codé avec ce format peut avoir les extensions suivantes: .aac, .mp4, .m4a, .m4b, .m4p, .m4r.

3.6. FLAC

C'est un format avec perte utilisé couramment. Il ne modifie pas le flux audio et le son codé avec ce format est identique à l'original. Il est le plus souvent utilisé pour la lecture audio sur les systèmes audio de pointe. Sa compatibilité de lecture sur les appareils et lecteurs est limitée. Dès lors, si souhaité, il est généralement converti en un autre format avant de le lire sur un lecteur.

A titre indicatif voici un tableau comparatif entre les différents formats de son :

Tableau 1:Comparatif entre la taille /qualité des principaux formats audio utilisés.

Format	Taille	Qualité
AAC	3,99 Mo	128Kbps
VORBIS	3,75 Mo	128Kbps
WMA	3,9 Mo	128Kbps
MP3	5,2 Mo	128Kbps
AIFF	43,5 Mo	1411Kbps(CD)
APPLE LOSSLESS	25,4 Mo	841Kbps
FLAC	26 Mo	CD
WAV	42,5 Mo	CD

Chapitre III
Théorie de la transformation en ondelettes

Introduction

D'abord, pourquoi a-t-on besoin des transformées, ou encore qu'est-ce une transformée ?

Les transformations mathématiques sont appliquées aux signaux bruts pour obtenir davantage d'informations qui sont disponibles dans ces signaux. Dans la suite, le signal dans le domaine temporel s'appellera le signal brut, et un signal transformé par une transformation mathématique sera un signal traité.

Il existe un grand nombre de transformations qui peuvent s'appliquer à un signal. Parmi celles-ci la transformée de Fourier. Dans la pratique, la plupart des signaux sont des signaux dépendant du temps (du domaine Temporel) sous leur format brut. La représentation du signal est une représentation temps - amplitude. Cette représentation n'est pas toujours la meilleure pour la plupart des applications de traitement du signal. Dans beaucoup de cas, l'information la plus pertinente est cachée dans la composante de fréquence du signal. Le SPECTRE de fréquence d'un signal est constitué par les composantes de fréquence de ce signal. Le spectre de fréquence d'un signal indique quelles sont les fréquences qui existent dans le signal.

1. Définition

1.1 De La Transformée de Fourier à la Transformée en Ondelette

La plupart des signaux ne sont pas stationnaires, et l'essentiel de l'information qu'ils contiennent réside dans ce non stationnarité. L'analyse de Fourier propose une approche globale du signal du signal. Toute notion temporelle dans l'espace de Fourier (espace fréquentiel) disparaît. Il faut trouver une transformation qui nous renseigne sur le contenu fréquentiel tout en préservant la localisation afin d'obtenir une représentation temps / fréquence. Plusieurs solutions ont été proposées. Ces solutions sont : la transformée de Fourier à fenêtre glissante, et la transformée de Gabor. Mais ces deux méthodes donnent une même résolution temporelle pour les hauts et les basses fréquences. Donc l'analyse n'est pas idéale.

C'est dans ce contexte qu'intervient la transformée en Ondelettes qui propose une solution de compromis entre la résolution temporelle et la résolution fréquentielle.

1.2-Historique des ondelettes

L'analyse par transformée en ondelettes est apparu dans la seconde moitié du vingtième siècle. C'est dans cette période que des équipes de chercheurs ont développé des outils d'analyse comparables à celui de Gabor qui a introduit l'analyse de Fourier à fenêtre en 1946- 1947. En 1975, Jean Morlet introduisait le nom d'ondelettes, pour un outil d'analyse appliqué aux recherches pétrolières. Ce n'est que dans les années 1980 que Ingrid Daubechies, Stéphane Mallat et Yves Meyer ont formalisé la théorie des ondelettes et développé des algorithmes présentant des propriétés

mathématiques intéressantes, qui les rendent très exploitables dans plusieurs applications du traitement du signal. En 1981, Meyer et **Grossman prouvent la conservation de l'énergie du signal dans sa représentation en ondelettes**. En 1984, Morlet et Grossman parviennent à reconstruire la fonction initiale à partir de sa transformée en ondelettes, sans erreur et avec une intégrale simple. En 1985, Meyer et Grossman créent une base d'ondelettes orthogonales. En 1986, Mallat et Meyer constatent que les algorithmes orthogonaux, le codage en sous bande, les filtres miroirs en quadratures et les ondelettes sont fondés sur le même principe d'analyse: la multirésolution.

En 1989, Meyer et Coiffman construisait le premier algorithme hybride temps – fréquence – échelle: les paquets d'ondelettes. En 1990, Meyer et Coiffman étendent leur découverte et créent les ondelettes de Malvar. En 1990 et après, Daubechies trouve des bases orthonormées d'ondelettes de régularité imposé, à support compact (en prolongement des travaux de Haar, 1910). Il est intéressant de noter que l'analyse par transformation en ondelettes a été en premier lieu pensée et utilisée par des membres de « la recherche appliquée » avant d'être synthétisée, développée et précisée par des acteurs d'une recherche plus fondamentale tel que le traitement des signaux.

1.3- La Transformée d'Ondelette

Une Ondelette mère Ψ est une fonction f de base que l'on va traduire et dilater '6' pour recouvrir le plan temps fréquences et analyser le signal .On peut définir la transformée en Ondelettes

d'un signal $f(t)$ comme une projection sur la base des fonctions Ondelettes $\psi\left(\frac{t-b}{a}\right)$

$$TO(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \cdot \psi\left(\frac{t-b}{a}\right) dt \quad \text{Avec } a, b \in \mathbb{R} \quad , \quad a \neq 0 \quad (3.2)$$

Dans cette expression, a est le facteur d'échelle, b est le paramètre de translation. En posant :

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right)$$

Où $(\psi_{a,b})$ sont une famille d'Ondelettes analysantes générales à partir d'une Ondelette « mère » $\psi(t)$ par dilatation (facteur a) et translation (paramètre b). L'équation devient :

[15, 18, 19].

$$TO(a, b) = \int_{-\infty}^{+\infty} f(t) \cdot \psi_{a,b}(t) \cdot dt = \langle f, \psi_{a,b} \rangle$$

1.4-Propriétés :

La fonction Ondelette doit vérifier un certain nombre de propriétés,

1. condition d'admissibilité :

Soit $\psi(t) \in L^2$ (ensemble des fonctions à carré sommable) [23] alors :

$$\int_{-\infty}^{+\infty} \frac{|\psi(\omega)|}{|\omega|} \cdot d\omega < \infty$$

Cette condition permet d'analyser signal, puis de le synthétiser sans perte d'information. La condition d'admissibilité implique en outre que la transformée de Fourier de l'Ondelette à la Fréquence continu (pour $\omega=0$) doit être nulle. Soit :

$$\psi(\omega) \Big|_{\omega=0} = 0$$

Ceci implique en particulier deux conséquences importantes :

- la première est que les Ondelettes doivent posséder un spectre de type passe-bande
- la seconde apparaît en réécrivant l'équation de façon équivalente sous la forme :

$$\int_{-\infty}^{+\infty} \psi(t) = 0$$

Donc $\psi(t)$ doit être à moyenne nulle. $\psi(t)$ est une fonction à largeur temporelle finie (fenêtre temporelle) possédant un caractère oscillatoire. On est alors bien en présence d'une petite onde : une Ondelette

Cette Ondelette agit comme un filtre passe bande, pour retrouver la partie du spectre éliminée par l'Ondelette (les basses fréquences), on utilise une autre fonction appelée fonction échelle $\varphi(t)$, sa moyenne est non nulle

$$\int_{-\infty}^{+\infty} \varphi(t) dt \neq 0$$

3 .La fonction fenêtre

Une fenêtre est une fonction régulière, qui varie lentement et qui est bien localisée, c.à.d. qui est nulle en dehors d'une certaine zone, qu'on appelle son support. Pour étudier le comportement spectral local d'un signal, il faut donc procéder par portion de temps en multipliant le signal par une fonction fenêtre qui s'annule en dehors de l'intervalle désiré. La fonction fenêtre la plus simple est la fonction rectangulaire qui possède une amplitude fixe .La figure (16) représente une telle fonction :

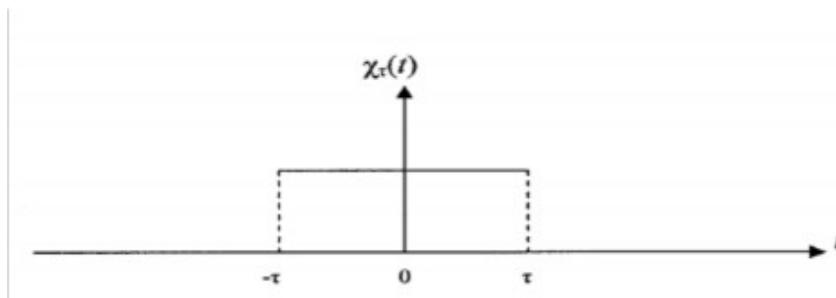


Figure 11: Fonction fenêtre

Soit $\varphi(t) \in L^2(\mathbb{R})$ une fonction fenêtre où $L^2(\mathbb{R})$ est l'espace vectoriel des signaux $f(t)$ continus à énergie finie où t appartient à l'ensemble des réels \mathbb{R} . On pourrait alors obtenir de l'information sur le signal $f(t)$ au voisinage de $t = b$ en considérant le produit $f(t) \varphi(t - b) = f_b(t)$. En particulier si $\varphi(t) = y_x(t)$ tel que représenté dans la figure (16), alors nous pouvons définir l'équation suivante :

$$f_b(t) = \begin{cases} f(t), & t \in [b - \tau, b + \tau] \\ 0 & \text{autrement} \end{cases} \quad \text{(III. 5)}$$

En faisant varier le paramètre b , il sera ainsi possible de déplacer la fenêtre tout au long de l'axe afin d'analyser le comportement local de la fonction $f(t)$ sur différents intervalles. Deux facteurs importants pour une fonction fenêtre sont son centre et sa longueur. Pour la figure précédente, 0 est le centre, 2τ est la longueur. On appellera $f_b(t)$ la fonction "fenêtrée" de $f(t)$.

