N° d'ordre : .....

N° de série : .....

République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Mémoire de Fin d'Étude

Présenté à

L'Université Echahid Hamma Lakhdar d'El Oued

Faculté de Technologie

Département de Génie Electrique

En vue de l'obtention du diplôme de

MASTER ACADEMIQUE

En Télécommunications

Présenté par

BEN NEDJOUE Yacine et MEKKAOUI Belkis

Thème

Débruitage de signal parole par les transformées en ondelettes

Soutenu le 28/05/2016. Devant le jury composé de :

Mr. MEDJOURI Abdelkader	Maitre assistant A	Président
Mr. BOULILA Mohamed	Maitre assistant A	Examinateur
Mr. AJGOU Riadh	Maitre de conférence B	Rapporteur

Année Universitaire 2015/2016


بسم الله الرحمن الرحيم

Dédicaces

À notre parents, frères et soeur,

Nous vous dédions ce mémoire en guise de remerciements pour vos encouragements et votre soutien. Nous vous souhaitons le plus radieux des avenirs.

À tous nos amis pour leurs encouragements et leur soutien, et à tous ceux qui nos ont aidons et soutenu le long de la réalisation de ce projet.

Nos pensées vont également à tous ceux qui nos ont aidons de près ou de loin à mener à bien ce travail.

MEKKAOUI Belkis
BEN NEDJOUE Yacine

Remerciements

La louange est à Allah, qui nous a facilité l'accomplisseme de ce travail de recherche chose ne peut étre qu'avec la volonté de Dieu à lui la tout puissance et la Majesté et que la louange initiale et finale appartieut à allah, Seigneur des mondes.

Je tiens à exprimer ma vive reconnaissance et mes sincères remerciements à Monsieur AJGOU Riadh*, pour avoir accepté de diriger mes recherches. Je le remercie également pour sa bienveillance, ses conseils judicieux et l'encadrement de qualité dont il m'a fait bénéficier aimablement. Je lui adresse, en signe de reconnaissance, toute ma gratitude et tout mon respect pour ses qualités humaines et scientifiques.*

Je tiens à remercier mes parents pour les nombreuses relectures de ce mémoire et leur soutien sans faille, depuis toujours.

Enfin, j'adresse un remerciement chaleureux à l'ensemble du personnel de Université d'El-Oued pour avoir accompagné mon évolution pendant ces cinq belles années.

A vous tous, MERCI !

MEKKAOUI Belkis
BEN NEDJOUE Yacine

Résumé

Amélioration de la parole vise à améliorer la qualité de la parole en utilisant divers algorithmes. L'objectif de débruitage de signal parole est l'amélioration de l'intelligibilité et la qualité de perception globale du signal dégradé en utilisant des techniques de traitement du signal audio. Ce travail examine l'utilisation de la transformée en ondelettes pour le débruitage des signaux de parole contaminés par des bruits. L'étude sera fondée sur le traitement d'une base de données dédiée de signaux de parole contaminés par divers bruits et niveaux SNR. Cette étude tend à être pour l'amélioration du signal de parole pour les fins de dispositifs d'appareils auditifs.

La qualité de la parole reconstruite est mesurée avec la technique de mesure de l'évaluation perceptuelle de la qualité de la parole PESQ (Perceptual Evaluation of Speech Quality). Ainsi l'évaluation peut être éffectué en termes du taux d'améliorations du SNR. Dont, une amélioration du SNR ou le PESQ implique une meilleure intelligibilité de la parole. Les méthodes discutés dans ce travail ont été évaluée en présence de différents types de bruit en utilisant la base de données TIMIT [x]et le corpus de NOIZEUS[xx]. Le bruit a été pris à partir de la base de données AURORA, dont, les bruits inclus sont le bruit de train, bavardage, voiture, salle d'exposition, restaurant, rue, l'aéroport et la gare.

Mot clés : Transformée d'ondelettes , PESQ, SNR.

ملخص

يتم تحسين الكلام بتحسين جودته وذلك باستخدام خوارزميات متعددة . فالهدف من نزع التشويش للإشارة الصوتية هو زيادة الوضوح وتحسين الجودة للإشارة المتدهورة بإستخدام تقنيات معالجة الإشارات الصوتية . نحقّق هذا العمل باستخدام تحويل المويجات من أجل فك تشويش إشارات الكلام المشوّشة. تعتمد الدراسة على المعالجة بقاعدة بيانات مخصصة لإشارات الصوت و مزودّة بمختلف أنواع التشويش و مستويات ال SNR. فهذه الدراسة تهدف لتحسين إشارة الكلام بالنسبة للأجهزة السمعية .

يتم قياس جودة الكلام المنظف من التشويش بتقنية قياس التقييم الحسي لنوعية الكلام (PESQ)، كما يمكن تقييمه من حيث معدل التحسين بال SNR. وذلك أخذا بالتقنية SNR أو PESQ التي تعطينا وضوح أحسن لإشارة الكلام ،كل طريقة تقيّم بوجود أنواع مختلفة من التشويش إعتمادا على قاعدة البيانات TIMIT ومدونة NOIZEUS أما لأنواع التشويش فمأخوذة من قاعدة البيانات AURORA ، و نجد منها التشويش المحتوي ل : ضجيج القطار ، الثرثرة ، السيارات ، صالة العرض ، المطاعم ، الشارع ، المطار و محطة السكك الحديدية .

كلمات مفتاحية : تحويل المويّجات ، معدل التحسين SNR التقنية ال PESQ

Abstract

The speech enhancement aims to improve the quality of the speech using different algorithms . Denoising of speech signal aims to the improvement in intelligibility and quality of overall perception of the degraded signal using audio signal processing techniques . This work investigates by using the Wavelet transformed for denoising speech signals contaminated with noise . The study will be based on the treatment of a dedicated data base of speech signals contaminated by different noises and SNR levels . This study tends to the improvement of a speech signal for the purposes of hearing devices.

Quality of the reconstructed speech is measured by the measurement of perceptual evaluation of speech quality technique PESQ. also the evaluation can be done in terms of the improvements rate SNR. Which, an improvement of SNR or the PESQ implies a better speech intelligibility. All method has been evaluated in the presence different types of noise using the TIMIT data base and the NOIZEUS corpus. The noise was taken from the AURORA data base, which included noises such as: the train noise, babble, car, showroom, restaurant, street, airport and railway station.

Key words : Wavelet Transform , PESQ, SNR

Table des matières

CHAPITRE II : TECHNIQUE DE DEBRUITAGE DE SIGNAL PAROLE PAR LES TRANSFORMEES D'ONDELETTES

CHAPITRE III : TECHNIQUES D'EVALUATION DES ALGORITHMES D'AMELIORATION DE SIGNAL PAROLE

CHAPITRE IV : RESULTATS DES SIMULATIONS ET DISCUSSION

LISTE DES FIGURES

LISTE DES TABLEAUX

.

LISTE DES ACRONYMES ET ABREVIATIONS

ACR	*Absolute Category Rating*
ANN	*Artificial Neural Networks*
BSD	*Bark Spectral Distortion*
CWT	*Continued Wavelet Transform*
DCR	*Degradation Category Rating*
DCT	*Discrete Cosine Transform*
DMOS	*Degradation Mean Opinion Score*
DWT	*Discrete Wavelet Transform*
FFT	*Fast Fourier Transform*
FT	*Fourier Transform*
IRS	*Intermediate Reference System*
IS	*itakura sait*
LAR	*Log-Area-Ratio*
MOS	*Mean Opinion Score*
MSE	*Mean Square Error*
PAMS	*Perceptual Analysis Measurement System*
PESQ	*Perceptual Evaluation of Speech Quality*
POLQA	*Perceptual Objective Listening Quality Analysis*
PSQM	*Perceptual Speech Quality Measure*
QoE	*Quality of Experience*
SNR	*Signal-to-Noise-Ratio*
SSNR	*Segmental Signal-to-Noise-Ratio*
STFT	*Short Time Fourier Transform*
UIT	*Union Internationale des Télécommunications*
WSS	*Weighted Spectral Slope*
WT	*Wavelet Transform*

INTRODUCTION GENERALE

Le domaine du traitement du signal connaît une progression importante liée à l'évolution des technologies de l'information et de la communication. Dans de nombreuses applications, il est fait appel aux techniques d'analyse et de synthèse pour passer d'un état complexe difficile à analyser ou à traiter à un état simple facile à interpréter ou à manipuler. Souvent, les signaux complexes subissent des transformations nécessaires à 1' augmentation de 1' efficacité du traitement. Le procédé d'analyse le plus classique et le plus populaire est la transformée de Fourier. Depuis plusieurs décennies, cette méthode dominait les techniques d'analyses et la plupart des méthodes de traitement du signal en faisait usage. De nos jours, même si d'autres méthodes ont fait leur apparition, elle est toujours très utilisée pour obtenir le contenu spectral d'un signal quelconque. Les besoins grandissants de la technologie et la complexité des applications en particulier dans des domaines tel que l'Internet et la téléphonie, sont tellement exigeants en matière d'extraction d'informations fiables tout en requérant une vitesse de traitement élevée qu'il a été nécessaire de développer d'autres méthodes pour répondre à ces besoins qui augmentent de jour en jour.

Au cours des dix dernières années, une de ces nouvelles techniques d'analyse du signal a acquis une reconnaissance grandissante dans la communauté scientifique internationale par ses multiples avantages, c'est la transformée en ondelettes. Cette méthode présente des caractéristiques intéressantes dans de nombreux domaines du traitement du signal et de l'image. Les qualités de la transformée en ondelettes lui ont valu une attention particulière de la part des scientifiques œuvrant dans tous les domaines : mathématiques, informatique, physique, géologie, microélectronique, ... etc , et d'une manière générale, presque toutes les sciences utilisant 1' analyse de signaux ont utilisée. La transformée en ondelettes est donc au carrefour des sciences et 1' on retrouve ses applications dans de nombreux domaines.

L'utilisation des ondelettes s'est avérée très performante pour trois problèmes généraux, le premier est l'analyse, le second est la compression, et le troisième est le débruitage.

Nous intéressons par le débruitage qui est un domaine dont, les recherches se sont multipliées ces dernières années, est un processus délicat en particulier dans le cas de signaux complexes tels que les signaux de parole. La non-stationnarité de ces derniers fait en sorte que bon nombre de chercheurs concentrent leur attention sur le problème de débruitage de la parole pour trouver des algorithmes efficaces. De nos jours un grand nombre de ces méthodes utilisent les ondelettes qui donnent des algorithmes de débruitage très simples, souvent plus performants que les méthodes traditionnelles basées sur 1' estimation fonctionnelle et cela grâce à leur adaptativité. Le dé bruitage par ondelette est basé sur un algorithme simple appelé algorithme de seuillage qui est souvent facile à exécuter. Son principe est d'estimer le niveau de bruit qui correspond souvent à la valeur du seuil, puis à partir des coefficients de la transformées en ondelettes, à effectuer l'opération de seuillage qui consiste dans la plupart du temps à soustraire des coefficients la valeur du seuil et éliminer carrément ces coefficients s'ils sont au-dessous du seuil. Les méthodes de seuillage les plus connues sont le seuillage doux et le seuillage dur. Par la suite, il suffit simplement d'appliquer la transformée en ondelettes inverse sur les coefficients seuillés pour récupérer le signal débruité.[xxx]

Dans le cadre de ce projet, qui consiste à étudier le débruitage d'un signal de parole corrompu par un bruit coloré et un bruit réel, la transformée en d'ondelettes de Daubechies sera appliquées sur le signal bruité et cela jusqu'aux cinquièmes niveaux. Deux types de seuillage seront utilisés à la fois, le seuillage doux, le seuillage dur et le seuillage dépendant de niveau pour le cas doux et dur. Une série de tests de simulation effectuée en utilisant le logiciel Matlab fournira les résultats qui permettront de faire une étude comparative des performances des méthodes de seuillage doux, dur ainsi seuillage dépendant de niveau.

Ce manuscrit se compose de quatre chapitres.

Dans le chapitre 1, nous donnerons un bref aperçu sur la parole, ses caractéristiques générales, la modélisation de signal parole et les différentes taxonomies des sons observables en parole, et les variations qui peuvent y être constatées.

Dans le chapitre 2, nous allons étudier la transformée en ondelette et les différents méthodes de débruitage par ondelettes .

Dans le chapitre 3, on explore les différentes techniques d'évaluation de la qualité de la parole dont, nous intéressons aux méthodes objectives temporel (rapport signal / bruit (SNR), segment SNR) et perceptuel (PESQ)).

Dans le chapitre 4, les méthodes de seuillages par ondelettes seront introduites. Nous décrirons également en détail les résultats de simulations effectuées. Une comparaison de ces résultats y sera également discutée et des mesures de performance y seront données en utilisant les différents niveaux de rapport signal/bruit (SNR) de chaque méthode présentée.

Une synthèse des ces travaux ainsi que quelques perspectives seront données en conclusion de ce manuscrit.

Chapitre I

Le signal parole

I.1 Introduction

La parole est le principal moyen de communication dans toute société humaine. D'un point de vue humain, la parole permet de se dégager de toute obligation de contact physique avec la machine, libérant ainsi l'utilisateur qui peut alors effectuer d'autres tâches.

Nous allons présenter, dans ce chapitre, la modélisation de la parole, ainsi que les différentes taxonomies des sons observables en parole, et les variations qui peuvent y être constatées. Nous allons cependant tout d'abord parler des notions qui se rattachent à l'étude des organes biologiques de production, et de compréhension. On distingue généralement plusieurs niveaux de description du signal parole : acoustique, phonétique et phonologique.

I.2 Aspect Anatomo-physiologique

Le système vocal humain peut être décomposé en quatre sous-systèmes élémentaires, en l'occurrence :

I.2.1 Poumons et conduit Trachéo-bronchique

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

I.2.2 Larynx

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

I. 2.3 Cordes vocales

En sortes de lèvres symétriques placées en travers du larynx en s'écartant, ils déterminent une ouverture triangulaire appelée glotte, les Sous voisés résultent d'une vibration périodique des cordes vocales dans un plan horizontale obéissant à un phénomène d'oscillation et de relaxation.

I.2.4 Conduit vocal

Et un ensemble de cavités situées entre la glotte et les lèvres. On peut distinguer aussi :

- la cavité nasale : formée des fosses nasales qui sont deux cavités de fixes.

- la cavité buccale
- la cavité pharyngienne

Les cavités buccale et pharyngienne forment le conduit oral, qui possède un volume et une géométrie extrêmement variable grâce à la grande mobilité de la langue. [1]

I.3 L'appareil phonatoire humain

I.3.1 Généralités

La production de la parole est assurée, chez l'homme, par plusieurs organes successifs. Les poumons sont indispensables dans ce processus puisqu'ils assurent la génération d'un composant incontournable: de l'air sous pression. Cet air, expulsé, traverse alors les cordes vocales qui entrent ou non en action pour produire un voisement. Ce voisement correspond à la fréquence fondamentale qui est le timbre de la voix. [1, 2]

Cette fréquence fondamentale étant produite, elle est propagée dans l'ensemble du conduit vocal. Ce conduit est de forme de volume variable. Plusieurs organes concourent à des possibles modifications du volume qui permettent de produire des sons différents. Parmi ces organes se trouve la langue, acteur principal des modifications qui peut agir par constriction ou occlusion du conduit vocal. Les dents et les lèvres agissent également par occlusion ou constriction, à des degrés cependant moindres. Le conduit vocal est, la plupart du temps, constitué du seul conduit buccal. La luette et son prolongement vers le palais, le vélum, assurent normalement la fermeture du conduit nasal pendant la production de parole. Le conduit nasal peut, dans certains cas, être connecté au conduit vocal. Cette connexion permet de générer des sons supplémentaires en modifiant le Volume de la caisse de résonance normalement constituée par le seul conduit buccal. Une coupe de l'appareil phonatoire humain est donnée par la figure 1.1.

Les différents organes de la production de la parole et leur agencement peuvent servir de base à des modélisations du conduit vocal.

I.3.2 *Modèles Articulatoires*

Pour mieux comprendre le fonctionnement de l'appareil phonatoire, il semble judicieux d'essayer de simuler physiquement cet organe faisant intervenir la mécanique et la dynamique des fluides. De telles études partent du principe que la parole est un ensemble de mouvements rendus audibles plutôt qu'un ensemble de sons produits par des mouvements.

Un des modèles les plus connus est le modèle de Maeda qui caractérise le conduit vocal grâce à un ensemble de mesures réalisées sur des images radiographiques par le biais d'une grille semi-polaire.

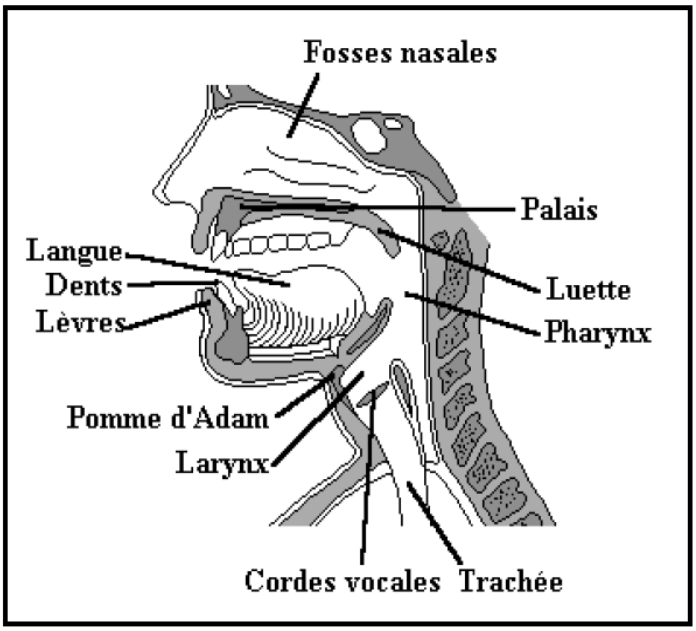


Fig. I.1 : Coupe de l'appareil phonatoire humain [2]

Ce modèle n'est pas à proprement parler articulatoire mais est plus simplement descriptif. Un ensemble de mesures détermine cependant une forme de conduit et permet donc de prévoir le son qui est y est associé. Certaines études vont actuellement dans le sens d'une exploration fonctionnelle du modèle de Maeda.

I.4 Mécanisme de la phonation

La parole est le résultat de l'action volontaire et coordonnée des appareils respiratoire et masticatoire sous le contrôle du système nerveux central qui reçoit en permanence des informations par rétroaction auditive et par sensation cénesthésiques. [3]

I.4.1 Modèle équivalent du Conduit vocal

La fonction vocale est difficile à cerner dans la mesure où il n'ya pas d'unité organique particulière exclusivement destinée à produire la voix, cette fonction se greffe sur différents systèmes analogiques.

On peut schématiser le système vocal humain de la manière suivante :

Un excitateur délivre un signal de source qui est modifié par la fonction de transfert d'un résonateur, le fonctionnement du système se déroule, selon la nature du signal excitateur

(signal périodique, signal de bruit continue, signal impulsionnel) Tout cela se passe dans un espace très restreint dans le larynx.

On sait qu'en mécanique des fluides, il suffit de créer une turbulence sur un écoulement gazeux pour créer un bruit, si cette turbulence persiste et devient régulière on obtient un son et l'émetteur vibre, c'est ce qui se passe au niveau des cordes vocales parce qu'il y a accord entre la force d'écoulement de l'air venant des poumons et la force de Bernoulli qui tend à fermer le larynx.

Pour que le son crée par les vibrations des cordes vocales devienne parole, il faut qu'il y'ait modulation par un résonateur. Ce qui consiste à transformer un phénomène continu en signal signifiant.

Cette modulation est attribuée au conduit vocal, qui être assimilée à un ensemble de résonateurs . [1]

Le conduit vocal peut êtres assimilé à des cellules élémentaires (Fig.1.2),

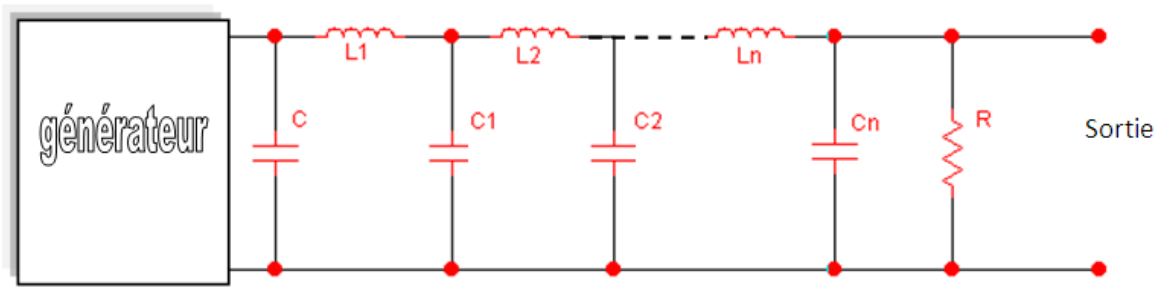


Fig. I.2 : Schéma électrique équivalent du conduit vocal. [1]

Chacune de ces cellules correspond à un schéma électrique (Fig. I.3). Au débit d'air du conduit et à une certaine pression, correspondent une intensité et une tension électrique (Fig. I.4).

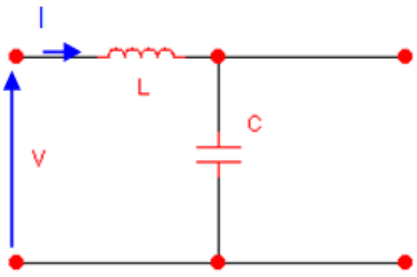


Fig.1.3 : Schéma électrique équivalent[1]

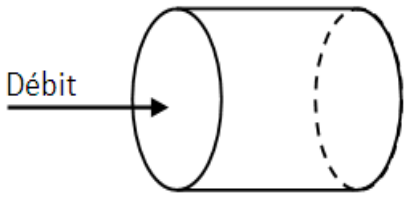


Fig.1.4 : Schéma acoustique équivalent[1]

Pour les sons voisés figure I.5. Le fonctionnement de la source périodique située à l'entrée de l'appareil phonatoire est caractérisé par le mécanisme de contraction (relaxation au niveau des cordes vocales), autrement dit, les sons voisés mettent les cordes vocales en vibrations quasi-périodiques, sous l'action de la pression de l'air et des muscles du larynx, le

spectre d'un son voisé de raies correspondantes aux harmoniques. L'enveloppe de ces raies représente les formants. Les trois premiers formants caractérise le spectre vocal .

I.4.2 Acoustique des Voyelles

Le Modèle « Source-Filtre » présenté sur la figure suivant :

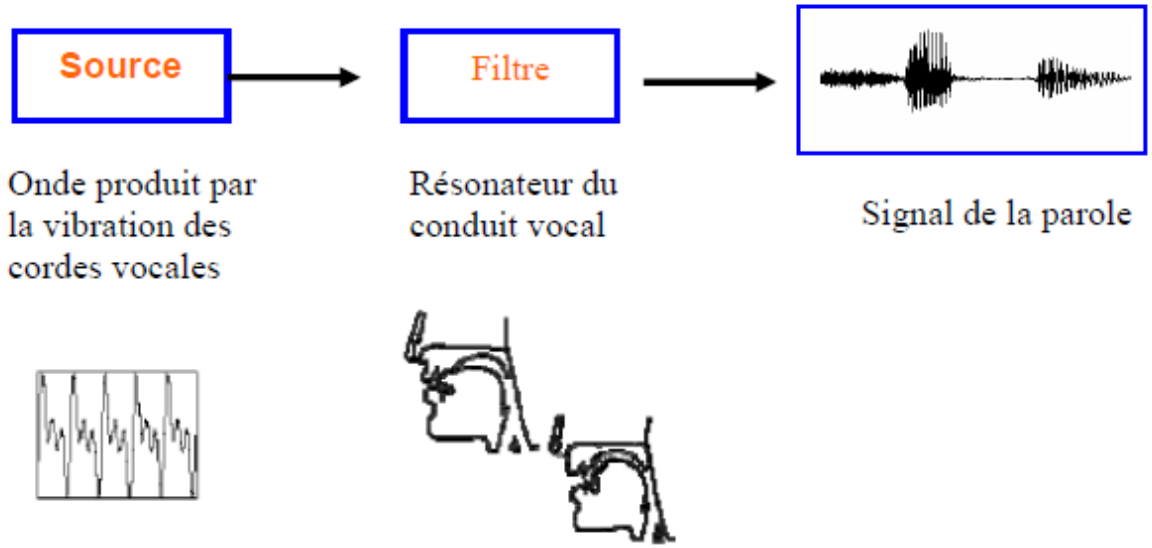


Fig. I.5 : Acoustique Des Voyelles [1].

