الجمهورية الجزائرية الديمقراطية الشعبية

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DR. YAHIA FARES DE MEDEA

FACULTE DES SCIENCES ET DE LA TECHNOLOGIE

Département du Génie Électrique
et de l'Informatique



Mémoire de Fin d'Études
Présenté pour l'obtention du Diplôme
de Master en Génie Électrique

Option : Systèmes de Télécommunication

THEME

Débruitage des signaux de la parole corrompu par un bruit coloré

Proposé et dirigé par,
M. Boucherit Ismaile

Préparé et présenté par,
MAZIGHI Ahmed
OUKRIF Sofiane

Année Universitaire 2014-2015

Dédicace

Je dédie ce modeste travail :

A mes très chers parents
Mon très cher père pour sa patience et tous ses efforts

A ma mère pour m'avoir épaulé, encouragé à reprendre les études et motivé dans les moments les plus difficiles

A mes freres.
Walid et Lokmane

A mes ancleset mes tantes.

Pour tous leurs encouragements

A mon encadreur Monsieur Boucherit pour l'excellence de son accompagnement et la confiance qu'il nous a accordé.

A tous mes professeurs ainsi que tous les étudiants

de la promotionde télécommunication 2014/2015, qui nous ont accueillis si généreusement et aidés tout au long années universitaire.

A tousmes amis.

Et toute personne que je connais et qui me sont chers et tous ceux
qui m'aiment.

Ahmed

Dédicace

Je dédie ce modeste travail :

A mes très chers parents
Mon très cher père pour sa patience et tous ses efforts

A ma mère pour m'avoir épaulé, encouragé à reprendre les études et motivé dans les moments les plus difficiles

A mes freres.
Sidali et Mohamed

A mes ancleset mes tantes.

Pour tous leurs encouragements

A mon encadreur Monsieur Boucherit pour l'excellence de son accompagnement et la confiance qu'il nous a accordé.

.

A tous mes professeurs ainsi que tous les étudiants

de la promotionde télécommunication 2014/2015, qui nous ont accueillis si généreusement et aidés tout au long années universitaire.

A tousmes amis.

Et toute personne que je connais et qui me sont chers et tous ceux qui m'aiment.

Sofiane



REMERCIEMENTS

En premier lieu, nous remercions Dieu qui nous a procuré ce succès.

Nous tenons à remercier vivement notre promoteur Mr Boucherit pour ses conseils précieux et pour toutes les commodités et aisances qu'il nous a apportées durant notre étude et réalisation de ce projet.

Nos remerciements les plus vifs s'adressent aussi aux messieurs le président et les membres de jury d'avoir accepté d'examiner et d'évaluer notre modeste travail.

Nous exprimons également notre gratitude à tous les professeurs et enseignants qui ont collaboré à notre formation depuis notre premier cycle d'étude jusqu'à la fin de notre cycle universitaire. Sans omettre bien sur de remercier profondément tous ceux qui ont contribué de prés ou de loin à la réalisation du présent travail.

Et enfin, que nos chers parents et familles, et bien avant tout, trouvent ici l'expression de nos remerciements les plus sincères et les plus profonds en reconnaissance de leurs sacrifices, aides, soutien et encouragement afin de nous assurer cette formation de master dans les meilleures conditions



Résumé

Le débruitage de la parole est un domaine de traitement du signal. En effet, dans un monde ou la télécommunication connait un véritable essor, les technologies se doivent être de plus en plus performantes afin de satisfaire au plus grand nombre d'applications. Les applications qui nécessitent un rehaussement de la parole sont très nombreuses, la plus évidente étant sans doute celle de la téléphonie mobile, où de nombreux bruits environnants peuvent gêner la qualité et l'intelligibilité du signal de parole transmis.