3. La transformation en ondelettes TOC:

La transformée en ondelettes continue (TOC) :

$$T_x^h(t, a) := \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(s) h^* \left(\frac{s-t}{a} \right) ds.$$

Cette représentation est une fonction du temps t et du paramètre d'échelle a , mais si l'ondelette analysante $h(t)$ vérifie certaines conditions peu contraignantes (elle doit être de valeur moyenne nulle, sa transformée de Fourier doit être uni modale et de fréquence centrale ω_0), elle peut également être vue comme une fonction du temps t et de la pulsation $\omega = \omega_0/a$.

4. Transformée en Ondelettes Discrète

Contrairement à la transformation de Fourier, la transformation en ondelettes continue ne se prête pas aisément à des calculs analytiques simples, même dans le cas où l'ondelette-mère et le signal à analyser sont de formes simples. Le calcul numérique devient indispensable et de ce fait on est mené à travailler avec une transformée en ondelettes discrète qui est générée par échantillonnage dans le plan temps-fréquence de la transformée en ondelette continue. En pratique, on a plus souvent affaire à des signaux discrets, mais même sans cela, on a intérêt à discrétiser les valeurs de a et b . Pour cela, on suggère une méthode de discrétisation exponentielle pour les échelles et pour le temps. Soit $a = a_0^m$ et $b = nb_0^n a_0^m$ avec a_0 et $b_0 \geq 0$. On obtient alors une transformée en ondelette discrète donnée par la formule suivante :

$$Wf(m, n) = a_0^{-m/2} \int_{-\infty}^{+\infty} f(t) \psi(a_0^{-m}t - nb_0) dt \quad m, n \in \mathbb{Z} \quad \text{(III.16)}$$

Si on choisit $a_0 = 2$ et $b_0 = 1$, on parle alors de transformée en ondelettes dyadique qui, conduit à une échelle dyadique formant un réseau appelé réseau d'échantillonnage dyadique. Ce réseau est représenté dans la figure suivante :

5. Transformée de Fourier à court terme TFCT:

La transformée de Fourier à court-terme (TFCT), que l'on définira dans ce chapitre par :

$$F_x^h(t, \omega) := \int_{-\infty}^{\infty} x(s) h^*(s-t) e^{-j\omega s} ds \cdot e^{j\omega t/2},$$

Où la localisation de l'analyse autour de l'instant courant t est obtenue par une fenêtre arbitraire $h(t)$. La présence du facteur $e^{j\omega t/2}$, qui distingue la définition proposée ici de celle utilisée dans le reste de cet ouvrage, et l'utilisation de la pulsation ω au lieu de la fréquence f seront justifiés par la suite, par des arguments de dualité temps-fréquence

5. La comparaison entre la TFCT et TOC :

L'utilisation de l'un ou l'autre de ces deux outils nécessite inmanquablement de trouver un compromis entre les résolutions temporelle et fréquentielle. Pour la TFCT par exemple, une fenêtre d'analyse $h(t)$ plus étroite fournira une meilleure résolution temporelle, mais la concentration autour de l'origine de sa transformée de Fourier $h\omega$ sera nécessairement moindre, ce qui implique une plus mauvaise résolution fréquentielle. La TOC conduit à un compromis similaire, à ceci près qu'il dépend de la fréquence d'analyse : plus la fréquence d'analyse augmente plus la résolution temporelle s'améliore, mais aux dépens d'une moindre résolution fréquentielle.

Les sinusoïdes à modulation linéaire de fréquence (ou "chirps" linéaires) sont des signaux pour lesquels un tel compromis est difficile à trouver, puisque ni la TFCT ni la TOC ne fournissent une localisation parfaite le long de la droite de fréquence instantanée. Ceci résulte bien entendu du fait que la fenêtre d'analyse $h(t)$ ou l'ondelette analysante $\psi(t)$ sont choisies arbitrairement, indépendamment du signal analysé. Si au contraire celles-ci sont déduites du signal analysé, on peut imaginer améliorer la situation. En s'appuyant sur le concept de filtre adapté, un choix naturel pour la TFCT est de prendre $h(t) = x(-t)$, c'est-à-dire la version retournée dans le temps du signal analysé

6. Les avantages de la transformé en ondelette :

La théorie de Fourier repose sur le fait que les fonctions montrant un certain degré de régularité peuvent être représentées par une combinaison linéaire de sinus et cosinus. Les coefficients de cette combinaison linéaire donnent une information au niveau des fréquences présentes dans le signal.

Considérons deux exemples.

Premièrement, les séries de Fourier sont habituellement utilisées pour estimer le spectre des fréquences d'un signal donné en fonction du temps. En médecine, l'électrocardiogramme d'un patient malade diffère de celui d'un patient sain. Cette différence, parfois très difficile à repérer lorsque l'électrocardiogramme est donné en fonction du temps, devient évidente lorsque celui-ci est donné en fonction des fréquences, c'est-à-dire lorsqu'on regarde ses coefficients de Fourier. Mais les séries de Fourier donnent la quantité de chaque fréquence présente dans le signal pour l'ensemble de la période d'observation. La théorie de Fourier devient donc inefficace pour un signal dont le spectre des fréquences varie considérablement dans le temps.

Deuxièmement, les séries de Fourier sont aussi utilisées pour approximer des fonctions. Certaines fonctions régulières ont une série de Fourier économique, autrement dit elles sont bien approximées avec peu de coefficients de Fourier. Mais comme les fonctions sinus et cosinus ont un support infini, la série de Fourier ne fonctionne pas bien quand elle doit décrire localement une fonction qui montre des discontinuités. Nous verrons plus tard que, contrairement à la transformée de

Fourier, l'analyse par ondelettes offre une large gamme de fonctions de base parmi lesquelles on peut choisir la plus appropriée pour une application donnée.

7. quelques types d'ondelettes :

7. 1. Ondelette de Haar:

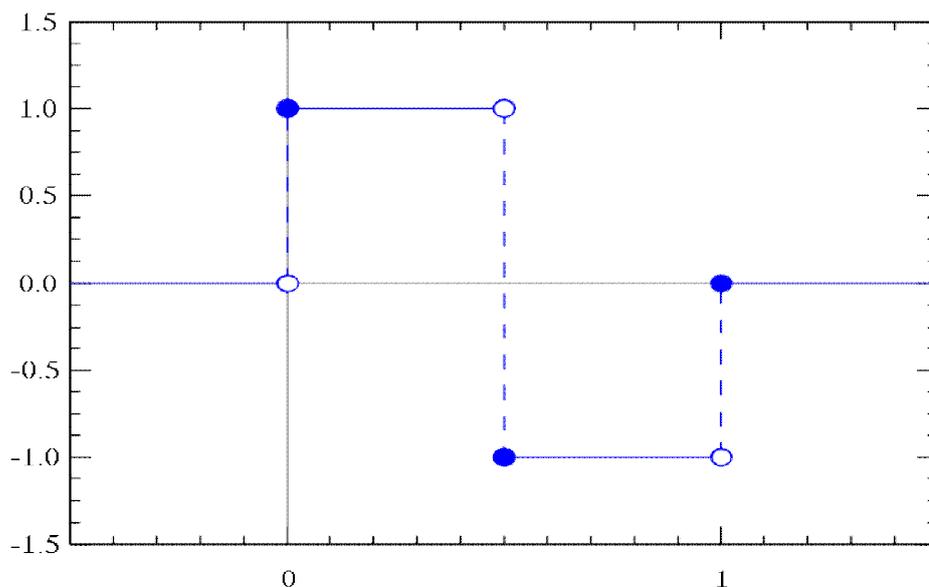
L'ondelette de Haar, ou fonction de Rademacher, est une ondelette créée par Alfréd Haar en 1909¹. On considère que c'est la première ondelette connue. Il s'agit d'une fonction constante par morceaux, ce qui en fait l'ondelette la plus simple à comprendre et à implémenter. L'ondelette de Haar peut être généralisée par ce qu'on appelle le système de Haar.

La fonction-mère des ondelettes de Haar est une fonction constante par morceaux :

$$\psi(t) = \begin{cases} 1 & \text{pour } 0 \leq t < \frac{1}{2}, \\ -1 & \text{pour } \frac{1}{2} \leq t < 1, \\ 0 & \text{sinon} \end{cases}$$

La fonction d'échelle associée est alors une [fonction porte](#) :

$$f(t) = \begin{cases} 1 & \text{pour } 0 \leq t < 1, \\ 0 & \text{sinon} \end{cases}$$

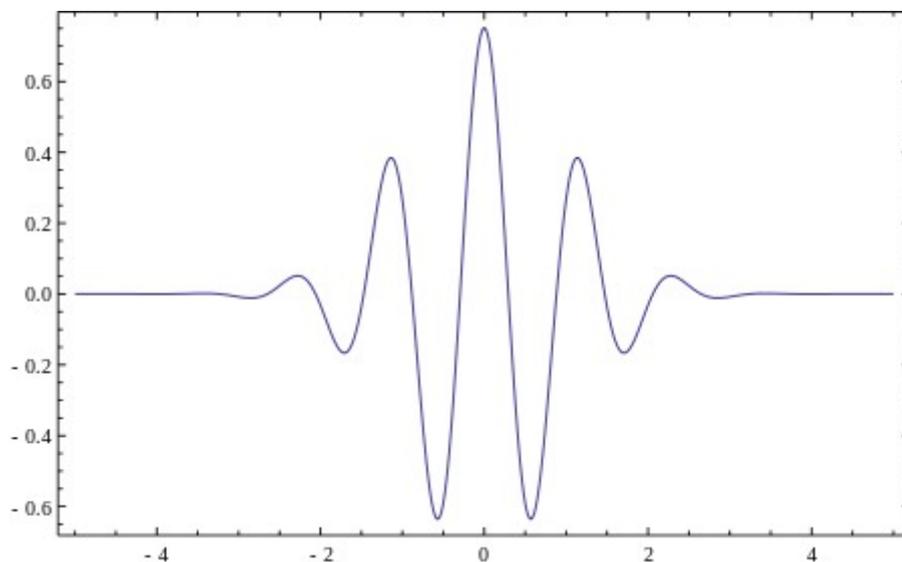


7.2. Ondelette de Morlet

En mathématiques, l'ondelette de Morlet (ou l'ondelette de Gabor) est une ondelette composée d'un exponentiel complexe (support) multiplié par une fenêtre gaussienne (enveloppe). Cette ondelette est étroitement liée à la perception humaine, à la fois l'ouïe et la vision.