Le rôle des cavités de résonance du conduit vocal est d'amplifier certaines des composantes de la source et d'amortir les autres. Les cavités de résonance ont un rôle de filtre.

Lorsque l'onde produite par la vibration de cordes vocales se propage dans les cavités, celles-ci se mettent à résonner à leur fréquence de résonance propre, ce qui va amplifier les composantes de la source qui sont à la même fréquence.

I.5 Son voisé et Non voisé

Les sons voisées sont produits lorsque les cordes vocales se mettent en vibration (Fig. I.6), Tandis que les sons non voisés, les cordes n'entrent pas en vibration, et le passage de l'air ne se fait pas librement, ce qui donne naissance à un bruit qui se propage sur les parois du conduit vocal, et leurs spectres ne présentent pas de structures de fondamental . [1]

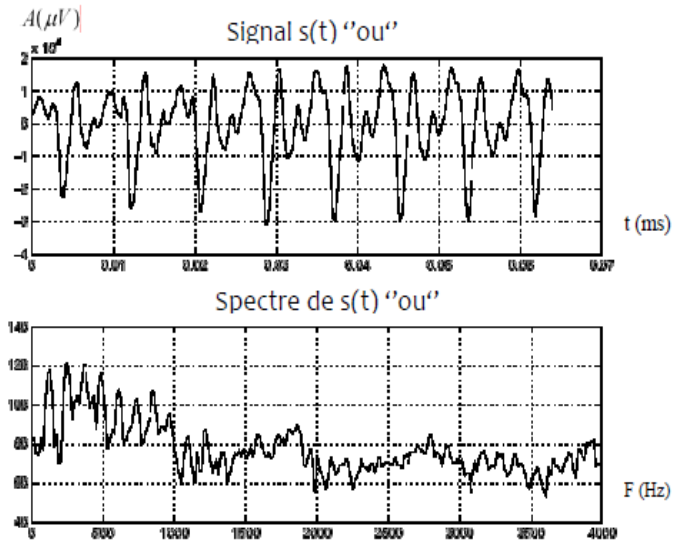


Fig. I.6 : Signal Voisé et Son Spectre[1]

Les sons voisés sont Quasi périodique et caractérisés par la fréquence fondamentale. Cette fréquence est un paramètre très important pour la synthèse de la parole car l'oreille est très sensible à ses variations. [3]

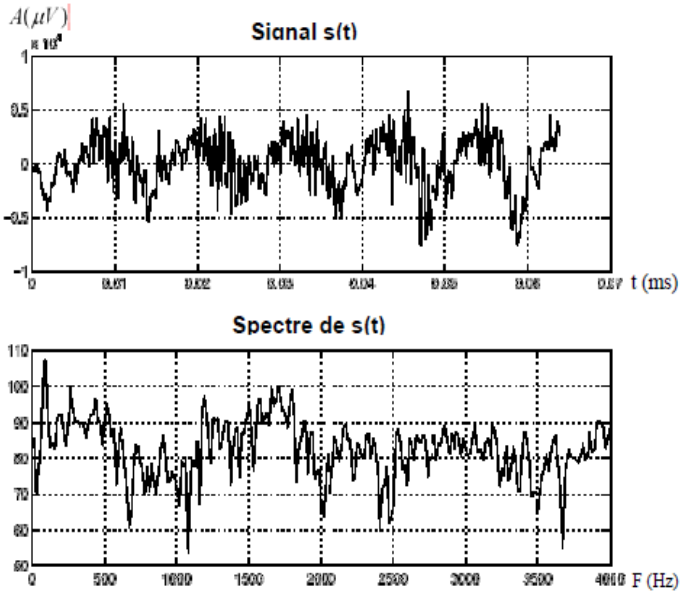


Fig. I.7 : Signal Non Voisé et Son Spectre[3]

I.6. Caractérisation Acoustique et Articulatoire des Voyelles

Bien évidemment, c'est la position des articulateurs qui va déterminer ce mode de propagation et donc les sons produits. Une différence fondamentale est faite entre les sons où l'air circule librement de la glotte aux lèvres (il s'agit des voyelles) et ceux où l'air rencontre des obstacles sur son trajet (il s'agit des consonnes et des semi-voyelles).

- les voyelles sont caractérisées acoustiquement par leurs formants (renforcement spectraux qui correspondent aux fréquences de résonance du conduit vocal),

- les voyelles sont caractérisées articulatoirement par le lieu de la plus forte constriction du conduit vocal.

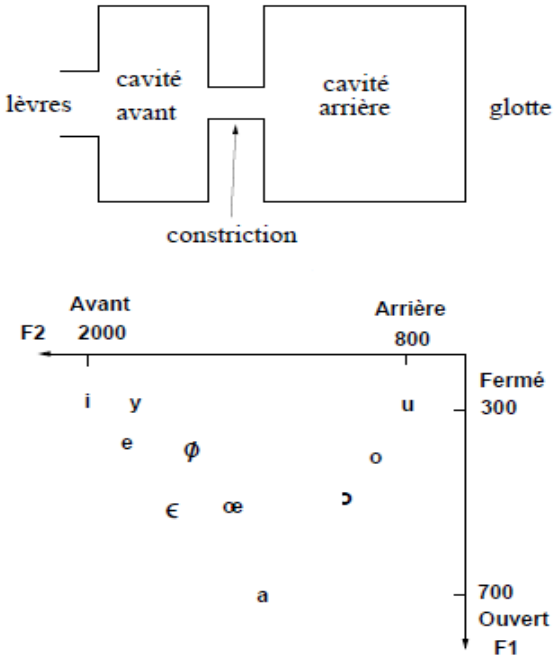


Fig. I.8 : Triangle Vocalique Pour le Français [5]

Les propriétés du conduit vocal varient dans le temps. D'un part, la forme du conduit vocal varie durant la production de la parole à cause des mouvements des articulations. D'autre part, des variations apparaissent au rythme du cycle glottique, à cause de la variation des corde vocales, en effet, les cordes vocales oscillent entre une phase fermée et une phase ouverte, ce qui modifie les caractéristique du système : pendant la phase fermée, le conduit vocal est fermé à la glotte et le signal de la parole résulte des résolution libres dans le conduit, tandis que pendant la phase ouverte, le conduit vocal est couplé acoustiquement avec la glotte et la trachée, ce qui modifie les résolution du conduit [1,3].

I.7 Classes phonétiques

I.7.1 Définition du phonème

La plupart des langues naturelles sont composées à partir de sons distinct, les phonèmes. Un phonème est la plus petite présente dans la parole et susceptible par sa présence de changer la signification d'un mot. Par exemple riz et rat. Pari et mari,…Le nombre de phonèmes est toujours très limité, normalement inférieur à cinquante ; on considéré

généralement que la longue française comprend 36 phonèmes, un système de phonèmes est un ensemble d'images acoustique Emmagasinées dans le cerveau du locuteur dans la mesure où celui-ci maîtrise de langue. La production d'un phonème donné laisse toutefois place à une certaine variabilité sur le plan acoustique. [4]

I.7.2 Différentes classes phonétiques

Les différents sons de la parole sont regroupés en classes phonétiques en fonction de leurs caractéristiques principales. Ces caractéristiques représentent des différences qui sont suffisamment importantes pour qu'il soit possible de classer les différents sons visibles sur un spectrogramme selon leur classe respective en très peu de temps et sans aucune écoute de la phrase correspondante. Les différentes classes phonétiques en français et en anglais sont :

I.7.2.1 Les voyelles

Cette classe correspond, à quelques nuances supplémentaires près, aux voyelles de l'écrit. Elles se caractérisent principalement par le voisement qui crée des formants. Ces formants, qui sont des zones fréquentielles de forte énergie, correspondent à une résonance dans le conduit vocal de la fréquence fondamentale produite par les cordes vocales. Ces formants peuvent s'élever jusqu'à des fréquences de 5 kHz mais c'est principalement les formants en basses fréquences qui caractérisent les voyelles. Cette caractéristique permet de distinguer grossièrement les voyelles en fonction de leur premier et deuxième formant.

I.7.2.2 Les Occlusives

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

I.7.2.3 Les Fricatives

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

I.7.2.4 Les sonantes

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

I.7.2.5 Les semi-consonnes (ou semi-voyelles ou glissantes)

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

I.7.2.6 Les liquides

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

I.7.2.7 Les nasales

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

I.7.2.8 Les affriquées

Cette classe est, elle aussi, propre à l'anglo-américain mais les affriquées peuvent également être observées dans le français québécois. Les affriquées sont composées d'un occlusive immédiatement suivie par une fricative de durée cependant plus faible que celle de la véritable fricative . [2]

I.8 Caractéristiques de phonation

Toute phrase d'une langue est perçue comme une suite d'éléments discrets phonèmes : voyelle et consonnes. Les 36 phonèmes de la longue française (42 en langue anglaise), suffiront pour représenter tous les mots de la même langue. [1]

L'alphabet phonétique international (IPA) associe des symboles phonétiques aux sons, de façon à permettre l'écriture compacte et universelle des prononciations

I.9 Les Méthodes d'analyse de la parole

Les deux problèmes principaux d'analyse de la parole présentée dans le schéma synoptique suivent figure I.9 [5].

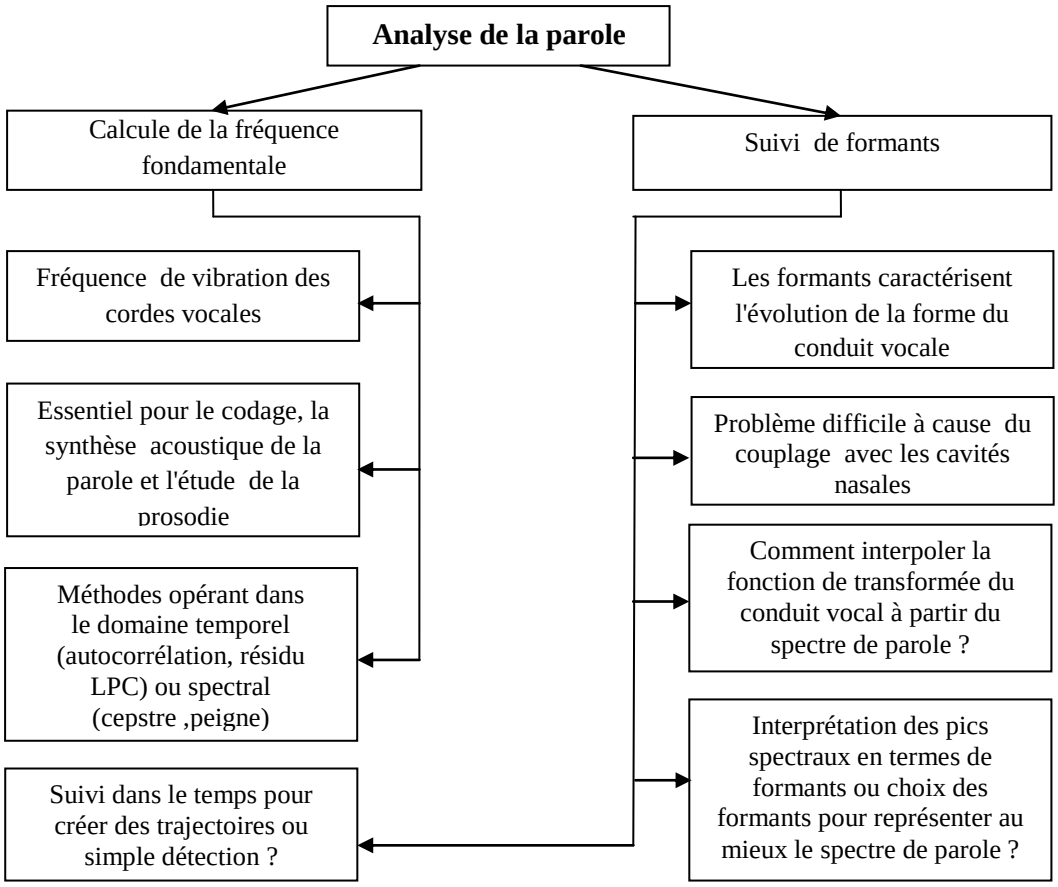


Fig I.9 : Schéma Synoptique de Différentes Méthodes d'analyse de la Parole [5]

I.10. Conclusion

Le traitement de la parole est aujourd'hui une composante fondamentale des sciences de l'ingénieur. Située au croisement du traitement du signal numérique et du traitement du langage, cette discipline scientifique a connu depuis les années 60 une expansion fulgurante, liée au développement des moyens et des techniques de télécommunications. L'importance particulière du traitement de la parole dans ce cadre plus général par la position privilégiée de la parole comme vecteur d'information dans notre société humaine.

Chapitre II

Technique de débruitage de signal parole par les transformées d'ondelettes

II.1 Introduction

D'abord, pourquoi a-t-on besoin des transformées, ou encore qu'est-ce 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 ce chapitre, nous allons étudier la transformée en ondelette et les différents méthodes de débruitage par ondelettes.

II.2 Un peu d'histoire

• 1805 : Analyse de Fourier
• 1965 : Transformée de Fourier rapide
• 1980 : Début des ondelettes *« ad hoc »*
pourquoi/quand cela marche (physique, vision, parole) ?
• 1983: Analyse d'image Multi-résolution (Burt)
• 1985: Transformée continue (Morlet & Grossman)
Reconstruction sans redondance ?
• 1986-87: Unification des travaux disparates (Mallat)
– analyse Multi-résolution et transformée discrète.
• 1988 : Classe d'ondelettes (Daubechies)
– compactes
– orthogonales
– nombre de moments quelconques
• 1990 : Les ondelettes attirent théoriciens et ingénieurs, le décollage !
• 1992 : Paquets d'ondelettes (Coifman)

II.3 Définition

II.3.1 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 signale. 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 [6,7].

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. [6, 8, 9].

II.3.2 La Transformée d'Ondelette

Une ondelette mère Ψ est une fonction *f* de base que l'on va translater et dilater pour recouvrir le plan temps – fréquences et analyser le signal [9, 10, 11]. On peu définir la transformée d'ondelettes d'un signal $f(t)$ comme une projection sur la base des fonctions ondelettes ($\Psi\left(\frac{t-b}{a}\right)$) :

$$TO(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t)\ \Psi\left(\frac{t-b}{a}\right) dt \qquad \text{avec} \qquad a,b \in R\ ,\ a \neq 0 \qquad \text{(II .1)}$$

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

En posant :

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right) \qquad \text{(II.2)}$$

Où : $\Psi_{a,b}(t)$ sont une famille d'ondelettes analysantes générales à partir d'une ondelette « mère » par dilatation (facteur a) et translation (paramètre b). L'équation (II.2) devient [9, 12, 13] :

$$TO(a,b) = \int_{-\infty}^{+\infty} f(t)\Psi_{a,b}(t)\ dt = \langle f, \Psi_{a,b} \rangle \qquad \text{(II .3)}$$

$\langle f, g \rangle$: est un produit scalaire entre deux fonctions f et g. La fonction ondelette doit vérifier un certain nombre de propriétés, la première d'entré elle se nomme condition d'admissibilité.

Soit $\psi(t) \in L^2$ (ensemble des fonctions à carré sommable) [6].

Alors : $\int_{-\infty}^{+\infty} \frac{|\Psi(w)|}{|w|} dw < \infty$ (II.4)

Cette condition permet d'analyser e 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 ω = 0) doit être nulles. Soit :

$$\Psi(w)|_{w=0} = 0 \quad \text{(II .5)}$$

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 (II.6) de façon équivalente sous la forme :

$$\int_{-\infty}^{+\infty} \Psi(t) = 0 \quad \text{(II .6)}$$

Donc ψ (*t*) doit être à moyenne nulle.ψ (*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 [9,10,15].

Cette ondelette agit comme un filtre passe bande, pour retrouver la partie du spectre éliminé par l'ondelette (les basses fréquences), on utilise une autre fonction appelée fonction échelle φ (*t*), sa moyenne est non nulle [6,10].

$$\int_{-\infty}^{+\infty} \varphi(t) \neq 0 \quad \text{(II .7)}$$

II.3.3 Illustration du changement d'échelle et de la translation

Le changement d'échelle sert à compresser ou à dilater l'onde mère, ce qui même a analysé les hautes ou basses fréquences continues dans un signal [9, 15].

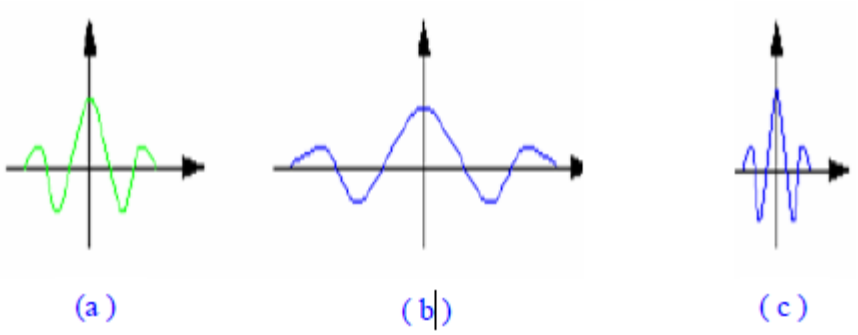


Fig. II.1 : illustration de la variation du facteur d'échelle, (a) l'Onde mère , (b) l'Onde mère délatté, (c) l'Onde mère compressé [15] .

II.4 L'Algorithme pour le calcul des coefficients d'ondelettes

La procédure de calcule des coefficients $C(s, u)\ f$ s'effectue comme suit :
on multiplie le signal et la fonction analysante et l'on calcule l'intégrale du produit, c'est un processus assez simple ; en fait il se déroule en cinq étapes [9, 15]:

1. On prend une ondelette et on l'a compare à une section au début du signal original.

2. On calcul le coefficient $C(a.b)\ x$ qui représente le degré de corrélation de l'ondelette avec cette portion du signal.

3. On translate l'ondelette vers la droite et on répète les étapes (1) et (2) jusqu'à ce que le signal soit couvert en entier.

4. On dilate l'ondelette et on répète les étapes de (1) à (3).

5. On recommence l'opération pour toutes les étapes à différentes échelles

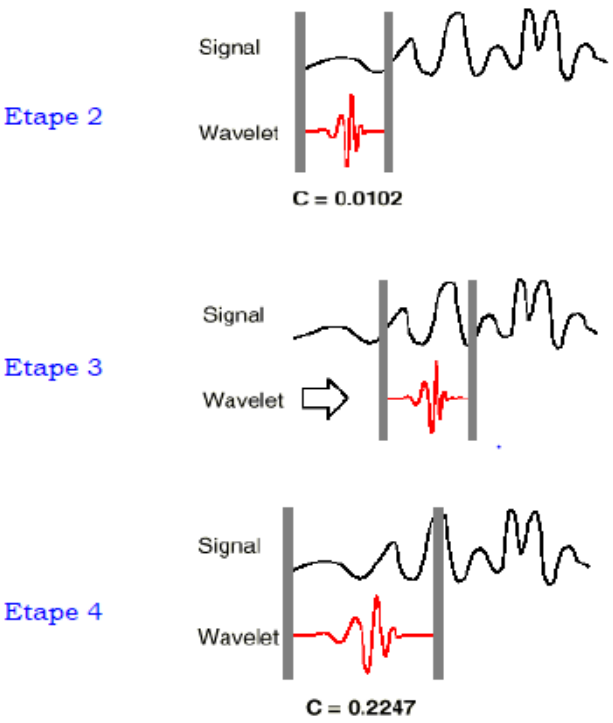


Fig. II.2 : les étapes de Calcul des coefficients d'ondelettes [9]

II.5 Transformée en Ondelettes Continue (CWT)

Une fois que l'ondelette mère est choisie le calcul commence par a = 1 et la CWT est calculée pour toutes les valeurs de a < 1 et a > 1. Cependant, selon le signal, une transformée complète n'est habituellement pas nécessaire. Pour tous les besoins pratiques, les signaux sont limités en largeur de la bande (band limite) et donc, calcul de la transformation pour un intervalle limité d'échelles est habituellement adéquat. Si le signal a une composante spectrale

qui correspond à la valeur courante de a, le produit de l'ondelette mère avec le signal à l'endroit où cette composante spectrale existe donne une valeur relativement grande. Autrement ce produit donne une valeur relativement petite ou nulle.

II.5.1 Initialisation de a

Pour la convenance, le procédé sera commencé à partir de l'échelle a = 1 et continuera pour les valeurs croissantes de a, i.e., l'analyse commencera à partir des haute fréquences et procédera vers les basses fréquences. Cette première valeur de « a » correspondra à l'ondelette la plus comprimée.

L'ondelette est placée au début du signal au point qui correspond à temps = 0. la fréquence d'ondelette à l'échelle 1 est multipliée par le signal et puis intégrée sur tout le temps. Le résultant de l'intégration est alors multiplié par le nombre constant 1/*a*. Cette multiplication est pour la normalisation d'énergie de sorte que le signal transformé ait le même énergie à chaque échelle. Le résultat final est la valeur de la transformation, i.e., la valeur de la CWT au temps zéro et à l'échelle a = 1. En d'autres termes, c'est la valeur qui correspond au point *b* =0, *a* =1 dans le plan temps-échelles

II.5.2 Incrémentation de b

L'ondelette à l'échelle a = 1 et ensuite transformée (ou décalée) vers la droite par une valeur τ à l'emplacement t = b, ce procédé est répété jusqu'à ce que l'ondelette atteigne l'extrémité du signal. Une rangée des points sur le plan temps-échelle pour l'échelle a =1 est maintenant accomplie. Puis, « a » est augmenté par une petite valeur. Notez qu'il s'agit d'une transformation continue, et donc, « b » et « a » doivent être incrémentés d'une façon continue. Cependant, si cette transformée a besoin d'être calculée par un ordinateur, alors les deux paramètre sont augmentés par un pas suffisamment petit. Ceci correspond à l'échantillonnage du plan temps-échelle. Le procédé ci-dessus est répété pour chaque valeur de « a ». chaque calcul pour une valeur donnée de « a » remplit une rangée simple correspondante du plan d'échelle de temps. Quand le processus est complété pour toutes les valeurs désirées de a, le CWT du signal a été calculé.

II.6 Transformée en Ondelettes Discrète

Dans le monde d'aujourd'hui, les ordinateurs sont presque tous les calculs. Il est évident que ni la WT, ni le STFT, ni le CWT ne puissent être pratiquement calculés en employant des

équations analytiques, des intégrales, …. etc. Il est donc nécessaire de discrétiser les transformées Comme dans la FT et le STFT, la manière la plus intuitive de faire ceci est simplement d'échantillonner le plan temps-fréquence (échelle). Encore intuitivement, l'échantillonnage du plan avec un taux d'échantillonnage uniforme semble le choix le plus normal. Cependant, dans le cas de la WT, le changement de l'échelle peut être employé, pour réduire le taux d'échantillonnage.

D'autre part cet aspect, la transformée en ondelettes telle qu'elle est définie est redondante, c'est-à dire que l'on obtient plus de coefficients d'ondelettes qu'il n'en est nécessaire pour décrire le signal de manière exhaustive. 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 ce rendre compte d'une part, de l'intérêt d'utiliser les ondelettes et d'autre part, de la manière dont on va discrétiser les valeurs de « a » et « b », on va regarder comment les ondelettes se déploient ou se répartissent en temps et en fréquence, par rapport, par exemple, à une transformée de Fourier à fenêtre glissante. Pour cela, on a représenté, en les juxtaposant, les supports temporels et fréquentiels des ondelettes dans le plan défini en abscisse par l'axe temporel et en ordonnées par l'axe fréquentiel. On visualise ainsi comment est découpé le plan Temps-Fréquence pour chaque type de transformée figure (II.3) [6, 16].