L'ondelette est définie comme une constante soustraite d'une onde plane, puis localisé par une gaussienne fenêtre :

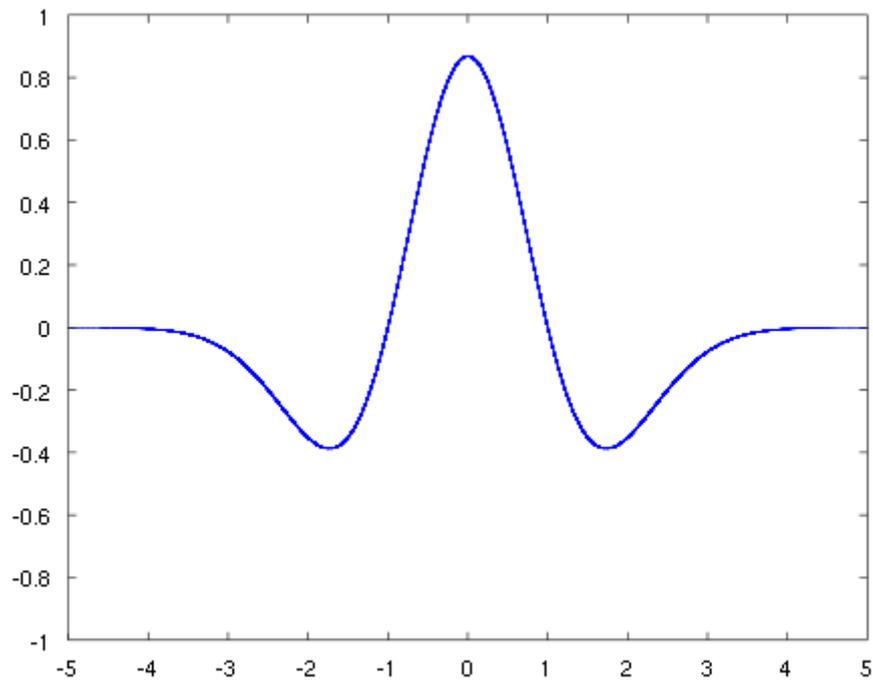
$$\Psi_{\sigma}(t) = c_{\sigma} \pi^{-\frac{1}{4}} e^{-\frac{1}{2}t^2} (e^{i\sigma t} - \kappa_{\sigma})$$



3. Ondelette chapeau mexicain

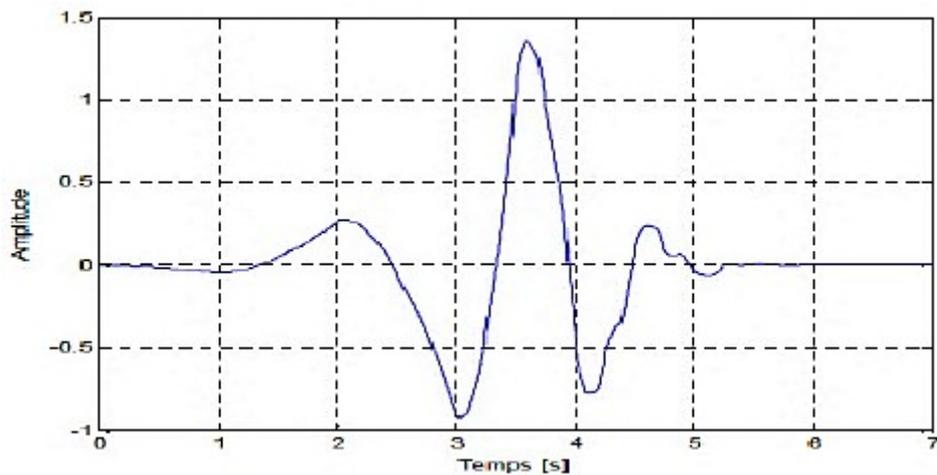
En [mathématiques](#) et en [analyse numérique](#), l'ondelette chapeau mexicain

$$\psi(t) = \frac{2}{\sqrt{3\sigma\pi^{\frac{1}{4}}}} \left(1 - \frac{t^2}{\sigma^2}\right) e^{-\frac{t^2}{2\sigma^2}}$$



7.3. Ondelette de Daubechies :

L'algorithme de transformée d'ondelette de Daubechies est composé d'une fonction d'ondelette et d'échelle. Les coefficients de la fonction d'échelle sont notés "h" et les coefficients de la fonction d'ondelette sont notés "g". L'application de la fonction d'échelle à un signal "y" donne un coefficient de type "e" et l'application de la fonction d'ondelette à ce même signal "y" donne un coefficient de type "o"(site)



Chapitre IV

Compression de signal de la parole par la transformée en ondelettes

IV. Compression du signal de la parole utilisant les transformées en ondelette:

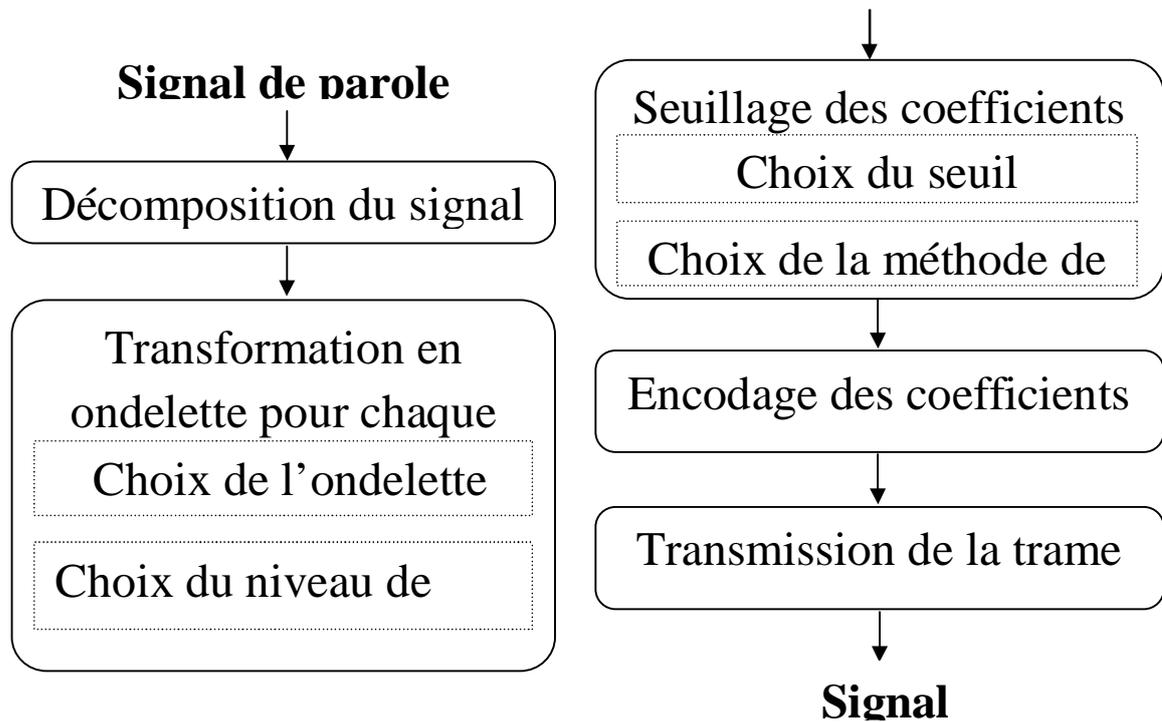
1- Introduction

la particularité de ce type de compression, est la représentation de signal d'entrée par une représentation temps/fréquence. Le changement d'espace permet la réduction de la redondance du signal par conséquent celle de débit de la transmission

Dans notre travail on s'intéresse plus particulièrement à la compression du signal de la parole. La finalité de la compression de la parole est de réduire le nombre de bits requis pour représenter les signaux vocaux (en réduisant la redondance) afin de minimiser l'exigence de la bande passante de transmission (par exemple, pour la transmission de la voix sur les canaux mobiles ayant une capacité limitée) ou pour réduire les coûts de stockage (par exemple, pour l'enregistrement de la parole). Le type de compression envisagé repose sur l'utilisation de l'ondelette comme outil mathématique. Dès lors on procède par un changement d'espace. En effet on transforme la représentation temporelle du signal d'entrée par une représentation temps-fréquence. Le but visé est de transformer le signal sous forme de coefficients moins corrélés et concentre l'énergie du signal sur un petit nombre de ces coefficients, cependant avec ce résultat on réduit le débit de transmission.

1.2 Méthode de compression par ondelettes :

La majorité des systèmes de compression utilisés en littérature sont fondés sur l'utilisation d'une transformation temps-fréquence telle que la transformation en ondelettes. Les différentes étapes de compression du signal de parole en utilisant la transformée en ondelettes discrète sont illustrées ci-dessous.



1.2. Choix des ondelettes

Le choix de l'ondelette analysante la plus adéquate est un point important dans les systèmes de compression des signaux de la parole par transformation en ondelette. En effet, la transformation en ondelette discrète n'est pas unique. Plusieurs types d'ondelettes analysante peuvent être considérés. Il n'y a pas une ondelette qui soit meilleure que les autres pour tous les cas. Tout dépend de l'application envisagée. Dans certains cas, l'ondelette la plus simple de Haar est optimale. Pour d'autres applications, elle sera le pire choix. Dans les applications pratiques, l'ondelette optimale peut être choisie en tenant compte de plusieurs critères tels que le nombre de moments nuls, la taille du support et la régularité. Pour une meilleure transformation du signal, il est légitime d'avoir le plus de moments nuls. Cependant, l'augmentation du nombre des moments nuls engendre un support plus grand. A titre d'exemple, dans le cas des ondelettes de Daubechies et de Symlets, si le nombre de moments nuls est N , alors la taille du support est $2N-1$, ceci signifie que le support s'élargit avec l'augmentation de N . L'élargissement du support entraîne l'apparition du phénomène de Gibbs, due aux discontinuités brusques du signal. D'où un compromis doit être effectué entre ces deux critères. Par ailleurs, Il est important aussi de choisir une ondelette à support compact. D'où la quantité d'énergie peut se concentrer au premier niveau de coefficients d'approximations. Dans, il a été montré que l'ondelette Battle-Lemarie et de

Daubechies (db20, db10, db8) concentre plus de 96% de l'énergie du signal dans les coefficients d'approximations. Dans le cadre de ce travail, une série de simulations a été effectuée afin de déterminer l'ondelette optimale la plus adaptée pour compresser les signaux de parole.

1.3 Choix du niveau de décomposition

Le choix du niveau de décomposition est un point clé dans l'analyse par ondelette. Il dépend de l'application considérée et du type du signal à traiter ou d'un autre critère approprié tel que l'entropie. Dans le traitement des signaux de la parole la décomposition en ondelette s'arrête au niveau 5, et aucun avantage supplémentaire n'est acquis dans le traitement au-delà de ce niveau.

2 .Calcul des coefficients de la transformation en ondelettes

Dans les applications de compression des signaux de parole, le calcul de la transformation en ondelettes se fait par bloc, Ceci est due d'une part à la non stationnarité du signal de parole, en effet le signal de parole est considéré comme quasi-stationnaire pour des durée de 10 à 30 ms, d'où la nécessité de le segmenter avant tout traitement.. D'autre part, lors du calcul de la transformée en ondelette discrète orthogonale les étapes d'analyse et de synthèse engendrent un retard très important. Le traitement par blocs permet de réduire ce retard.

La décomposition en ondelettes vise à transformer le signal sous formes de coefficients moins corrélés que le signal lui-même et à concentrer l'énergie du signal sur un petit nombre de ces coefficients. Dans ce travail, en vue de calculer les coefficients transformés, nous avons considéré la transformation en ondelettes discrète à support compact. L'importance de cette transformation dans la compression de la parole réside dans le fait qu'elle conserve l'énergie, les fonctions de base sont identiques pour l'analyse et la synthèse, et l'implémentation des algorithmes est possible et aisé .Le principe de base de la transformation en ondelettes discrète, consiste à séparer le signal en deux composantes, une représente l'allure générale du signal et l'autre représente ses détails. L'allure générale d'une fonction est représentée par ses basses fréquences et les détails sont représentés par ses hautes fréquences. Dans le but de faire la séparation entre les deux, on a besoin alors d'une paire de filtres: un filtre passe-bas qui a pour rôle d'obtenir l'allure générale que l'on appelle aussi approximation ou moyenne, et un filtre passe-haut qui a pour rôle d'estimer ses détails, c'est-

à-dire les éléments qui varient rapidement. Il est nécessaire que ces deux filtres soient complémentaires. Les deux filtres forment une paire de filtres miroirs en quadrant

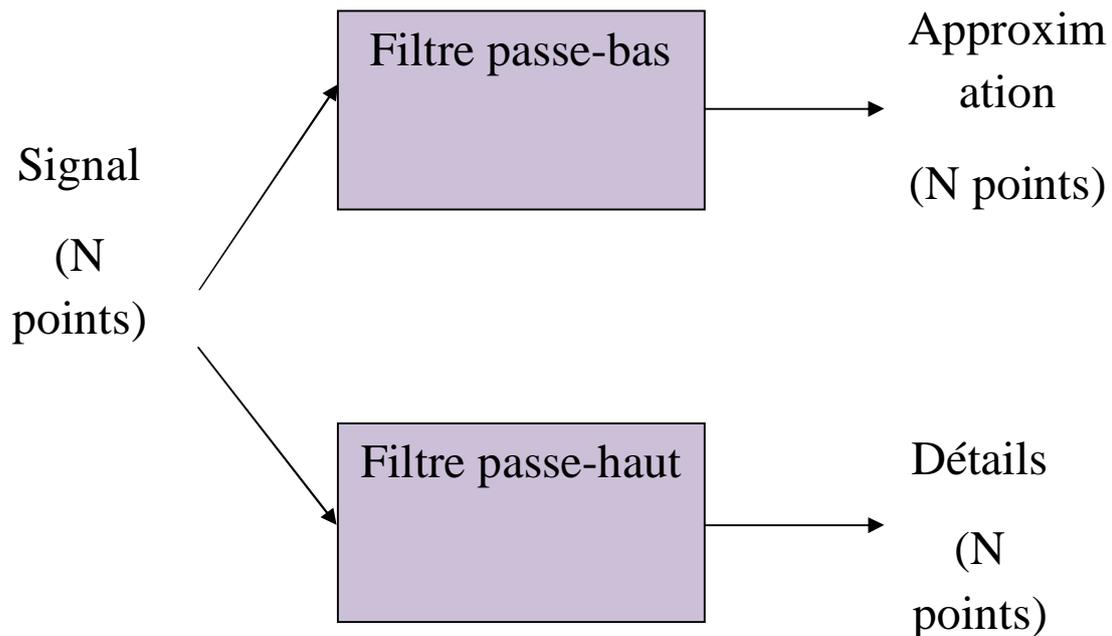


Figure 1: Décomposition d'un signal en approximation et détails

Ainsi l'illustration ci-dessus représente la décomposition d'un signal à N échantillons en deux signaux, l'un d'approximation et l'autre de détails de même taille. Nous remarquons que la quantité d'information c'est-à-dire le nombre de coefficients est multipliée par 2. En effet, si le signal à traiter admet N points alors chacun des signaux d'approximation et de détails fera également N points donc en tout soit $2N$ points. Pour y remédier à ca, le filtre passe-bas est choisi de telle sorte que sa fréquence de coupure soit $F_e/4$, où F_e est la fréquence d'échantillonnage du signal. Un filtre passe-haut orthogonal au filtre passe-bas, peut être calculé facilement. En sous-échantillonnant (décimation) d'un facteur égal à 2 l'approximation et les détails, on aura deux signaux de longueur $N/2$, soit N points en tout.

L'opération de sous-échantillonnage ou de décimation consiste à ne retenir qu'un seul coefficient pour deux (un point du signal sur deux).

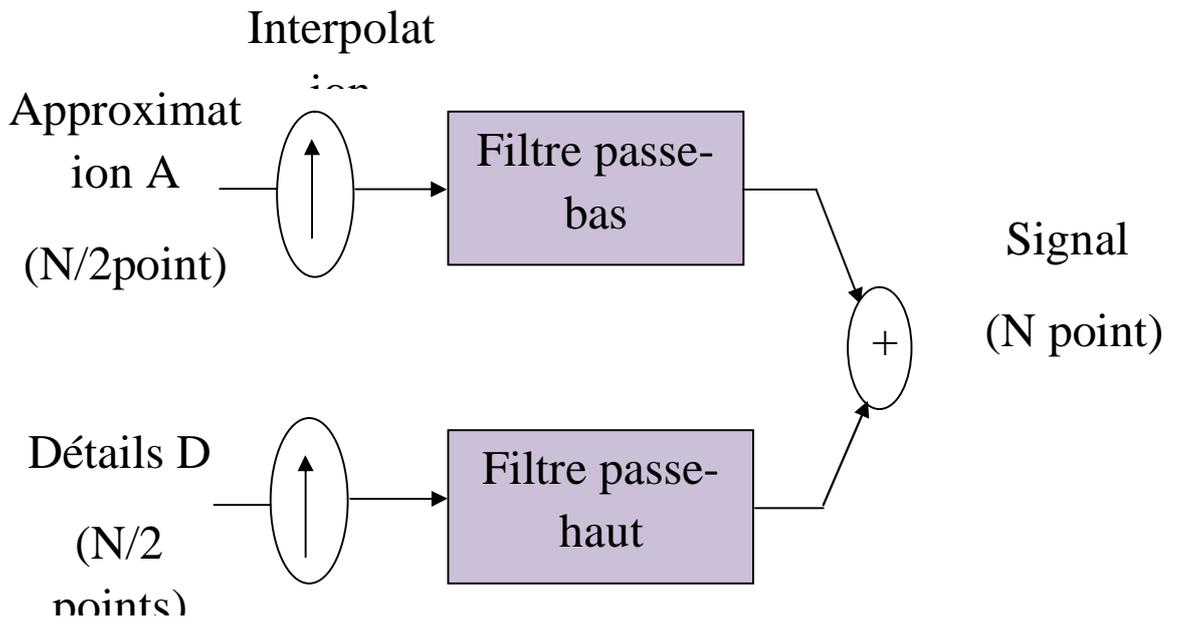


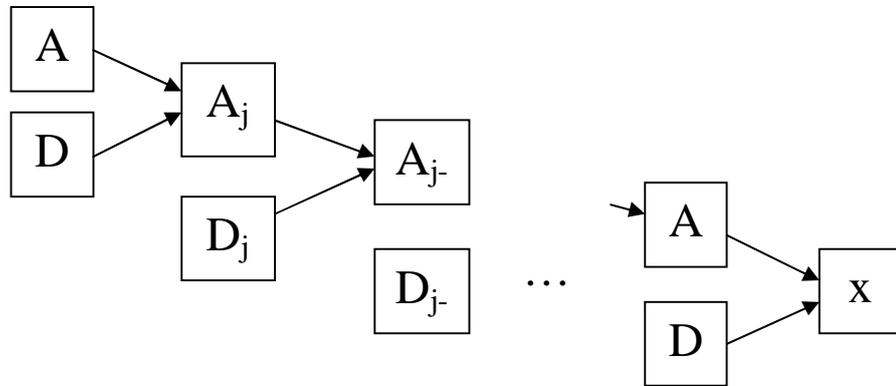
Figure 2: phase de synthèse au niveau 1

Comme pour le cas de la transformation en ondelettes continue, il est intéressant dans de nombreuses applications de pouvoir reconstruire le signal à partir de ces coefficients d'ondelettes (i.e. les signaux d'approximation et de détails). Cette opération est appelée reconstruction ou synthèse. Pour revenir à un signal de longueur N , on effectue un sur-échantillonnage ou interpolation du signal d'un facteur 2 à chaque itération, avant l'opération de filtrage. Pour cela, il suffit de doubler à chaque pas la longueur de l'approximation et du détail en introduisant un zéro entre chaque échantillon. Cette opération est présentée dans B.