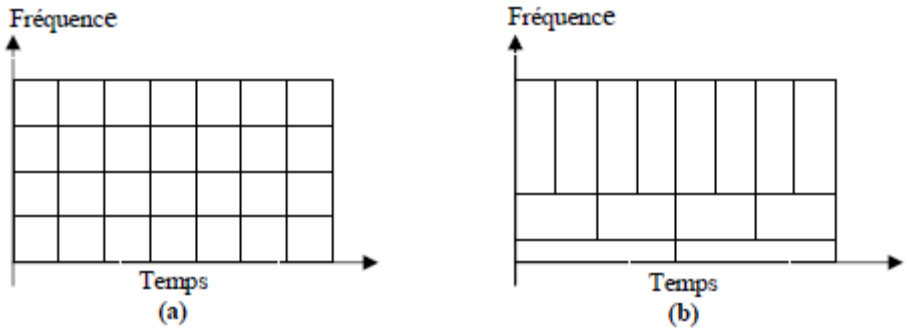


Fig. II.3 : Plan Temps-Fréquence : a) transformée de Fourier à fenêtre Glissante, b) transformée en Ondelette [16].

On remarque que les résolutions temporelle et fréquentielle pour la transformée de Fourier à fenêtre glissant sont constants, alors que pour la transformée en ondelette les résolutions varient en sens inverse [7, 15].

Le pavage Temps-Fréquence utilisé sur la figure (II.3,a) précédent suggère une méthode de discrétisation exponentielle pour les échelles et pour le temps.

L'une des caractéristiques fondamentales de la transformée en ondelettes continue est sa redondance. En effet, l'information contenue dans un signal est représentée dans un espace à deux dimensions, il s'agit du plan temps-échelle *(t ,a).* Des coefficients d'ondelettes voisins contiennent des informations communes. Il est toutefois possible de réduire cette redondance

en remplaçant la famille continue d'ondelettes par une famille indexée par des variables de temps et d'échelle discrètes, et les intégrales par des sommes discrètes. Il est préférable de réduire au maximum cette redondance en fixant $a=2^{-j}$ et $b= k2^{-j}$, la famille d'ondelettes correspondantes devient:

$$\psi_{j,k}(t) = 2^{\frac{j}{2}} \psi(2^j t - k) \quad \text{(II .8)}$$

Cette transformée est appelée transformée dyadique. En choisissant Adéquatement ψ, la famille $\psi_{j.k}$ constitue une base orthonormée, on pourra dès lors récupérer le signal original par la transformée inverse qui s'écrit alors :

$$f(t) = \sum_j \sum_k c_{j,k} \psi_{j,k} \quad \text{(II .9)}$$

Avec:

$$c_{j,k} = \int_{-\infty}^{+\infty} f(t) \psi_{j,k}(t) dt \quad \text{(II .10)}$$

Cette équation (II.10) représente les coefficients d'ondelettes qui fournissent donc une représentation alternative de ces fonctions, sans perte d'information ni redondance.

Soit $a = a_0^j \quad et\ b = n.b_0 a_0^j$ avec: $a_0 \,\&\, b_0 \in Z$ on obtient alors une transformée en ondelettes discrète :

$$TO(n,j) = a_0^{\frac{-j}{2}} \int_{-\infty}^{+\infty} f(t).\Psi(a_0^{-j}\, t - n b_0)\, dt \quad \text{(II .11)}$$

Si on choisi $a_0 = 2$ & $b_0 = 1$, on parle alors d'une Transformée en Ondelette dyadique.

$$TO(n,j) = \sqrt{2}^{-j} \int_{-\infty}^{+\infty} f(t).\Psi(2^j\, t - n) dt \quad \text{(II .12)}$$

Remarque : il faut noter dans ce cas que c'est la transformée que est discrète, et non l'Ondelette qui reste toujours une fonction continue [9].

II.7 Analyse Multi-résolution

On construira une analyse multi – résolutions à l'aide du son espace d'approximation *Vj* (généré par la fonction échelle) emboîtés les uns dans autres, tel que le passage de l'un à l'autre soit le résultat d'un changement d'échelle (zoom) [17].

L'espace des détails *Wj* (généré par la fonction Ondelette) vient compléter l'analyse. On peut définir pour *Vj* son complément orthogonal *Wj* dans V_{j-1} tel que :

$$V_{j-1} = V_j \oplus W_j \qquad \text{(II .13)}$$

Le schéma de la décomposition est représenté symboliquement sur la figure (II.4) dans la quelle la largeur des rectangles symbolisant les sous espaces est proportionnelle à la densité de l'échantillonnage réalisé par la projection du signal dans le sous espace considéré [9, 12, 18].

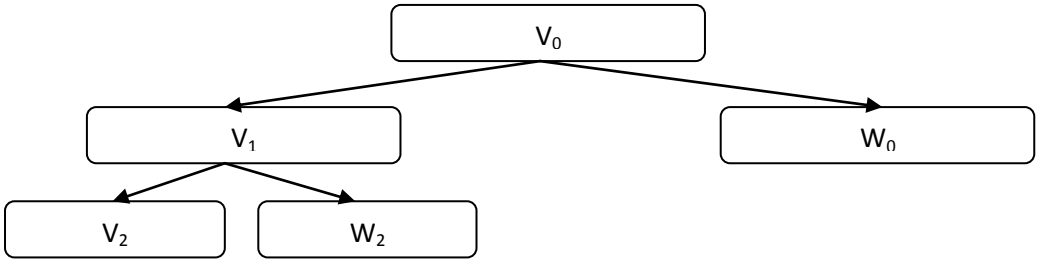


Fig. II.4: Schéma d'analyse Multi-résolution [18]

La décomposition en Ondelettes est similaire à la décomposition de Gabor, un signal s'écrit sous la forme d'une superposition de telles ondelettes décalées et dilatée. Les poids de ces ondelettes dans la décomposition (appelés les coefficients d'ondelette) forment la Transformée en Ondelettes, qui est donc une fonction de deux variable; le temps et l'échelle (ou dilatation) [11, 15, 16]. Avec l'analyse par ondelettes, le signal est décomposé en fonction élémentaires, engendrée par des Transformations simples d'une fonction de base qui est translatés et dilatée (fig II.5).

Les Ondelettes sont très étendues et dilatée pour étudier les basses fréquences (les grandes échelles) et très fines pour étudier des phénomènes plus transitoires (hautes fréquences, ou petites échelles). Cette procédure, développée par S. Mallat et systématisée par I. Daubechies, pour le nom de Multi-résolution, et suggère une interprétation différente de l'analyse par ondelettes, basée sur les idées de lissage, ou d'approximation des fonctions. Les décompositions en ondelettes existent dans plusieurs versions, que l'on choisit en on fonction de l'application visée [6,19].

La décomposition du signal en ondelettes permet à l'utilisateur de s'adapter plus au signal selon son contenu fréquentiel pour extraire les informations utilisées. Dans ce cas tout changement de fréquence se traduira par un changement sur plusieurs niveaux d'échelles.

La classification sera appliquée sur certains signaux de détails et non sur le signal original [20].

$$Détail_m\,(t) = approx_{m-1}\,(t) - approx_m\,(t)$$

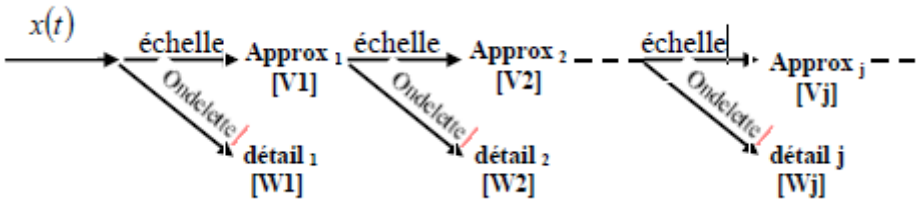


Fig. II.5 : Analyse Multi-résolution : décomposition successive en approximation et détails [20].

La transformée en ondelettes mesure la similitude entre le signal et l'ondelette pour ses différentes dilatations (résolution) et translation (localisation) [16, 17, 21].

$$V_{j+1} = V_j + W_j \qquad \text{(II .14)}$$

Echantillonnage de départ Approximation Détails

II.8 L'algorithme pyramidal de Mallat

L'algorithme pyramidal [22,23, 24] est une méthode récursive basée sur une succession de convolutions. En effet, Mallat a montré que les coefficients d'ondelettes peuvent être calculés à partir d'une transformée pyramidale mise en œuvre à l'aide de filtres numériques, récursifs ou non. Le principe de la transformée pyramidale consiste dans la décomposition en banc de filtre (figure II.6) du signal à analyser à l'aide d'une paire de filtres miroirs en quadratures. L'un de ces filtres fournira les coefficients d'ondelettes (ou détails), le second les coefficients d'approximation. L'approximation est elle-même à son tour décomposée par une seconde paire de filtres, 1'ensemble constituant une pyramide de filtres. Cet algorithme est par ailleurs inversible, la reconstruction s'obtient simplement par inversion des filtres dans le cas de bases orthogonales (figure II.6).

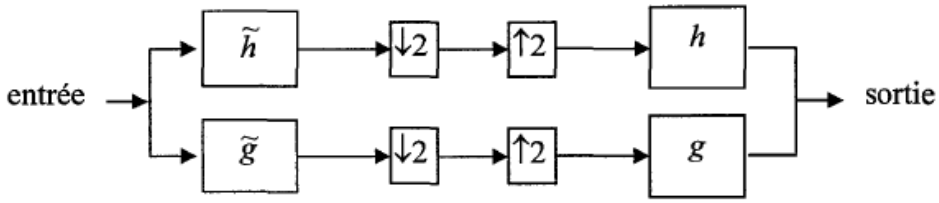


Fig. II.6 :Décomposition en banc de Filtre d'une analyse multi-résolution [24]

Dans la figure II.6, $\tilde{h}$ est le filtre conjugué en quadrature de *h*, et $\tilde{g}$ est le filtre conjugué en quadrature de *g*. Mallat [24] a déduit les formules de décomposition et de reconstruction suivantes :

Formules de décomposition (II.14) :

$$\boldsymbol{a_n^j = \sum_i \tilde{h}[2n-1]\, a_i^{j-1}} \quad \text{(II.15)}$$

$$\boldsymbol{d_n^j = \sum_i \tilde{g}[2n-1]\, a_i^{j-1}}$$

Formules de reconstruction (II.15) :

$$\boldsymbol{a_n^{j-1} = \sum_k a_k^j\, h\,[n-2k] + \sum_k g[n-2k]\, d_k^j} \quad \text{(II.16)}$$

Les filtres conjugués respectent les relations suivantes (II.15) :

$$\boldsymbol{\tilde{h}[n] = h[-n]} \quad \text{(II.17)}$$

$$\boldsymbol{\tilde{g}[n] = g[-n]}$$

De même, Les filtres h, et g peuvent se déduire à partir des fonctions d'ondelettes et d'échelle par produit scalaire comme suit:

$$h[n] = \langle \phi, \phi_{-1,n} \rangle \quad \text{(II.18)}$$

$$h[n] = \langle \psi, \phi_{-1,n} \rangle \qquad \text{avec :} \qquad \phi_{-1,n}(t) = \overline{2^{-12}}\phi(2x-n)$$

II.9 Débruitage par ondelette :

L'un des plus grand succès des ondelettes est le débruitage [25]. En effet, cette technique repose essentiellement sur des algorithmes simples et performants et s'est avérée souvent beaucoup plus efficaces que les techniques traditionnelles souvent plus lourdes et moins efficaces. Cette approche se base sur la construction d'estimateurs statistiques à base d'ondelettes et nécessite essentiellement le calcul d'un seuil qui correspond à 1'amplitude maximale du bruit et dépend de 1' énergie du signal et du bruit. Ces méthodes reposent sur le fait que la représentation de plusieurs type de signaux dans le domaine de la transformée en

ondelettes est creuse et l'on n'a donc besoin d'estimer que quelques grands coefficients pour obtenir une bonne estimation de la fonction.

II.9.1 Algorithme de débruitage par ondelettes :

L'algorithme de base de débruitage par ondelettes peut être décomposé en trois étapes essentielles [26] :

- La décomposition par la transformée en ondelettes.
- Le seuillage des coefficients issus par la décomposition.
- La reconstruction par la transformée en ondelettes inverse.

En effet, à partir du signal à débruiter, on décompose le signal sur une base orthogonale d'ondelettes. On effectue ensuite une opération de seuillage qui consiste à éliminer les coefficients qu'on considère comme du bruit ou à les réduire en fonction du seuil calculé. En dernier lieu, on applique la transformée en ondelettes inverse sur les coefficients seuillés et on récupère le signal débruité.

Formulons le problème par le modèle mathématique. Soit *y* un signal corrompu par un bruit *e.* On peut ainsi écrire :

$$y(i) = x(i) + e(i) \qquad (i = 0, 1, \ldots, N-1) \tag{II.19}$$

où *N* est la taille du signal *y.*

La transformée en ondelettes étant une fonction linéaire, on l'applique sur l'équation (II.10) et on obtient :

$$Wy = Wx + We \tag{II.20}$$

W étant la transformée en ondelettes. Soit *TO* la fonction de seuillage par ondelettes, alors le schéma de débruitage par ondelettes peut être exprimé selon l'équation :

$$X = w^{-1}\,(T(Wy)) \tag{II.21}$$

où *T(Wy)* est le vecteur des coefficients de la TOD seuillés et $\hat{x}$ le signal débruité. Ceci s'interprète simplement par l'application de TOD sur le signal bruité *y,* on obtient ainsi le vecteur des coefficients de la TOD *(Wy)* sur lequel on applique la fonction de seuillage *(T(Wy)).* Le signal débruité $\hat{x}$ quand à lui, est obtenu en appliquant la transformée en ondelettes inverse sur le vecteur des coefficients seuillés $(W^{-1}(T(Wy)))$.

II.9.2 Les méthodes de seuillage

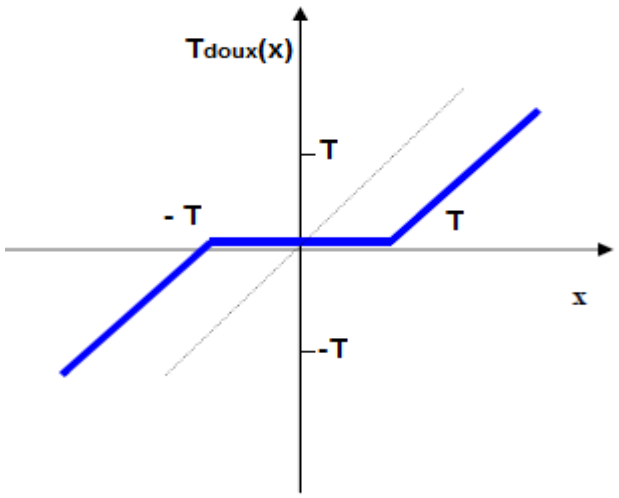
Les méthodes de seuillage les plus connues, introduites par Donoho [27], sont le seuillage doux et le seuillage dur. Des variantes de ces types de seuillage [28,29] ont été développées pour tenter d'améliorer ces méthodes.

II.9.2.1 Seuillage doux **ou "soft thresholding"**

Le seuillage doux [26,27] consiste à éliminer tout coefficient au dessous du seuil et à soustraire ce seuil des autres coefficients. Soit un vecteur *x* quelconque. La fonction de seuillage doux T_{doux} (') appliquée sur *x* est donnée par 1' équation :

$$T_{doux}(x) = \begin{cases} sign(x)(|x| - \lambda) & |x| \geq T \\ 0 & |x| < T \end{cases} \quad \text{(II.22)}$$

Où T est la valeur du seuil.



Le seuil T est généralement choisi pour être avec une probabilité juste au dessus du niveau maximum des coefficients *WB*[*m*] du bruit. La réduction de T de l'amplitude de tous les coefficients bruités permet de s'assurer que les amplitudes des coefficients estimés sont plus petites que celles des coefficients d'origine. Dans une base d'ondelettes où les coefficients de grande amplitude correspondent à des variations transitoires du signal, cela signifie que l'estimation ne conserve que des transitions du signal d'origine, sans en ajouter d'autres dues au bruit. Le coefficient seuillé sera donc plus petit que le coefficient du signal.

Ce type de seuillage garantit que le signal obtenu sera toujours plus régulier que le signal de départ. Ces techniques donnent des résultats qui dépendent des seuils choisis et des images que l'on prend.

En effet, quand on seuille les coefficients d'ondelettes pour enlever le bruit, on perd aussi quelques caractéristiques de signal parole. Il arrive après le débruitage que :

- Les coefficients gardés contiennent du bruit.
- On a supprimé les petits coefficients.

L'erreur en seuillage doux est, en général, plus grande qu'en seuillage dur. Il existe donc des méthodes pour tenter de trouver un seuil adapté à chaque signal parole et c'est ce que nous allons voir par la suite.

Le graphe de la figure II.7*:* est un exemple de seuillage doux appliqué à la fonction $y = t$ avec un seuil: $T = 0{,}5$.

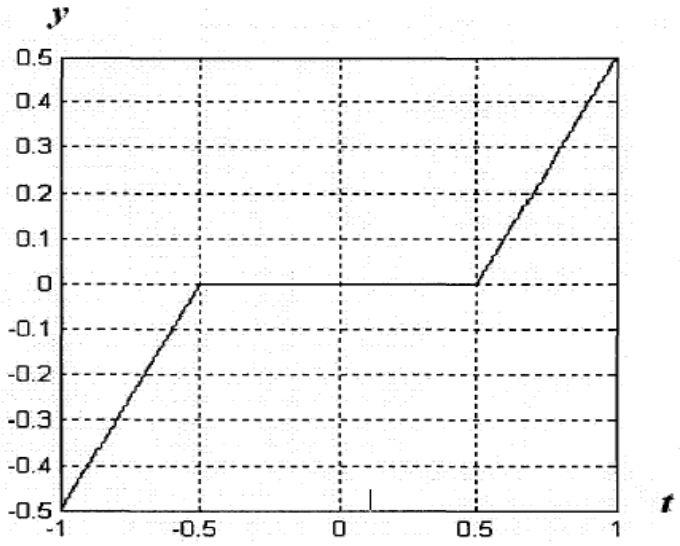


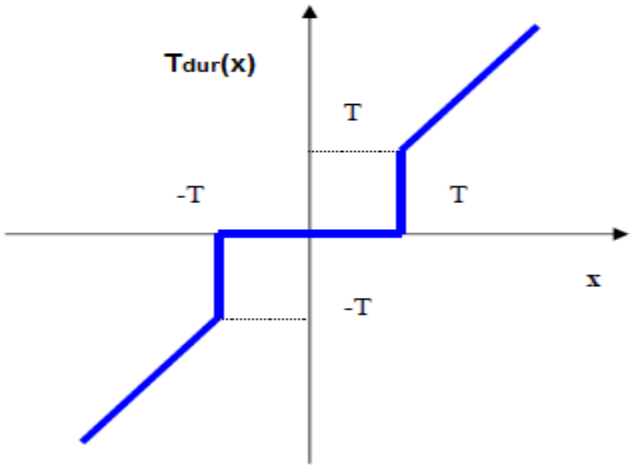
Fig. II.7 : Graphe de la fonction du seuillage doux [26]

II.9.2.2 Seuillage dur ou "hard thresholding"

Le seuillage dur [26] est plus catégorique que le seuillage doux du fait qu'on considère un coefficient donné soit comme représentant totalement un bruit pur donc à éliminer, ou comme un coefficient représentant une portion du signal donc à conserver. La fonction de seuillage dur $T_{dur}(')$ appliquée sur x est donnée par l'équation :

$$T_{dur}(x) = \begin{cases} x & |x| > T \\ 0 & |x| \leq T \end{cases} \qquad \text{(II.23)}$$

Où: T est la valeur du seuil.



Le graphe de la figure II.5 est un exemple de seuillage dur appliqué à la fonction *y= t* avec un seuil λ = 0,5.

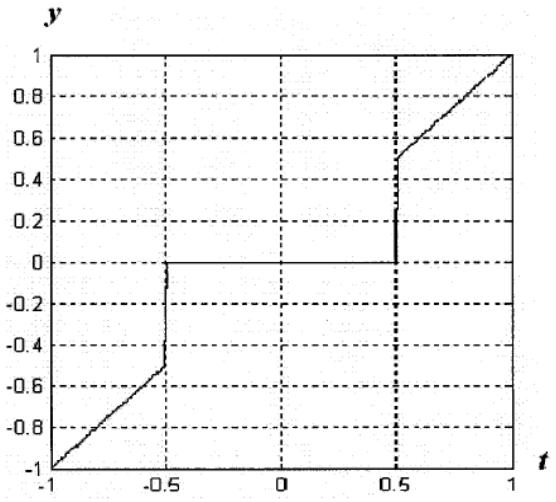


Fig. II.8 : Graphe de la fonction du seuillage dur [26]

Le seuillage dur est caractérisé par sa fonction qui est très simple à mettre en œuvre, et sa propriété un peu trop brutale à écarter le bruit, ce qui implique souvent une altération excessive du signal original.

II. 9.2.3 Le seuillage dur modifié

Pour corriger le problème de la discontinuité fréquentielle, K won a proposé une version modifiée du seuillage dur [31]. La formule (II.21) représente la fonction du seuillage de Kwon.

$$T_{dur-Mod}(x) = \begin{cases} x & |x| > \lambda \\ \frac{1}{\mu} sign(x).\lambda.\left((1+\mu)^{\left|\frac{x}{\lambda}\right|} - 1\right) & |x| < \lambda \end{cases} \quad \text{(II.24)}$$

Où λ est la valeur du seuil et *Il* est une constante.

La figure II.6 montre les trois types seuillage (doux, dur, dur modifié) superposés et appliqués à la fonction $y = t$ avec un seuil $\lambda = 0{,}5$.

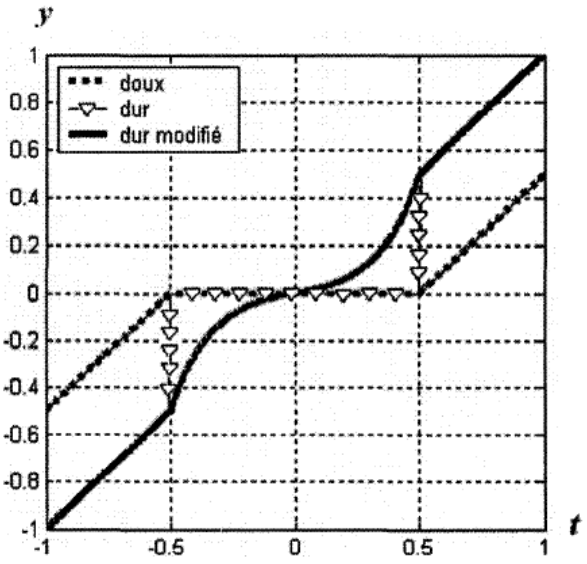


Fig. II.9: Graphe de la fonction du seuillage dur modifié (seuillage de Kwon) [31]

Cette technique de seuillage procure des résultats très satisfaisant du fait qu'elle est très performante quand à l'élimination d'une grand partie du bruit et qu'elle altère moins le signal par rapport aux deux autres méthodes de seuillage.

II.9.3 Sélection du seuil

Le calcul du seuil est une tâche très importante dans la mesure où on estime le niveau du bruit en fonction de sa nature et la nature du signal. Avec l'hypothèse que le bruit est un bruit blanc gaussien de variance σ^2, Donoho et Johnstone ont alors montré [27] que le risque induit par un seuillage (dur ou doux) sur les coefficients d'ondelettes pouvait être encadré par des valeurs proches de la borne inférieure obtenue avec des estimateurs d'oracle. Un estimateur d'oracle est un estimateur construit en connaissant le signal recherché. Le risque est défini par 1'équation :

$$R(x - \hat{x}) = E\left(\|x - \hat{x}\|^2\right) = E\left(\ \sum_{i=0}^{N-1}\left(x - \hat{x}(i)\right)^2\right) \qquad \text{(II.25)}$$

où N est la taille du signal x.