On peut répéter l'opération de filtrage sur le signal d'approximation, dans le but d'accéder à une résolution encore inférieure. Notez bien que, pour la TOD, seuls les signaux d'approximation sont de nouveau décomposés. Les signaux de détails issus du filtrage passe-haut sont laissés de côté à chaque pas. A chaque itération, on divise la résolution par 2.

Pour reconstruire le signal original x à partir d'un niveau de décomposition donné, il suffit d'itérer cette suite d'opérations de sur-échantillonnage/filtrage. On reconstruit ainsi

récurivement tous les signaux d'approximation à partir du signal d'approximation à la résolution inférieure et du signal de détails. Cette phase de reconstruction est schématisée dans le schéma suivant :



2.1 Troncature des coefficients

Après le calcul des coefficients de la transformation en ondelette nous passons à l'étape de seuillage des coefficients qui consiste à éliminer les coefficients en dessous d'un seuil. D'où on ne tient pas compte des coefficients d'ondelettes de très faible valeur. Ces valeurs, apportent du bruit qui cache l'information importante du signal. L'opération de seuillage est très exploitée aussi dans le débruitage de la parole.

2.1.1-Sélection du seuil :

Le calcul du seuil est une tâche très importante. En effet, un mauvais choix du seuil peut déformer le signal original. Il existe en littérature plusieurs méthodes pour calculer le seuil tel que le seuillage global et le et le seuillage dépendant du niveau de décomposition.

VI-2-4-1-1-Seuillage global

Le principe du seuillage global consiste à estimer un seuil unique quelque soit le niveau de décomposition considéré. Il est calculé en utilisant en utilisant l'équation suivante:

$$T = \begin{cases} 0.05 \times \max(\text{abs}(c_1)) & \text{si } \delta = 0 \\ \delta & \text{sinon} \end{cases} \quad (\text{IV.1})$$

Où c_1 représente les coefficients de détails du premier niveau et δ la médiane de la valeur absolue des coefficients de détails du premier niveau:

$$\delta = \text{médiane} (\text{abs} (c1)) \quad (\text{IV.2})$$

La figure suivante montre le réglage du seuil global d'un signal de parole échantillonné sur 8 kHz et codé sur 8 bits effectué sur plateforme Matlab. L'axe des abscisses représente les valeurs des coefficients qui sont normalisé à un. La ligne verticale noire se déplace à droite et à gauche pour choisir le seuil. L'intersection de cette ligne avec la ligne verte donne le pourcentage de coefficients mis à zéro en dessous de ce seuil. Son intersection avec ligne rouge indique le pourcentage de l'énergie du signal retenu après avoir tronqué les coefficients nuls.

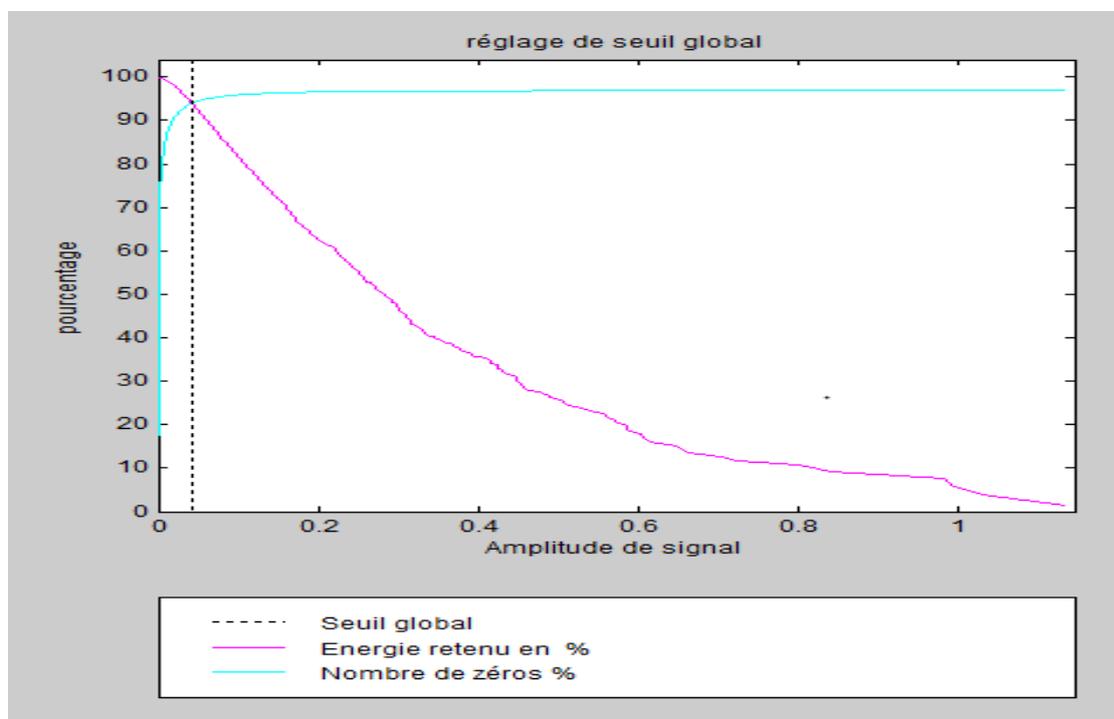


Figure 3: Réglage du seuil global

2-1.2-Seuillage dépendant du niveau

Les seuils dépendant du niveau sont calculé en utilisant la stratégie de Birge-Massart. Cette stratégie conserve l'ensemble des coefficients d'approximation au niveau de décomposition J. Le nombre de coefficients de détail conservé au niveau i en partant du niveau 1 à J sont donné par la formule suivant :

$$n_i = \frac{M}{(J+2-i)^\alpha} \quad (\text{IV.3})$$

Où α désigne le paramètre de compression, qui est égal à 1.5. M désigne comment les coefficients sont distribués dans le vecteur transformée. Les valeurs de M varient entre L et $2L$, où L désigne la longueur des coefficients d'approximation du premier niveau de décomposition. Nous présentons dans la figure 31 un exemple de calcul du seuil selon la méthode de seuillage dépendant du niveau. Pour cela nous décomposons un signal de parole par transformation en ondelettes discrète au niveau 5 en considérant l'ondelette de Daubechies à 10 moments nuls db10.

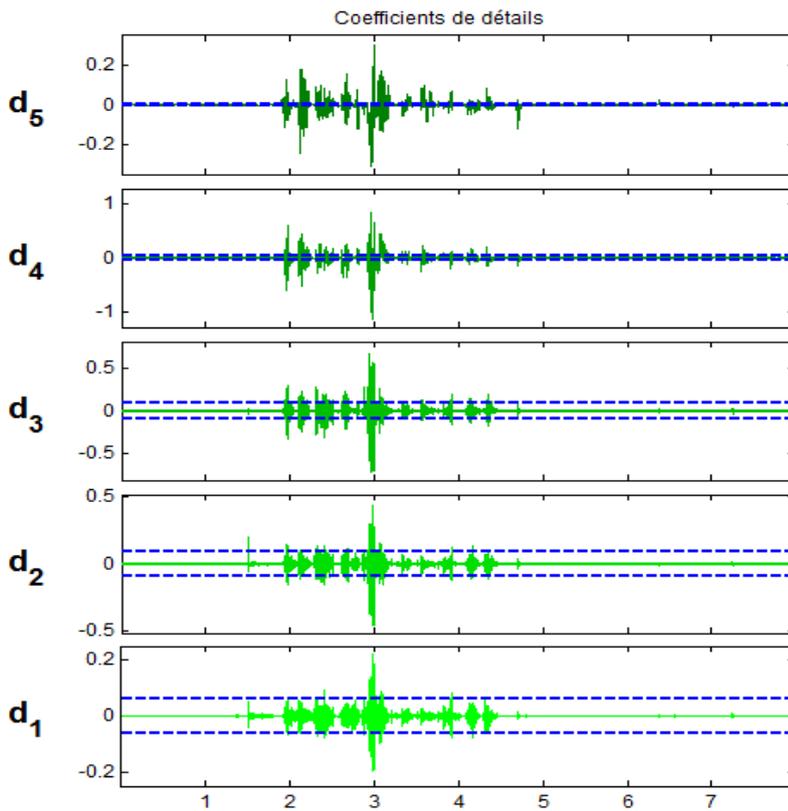


Figure 4: Exemple de calcul de seuil dépendant du niveau de décomposition

2-1.3 Les méthodes de seuillage

Les méthodes de seuillage les plus connues ont été introduites par Donoho [36]. Ils sont le seuillage doux et le seuillage dur.