Le théorème de Donoho et Johnstone définit un seuil T pour les coefficients de la transformée en ondelettes [10] et dont la valeur est:

$$T = \sigma\sqrt{2\log\ (N)} \qquad \text{(II.26)}$$

avec un risque $r_d(x)$ qui est encadré par deux valeurs comme le montre l'expression:

$$r_0(x) \leq r_d(x) \leq (2\log(N) + 1)(\sigma^2 + r_0(x)) \quad \text{(II.27)}$$

où $r_0(x)$ est défini comme suit :

$$r_0(x) = \sum_{i=0}^{N-1} \min(d_i^2, \sigma^2) \quad \text{(II.28)}$$

où les *d;* représentent les coefficients de la transformée en ondelettes de *x.*

Lorsqu'on utilise le seuil de Donoho et Johnstone, il est nécessaire d'avoir une estimation de la variance σ^2 du bruit. Pour l'estimer à partir des données *x(i),* il faut supprimer l'influence de *x(i).* L'estimateur utilisé est celui de la médiane des coefficients d' ondelettes proposé par Donoho. Cette estimation nous ramène à discuter du niveau de composition à considérer pour appliquer le seuillage. Une classification du seuillage a été faite dans ce sens, on distingue ainsi les différents cas suivants :

- seuillage global.
- seuillage dépendant du niveau.
- seuillage dépendant du nœud.

II.9.3.1 Seuillage global

Appliquer un seuillage global revient à estimer un seuil unique quelque soit le niveau de décomposition. Le seuil [26,27] est calculé en utilisant l'équation:

$$T = \sigma\sqrt{2\log\ (N)} \quad \text{(II.29)}$$

Où σ est l'écart-type ou le niveau du bruit qui est estimé par l'équation (II.31):

$$\sigma = \frac{mediane\ (|d_1|)}{0.6745} \quad \text{(II.30)}$$

et où d_1 représente les coefficients des détails obtenus au niveau 1 correspondant au nœud de la plus haute résolution.

II.9.3.2 Seuillage dépendant du niveau

Le calcul d'un seuil global n'étant plus adéquat dans le cas d'un bruit coloré, Johnstone et Silverman [30] ont proposé un seuil dépendant du niveau donné par:

$$T_j = \sigma_j \sqrt{2\log(N)} \quad \text{(II.31)}$$

Où j représente le niveau de décomposition et σ_j l'écart-type du bruit calculé dépendamment du niveau en question qui est donné par :

$$\sigma_j = \frac{mediane\ (|d_1^j|)}{0.6745} \quad \text{(II.32)}$$

où d_1^j représente les coefficients des détails obtenus au niveau j correspondant au nœud de la plus haute résolution.

II.9.3.3 Seuillage dépendant du nœud

Une extension du seuillage dépendant du niveau est celui du seuillage dépendant du nœud qui consiste à effectuer un calcul du seuil pour chaque noeud [28]. Le seuil est calculé en utilisant :

$$T_{j,k} = \sigma_{j,k} \sqrt{2\log(N)} \quad \text{(II.33)}$$

Où j représente le niveau de décomposition , k la sous-bande, et $\sigma_{j,k}$ l'écart-type du bruit calculé dépendamment du niveau et la sous-bande en question qui est donné par :

$$\sigma_{j,k} = \frac{mediane\ (|d_k^j|)}{0.6745} \quad \text{(II.34)}$$

Où d_k^j représentent les coefficients obtenus au niveau j et la sous bande k. Notons qu'une valeur de (j,k) identifie bien un nœud spécifique d'une décomposition, avec j = 1,2 ... , et k= $1,2^j$ (figure II.10)

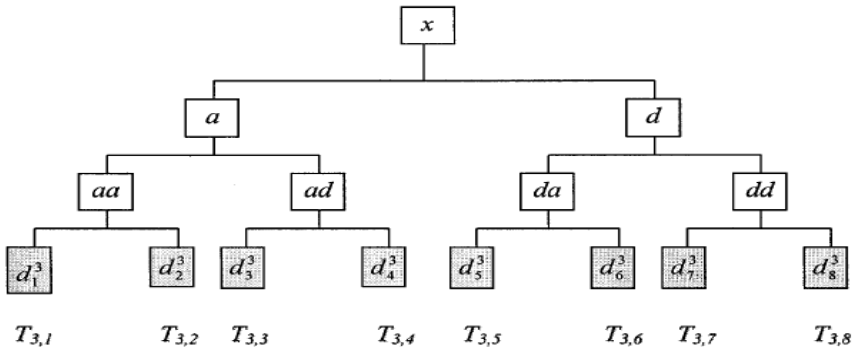


Fig. II.10 : Exemple d'un arbre de décomposition en paquets d'ondelettes à trois niveaux et désignation des seuils requis dans le cas d'un seuillage dépendant du nœud [28]

II. 10 Exemple de décomposition par ondelettes

Cet exemple qui est obtenue à l'aide du logiciel MATLAB, représente une décomposition d'un signal sinusoïdal qui comporte deux sinus côte à côte et qui représente une rupture de fréquence à l'instant t = 500. Ce signal est donné par 1' équation :

$$S(t) = \begin{cases} \sin(0.03t) & si \quad t \leq 500 \\ \sin(0.3t) & si \quad t > 500 \end{cases} \tag{II.35}$$

Le signal *s(t)* est constitué d'un signal 'lent' et d'un signal 'moyen' de part et d'autre de l'instant $t = 500$ où on remarque un changement brusque ou rupture de fréquence. La figure 5 montre que les détails d_1 et d_2 permettent de détecter cette discontinuité qui est bien localisée à cet instant. On peut remarquer également que les détails sont nuls ou presque nuls dans la partie lente du signal. De même, le signal lent apparaît dans 1' approximation a_5 où les hautes fréquences ne sont pas représentées.

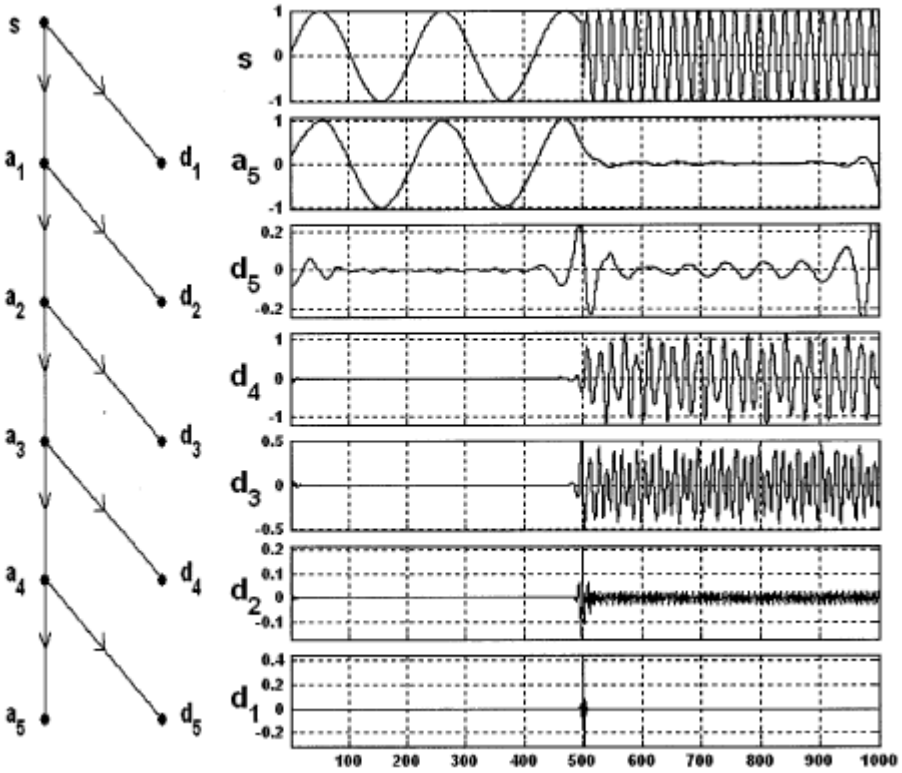


Fig. II.11: Exemple d'une décomposition d'un signal par la TO [28]

II.10 Conclusion :

Nous avons défini l'analyse multi-résolution qui a permis une organisation hiérarchique pour une décomposition donnant naissance à une meilleure résolution en fréquence. Nous avons également présenté les formules de décomposition et de reconstruction de Mallat qui rendent possible un calcul facile des coefficients via une méthode récursive utilisant des filtres conjugués.

Chapitre III

Techniques d'évaluation des algorithmes de rehaussement de signal parole

III.1 Introduction

Le jugement et l'évaluation de la qualité du signal de parole après un certain traitement (débruitage, codage, compression, transmission,...) ne peut se faire d'une manière satisfaisante qu' à partir de tests subjectifs. En effet, ce qui intéresse l'auditeur, c'est la qualité perceptuelle du signal. Il est tout à fait évident qu'on ne peut pas se contenter d'une ou deux personnes pour évaluer la qualité du signal de parole. En effet, une telle évaluation risque d'être trop subjective. Afin d'éviter ce genre de problème, tout test subjectif doit être réalisé avec plusieurs personnes d'âge et de sexe différents et dans des conditions expérimentales bien déterminées. Dans ce cadre, l'Union International des Télécommunications (UIT) a développé des normes spécifiant les procédures expérimentales à suivre pour évaluer la qualité subjective du signal vocal.

Dans ce chapitre, on va étudier les différentes techniques d'évaluation de la qualité de la parole dont nous intéressons aux méthodes objectives temporel (rapport signal / bruit (SNR) et segment SNR) et perceptuel (PESQ) .

III.2 Classification des méthodes d'évaluation de la qualité de la parole

La mesure de la qualité de la parole peut être effectuée en utilisant soit des méthodes subjectives ou objectives comme le montre la figure III.1. La note moyenne d'opinion MOS est la mesure subjective de la qualité vocale la plus largement utilisée et recommandée par l'ITU [31]. Nous rappelons que la valeur MOS est obtenue en moyennant les scores donnés par les observateurs pour évaluer la qualité vocale.

Comme le cas d'évaluation de la qualité de la parole, le problème inévitable des méthodes de mesure subjective telle que le MOS, consiste qu'elles sont coûteuses, manquent d'automatisation et ne peuvent pas être utilisées pour la surveillance à grande échelle de la qualité de la voix dans une infrastructure réseau. Ces inconvénients ont rendu très attractives les méthodes objectives pour l'évaluation de la qualité afin de répondre à la demande pour la mesure de qualité de la voix dans les réseaux de communication.

La mesure objective de la qualité de la voix, dans les réseaux de communication modernes, peut être intrusive ou non intrusive. Les méthodes intrusives analysent les signaux vocaux transmis et reçus. Ces dernières comparent le signal vocal de référence avec le signal correspondant déformé. Les méthodes non intrusives utilisent uniquement le signal de parole dégradé pour estimer la qualité vocale correspondante.

Les méthodes intrusives sont plus précises que les non intrusives, mais elles ne sont pas adaptées pour le suivi en temps réel du trafic à cause de la nécessité de disposer des données

de référence. Une méthode typique intrusive est basée sur la dernière norme ITU P.862, *Perceptual Evaluation of Speech Quality* (PESQ) [32].

Les méthodes non intrusives sont plus appropriées pour la surveillance du trafic en temps réel vu qu'elles n'ont pas besoin du signal de référence. Il y a deux catégories de méthodes non intrusives : celles qui sont basées sur le signal dégradé reçu (*signal-based method*) et celles qui sont basées sur des paramètres du réseau ou de la voix (*parameter-based method*). Un exemple de méthode non-intrusive basée sur le signal est le modèle d'appareil vocal : *vocal tract model* [52], qui vise à prédire la qualité vocale en analysant directement le signal de parole en cours d'écoute (un signal dégradé) sans le signal de référence. L'exemple le plus connu et le plus largement utilisé des méthodes non intrusives paramétrique, est le ITU T *E-model* [53].

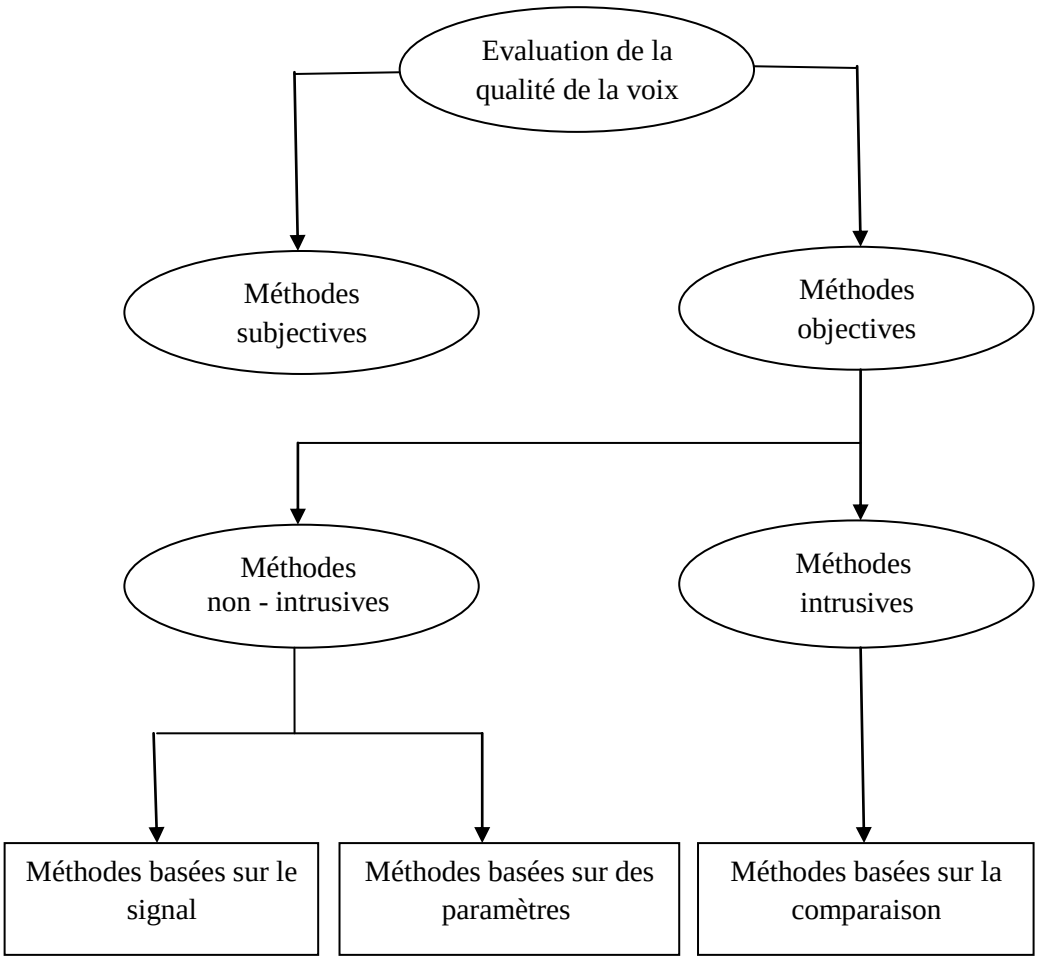


Fig. III.1: Classification des méthodes d'évaluation de la qualité de la parole [31]

III.3 Mesure subjective de la qualité vocale

Comme pour la vidéo, les méthodes subjectives sont cruciales pour l'analyse comparative des méthodes objectives. L'ITU P.800 [50] décrit plusieurs méthodes et

procédures pour la conduite des évaluations subjectives de la qualité de transmission. La méthode la plus couramment utilisée est l'évaluation par catégories absolues (*Absolute Category Rating* ACR) qui donne la note moyenne d'opinion (MOS). La méthode d'évaluation par catégories de dégradation (*Degradation Category Rating* DCR) est également utilisée dans certaines occasions. Cette méthode donne la dégradation de la note moyenne d'opinion : *Degradation Mean Opinion Score* (DMOS).

L'évaluation subjective de MOS est habituellement effectuée dans des conditions contrôlées en laboratoire.

III.3.1 Evaluation par catégories absolues (ACR)

Les tests ACR sont le plus couramment utilisés pour évaluer la qualité intégrale de la parole (ITU-T Rec.P.800 [50]). Dans ce genre de test, un groupe d'auditeurs évalue une série de fichiers audio (voix) à l'aide d'une échelle d'appréciation à cinq niveaux de valeur (Tableau III.1), sans avoir à écouter la séquence originale. La quantité évaluée d'après les notes (note moyenne d'appréciation de qualité d'écoute ou simplement note moyenne d'appréciation) est représentée par le symbole MOS.

Qualité de la parole	*Score*
Mauvaise	**1**
Médiocre	**2**
Passable	**3**
Bonne	**4**
Excellente	**5**

Tab III.1: Echelle d'appréciation de la qualité d'écoute [50]

III.3.2 Evaluation par catégories de dégradation (DCR)

Lorsque des échantillons de parole de bonne qualité sont évalués, l'ACR a tendance à être insensible aux petites dégradations de qualité. La procédure d'évaluation par catégories de dégradation DCR (Degradation Category Rating), qui repose en particulier sur une échelle de perturbation et sur une référence de qualité élevée avant chaque configuration à évaluer, semble convenir pour évaluer la parole de bonne qualité.

Les sujets notent le niveau de dégradation et de gêne en comparant avec le signal vocal d'origine (de référence). Les échelles de notation (les niveaux de dégradation) sont présentées dans le Tableau III-2.

La quantité de dégradation évaluée d'après les notes (notes d'appréciation moyenne de la dégradation) est représentée par le symbole DMOS.

Niveau de dégradation	Score
Dégradation inaudible	5
Dégradation audible mais pas gênante	4
Dégradation peu gênante	3
Dégradation gênante	2
Dégradation très gênante	1

Tab III.2: Echelle de notation pour les tests DCR [50].

Afin de normaliser les tests subjectifs, ITU P.800 défini des conditions détaillées des essais subjectifs tels que les caractéristiques des matériaux d'essai et l'environnement de test et d'essai. Les tests subjectifs sont normalement effectués dans une zone réglementée, à double paroi, chambre insonorisée.

Les principaux inconvénients de l'évaluation subjective (auditive) sont :

(i) la nécessité des méthodes spécialisées,

(ii) le besoin de personnel ayant une expertise spécifique,

(iii) la nécessité de beaucoup de temps,

(iv) l'impossibilité de les utiliser pour une mesure en temps réel. Ces limitations ont motivé le développement de méthodes objectives (appelées aussi instrumentales) qui mesurent la qualité vocale d'une manière systématique.

III.4 Mesure objective (instrumentale) intrusive de la qualité vocale

Afin de réduire le temps et les coûts nécessaires pour mesurer la qualité de la voix, des mesures objectives (appelées aussi instrumentales) ont été proposées. Les méthodes objectives sont des algorithmes ou des dispositifs, qui agissent comme un instrument de mesure, et génèrent une mesure quantitative à partir des échantillons de la voix et la faire correspondre à une prédiction de qualité. La qualité prédite doit être très proche de la valeur réelle de la qualité perçue de la voix.

III.4.1 Mesures dans le domaine temporel

Les mesures dans le domaine temporel sont surtout destinées à des systèmes de codage/ transmission qui tentent de reproduire la forme d'onde d'origine (par exemple, dans le cas des techniques de codage de forme d'onde).

Les mesures dans le domaine temporel, les plus simples et les plus communes, pour l'évaluation de la qualité du signal sont *Signal-to-Noise-Ratio* (SNR) et *Segmental Signal -to-Noise-Ratio* (SSNR).

Soient *x(i)* la parole d'origine et *y(i)* la version déformée du signal, le SNR est défini par :

$$SNR = 10 \log_{10} \frac{\sum_{i=1}^{N} x^2(i)}{\sum_{i=1}^{N} (x(i) - y(i))^2} \quad \text{(III.1)}$$

Avec : i l'indice dans le temps couvrant la période de mesure.

SSNR est une variation du SNR qui fonctionne sur des segments courts (15 à 20 ms) de flux. Cela permet de faciliter l'alignement temporel et il fournit des résultats qui sont légèrement mieux (en termes de corrélation avec l'évaluation subjective) que ceux du SNR. Il est calculé comme suit :

$$SSNR = \frac{1}{N} \sum_{m=1}^{N} SNR_m \quad \text{(III.2)}$$

Il s'agit d'une moyenne des valeurs SNR obtenues pour des trames isolées. Les trames étant un bloc d'échantillons. Les mesures dans le domaine temporel sont faciles à mettre en œuvre, et peuvent être utiles pour détecter les distorsions introduites par un bruit additif. Cependant, ils montrent des limitations lorsqu'elles sont utilisées dans un contexte général surtout quand il y a des dégradations telles que le filtrage.

III.4.2 Mesures dans le domaine spectrales

Les mesures dans le domaine spectrales sont généralement appliquées à des segments courts, typiquement de l'ordre de 15 à 30 ms. Ces techniques fournissent en général de meilleurs résultats. Elles sont, cependant, très dépendantes du codec utilisé et ne conviennent pas pour une utilisation dans les réseaux.

Les méthodes les plus connues dans le domaine fréquentiel sont *Itakura-Saito* (IS), *Log-Area-Ratio* (LAR) [36] et la méthode de la pente spectrale de mesure de distance pondérée (Weighted Spectral Slope Measures -WSS). Dans cette étude nous intéressons à cette dernière technique.

- **Weighted Spectral Slope Measures (WSS)**

La technique (WSS) est une mesure directe de la distance spectrale basée sur la comparaison des spectres lissée de signal propre et déformée. Le spectre lissé peut être

obtenue à partir de chaque analyse LP (liear Prediction), filtrage Cepstral. La technique de WSS peut être définie comme suit:

$$d_{WSS} = \frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{j=1}^{K} W(j,m)\left(S_c(j,m)-S_d(j,m)\right)^2}{\sum_{j=1}^{K} W(j,m)} \quad \text{(III-3)}$$

Où K est le nombre de bandes, M est le nombre total de trames, et S_c(j, m) et S_d(j, m) sont les pentes spectrales (typiquement les différences spectrales entre bandes voisines) de la bande j dans la trame « m » pour la parole propre et déformée, respectivement. W (j, m) sont des poids, qui peuvent être calculés comme indiqué par Klatts [37]. WSS a été largement étudiée au cours des dernières années, et a bénéficié d'une large acceptation [37].

III.4.3 Mesures dans le domaine perceptuel

Les mesures dans le domaine perceptuel sont basées sur une certaine forme de modèle psychoacoustique. Elles sont généralement plus précises que les autres types de mesure.

Plusieurs méthodes perceptives pour l'estimation de la qualité vocale ont été développées dans la littérature: BSD, TOSQA, PAMS, PSQM, PSQM+, PESQ, POLQA..

III.4.3.1 Perceptual Speech Quality Measure (PSQM)

PSQM, développé par *PTT research* en 1994, est une version modifiée de *Perceptual Audio Quality Measure* (PAQM), qui est une mesure objective de la qualité de la voix. C'était la première mesure intrusive pour la qualité de la parole à bande étroite elle est normalisée par l'ITU-T sous la recommandation P.861. [38]

PSQM est un algorithme d'évaluation perceptuelle, conçu pour évaluer la performance des codecs vocaux et des distorsions rencontrées dans les réseaux. Il utilise les résultats de calcul psycho-acoustiques de l'intensité vocale pour transformer la parole dans le domaine perceptuel correspondant. Le modèle PSQM donne des estimations fiables dans le cas des codecs vocaux à faible débit binaire, mais il peut ne pas être suffisamment robuste pour être appliquée à un plus grand éventail de distorsions. Une version améliorée du modèle PSQM, appelé PSQM+, a été développée. PSQM+ prédit les distorsions dues aux erreurs de canal de transmission, telle que la perte de paquets [39].

III.4.3.2 Perceptual Analysis Measurement System (PAMS)

British Telecom (BT) a élaboré en 1998 le système d'analyse de mesure perceptuelle *Perceptual Analysis Measurement System* (PAMS) [40]. PAMS compare le signal original avec le signal dégradé en transformant les distorsions en paramètres et les faire correspondre à une mesure objective de la qualité.

III.4.3.3 Perceptual Evaluation of Speech Quality (PESQ)

Les mesures les plus efficaces, évalués par l'ITU dans les années 1990, ont été combinés en un modèle amélioré *Perceptual Evaluation of Speech Quality* (PESQ), qui a été accepté en tant que recommandation de l'ITU en 2001 [32].

PESQ est considéré comme une combinaison optimale de deux algorithmes PSQM et PAMS. PESQ a hérité les deux composants suivant : l'algorithme d'alignement du modèle PAMS [41] et la transformation dans le domaine perceptuel du modèle PSQM99 [42]. L'alignement temporel consiste à calculer l'ensemble des retards entre le signal original et le signal déformé. La transformation perceptive est utilisée pour transformer, à la fois le signal original et dégradé, à une représentation psychophysique dans le système auditif humain, en tenant compte de l'intensité et de la fréquence perceptuelle.

PESQ est un outil puissant pour mesurer une qualité vocale lorsqu'il est utilisé judicieusement. La mesure objective de la qualité de la voix par PESQ est fortement corrélée avec les mesures subjectives.

Notons que le résultat de PESQ est un score qui varie entre -0,5 (une très mauvaise qualité) et 4,5 (une excellente qualité). Une version large bande de PESQ (WB-PESQ) a été définie dans la Recommandation ITU-T P.862.2 (2005). WB-PESQ utilise exclusivement des signaux de parole avec une fréquence d'échantillonnage de 16 kHz. [43]. Le modèle PESQ intègre principalement deux modules : le module d'échelonnement, d'alignement temporel et le module psychoacoustique.

a. Échelonnement et alignement temporel

Afin d'être robuste vis-à-vis du délai pouvant exister entre le signal initial et le signal dégradé, le modèle PESQ intègre un module d'alignement temporel. Les signaux initial et dégradé sont alignés temporellement d'après les étapes fournies dans la Figure III.2.

a.1 Échelonnement du niveau

Les niveaux des deux signaux initial X(t) et dégradé Y (t) sont échelonnés sur le même niveau de puissance constant. En effet, chaque système testé peut avoir un gain différent. De plus, le niveau auquel le signal dégradé est écouté par les utilisateurs varie selon les systèmes, les téléphones, etc... Le modèle ne doit donc pas prendre en compte la différence de niveau entre le signal initial et le signal dégradé. Leurs niveaux sont donc échelonnés à une valeur constante de 79 dB SPL au point de référence oreille. De cette opération découlent les signaux échelonnés $X_S(t)$ et $Y_S(t)$.

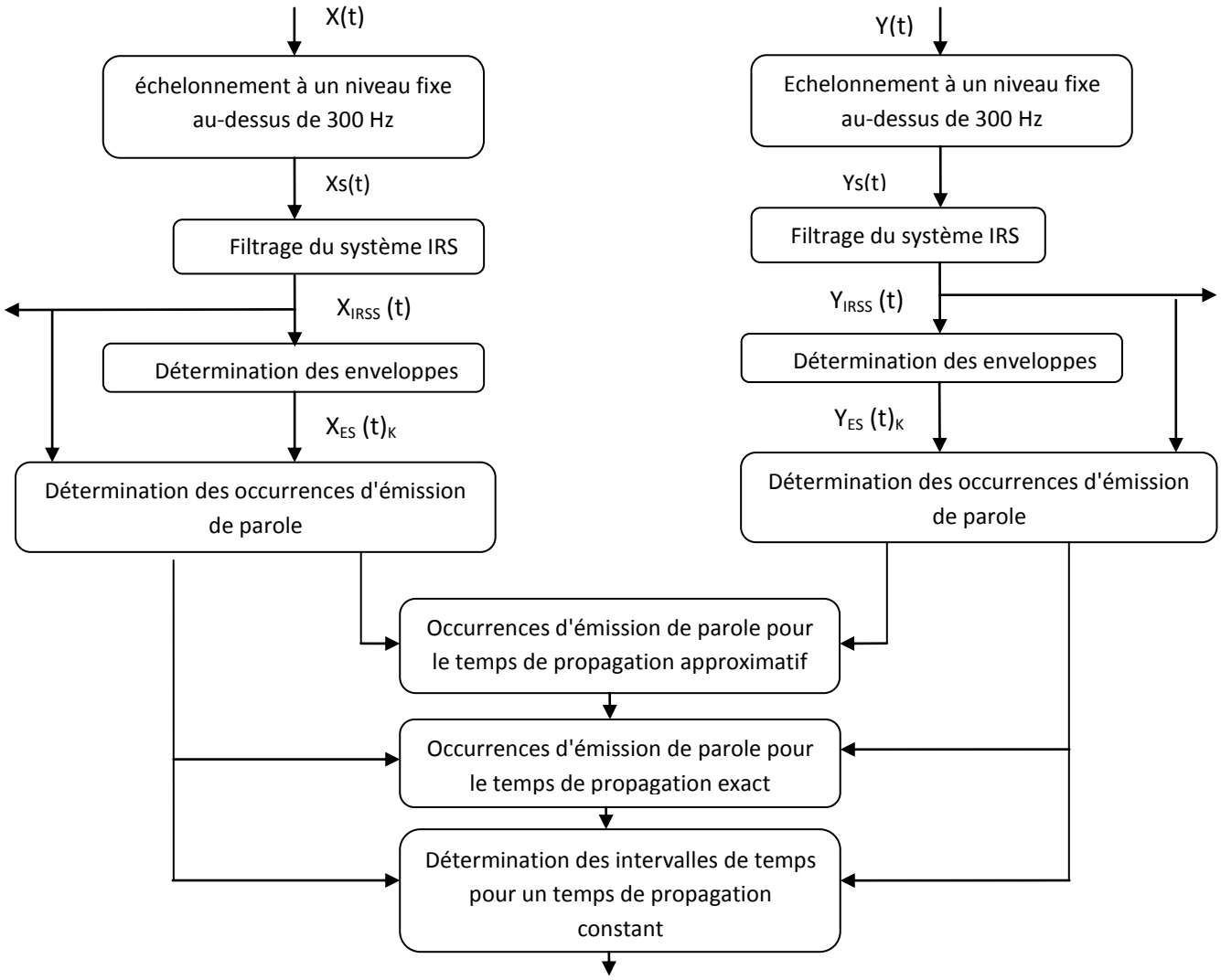


Fig. III.2 : Principe de fonctionnement du module d'alignement temporel utilisé dans le modèle PESQ pour déterminer le temps de propagation par intervalle de temps d_i [44].

a.2 Filtrage du système IRS

Le filtrage IRS (Intermediate Reference System) permet de simuler les caractéristiques fréquentielles d'un signal émis par un combiné téléphonique. Le modèle PESQ applique le filtre IRS aux FFT des signaux $X_S(t)$ et $Y_S(t)$. Les signaux filtrés $X_{IRSS}(t)$ et $Y_{IRSS}(t)$ sont ensuite obtenus grâce à une FFT inverse.

a.3 Alignement temporel

Le module d'alignement temporel comprend plusieurs étapes: [45]

- le temps de propagation approximatif est estimé grâce à l'intercorrélation des enveloppes $X_{ES}(t)k$ et $Y_{ES}(t)k$ des signaux $X_S(t)$ et $Y_S(t)$. L'enveloppe est définie par la formule: log(max(E(k)/Ethresh, 1)), où E(k) est l'énergie dans une trame k de 4 ms et Ethresh est le seuil de conversation déterminé par un détecteur d'activité vocale.

- le signal initial est divisé en occurrences d'émission de parole (ou paroles émises).
- le temps de propagation de groupe d'après les occurrences d'émission de parole est estimé.
- le temps de propagation fin est identifié d'après les occurrences de parole, par corrélation fine ou par histogramme.
- les occurrences de parole sont coupées et les intervalles de temps sont réalignés pour repérer les variations du temps de propagation en cours de conversation.
- les longs passages comportant des erreurs importantes, identifiés par le module perceptuel de PESQ, sont réalignés.

Le calcul du temps de propagation fin permet de déterminer une valeur de temps de propagation exact par échantillon, selon les étapes suivantes :

- les signaux initial et dégradé sont divisés en trames de 64 ms (se chevauchant à 75%) par fenêtrage de Hanning.
- l'intercorrélation entre chaque trame du signal initial et chaque trame du signal dégradé est calculée, après alignement selon les enveloppes.
- le maximum de la corrélation, à la puissance 0.125, est utilisé pour mesurer le niveau de confiance de l'alignement dans chaque trame. L'indice du maximum indique la valeur estimative du temps de propagation pour chaque trame.
- un histogramme de ces valeurs estimatives du temps de propagation, pondérées par le niveau de confiance mesuré, est calculé. L'histogramme est ensuite lissé par convolution avec un noyau triangulaire symétrique d'une largeur de 1 ms.
- l'indice du maximum dans l'histogramme, conjugué à la valeur estimative du temps de propagation précédent, indique la valeur estimative du temps de propagation final.
- le maximum de l'histogramme, divisé par la somme de l'histogramme avant convolution avec le noyau, indique un niveau de confiance compris entre 0 (aucune confiance) et 1 (confiance maximale).

b. Modèle psychoacoustique

Une fois alignés, les deux signaux sont envoyés dans le modèle psychoacoustique, présenté dans les figures III.2 et III.3.

b.1 Initialisations et calibrations

Les échantillons de début et de fin du signal de référence sont déterminés. Le premier échantillon déclaré actif est celui dont l'amplitude (i.e. valeur absolue) plus les amplitudes des quatre échantillons précédents vaut 500 ou plus. Le dernier échantillon déclaré actif est le dernier échantillon dont l'amplitude plus les amplitudes des quatre échantillons suivants vaut 500 ou plus.

Un étalonnage entre le niveau d'écoute et la sonie comprimée est également nécessaire. Pour cela, une sinusoïde de référence avec une fréquence de 1 kHz et une amplitude de 40 dB SPL (c'est-à-dire -64 dBov, et une amplitude zéro-à-crête de 29.54) est générée1. Le premier étalonnage consiste à échelonner à 104 la valeur maximale de la représentation de puissance fondamentale contenue dans la sinusoïde d'étalonnage. Cette valeur est obtenue par une post multiplication avec une constante, le facteur d'échelonnement de puissance Sp.

Le deuxième étalonnage fixe à la valeur de 1 Sone la sonie comprimée de la sinusoïde de référence. Cette valeur est obtenue par une post-multiplication avec une constante, le facteur d'échelonnement de la sonie Sl.

b.2 Transformation temps-fréquence

Un fenêtrage de Hanning est appliqué aux signaux échelonnés et filtrés $X_{IRSS}(t)$ et $Y_{IRSS}(t)$. Les trames temporelles durent 32 ms et deux trames successives se chevauchent à 50%. Les signaux fenêtrés $X_{WIRSS}(t)$ et $Y_{WIRSS}(t)$ sont ensuite transformés par FFT.

Leurs densités spectrales de puissance sont notées $PX_{WIRSS}(f)_n$ et $PY_{WIRSS}(f)_n$, où n est l'indice de trame et f est l'indice de fréquence.

b.3 Prédistorsion et densité de puissance fondamentale

La prédistorsion permet de passer de l'échelle des fréquences en Hertz à l'échelle psychophysique des tonies dans le domaine des bandes critiques en Bark, pour obtenir une représentation de la densité de puissance fondamentale trame par trame.

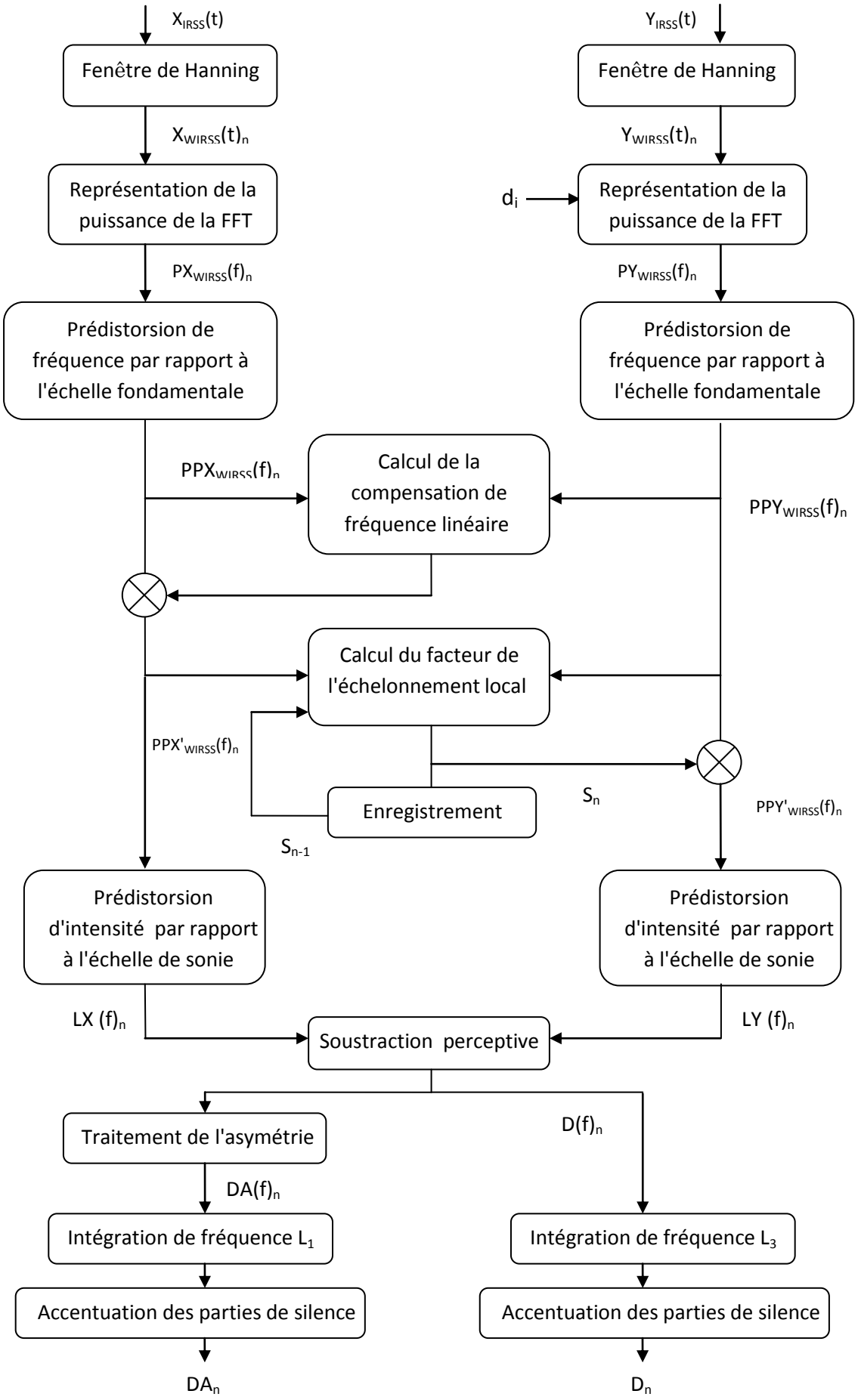


Fig. III.3 : Principe de fonctionnement du modèle psychoacoustique utilisé dans le modèle PESQ (Perceptual Evaluation of Speech Quality) [44].

L'échelle des bandes critiques est d'abord subdivisée en bandes d'intervalle égal et, pour chaque bande, une valeur (échantillon) de densité de puissance fondamentale est calculée à partir des échantillons de densité de puissance spectrale dans la bande correspondante sur l'échelle en Hertz. Les densités de puissance fondamentale sont notées $PPX_{WIRSS}(f)_n$ et $PPY_{WIRSS}(f)_n$.

b.4 Compensations

L'effet du filtrage et des variations de gain de courte durée sont partiellement compensées par le traitement des densités de puissance fondamentale trame par trame. Les densités de puissance fondamentale avec compensation partielle $PPX'_{WIRSS}(f)_n$ et $PPY'_{WIRSS}(f)_n$ sont alors obtenues.

b.5 Densité de sonie

Les densités de sonie $LX(f)_n$ et $LY(f)_n$ sont calculées à partir d'une fonction de compression donnée par Zwicker [46]

$$LX(f)_n = S_l \left(\frac{P_0(f)}{0.5}\right)^{\gamma} \left[\left(0.5 + 0.5\frac{PPX'_{WIRSS}(f)_n}{P_0(f)}\right)^{\gamma} - 1\right] \quad \text{(III.4)}$$

où $P_0(f)$ est le seuil absolu d'audition et S_l est le facteur d'échelonnement en sonie. Au-dessus de 4 Bark, la puissance de Zwicker " γ" est de 0.23. Au-dessous de 4 Bark, la puissance de Zwicker croît légèrement pour tenir compte de l'effet dit de recrutement.

b.6 Densité de perturbation

La différence entre les densités de sonie $LX(f)_n$ et $LY(f)_n$ est ensuite calculée. Celle-ci est corrigée de telle sorte que les petites différences de sonie soient moins perçues en présence de signaux ayant une valeur élevée de sonie (masquage). Il en résulte une densité de perturbation en fonction du temps (nombre de fenêtres n) et de la fréquence f, $D(f)_n$.

b.7 Traitement de l'asymétrie

Le phénomène d'asymétrie traduit le fait qu'une dégradation additive est plus audible et gênante qu'une dégradation soustractive pour le cerveau humain. Le modèle de PESQ prend en compte cette asymétrie en donnant plus de poids aux dégradations additives qu'aux dégradations soustractives. Cet effet est modélisé en calculant la densité de perturbation asymétrique $DA(f)_n$ par trame et en multipliant la densité de perturbation $D(f)_n$ par un facteur d'asymétrie. Ce facteur d'asymétrie est égal au rapport des densités de puissance fondamentale

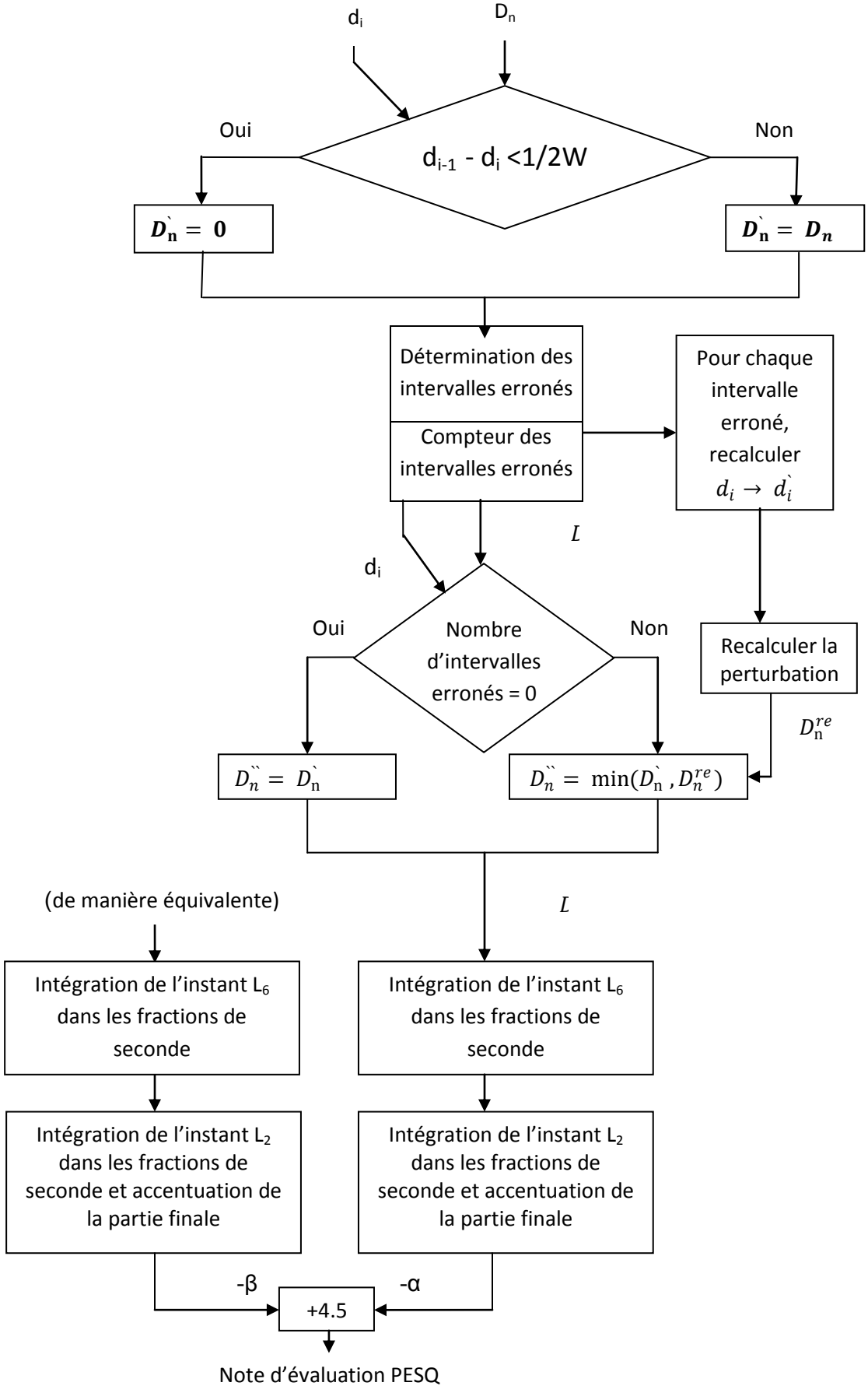


Fig. III.4 : Principe de fonctionnement du modèle psychoacoustique utilisé dans le modèle PESQ [44].

du signal dégradé et du signal initial élevé à la puissance de 1.2. S'il est inférieur à 3, le facteur d'asymétrie est mis à zéro. S'il est supérieur à 12, il est écrêté à cette valeur. Ainsi, seules demeurent les cellules temps-fréquence, sous forme de valeurs non nulles, pour lesquelles la densité de puissance fondamentale du signal dégradé est supérieure à la densité de puissance fondamentale du signal initial.

b.8 Accentuation des parties de silence

Les dégradations intervenant dans les segments de non-activité vocale sont également prises en compte mais avec une pondération moindre.

b.9 Intégration en temps et fréquence

Enfin, l'intégration en temps et fréquence des différences audibles est non linéaire afin de modéliser le fait que des erreurs isolées (en temps et/ou fréquence) ont un impact perceptif plus fort que des erreurs uniformément réparties.

b. 10 Calcul du score PESQ

La note d'évaluation PESQ finale est une combinaison linéaire de la valeur de perturbation moyenne et de la valeur de perturbation asymétrique moyenne.

c. Performances

Pour les 22 tests subjectifs utilisés à l'UIT pour évaluer la performance de PESQ, le coefficient de corrélation moyen s'est établi à 0.935 et la distribution moyenne des erreurs résiduelles a montré que l'erreur absolue était inférieure à une note MOS de 0.25 (0.25 sur une échelle de 5 points) pour 69.2% des conditions et inférieure à une note MOS de 0.5.

Pour les 8 tests utilisés pour la validation finale (inconnus pendant l'élaboration de PESQ), le coefficient de corrélation moyen s'est également établi à 0.935 et l'erreur absolue était inférieure à une note MOS de 0.25 pour 72.3% des conditions et inférieure à une note MOS de 0.5 pour 91.1% des conditions.

III.3.3.4 Perceptual Objective Listening Quality Analysis (POLQA)

POLQA est la technologie de la prochaine génération de mesure de la qualité de la voix pour la téléphonie fixe, mobile et IP. POLQA est considéré comme le successeur de PESQ. POLQA a été normalisé par l'Union Internationale des télécommunications ITU-T en tant que

nouvelle Recommandation P.863, et peut être appliquée à l'analyse de la qualité de la voix dans le contexte HD Voice, 3G et réseaux 4G/LTE. Il s'agit d'un modèle intrusif pour la mesure de la qualité vocal, convenable pour les interfaces électro-acoustiques et les connexions à bande étroite (*Narrow Band*) et à bande super large (*S-Wide Band*). POLQA utilise un modèle psycho-acoustique pour émuler la perception humaine. [47]

III.5 Mesure objective non-intrusive de la qualité vocale

Contrairement aux méthodes intrusives, dans lesquelles un signal de référence est nécessaire pour l'évaluation de la qualité, les méthodes de mesure non intrusives de la qualité de la parole n'ont pas besoin du signal de référence et sont plus appropriées pour la surveillance du trafic en temps réel. Il existe deux catégories de méthodes non intrusives pour la prévision de la qualité de parole.

La première consiste à prédire la qualité de la parole directement à partir de divers valeurs de paramètres du réseau (par exemple, la perte de paquets, la gigue et le retard) et les paramètres non liés au réseau (par exemple, le codec, écho, la langue, etc.). Le but est d'établir la relation entre la qualité vocale perçue et les paramètres associés du réseau ou hors réseau. Les méthodes typiques sont le E-model et les réseaux de neurones artificiels (*Artificial Neural Network* ANN).