Seuillage doux

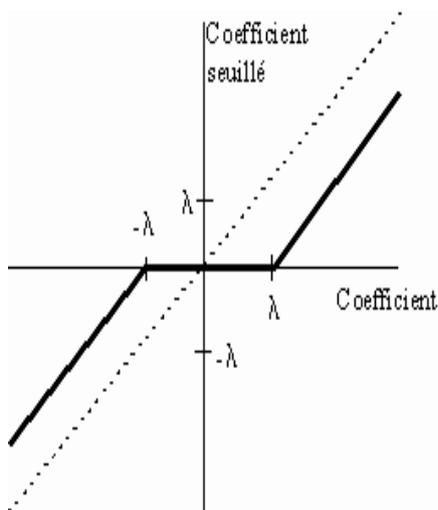
Le seuillage doux consiste à éliminer tout coefficient au-dessous du seuil et à soustraire ce seuil des autres coefficients, en vue de réduire le nombre de discontinuités dans le signal compressé. Pour un vecteur x quelconque, la fonction de seuillage doux T_{doux} est donnée par l'équation suivante, où λ est la valeur du seuil:

$$T_{\text{doux}} = \begin{cases} \text{sign}(x)(|x| - \lambda) & |x| > \lambda \\ 0 & |x| < \lambda \end{cases} \quad (\text{IV.4})$$

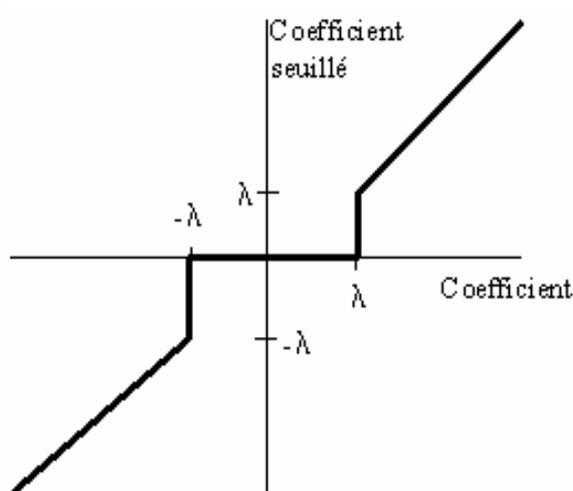
2.1.4. Seuillage dur

Le seuillage dur est celui qui est le plus "intuitif". On se fixe un seuil $T > 0$. On ne conserve que les coefficients d'ondelettes supérieurs à T et on met à zéro les autres. La fonction de seuillage dur T_{dur} est donnée par l'équation suivante:

$$T_{\text{dur}} = \begin{cases} x & |x| > \lambda \\ 0 & |x| < \lambda \end{cases} \quad (\text{IV.5})$$



Exemple de seuillage doux



Graphe de la fonction du seuillage dur

Les figures ci-dessous représentent le scalogramme des coefficients d'ondelettes pour les niveaux allant de 1 à 5, et ceci avant et après seuillage. Il est à noter que la densité spectrale d'énergie des coefficients d'ondelettes est diminuée après seuillage.

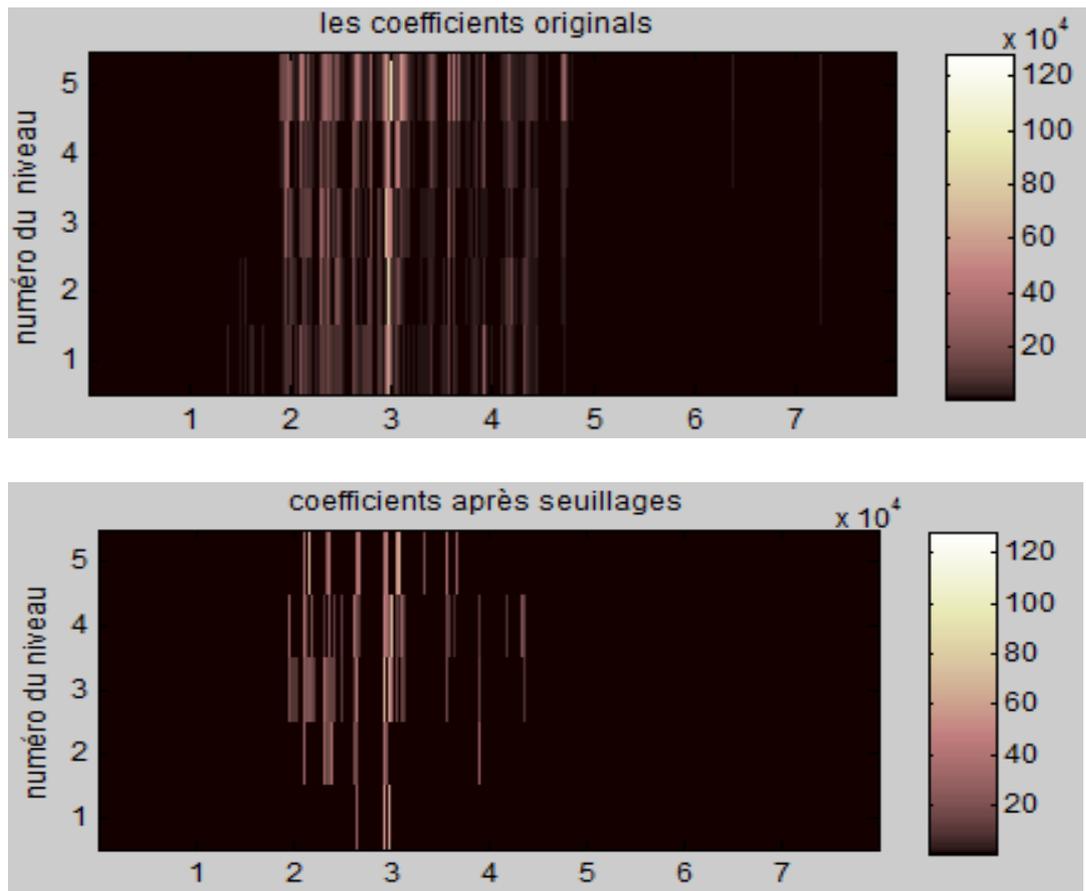


Figure 5: Scalogramme des coefficients d'ondelettes avant et après seuillage.

2.5 Encodage des coefficients d'ondelettes tronqués :

Après la troncature des coefficients on aura un vecteur contenant les coefficients d'ondelettes de hautes amplitudes, les autres coefficients sont mis à zéro. Plusieurs méthodes peuvent être envisagées pour coder ces coefficients. Une de ces méthodes consiste à parcourir tous les vecteurs des coefficients d'ondelettes tronqués et de remplacer les zéros consécutifs par deux octets, un octet pour indiquer la séquence de zéros dans le vecteur et l'autre représente le nombre de zéros de la séquence. Dans notre travail nous avons utilisé une approche un peu différente, dans laquelle le vecteur des coefficients d'ondelettes après troncature, est remplacé par deux vecteurs, un vecteur ne contenant que les coefficients significatifs sans valeur nuls, l'autre vecteur stocke la position de départ d'une chaîne de zéros ainsi que le nombre de zéros consécutifs. Supposons par exemple le vecteur C suivant : $C = [1, 0, 0, 0, 0, 0, 2, 3, 0, 0, 6, 0, 0, 0, 7]$. Après seuillage, ce vecteur sera codé en 2 vecteurs D et E de la manière suivante, $D = [1, 2, 3, 6, 7]$ et $E = [2, 5, 9, 2, 12, 3]$.

2.6 Critères d'évaluation des performances

L'évaluation de la qualité de la parole compressée est basée sur des mesures subjectives et des mesures objectives. Les mesures subjectives exigent un jugement humain à partir des tests d'écoutes. Ces tests subjectifs constituent une méthode d'évaluation la plus convaincu en codage ou compression de la parole. Cependant, ils nécessitent beaucoup de temps et un nombre important d'auditeurs. De ce fait, beaucoup de chercheurs ont recours aux mesures objectives. Nous résumons dans ce qui suit les différentes mesures retenue pour évaluer notre travail.

3. Le rapport signal sur bruit

Le rapport signal sur bruit est un critère permettant de juger objectivement la qualité du signal reconstitué, il est défini par l'expression suivante:

$$\text{SNR} = 10 \cdot \log_{10} \left[\frac{\sum_n s^2(n)}{\sum_n [s(n) - \hat{s}(n)]^2} \right] \quad (\text{IV.6})$$

Où s et \hat{s} sont respectivement le signal original et le signal reconstitué.

3.1 Le rapport signal sur bruit de crête (PSNR)

Le rapport signal sur bruit de crête est défini par la formule suivante :

$$\text{PSNR} = 10 \log_{10} \frac{N X^2}{\|x - r\|^2} \quad (\text{IV.7})$$

N : la longueur du signal reconstitué.

X : la valeur quadratique absolue maximale du signal x .

x : le signal original

r : le signal reconstitué

3.1 L'erreur quadratique moyenne à racine normalisée (NRMSE)

L'erreur quadratique moyenne à racine normalisée est définie par la formule suivant :

$$\text{NRMSE} = \sqrt{\frac{(x(n)-r(n))^2}{(x(n)-\mu_x(n))^2}} \quad (\text{IV.8})$$

Où

$x(n)$: le signal original.

$r(n)$: le signal reconstitué.

$\mu_x(n)$: la moyenne du signal.

3.2 Taux de compression(CR)

C'est le rapport du volume initial de données (nombre de bits) avant compression sur le volume final des données après compression, il est donnée par la formule suivante:

$$\text{CR} = \frac{\text{taille de signal initial}}{\text{taille de signal après compression}} \quad (\text{IV.9})$$

Nous proposons comme travail à faire la comparaison de deux ondelettes utilisées comme outils pour la compression du signal de la parole et d'en tirer des conclusions à partir des critères d'évaluation tels que le SNR, taux de compression etc. Parmi elles, l'ondelette de Haar et l'ondelette de Daubechies dans le but de choisir l'ondelette orthogonale la plus optimale en compression des signaux de la parole. Pour cela il faut fixer le nombre de niveau de décomposition, la fréquence d'échantillonnage et le nombre de trames.

Chapitre V
Simulations – résultats et Perspectives

1. première étape de la simulation :

Vu la non stationnarité du signal de la parole on le décompose en trames de durée égale à 20 ms.