III.5.1 E-model

L'ETSI a développé un modèle de calcul de la qualité de transport de la voix de bout en bout, de la bouche de l'émetteur à l'oreille du récepteur, connu sous le nom de E-model [48]. Ce modèle a été standardisé par l'ITU sous la référence G.107 [34]. Le principe de l'E-model consiste à calculer une grandeur unique *R*, qui reflète l'état de la conversation, en fonction des paramètres et caractéristiques des équipements ainsi que celles du canal de transmission du signal.

La formule simplifiée du calcul de *R* est la suivante :

$$R = R_0 - I_s - I_d - I_{e-eff} + A \qquad \text{(III.5)}$$

Avec :

- R_0: coefficient initial signal sur bruit. R_0 est la valeur que l'on obtiendrait si la transmission était Parfaite
- I_s: coefficient de dommages simultanés avec l'émission de la voix
- I_d : coefficient de dommages dus au délai de transmission et de transport
- $I_{e\text{-}eff}$: coefficient de dommages de distorsion causés par les équipements

- A: coefficient de prise en compte de facteurs d'amélioration du réseau.

  Par exemple, les utilisateurs de services mobiles sont plus tolérants à l'égard d'un canal dégradé que les utilisateurs de réseaux câblés.

Le facteur de transmission de R estimé par E-model peut être converti en un MOS [34]:

$$MOS = \begin{cases} 1 & R < 00 \\ 1 + 0.035.R + 7.10^{-6} R(R - 60)(100 - R) & < R < 100 \\ 4.5 & R > 100 \end{cases} \quad \text{(III.6)}$$

E-model est un modèle attractif pour la prédiction non-intrusive de la qualité de voix, mais il a un certain nombre de limitations. Par exemple, il est basé sur un ensemble complexe de formules fixes, empiriques et applicables à un nombre limité de codecs et de conditions de réseau (parce que les tests subjectifs sont nécessaires pour obtenir les paramètres du modèle).

III.5.2 IQX Hypothesis

IQX Hypothesis vise à décrire la qualité perceptuelle ou plus connue sous le nom de *Quality of Experience* en fonction d'un facteur unique qui détermine la qualité de service du réseau (par exemple taux de perte de paquets). Cette hypothèse exprime la QoE comme une fonction exponentielle de la dégradation de qualité de service :

$$QoE = \alpha.e^{-B.QoS} + \gamma \quad \text{(III.7)}$$

Pour valider le modèle mathématique, des échantillons de parole, codées avec le codec iLBC, ont été transmis sur un émulateur de réseau avec des taux de perte de paquets variables, et avec leurs sources correspondantes ont servi de base pour l'application de l'algorithme PESQ. Les résultats ont prouvé que, dans le cas du codec iLBC, la relation entre la QoE et le facteur de dépréciation (perte de paquets) est comme suit :

$$QoE = 3.010 * \exp(-4.473 * P_{loss}) + 1.065 \quad \text{(III.8)}$$

III.5.3 Les réseaux de neurones artificiels

Contrairement à la méthode E-model qui est un modèle de calcul mathématique et qui est statique, les modèles basés sur les réseaux de neurones artificiels (*Artificial Neural Networks* ANN) peuvent s'adapter à l'environnement dynamique des réseaux IP, en raison de sa capacité d'apprentissage. Un modèle ANN peut être construit par l'apprentissage des

relations non linéaires entre la qualité vocale perçue (par exemple, note MOS) et une variété de paramètres réseau ou liés à la parole. [34]

III.6 Conclusion

Les mesures objectives de qualité qui reposent sur des notions de psychoacoustique permettent de prévoir les notes de qualité de perception qu'attribueraient au signal testé les sujets participant à un essai d'´ecoute subjective. Elles permettent d'automatiser le processus d'évaluation de la qualité et se prêtent plus à une éventuelle application en temps réel. Elles sont donc indispensables pour les systèmes ou l'homme fait partie intégrante du processus de réception. Cependant, leur corrélation insuffisante avec les résultats des tests subjectifs limite encore leur substitution complète aux méthodes subjectives .

Chapitre IV

Résultats de simulation et discussion

IV.1 Introduction

Enfin, nous nous focalisons sur les dégradations dues à la prise de son (bruits ambiantes et variations du canal acoustique), et afin d'améliorer la qualité de signal parole en présence du bruit, une étude comparative des techniques de rehaussement (d'élimination de bruits) de signal parole est réalisée dans ce chapitre.

IV.2 Étude de différentes techniques d'élimination de bruit additif au signal parole

Le rehaussement de la parole améliore la qualité et l'intelligibilité de la communication vocale pour une gamme d'application, notamment les téléphones mobiles, les systèmes de téléconférence, les prothèses auditives, des codeurs de la parole et la reconnaissance automatique du locuteur/parole. Un signal peut être corrompu par le bruit dans diverses situations, comme les trains, les voitures, l'aéroport, bavardage, usines, rues....etc.

L'étude comparative était en termes de qualité de la parole. La qualité de la parole reconstruite est évaluée avec le PESQ (Perceptual Evaluation of Speech Quality), segSNR(SNR segmental) et WSS (Weighted Spectral Slope Measures). Les méthodes que nous avons évaluées sont les techniques de dé-bruitage par les transformées d'ondelettes dont le type d'ondelette de mère choisit est l'ondelette de Daubechies dordre 1 avec cinque niveaux de décomposition basant sur un:

- ✓ seuillage doux (soft thresholding) c-à-d le seuillage des coefficients issus par la décomposition en plusieurs niveaux (dans notre cas la décomposition est procédée sur 5 niveaux). Dont, le seuillage consiste à éliminer les coefficients qu'on considère comme du bruit ou à les réduire en fonction du seuil calculé. En dernier lieu, on applique la transformée en ondelettes inverse sur les coefficients seuillés et on récupère le signal débruité (rehaussé).
- ✓ seuillage dur (hard thresholding).
- ✓ seuillage doux dépendant du niveau (level-dependent thresholding soft).
- ✓ seuillage dur dépendant du niveau (level-dependent thresholding hard).

IV.3 Résultats et discussion

Quatre méthodes ont été discutées et évaluées en termes de robustesse contre les bruits réels (bruit de: voiture, bavardage, salle d'exposition, restaurant, rue et l'aéroport) en utilisant la base de donnée NOIZEUS et artificiels (bruit blanc gaussien) et TIMIT.

IV.3.1 Comparaison en présence d'un bruit blanc gaussien

IV.3.1.1 Evaluation par PESQ

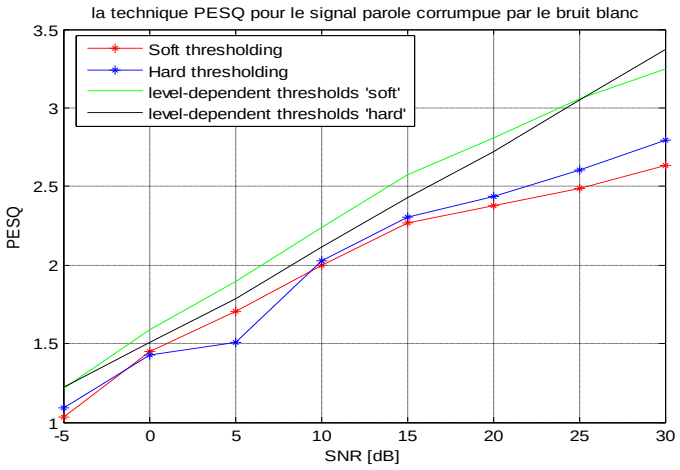


Fig. IV.1: Comparaison des performances en termes de PESQ en présence du bruit blanc (SNR=-5 à 30 dB par un pas de 5 dB)

Les Quatre méthodes mentionnées ci-dessus sont évaluées en termes de PESQ et en présence du bruit blanc gaussien.

La figure IV.1 illustre une comparaison des performances de quatre techniques de rehaussement en termes de scores de PESQ en présence du bruit blanc pour différents niveaux d' SNR, de -5 à 30 dB par un pas de 5 dB. De cette figure (figure IV.1) on peut conclure que la méthode de transformé d'ondelette basé sur un seuillage doux dépendant du niveau présente le meilleur score de PESQ, tandis que la technique fondé sur le Seuillage dur dépendant du niveau est la seconde en terme de PESQ, par contre la méthode fondé sur le seuillage doux présente le mauvais score de PESQ.

IV.3.1.2 Evaluation par segSNR

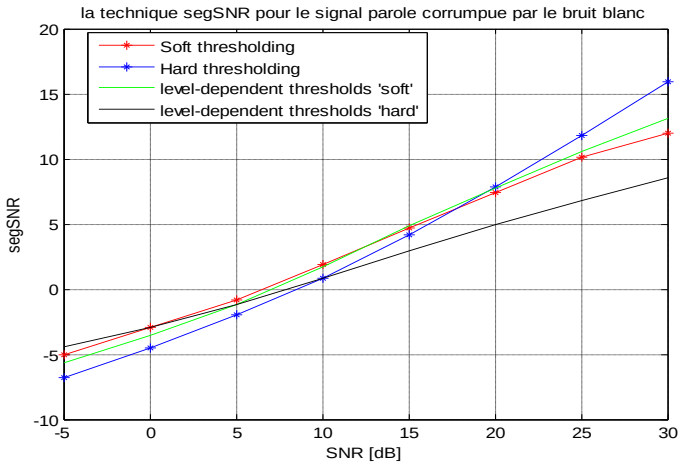


Fig. IV.2 : Comparaison des performances en termes de segSNR en présence du bruit blanc.(SNR=-5 à 30 dB par pas de 5 dB).

Cette figure montre que les méthodes basées sur le seuillage doux dépendant du niveau, seuillage dur dépendant du niveau et seuillage doux ont presque le même segSNR pour les SNR allant de -5 dB à 5 dB, tandis que, la méthode de seuillage dur a présenté le mauvais score de segSNR. En d'autre terme, on remarque que pour des niveaux de SNR de 5 dB à 15 dB le Seuillage doux dépendant du niveau et seuillage dur fournirent le meilleur résultat.

Pour des niveaux de SNR de 15 dB à 30 dB le seuillage doux fournit le meilleur résultat, en outre, la méthode de seuillage dur dépendant du niveau a présenté le mauvais score de segSNR

IV.3.1.3 Evaluation par WSS

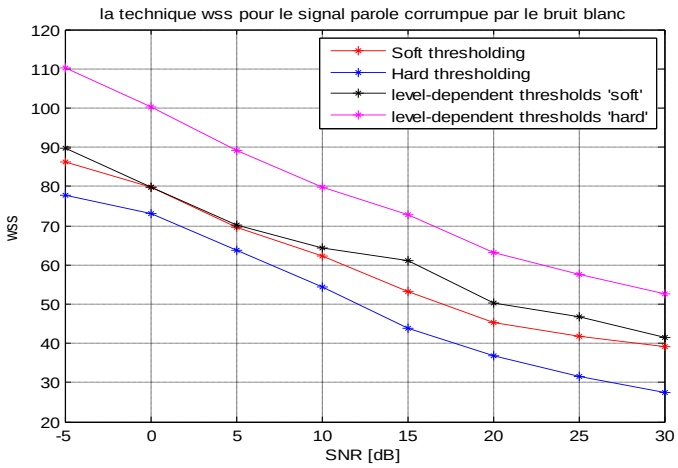


Fig. IV.3 :Comparaison des performances en termes de WSS en présence du bruit blanc.(SNR=-5 à 30 dB par pas de 5 dB) .

Cette figure montre que la méthode de transformée d'ondelette basé sur le seuillage dur dépendant du niveau a fourni de meilleurs résultats de WSS. En outre, la méthode de seuillage dur a présenté le mauvais score de WSS. On remarque que la méthode de seuillage dépendant du niveau doux et seuillage doux ont presque le même WSS pour SNR (-5 dB à 10 dB), mais, pour des niveaux de SNR (10 dB à 30 dB) le Seuillage dépendant du niveau doux fournit de bon résultat par apport au le seuillage doux.

IV.3.2 Comparaison par présence des bruits de: bavardage, aéroport, voiture, rue, restaurant et salle d'exposition en utilisant la technique ' PESQ '

Les approches de rehaussement de la parole ont également été évaluées en présence de plusieurs bruits réels extraits de la base de données NOIZEUS, le bruit de : bavardage, aéroport, voiture, rue, restaurant et salle d'exposition, sous différents niveaux de bruit (SNR varie entre 0 dB et 15 dB par un pas de 5 dB).

IV. 3.2.1 Présence de bruit de rue

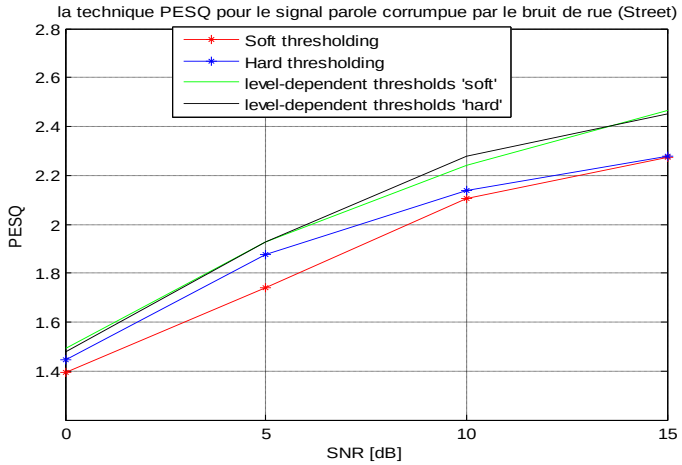


Fig. IV.4: Comparaison des performances en termes de PESQ en présence du bruit de rue. (SNR=0 dB à 15 dB par pas de 5 dB).

La Figure IV.4 présente une mesure de PESQ pour les quatre approches mentionnées plus haut en présence du bruit de la rue. On remarque que la méthode de dédruitage par seuillage dur dépendant du niveau et seuillage doux dépendant du niveau ont presque le même PESQ pour SNR plus de 5 dB, mais, pour de faibles niveaux de SNR (0 dB à 5 dB) le seuillage dépendant du niveau doux fournit le meilleur résultat. En outre, la méthode de débruitage par la transformée d'ondelette basés sur le seuillage doux présente le mauvais score de PESQ au-dessous de 0 dB à 15 dB.

IV.3.2.2 Présence de bruit d'aéroport

La figure IV.5 représente les mesures de PESQ en présence du bruit d'aéroport. Dont cette figure montre que la méthode Seuillage dépendant du niveau dur fournit de meilleurs résultats de PESQ.

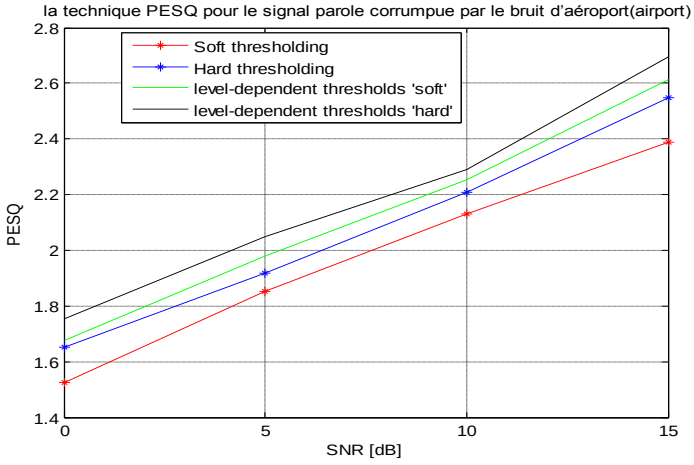


Fig. IV.5 :Comparaison des performances en termes de PESQ en présence du bruit d'aéroport. (SNR=0 dB à 15 dB par pas de 5 dB).

IV.3.2.3 Présence de bruit de voiture

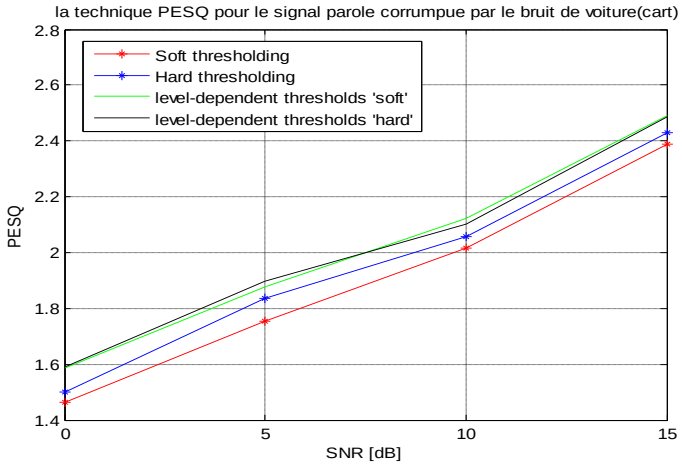


Fig. IV.6 : Comparaison des performances en termes de PESQ en présence du bruit de voiture. (SNR= 0 dB à 15 dB par pas de 5 dB).

La figure IV.6 représente la mesure de PESQ en présence du bruit de voitures. De cette figure, on peut conclure que les approches des méthodes de seuillage dépendants du niveau (dur et doux) ont fournit de meilleurs résultats de PESQ. En outre, les méthodes basées sur le seuillage dur et doux ont présenté le mauvais score de PESQ.

IV.3.2.4 Présence de bruit de bavardage

La Figure IV.7 représente la mesure de PESQ en présence du bruit de bavardage (babble). La technique de seuillage dur dépendant du niveau montre le meilleur résultat de PESQ.

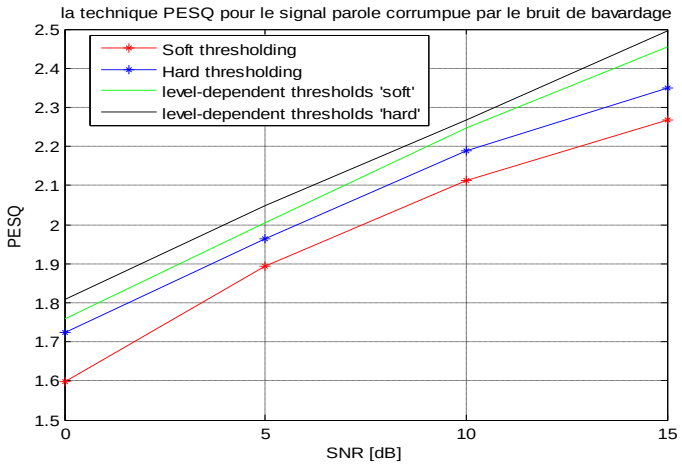


Fig. IV.7 : Comparaison des performances en termes de PESQ en présence du bruit de bavardage. (SNR= 0 dB à 15 dB par pas de 5 dB).

IV.3.2 .5 Présence de bruit de restaurant

La figure IV.8 représente la mesure de PESQ en présence du bruit de restaurant. De cette figure, nous pouvons conclure que la technique de seuillage dur dépendant du niveau montre le meilleur résultat de PESQ.

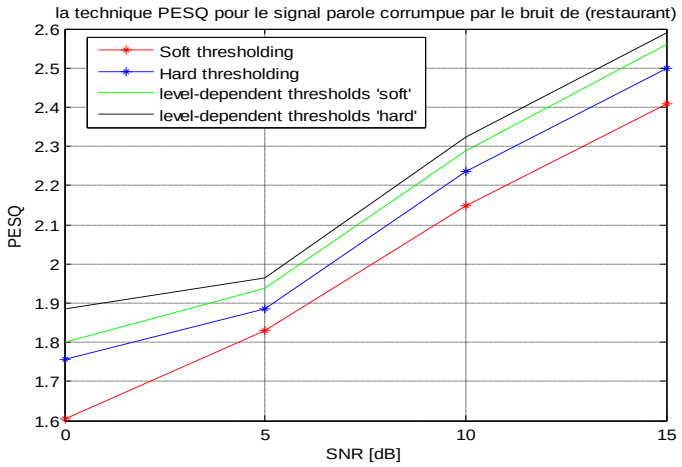


Fig .IV.8: Comparaison des performances en termes de PESQ en présence du bruit du restaurant. (SNR= 0 dB à 15 dB par pas de 5)

IV.3.2 .6 Présence de bruit de salle d'exposition

La Figure IV.9 représente la mesure de PESQ en présence du bruit de salle d'exposition. De cette figure, nous pouvons conclure que la méthode de seuillage doux dépendant du niveau était le premier en termes de PESQ.

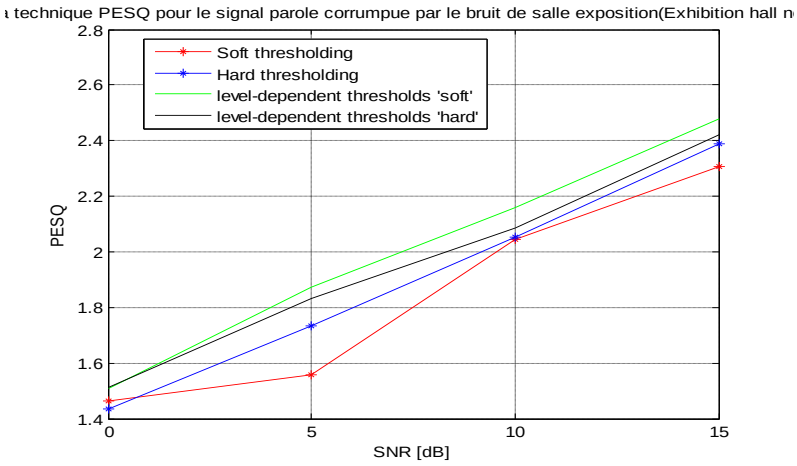


Fig. IV.9: Comparaison des performances en termes de PESQ en présence du bruit de salle d'exposition.

IV.3.3 Comparaison en termes des valeurs moyennes de PESQ pour les méthodes mentionnées précédemment

Méthode	**SNR [dB]**	**Types de bruit**					
		bavardage	**aéroport**	**Voiture**	**rue**	**restaurant**	**Salle d'exposition**
Seuillage dépendant du niveau doux	0	2.071	2.470	2.131	2.030	2.155	2.038
	5						
	10						
	15						
Seuillage dépendant du niveau dur	0	2.117	2.223	2.196	2.033	2.190	1.926
	5						
	10						
	15						
Seuillage dur	0	2.028	1.810	2.080	1.933	2.094	1.901
	5						
	10						
	15						
Seuillage doux	0	1.992	1.933	1.973	1.877	2.080	1.843
	5						
	10						
	15						

Tab. IV.1 :Mesures moyennes de PESQ pour les méthodes mentionnées précédemment pour la parole contaminée par les bruits de : bavardage, aéroport, voiture, rue et restaurant.

Le tableau IV.1 représente les mesures moyennes de PESQ pour les méthodes mentionnées précédemment pour la parole contaminée par les bruits de : bavardage, aéroport, voiture, rue et le bruit du restaurant. À partir de ce tableau, il est aisément de conclure que la méthode de seuillage dur dépendant du niveau a le meilleur score de PESQ.

IV.3.4 Comparaison en termes du temps d'exécution

Technique de rehaussement de la parole	*Temps d'exécution [sec]*
Seuillage dur	0.644261
Seuillage doux	0.667097
Seuillage dépendant du niveau dur	0.679753
Seuillage dépendant du niveau doux	3.049877

Tab IV.2: Résultats de simulation en termes de temps d'exécution.

Nous comparons les méthodes mentionnées plus haut en termes du temps d'exécution. Le tableau IV.2 montre les résultats des simulations en matière de temps d'exécution dont, nous pouvons observer que l'approche de débruitage basée sur le seuillage doux dépendant du niveau a plus de temps d'exécution que les autres techniques (seuillage dur, seuillage doux, seuillage dépendant du niveau dur).

IV.3.5 Comparaison par présence des bruits de: bavardage, aéroport, voiture, rue, restaurant et salle d'exposition en utilisant la technique ' segSNR '

IV.3.5.1 Présence de bruit de rue

La Figure IV.8 présente une mesure de segSNR pour les quatres approches mentionnées plus haut, en présence du bruit de la rue (street). On remarque que les méthodes basées sur le seuillage dur dépendant du niveau et seuillage doux ont presque le même segSNR. Les méthodes de seuillage doux dépendant du niveau et de seuillage dur pour SNR de 0 dB à 10 dB ont presque le même score de PESQ, mais pour des niveaux de SNR plus que 10 dB le seuillage dur fournit le meilleur résultat. En outre, la méthode de seuillage dur dépendant du niveau est la mauvaise en termes de PESQ.