Les résultats de la décomposition nous donnent une matrice qui contient les sous vecteur de signal partitionné de vecteur original (le signal de la parole)

2. deuxième étape :

La deuxième étape consiste à appliquer la transformée discrète des différentes ondelettes à chaque trame. Dans notre cas nous avons choisi respectivement les ondelettes de Haar et de Daubechies et de Haar.

3. Troisième étape :

Nous avons finalement opté dans notre simulation pour les ondelettes de **Haar**, **Daubechies 4**, **Daubechies 8**.

4. Quatrième étape :

Cette étape est consacrée à la décomposition. Nous avons une décomposition sur 5 niveaux ; ce qui est largement suffisant pour le signal de la parole.

5. Cinquième étape :

Cette étape consiste à prendre les coefficients de détails des résultats de décomposition faite par la transformée d'ondelettes discrète de chaque ondelette auxquels on applique un seuillage.

6. Sixième étape :

Pour le seuillage des coefficients de détail On a choisi un seuillage doux car il nous donne un résultat proche de 75 % du point de vue qualité.

7. Septième étape ;

Encodage des coefficients de détail avant la transmission du signal sur les supports qu'ils lui sont destinés.

La simulation :**Le signal de parole utilisé :**

Un ensemble de mots choisi parmi la liste de LAFON (base de données dyssilabique)

[Le bouchon- le soupe- le rondin - le glaçons- le réchaud- le coffret- le gamin - le clavier]

Avec un bruit additif de t 2 db.

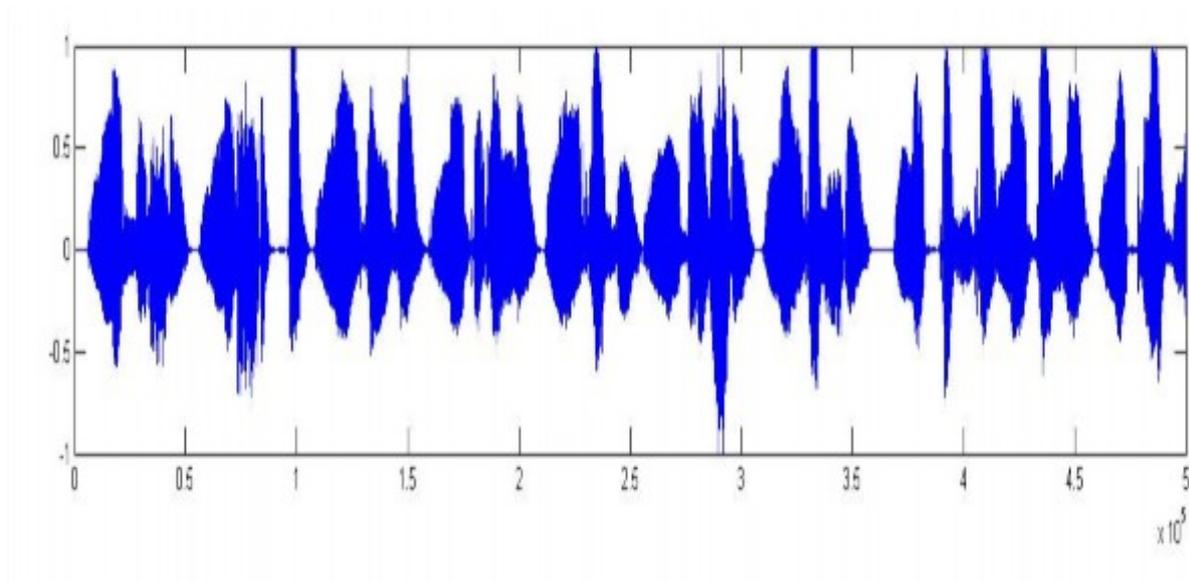


Figure16:signal de parole de 2db

Après la décomposition et l'application de l'ondelette de haar,db4 et db8 on obtient les coefficients de détails suivant :

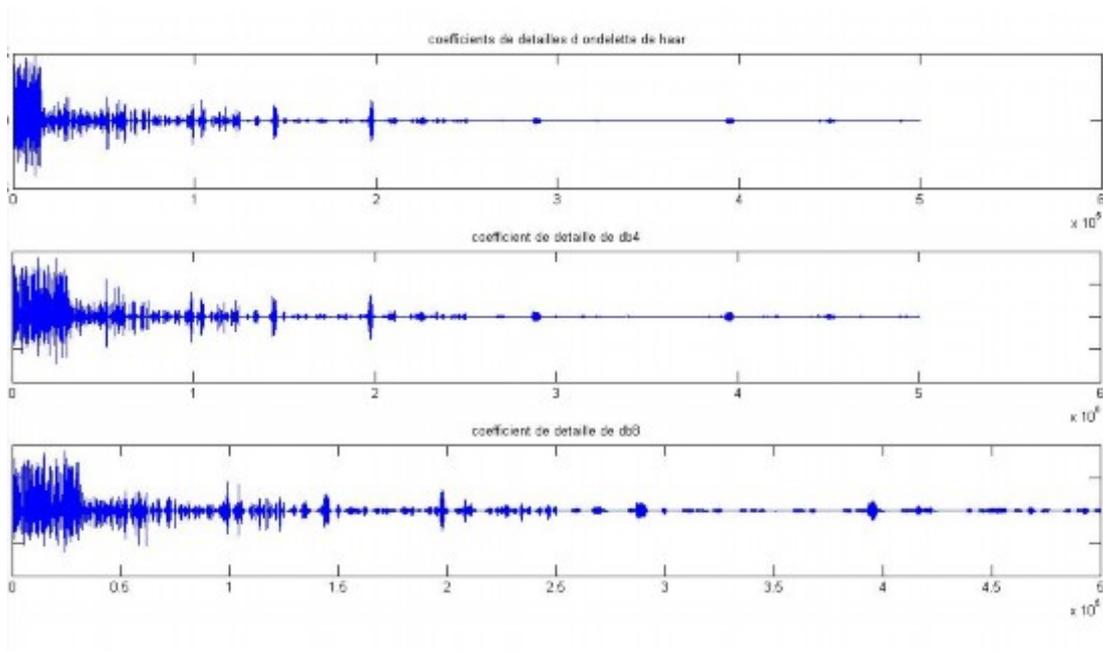


Figure 17:coefficient de détails de décomposition des ondelettes haar et db4 et db8 de 5eme niveau de décomposition

L'application de seuillage a donné des résultats comme le suivant

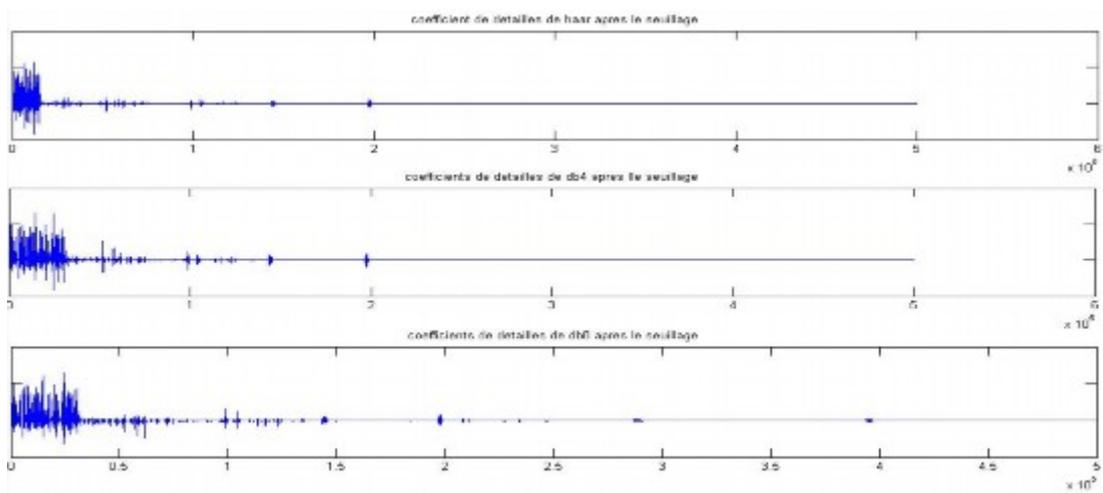


Figure 18:coefficient de détails de haar et db4 et db8 après l'application de seuillage de valeur 0.3

$$\text{Coef(Haar)} < \text{Coef(Db4)} < \text{Coef(DB8)}$$

L'ensemble des résultats sont illustrés dans le tableau suivant :

Les ondelettes et le signal original	Signal de parole original	Compression avec haar	Compression avec db4	Compression avec db8
Snr	2	4.5959	5.7798	6.1336
Taux de compression	1	475417/428206 =1.1103	475417/361352 =1.3157	475417 /215124 =2.2100

les illustrations suivantes représentent l'effet de la compression effectuée par ces ondelettes

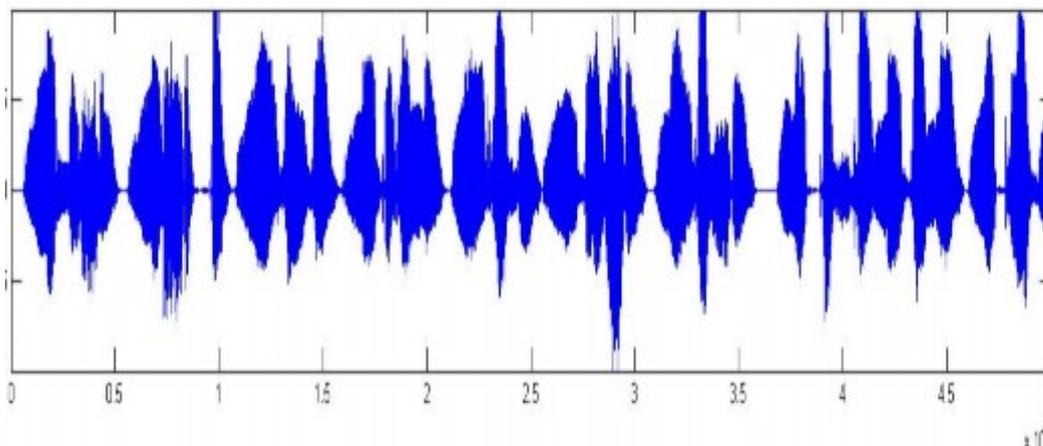


Figure 19: signal de parole original sans compression

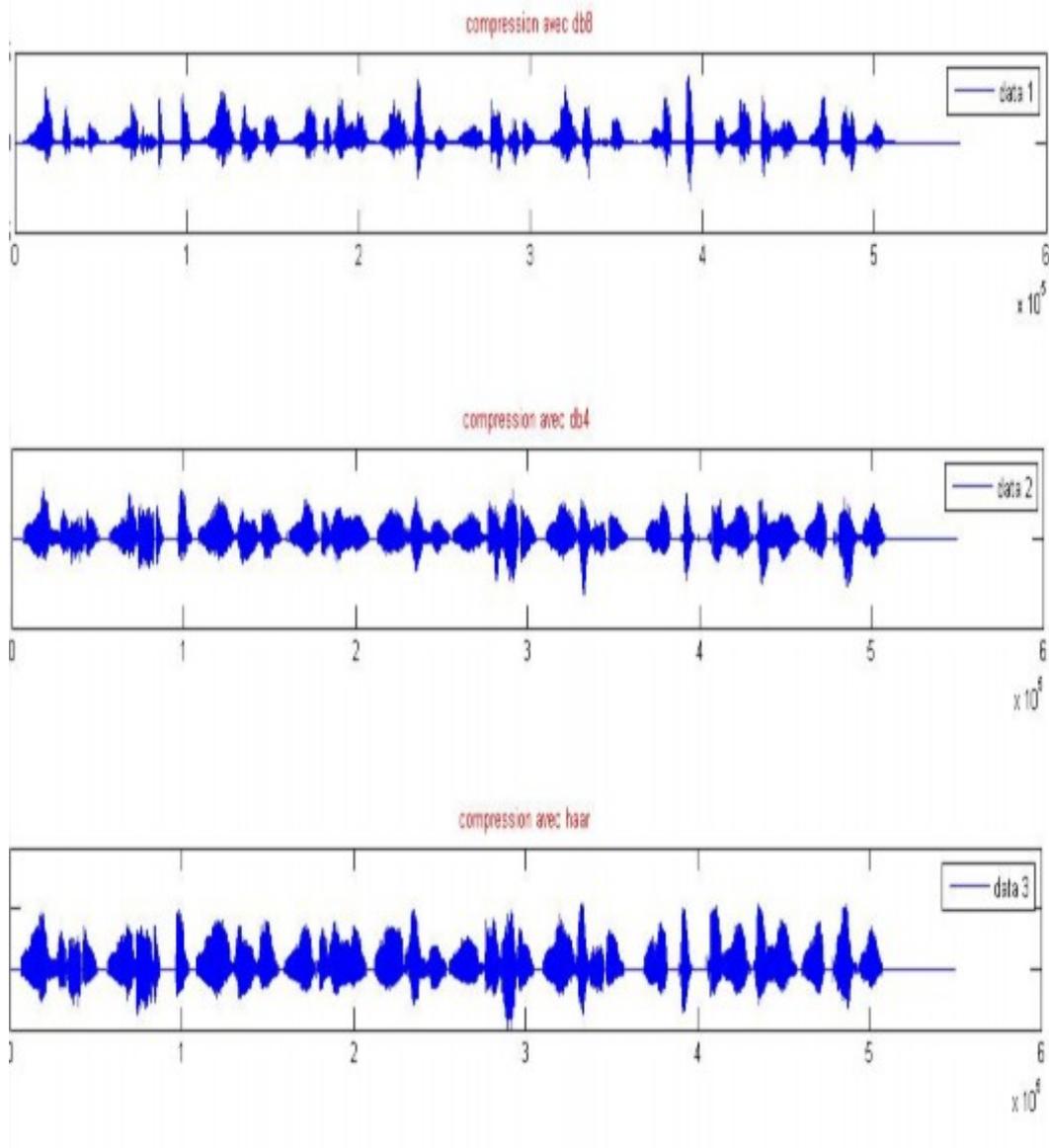


Figure 20:signal compressé

Remarque:

Suite aux valeurs du rapport signal /bruit (tableau comparatif) et les illustrations des signaux de parole ci-dessus on conclue que l'ondelette optimale pour effectuer la compression est daubechis 8.

2. Conclusion et Perspectives:

Ce projet nous a permis de découvrir le domaine qui se rapporte à la production, la réception, le traitement du signal de la parole et sa compression en vue de le transmettre sur un support avec le minimum de perte possible ; nous avons ensuite saisi cette occasion pour mettre en évidence les limites de la transformée de Fourier et proposer un algorithme classique de compression du signal vocal basé sur la transformée en ondelettes discrète (DWT).

Nous avons à cet effet déduit que le calcul des coefficients d'ondelettes dépend du niveau de décomposition et du choix de l'ondelette.. Nous avons aussi remarqué que l'augmentation du niveau de décomposition ne présente pas d'intérêt au-delà du niveau 5.

D'autres critères objectifs d'évaluation tels que le rapport signal sur bruit de crête et l'erreur quadratique moyenne auraient pu être utilisés pour mieux décider des performances de chaque type d'ondelettes.

Par ailleurs nous pensons que de nombreuses perspectives sont envisageables : On cite à cet effet l'utilisation de la transformée en paquet d'ondelettes et l'implémentation de l'algorithme sur processeur DSP ou FPGA .

Bibliographie

[En ligne]. - http://www.epm6604b.be/ondelette_daubechies.html.

ABDELAZIZ AMRANE ANIS Detection de l onde Pde L electrocardiogramme par des algorithmes basé sur la transformée en Ondelette et Modele Markov caché.

Alain MAGUER Pierre ALINAT,George GOULLET Etude de Bruit impulsives par analyses par banc de filtre Q constant [Conférence]. - 1989.

Ani Tarik Al Introduction aux ondelettes ,Deuxieme partie:Quelques concept Generaux de la theorie des ondelettes [Livre]. - Paris : Departement D informatique ESIEE.

Ashmead John Morlet Waveles in Quantum mechanics [Revue] // Quanta . - 2012. - pp. 58-70.

Bekesy G.Von Experiments in Hearing [Livre]. - NewYork : McGraw-Hill, 1960.

BENTALEB Youssef Analyse par ondelettes des Signaux Sismique :Application Aux ondes de surface // These de Doctorats .

BENZID Redha Ondelettes et statistiques d ordre superieur aus signaux uni et bidimensionnels . - Batna : Université de Batna , 2005.

Boite R,Boulard.H,Dutoit,T.,Hancq ,J.et Leich ,H Traitement de la parole . - [s.l.] : université Romance , 2000.

Boussalmi.S Ounik .K etude pour l amelioration de la parole codé par transformation en paquette de framelette serré. - Grenoble : [s.n.], 4-8 juin 2012. - pp. 473-480.

Calliope La parole et son traitement automatique . - Masson : [s.n.], 1989.

D.Salomon Data compression . - New-York : Springer Verlag, 1991.

Daugman J.G Uncertainty relation for resolution in space , spatial frequency , and orientation optimized by two-dimensional visual cortical [Revue] // Optical society of America A. - 1985. - pp. 1160-1169.

Demayer Jonathan Bebronne Michael et Forhomme Sébastien Les ondelettes [Livre]. - [s.l.] : Université Libre de Bruxelles.

E.Zwicker R.Feldetkeller Psychoacoustique [Livre]. - Paris : CENT-ENST,Collection technique et scientifique des telecommunications, 1981.

Elayed Hatem Speech compression using wavelets. - Ghaza Palestine : [s.n.].

El-Bahlul Fgee W.J,Phillips and W.Robertson comparing audio compression using wavelets [Conférence] // IEEE Canadian Conference . - 1999.

fac-mathinfo [En ligne] // univ-usto. - 2014. - www.univ-usto.dz/faculte/fac-mathinfo/Cours/Traitement_Parole2014.pd..

J.I.Agbenya Discret Wavelets Transform Techniques in Speech Processing [Revue]. - New-York : IEEE New-york, 1996. - pp. 514-519.

Mallat S A wavelet tour of signal Processing. - New york : [s.n.], 1998. - p. 1998.

Mallat Stephane Une exploration des signaux en ondelettes [Section du livre].

Mallat Time-frequency dictionaries [Livre].

N.ELLOUZE A.BOUZID et Caracterisation des singularités du signal de parole [Livre]. - Sfax : ISET , Ecole National d ingenieurs 88A,3099 .

P.Haton&all J Reconnaissance Automatique de la parole [Livre]. - [s.l.] : Dunod.

Pesquet Beatrice Pesquest-Popescu et Jean -Christophe Ondelettes et Application [Livre].

Pollak Ilya Audio Compression using wavelet Techniques [Rapport]. - 2005.

POUGAZA Doriano-Boris Compression sans perte par ondelettes [Livre]. - 2008.

Quinquis A. Cornel L Representation temps frequence et temps echelle. - [s.l.] : ENSIESTA , janvier 2002.

R.Boite&all Traitement de la parole [Revue] // Presses polytechnique et universitaire Romandes. - [s.l.] : Presses polytechnique et universitaire, 1999.

Rao Nikhil Speech compression using wavelet [Compte rendu d'audience]. - 2001.

Shiv Kumar Vijay K Chaudhari ,R.K.SINGh, Dinesh Varshney A new algorithm for voice signal compression (VSC)and analysis suitable for limited storage devices using matlab [Revue] // International Journal of computer and electrical ingenioring . - 2009.

Togneri Roberto Information theory and coding // cours . - Western Australia : Departement E&E engenieering .