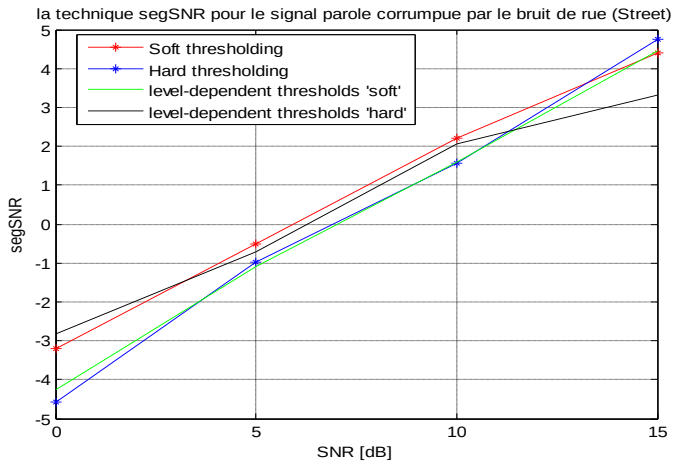


Fig. IV.10: Comparaison des performances en termes de segSNR en présence du bruit de rue(street). (SNR=0 dB à 15 dB par pas de 5 dB).

IV.3.5.2 Présence de bruit d'aéroport

La figure IV.11 présente une mesure de segSNR pour les quatres approches mentionnées plus haut, en présence du bruit d'aéroport (airport). On remarque que les méthodes de seuillage dur dépendant du niveau et seuillage doux ont presque le même segSNR. Pour les méthodes de seuillage doux dépendant du niveau et seuillage dur pour SNR

allant de 0 dB à 10 dB, tandis que pour des niveaux de SNR plus de 10 dB la méthode basée sur le seuillage dur fournit le meilleur résultat. En outre, la méthode de seuillage dur dépendant du niveau est la mauvaise en termes de PESQ.

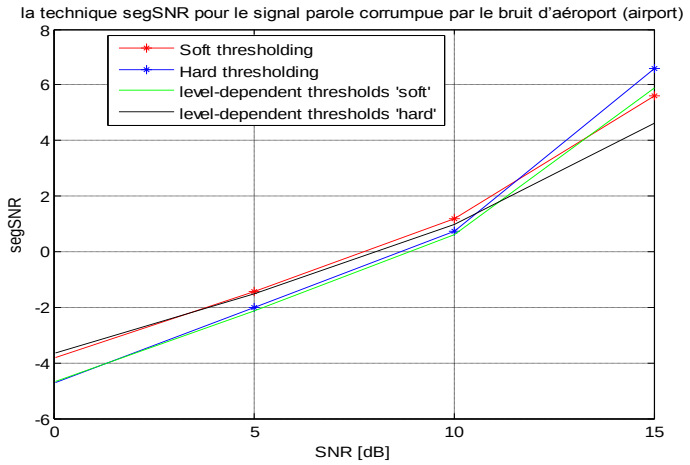


Fig. IV.11: Comparaison des performances en termes de segSNR en présence du bruit d'aéroport. (SNR=0 dB à 15 dB par pas de 5 dB).

IV.3.5.3 Présence de bruit de voiture

La figure IV.12 représente la mesure de segSNR en présence du bruit de voitures. De cette figure, on peut conclure que l'approche de la méthode de seuillage doux fournit de meilleurs résultats de segSNR. En outre, les méthodes de seuillage dur dépendant du niveau et seuillage dur ont présenté le mauvais score de segSNR.

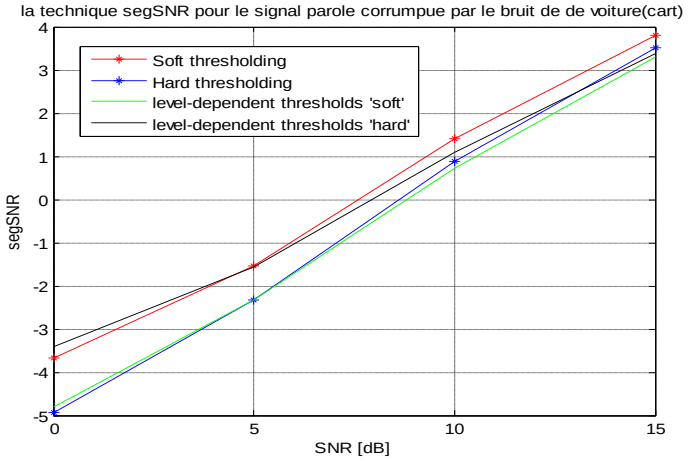


Fig. IV.12: Comparaison des performances en termes de segSNR en présence du bruit de voiture. (SNR= 0 dB à 15 dB par pas de 5 dB).

IV.3.5.4 Présence de bruit de bavardage

La Figure IV.13 représente la mesure de segSNR en présence du bruit de bavardage (babble). La technique de Seuillage dur montre le meilleur résultat de segSNR.

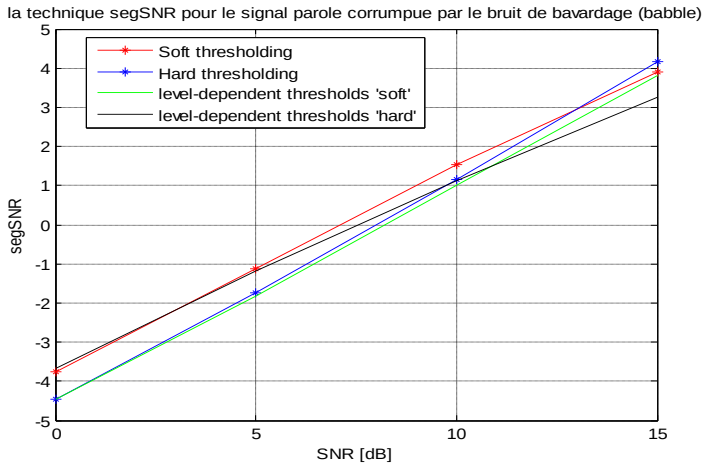


Fig. IV.13: Comparaison des performances en termes de segSNR en présence du bruit de bavardage. (SNR= 0 dB à 15 dB par pas de 5 dB).

IV.3.5.5 Présence de bruit de restaurant

La figure IV.14 représente la mesure de segSNR en présence du bruit de restaurant. De cette figure, nous pouvons conclure que La technique de Seuillage dur et Seuillage dépendant du niveau dur montre le meilleur résultat de segSNR.

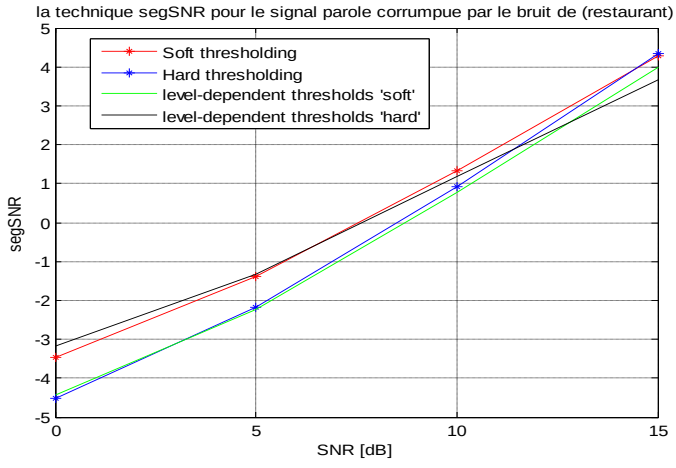


Fig. IV.14: Comparaison des performances en termes de segSNR en présence du bruit du restaurant. (SNR= 0 dB à 15 dB par pas de 5)

IV.3.5.6 Présence de bruit de salle d'exposition

La Figure IV.15 représente les mesures de segSNR en présence du bruit de salle d'exposition. De cette figure, nous pouvons conclure que la méthode de seuillage doux était le premier en matière de segSNR.

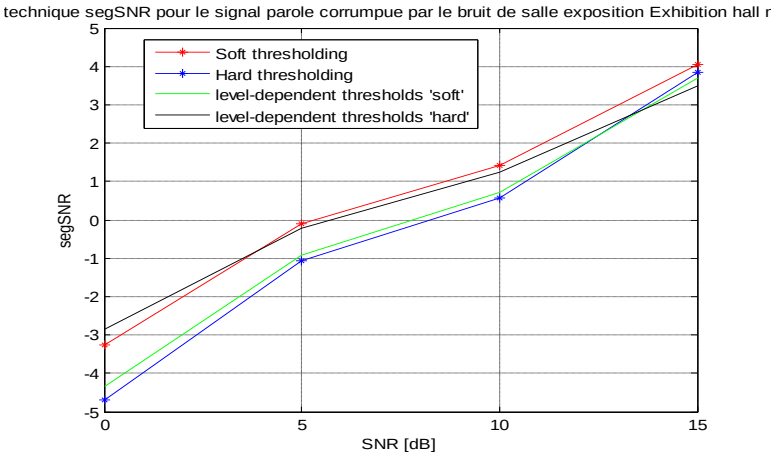


Fig. IV.15 : Comparaison des performances en termes de segSNR en présence du bruit de salle d'exposition.

IV.3.6 Comparaison en termes des valeurs moyennes de segSNR pour les méthodes mentionnées précédemment

Tableau IV.3 représente les mesures moyennes de segSNR pour les méthodes mentionnées précédemment pour un signal parole contaminé par les bruits de bavardage: aéroport, voiture, rue et le bruit du restaurant. À partir de ce tableau, il est aisément de conclure que la méthode de seuillage dur dépendant du niveau a le meilleur score de segSNR.

Méthode	**Types de bruit**					
	bavardage	**aéroport**	**Voiture**	**rue**	**restaurant**	**Salle d'exposition**
Seuillage dépendant du niveau doux	-0.324	0.253	0.384	0.726	0.186	0.530
Seuillage dépendant du niveau dur	-0.679	0.122	0.140	0.192	-0.355	-0.336
Seuillage dur	-0.820	1.810	2.262	0.171	-0.476	-0.213
Seuillage doux	0.232	0.933	0.112	0.457	0.091	0.424

Tab. IV.3 : Mesures moyennes de segSNR pour les méthodes mentionnées précédemment pour la parole contaminée par les bruits de : bavardage, aéroport, voiture, rue et restaurant.

IV.3.7 Comparaison par présence des bruits de: bavardage, aéroport, voiture, rue, restaurant et salle d'exposition en utilisant la technique ' WSS '

IV.3.7.1 Présence de bruit de rue

La figure IV.14 représente les mesures de WSS en présence du bruit de rue (street). De cette figure, nous pouvons conclure que la méthode basée sur le seuillage dur dépendant du niveau était le premier en matière de WSS, la méthode basée sur le seuillage doux est la deuxième, la méthode basé sur le seuillage doux dépendant du niveau est la troisième et seuillage dur à le mouvais score de WSS.

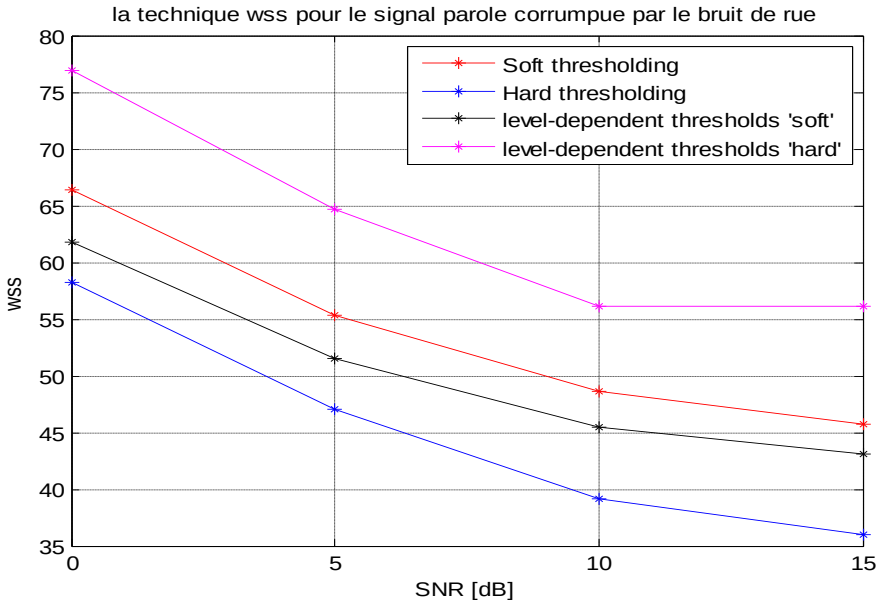


Fig. IV.16 : Comparaison des performances en termes de WSS en présence du bruit de rue (street). (SNR=0 dB à 15 dB par pas de 5 dB).

IV.3.7.2 Présence de bruit d'aéroport

La figure IV.17 représente les mesures de WSS en présence du bruit de rue (street). De cette figure, nous pouvons conclure que : la méthode de seuillage dur dépendant du niveau était le premier en matière de WSS, la méthode de seuillage doux la deuxième, la méthode de seuillage doux dépendant du niveau la troisième et la méthode de seuillage dur à le mouvais score de WSS.

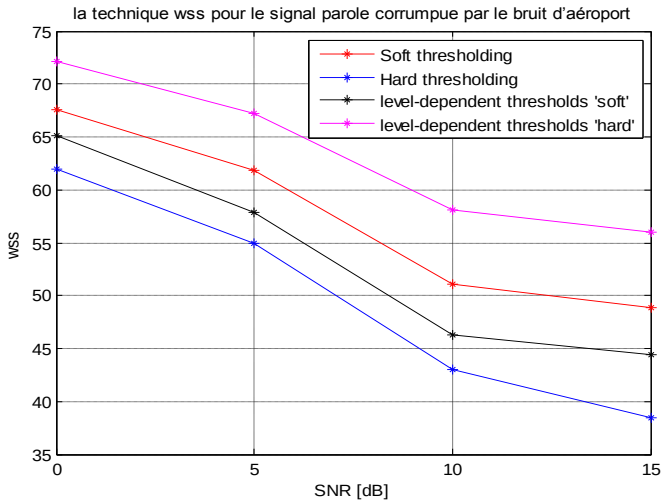


Fig. IV.17: Comparaison des performances en termes de WSS en présence du bruit d'aéroport. (SNR=0 dB à 15 dB par pas de 5 dB).

IV.3.7.3 Présence de bruit de voiture

La Figure IV.18 représente les mesures de WSS en présence du bruit de voiture. De cette figure, nous pouvons conclure que la méthode d'ondelette basée sur : seuillage dépendant du niveau dur était le premier en matière de WSS, seuillage doux la deuxième, seuillage doux dépendant du niveau la troisième et seuillage dur a le mouvais score de WSS.

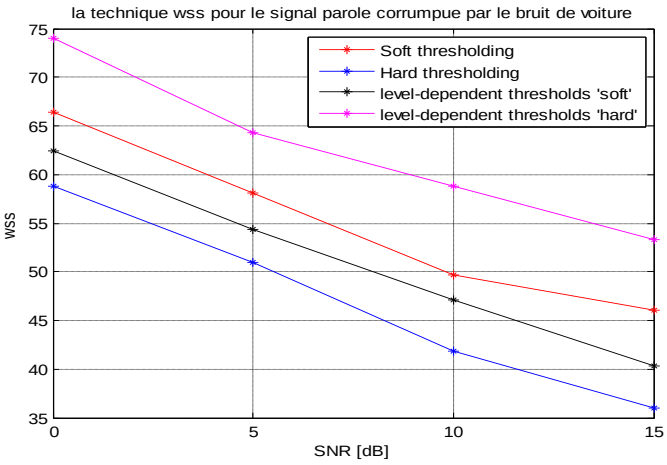


Fig. IV.18 : Comparaison des performances en termes de WSS en présence du bruit de voiture. (SNR=0 dB à 15 dB par pas de 5 dB).

IV.3.7.4 Présence de bruit de bavardage

La figure IV.19 représente les mesures de WSS en présence du bruit de bavardage. De cette figure, nous pouvons conclure que la méthode d'ondelette basée sur un: seuillage dur dépendant du niveau était le premier en matières de WSS, seuillage doux la deuxième , seuillage doux dépendant du niveau la troisième et seuillage dur à le mouvais score de WSS.

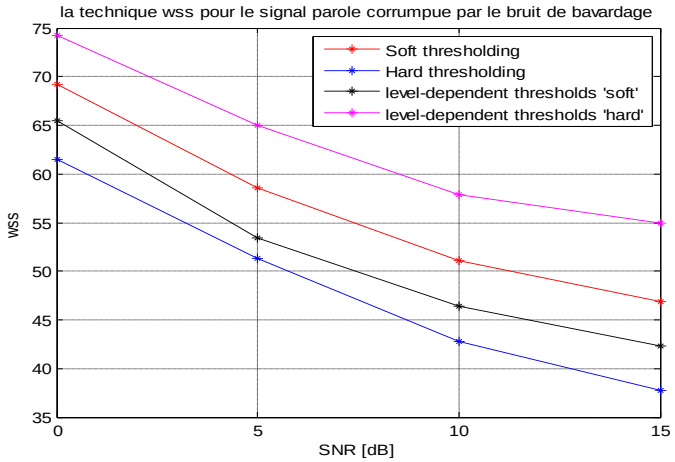


Fig. IV.19 : Comparaison des performances en termes de WSS en présence du bruit de bavardage. (SNR=0 dB à 15 dB par pas de 5 dB).

IV.3.7.5 Présence de bruit de restaurant

La figure IV.20 représente les mesures de WSS en présence du bruit de restaurant. De cette figure, nous pouvons conclure que la méthode d'ondelette basée sur un : seuillage dépendant du niveau dur était le premier en matière de WSS, seuillage doux la deuxième, Seuillage dépendant du niveau doux la troisième et Seuillage dur à le mouvais score de WSS.

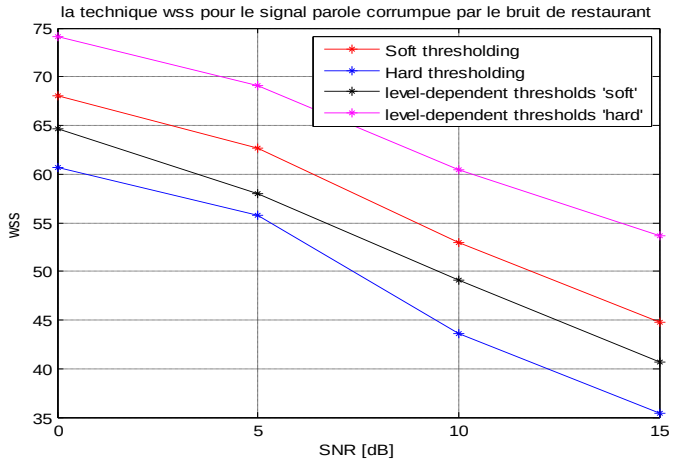


Fig. IV.20 : Comparaison des performances en termes de WSS en présence du bruit de restaurant. (SNR=0 dB à 15 dB par pas de 5 dB).

IV.3.7.6 Présence de bruit de salle d'exposition

La figure IV.21 représente les mesures de WSS en présence du bruit de salle d'exposition. De cette figure, nous pouvons conclure que la méthode d'ondelette basée sur un: seuillage dépendant du niveau dur était le premier en matière de WSS, seuillage doux la deuxième, seuillage dépendant du niveau doux la troisième et seuillage dur à le mouvais score de WSS.

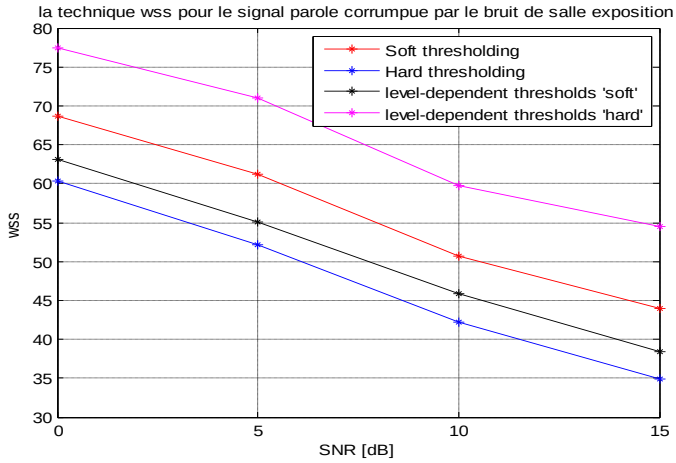


Fig. IV.21: Comparaison des performances en termes de WSS en présence du bruit de salle d'exposition. (SNR=0 dB à 15 dB par pas de 5 dB)

IV.4 Comparaison visuel entres les résultats des méthodes

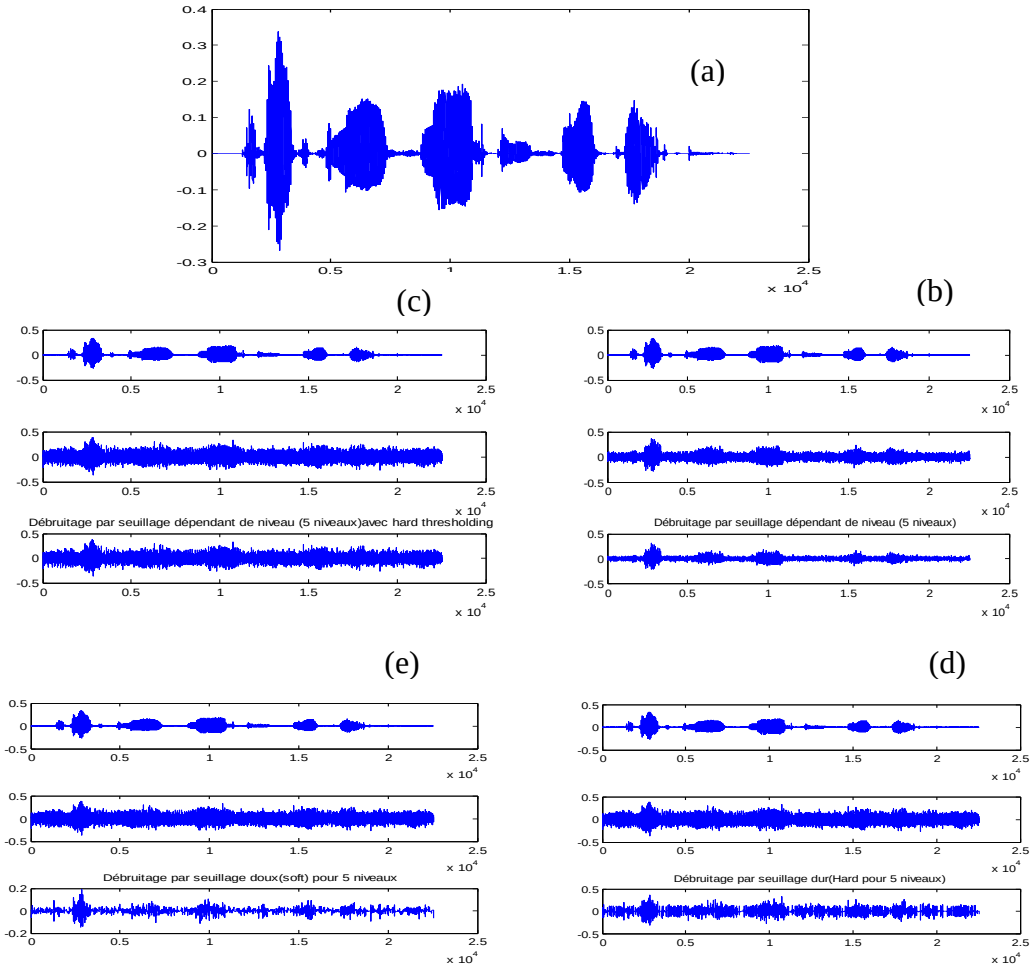


Fig.IV.22 : les résultats des simulation (a) signal original , le débruitage par seuillage: (b) Dépendant de niveau (5 niveaux) avec seuillage doux, (c) Dépendant de niveau (5 niveaux)avec seuillage dur , (d) Débruitage par seuillage dur pour 5 niveaux , (e) Débruitage par seuillage doux pour 5 niveaux

La figure *IV.22* présenté le signal originale, le signale corrompu par un bruit blanc et le signale débruiter.

On observe que la méthode Seuillage dépendant du niveau doux (*Fig.IV.22* .b) présente le meilleur méthode , la méthode de Seuillage doux (*Fig.IV.22* .e) la deuxième, la méthode de Seuillage dur la troisième (*Fig.IV.22* .d) et la plus mauvaise par apport que les autres méthodes est le Seuillage dépendant du niveau dur (*Fig.IV.22* .a).

IV.5 Conclusion

On a fait une étude comparative de différentes méthodes de rehaussement de la parole (élimination de bruit) en présence de plusieurs types de bruit basées sur les transformée d'ondelette discrète. La première évaluation a été effectuée par l'évaluation perceptive des scores de la qualité de la parole PESQ. L'approche par un seuillage dur dépendant du niveau a donné un PESQ significatives pour les différents niveaux d'SNR. En termes du temps d'exécution, la technique de la transformée d'ondelette basée sur un seuillage doux dépendant du niveau nécessite moins de temps d'exécution que les autres approches.

La deuxième évaluation a été effectuée par le segSNR qui convient bien à la qualité subjective. Une valeur typique pour la partie supérieure et la limite de rapport inférieur est de 35 et -10 dB. Dont, nous pouvons conclure que les transformée d'ondelette basées sur un seuillage dur et doux étaient le premier en matière de segSNR tandis que les transformées d'ondelette basées sur le seuillage dépendant du niveau (doux et dur) étaient la deuxième.

La troisième évaluation a été effectuée par la pente spectrale (WSS) mesure de distance pondérée. La technique de WSS est une mesure directe de distance spectrale fondée sur la comparaison des spectres lissés de signal parole propre et déformé des échantillons de signal parole. Dont, nous pouvons conclure que le débruitage par les transformée d'ondelettes basées sur un : seuillage dur dépendant du niveau était le premier en matière de **WSS**, seuillage doux la deuxième, seuillage doux dépendant du niveau la troisième et seuillage dur présente le mouvais score de WSS.

CONCLUSION GENERALE

Dans ce travail, nous avons présenté une étude sur le débruitage de la parole en utilisant la transformée en ondelettes. Une recherche bibliographique a été effectuée pour étudier les différentes méthodes de seuillages. Les méthodes classiques utilisant essentiellement le spectre fréquentiel du signal à débruiter ont tout d'abord été étudiées pour comprendre leur forces et leurs faiblesses. Un survol global des méthodes qui s'appuient sur le filtrage adaptatif a aussi été effectué. Les méthodes de seuillages qui sont en générale très simple à utiliser ont été testées puis comparées pour tenter de déduire la meilleure combinaison.

Trois éléments essentiels ont été considérés, le choix du type d'ondelettes, la méthode de seuillage à utiliser, et enfin le critère à considérer pour le calcul du seuil.

Pour le choix du type d'ondelettes, il fallait trouver une ondelette qui soit bien adaptée au débruitage de la parole, ce qui n'est tout à fait pas évident. Nous avons procédé à quelques essais avec plusieurs types. L'ondelette de Daubechies s'est avérée intéressante dans notre cas. La encore il fallait décider de l'ordre à choisir, d'où la nécessité de faire entrer ce paramètre dans nos tests d'essai.

Le choix de la méthode de seuillage est complexe. Nous avons tenté de combiner plusieurs méthodes. Le procédé qui a abouti à des résultats satisfaisants était de combiner le seuillage dur et le seuillage doux.

Le choix du critère est lui aussi très important, en particulier si le bruit additif n'est pas stationnaire tel que le bruit blanc considéré dans ce travail.

Les résultats de 1' approche utilisée sont satisfaisant dans la mesure où elle améliore quatre méthodes à la fois : le seuillage doux , le seuillage dur, le seuillage dépondant de niveau doux et le seuillage dépondant de niveau dur. ont été obtenus en effectuant plusieurs combinaisons pour déterminer les noeuds dont les coefficients doivent êtres seuillés par le seuillage doux , et ceux dont les coefficients doivent l'être par le seuillage dur. À chaque niveau de décomposition, nous avons également pris en compte le paramètre SNR et 1' ordre de l'ondelette de Daubechies .

La programmation du projet sur matlab a été menée à terme, ce logiciel permet de visualiser graphiquement les résultats et de charger automatiquement des données stockées sur ordinateur, ce qui procure une bonne représentation visuelle des signaux bruités et débruités et permet de simuler les sorties quasiment en temps réel.

Cependant, la programmation en temps réel proprement dite n'a malheureusement pas été effectuée, ce qui devrait être très intéressant à mettre en œuvre. Nous avons tenté de programmer une version qui supporte le traitement à temps réel. Par manque de temps, cette version n'a pas été finalisée. Nous espérons que ce travail a contribué à l'élaboration de méthodes efficaces de débruitage de la parole, et qu'il sera utile aux personnes désireuses d'effectuer d'autres recherches dans ce domaine.

Nous sommes concevoir une interface graphique pour afficher Les méthodes discutés dans ce travail d'une façon simple pour l'utilisateur ,ont été évaluée en présence de différents types de bruit en utilisant la base de données TIMIT [x]et le corpus de NOIZEUS[xx].

Bibliographies

[x] arofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., and Dahlgren, N. L., "DARPA TIMIT Acoustic Phonetic Continuous Speech Corpus CDROM," NIST, 1993.

[xx] LOIZOU, P. C. Subjective evaluation and comparison of speech enhancement algorithms. Speech Commun, 2007, vol. 49, p. 588-601.

[xxx] ABDELKRIM LALLOUANI, "Débruitage d'un signal de la parole corrompu par un bruit coloré en utilisant la transformée en ondelettes et implantation sur un processeur de traitement numérique des signaux", thèse Magister de l' université du QUÉBEC, 2004.

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

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

[3] FREDDY MUDRY, " Traitement des signaux et quelques applications " http://www.yopdf.com/freddy-mudry .2011

[4] BOITE. R & KUNT. M, " Complément au traité d'électricité, Traitement de la Parole" 2007

[5] LAPRIE Yves, " analyse spectrale de la parole " 16 octobre 2009

[6] AMRANE ANIS ABDE EL AZIZ, " Détection de l'onde P de L'électrocardiogramme Par des algorithmes basés sur la transformée en Ondelette et modèle Markov caché " Thèse magistère en électronique Université de Skikda

[7] Doriano-Boris POUGAZA , "Compression Sans Perte Par Ondelettes " , Mars 2008.

[8] Demaeyer Jonathan, Bebronne Michael et Forthomme Sébastien, " les Ondelettes " , Université Libre de Bruxelles ,2006

[9] Redha BENZID, " Ondelettes et statistiques d'Ordre supérieur aux signaux uni Et Bidimensionnels ", thèse de doctorat science en électronique , Université de Batna, 2005

[10] Yves Meyer, Stéphane Jaffard et Olivier Rioul, " L'analyse par Ondelettes " Pour la science, Septembre 1987

[11] A. BOUZID1 et N.ELLOUZE 2 , "Caractérisation des singularités du signal de Parole" , 1ISET de Sfax, Rte Mahdia Km 2.5, BP 88A, 3099 Elbustan Sfax, Tunisie, 2ENIT, Ecole Nationale d'Ingénieurs de Tunis, BP. 37, Le Belvédère 1002 Tunis, Tunisie.

[12] Y. LAKSARI, H. AUBERT et J.Y. TOURNERET " Analyse en Ondelettes de la Réponse Impulsionnelle Bruitée de Structures Multicouches Fractales ", Laboratoire D'Electronique, 31071 Toulouse, France 2003

[13] Alain MAGUER - Pierre ALINAT - Georges GOULLET ," Etude de bruits Impulsifs Par analyse par banc de filtres a Q constant " , Douzième Colloque Gretsi – Juan-les-Pins 12 Au 16 Juin 1989.

[14] Aicha BOUZID & Noureddine ELLOUZE , " production Multiéchelle pour la Détection Des Instants d'ouverture et de Fermeture de la Glotte Sur le Signal de Parole" , Laboratoir Signal, Image et reconnaissance de formes (LSTS-ENIT), Tunis,1997

[15] YOUSSEF BENTALEB. " Analyse par Ondelettes des Signaux Sismiques : Applications Aux ondes de surface ". Thèse de Doctorats Université *Mohamed VAGDAL Rabat*

[16] Tarik AL ANI. " Introduction aux ondelettes, Deuxième partie : Quelques concepts Généraux de la théorie des ondelettes". Département Informatique ESIEE Paris ,2005

[17] JAFFARD Stéphane ," Décompositions en ondelettes ", 2003

[18] ALAIN YGER , " traitement du signal et Ondelettes ", Master ingénier Mathématique, 18 octobre 2006

[19] Jean-Yves ANTOINE, " Outils informatiques d'analyse de corpus ", LI – Université Rabelais de Tours.

[20] Talbi MOURAD, Cherif ADNEN ," Débruitage de la parole par paquet D'ondelettes ", Conférence International IEEE. SETIT 2007 .TUNISIA.

[21] Yves Meyer, " Ondelettes, Filtres Miroirs en Quadrature et Traitement Numérique ", Exposé n ° 2 .2002

[22] Mallat S., (1989), *A theory for multiresolution signal decomposition : The wavelet representation,* IEEE Trans. Patt. Anal Machine Intell.

[23] Mallat, S., (1989), *Multiresolution approximations and wavelet orthonormal bases of L2(R),* Trans. Amer. Math.

[24] Mallat, S. (1998), *A wavelet tour of signal processing.* New York: Academie Press.

[25] Mitisi M., Y. Mitisi, (2003), G. Oppenheim et J. Poggi, *"Les ondelettes et leur applications",* Paris :Hermès Science Publications.

[26] Donoho, D.L., (1995), *Denoising by soft thresholding,* IEEE Trans. Information Theory, vol. 41.

[27] Donoho, D.L., and Johnstone, LM., (1994), *Ideal spatial adaptation by wavelet shrinkage,* Biometrika, vol. 81.

[28] Sungwook Chang, Y. Kwon, Sung-il Yang, and I-Jae Kim, *Speech Enhancement for non-stationnary noise environment by adaptive wavelet packet,* Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '02), vol. 1 , 2002.

[29] Texas Instruments, (1999), *" TMS320C6711, TMS320C6711B, TMS320C6711C Floating-Point Digital Signal Processors",SPRS088H,* Texas Instruments, Dallas,TX.

[30] Johnstone, I. M., and Silverman, B.W., (1997), *Wavelet threshold estimators for data with correlated noise,* J. Roy. Statist. Soc. B.

[31] ITU-T Recommendation G.114. "One-way transmission time", 2003.

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

[33] P. Gray, M. P. Hollier, and R. E. Massara, "Non-intrusive Speech Quality Assessment using Vocal-tract Models," IEE Proceedings - Vision, Image and Signal Processing, vol. 147, pp. 493 - 501, Dec. 2000.

[34] ITU-T Recommendation G.107. "The E-model, a computational model for use in transmission planning", International Telecommunication Union CH-Geneva. 2005.

[35] ITU Recommendation P.800, "Methods for Subjective Determination of Transmission Quality", August 1996.

[36] N. Nocerino, F.K. Soong, L.R. Rabiner, D.H. Klatt, "Comparative study of several distorsion measures for speech recognition" Proc. IEEE Internat. Conf. Acoust. Speech Signal Process., 1985 (1985), pp. 25–28

[37] Klatt, D. Prediction of perceived phonetic distances from critical band spectra. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 7, pp.1278–1281. Paris, France (1982)

[38] ITU-T Recommendation P.861, "Objective quality measurement of telephone-band (300 3400 Hz) speech codecs" 1998.

[39] ITU–T Contrib. COM 12–20. "Improvement of the P.861 Perceptual Speech Quality Measure". International Telecommunication Union, CH–Geneva. 1997.

[40] M. P. Hollier, M. O. Hawksford and D. R. Guard, D. R. "Error Activity and Error Entropy as a Measure of Psychoacoustic Significance in the Perceptual Domain". IEE Proceedings-Vision, Image and Signal Processing, 141(3):203–208. 1994.

[41] A. Rix, M. Hollier, A. Hekstra, and J. Beerends, "Perceptual Evaluation of Speech Quality (PESQ), the New ITU Standard for End-to-End Speech Quality Assessment, Part I-Time Alignment". Journal of the Audio Engineering Society, 50(10):755. 2002.

[42] J. Beerends, A. Hekstra, A. Rix, and M. Hollier, "Perceptual Evaluation of Speech Quality (PESQ): The New ITU Standard for End-to-End Speech Quality Assessment Part II -Psychoacoustic Model". Journal of the Audio Engineering Society, 50(10):765–778. 2002.

[43] ITU–T Rec. P.862.2 “Wideband Extension to Recommendation P.862 for the Assessment of Wideband Telephone Networks and Speech Codecs”, International Telecommunication Union, CH–Geneva. 2005.

[44] UIT-T Rec. P.862 (2001). Évaluation de la qualité vocale perçue : méthode objective d'évaluation de la qualité vocale de bout en bout des codecs vocaux et des réseaux téléphoniques à bande étroite.

[45] Rix, A. W., Hollier, M. P., Hekstra, A. P. et Beerends, J. G. (2002). Perceptual Evaluation of Speech Quality (PESQ), the new ITU standard for end-to-end speech quality assessment - Part I : time-delay compensation. Journal of Audio Engineering Society, 50(10):755_764.

[46] Zwicker, E. et Feldtkeller, R. (1981). Psychoacoustique : l'oreille récepteur d'information. Collection technique et scientifique des télécommunications -CNET - ENST. Masson, Paris, France.

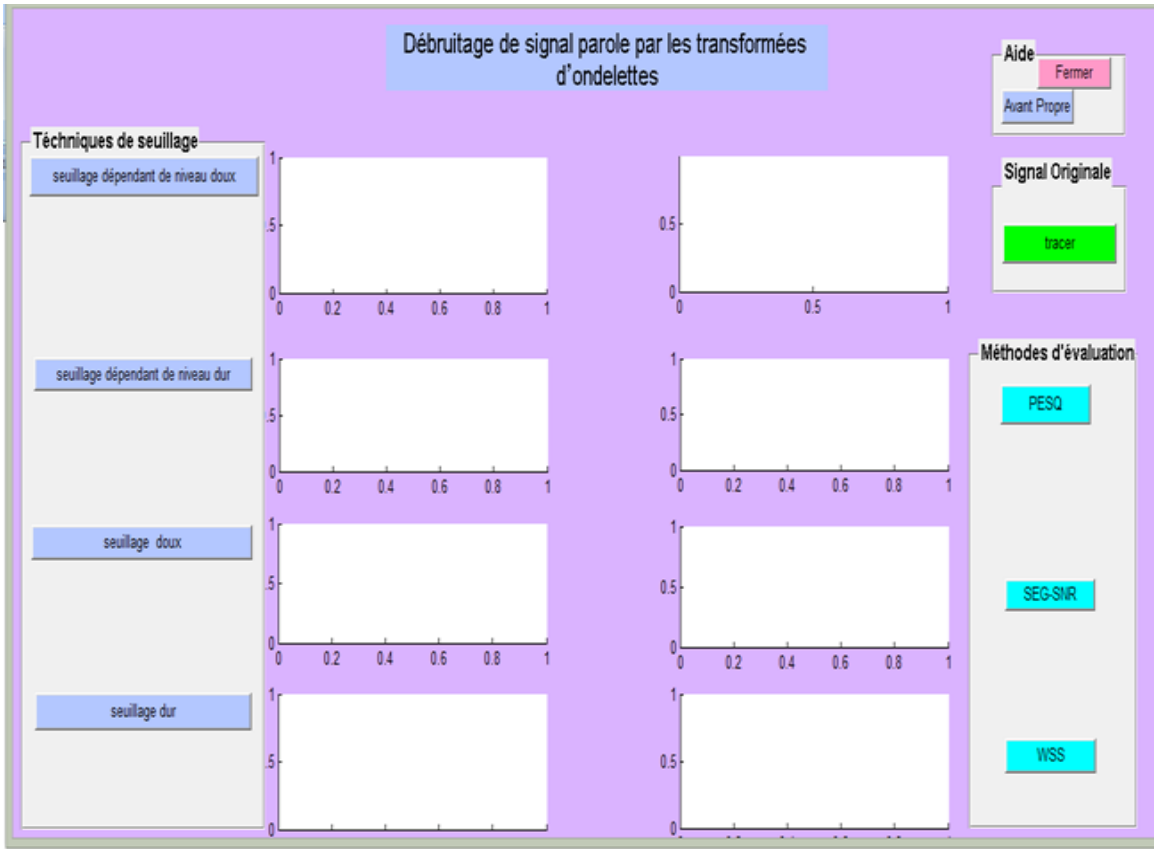
[47] ITU-T P.863, “Perceptual Objective Listening Quality Assessment (POLQA)”, Geneva, January 2011.

[48] ETSI ETR 250, “Speech Communication Quality from Mouth to Ear for 3,1 kHz Handset Telephony across Networks”, European Telecommunications Standards Institute, FR–Sophia Antipolis. 1996.

ANNEXE

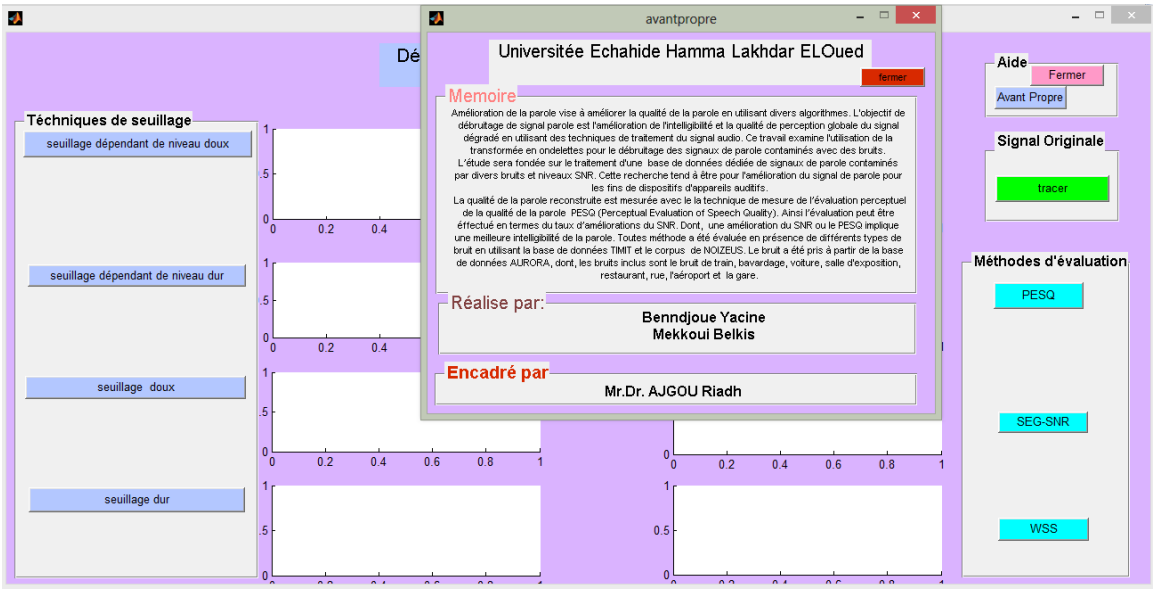
INTERFACE GRAPHIQUE

Dans cette interface, nous avons présenté les techniques de débruitage de la parole en utilisant la transformée en ondelettes et les méthodes d'évaluation des cette techniques comme suit:

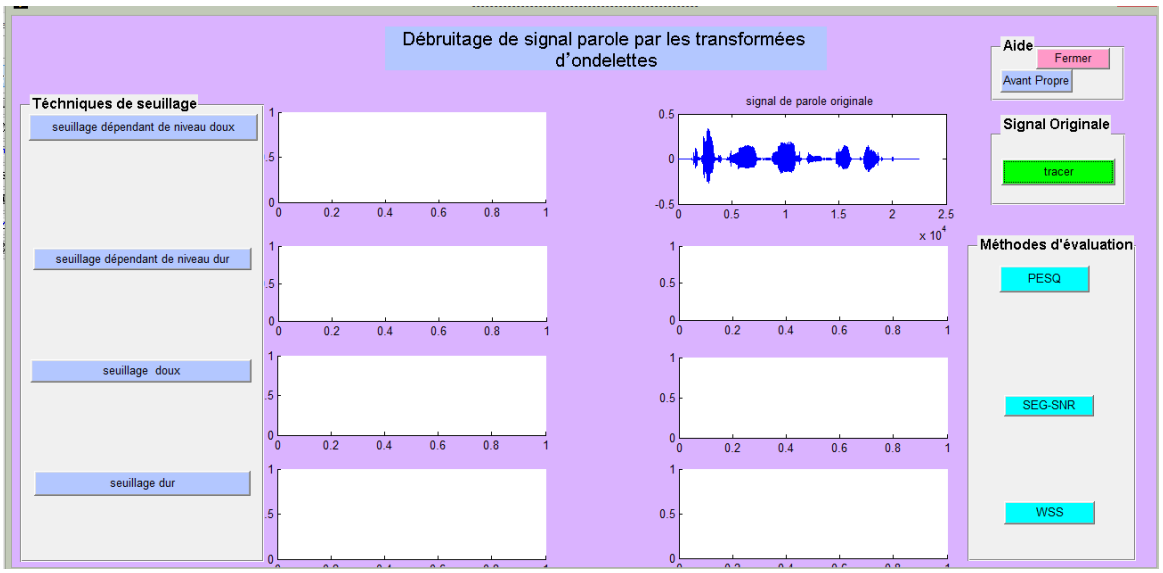


✓ Exemple d'utilisation de cette interface pour le bruit de bavardage (babble):

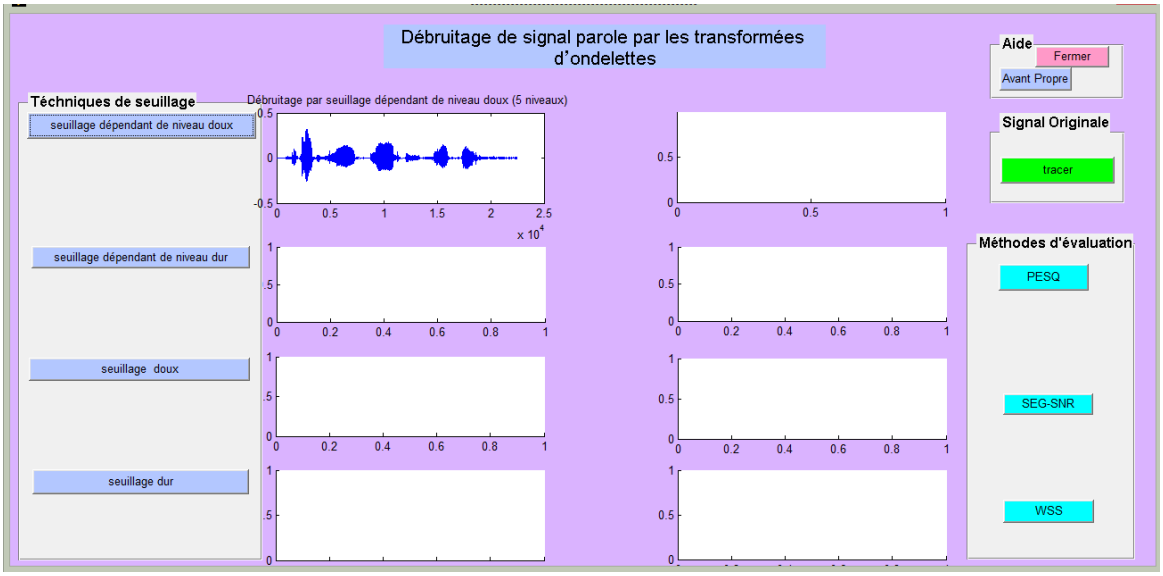
1- Cliquer sur le bouton 'avant propre' on obtiens:



2- par exemple je veux affiché le signal parole originale , Cliquer sur le bouton 'tracer' on obtiens:



3- par exemple je veux affiché le technique de seuillage dépendant de niveau doux, Cliquer sur le bouton ' seuillage dépendant de niveau doux' on obtiens:



Et le même chose pour les autres bouton.