Dans ce travail ,on commence par les notions et les caractéristique du bruit et la parole puis différentes techniques de débruitage du signal de parole telques les méthodes de la soustraction spectrale, le filtrage adaptatif de Wiener et la méthode de sous-espace sont présentées et finalement une étude comparative entre les techniques (soustraction spectrale multi-bande et la méthode classique) en basant sur des critères d'évaluation objectifs . En effet, meilleur résultats implique la meilleure méthode pour le rehaussement de la qualité de la parole.

Mot clés : Parole et bruit, Qualité et intelligibilité, Evaluation de la qualité vocale, Algorithmes de débruitage.

Abstract

The Enhancing Speech is an area of the treatment of the signal . Indeed, in a world or telecommunication makes true great strides, technologies must be increasingly powerful in order to satisfy greatest number. The applications which require a noise suppression are very many, most obvious being undoubtedly that of mobile telephony, where many surrounding noises can obstruct the quality and the intelligibility of the signal transmitted.

In this work, we starts with the concepts and the characteristic of the noise and the speech then different techniques of noise suppression as the methods of the spectral subtraction, the adaptive filtering of Wiener and the method of subspace are presented and finally a comparative study of two algorithms (spectral subtraction multi-band and classical method) while basing on objective evaluation . Indeed, better results implies the best method for Enhancing speech.

Key words : speech and noise, Quality and intelligibility, Evaluation of speech quality, Noise reduction algorithm.

ملخص

نزع الضوضاء من الصوت هو مجال معالجة.في عالم أصبحت فيه الاتصالات تشهدطفرة حقيقية ؛فإن التكنولوجيا تكون فعالة على نحو متزايد من أجل إرضاء أكبر عدد من التطبيقات. التطبيقات التي تتطلب تحسين الصوت كثيرة جدا، و الأكثر تطلبا دون شك هو مجال الهواتف القالة، حيث الكثير من التشويشات يمكن أن تتداخل مع جودة و وضوح الإشارة الصوتية المرسلة.

في هذا العمل نبدأ مع مفاهيم ومميزات الضوضاء والصوت ثم نتطرق الى تقنيات مختلفة من معسنات الصوت مثل أساليب الطرح الطيفي، وطريقة وينر للتصفية وطريقة الفضاء الجزئي و في الأخير مقارنة بين مختلف التقنيات(الطرح الطيفي متعدد الفرق و الطريقة التقليدية) بناءا على معايير تقييم موضوعية و أفضل النتائج تدل على أفضل وسيلة لتحسين نوعية الصوت.

كلمات مفتاحية : الصوت والضوضاء ، الجودة و الوضوح، تقييم جودة الكلام، خوارزمية نزع الضوضاء.

Liste des Figures

Figure I.1. M*odèle simple de production de la prole* …… 3

Figure I.2. *Différence entre les différents types de fenetres* ……..... 6

Figure I.3. *Seuil d'audition absolu* …… 7

Figure I.4. *Onde triangulaire 440 Hz (forme d'onde à gauche, FFT à droite).…..* 13

Figure I.5. *Onde en dents-de-scie 440 Hz (forme d'onde à gauche, FFT droite).……..…* 14

Figure I.6. *Représentation schématique de différentes sources de variabilité qui peuvent dégrader les performances d'un système de reconnaissance de la parole.* 16

Figure I.7. *Spectre d'un bruit blanc* ……... 17

Figure I.8 *Les radiations qui compose la lumière blanche.……..... 17*

Figure I.9.*Spectre d'un bruit blanc et d'un bruit rose dans la bande audio.……* 18

Figure II.1. *Modèle de débruitage utilisé dans ce document.…….* 22

Figure II.2. *Schéma-bloc d'un réducteur de bruit par la soustraction spectrale.……..* 24

Figure III.1. Le RSB segmentaire de quatre bandes de fréquences de la parole corrompues……... 33

Figure III.2.*Diagramme de la réduction du bruit utilisant un analyse multi-bande.…….....* 36

Figure.III.3 *Spectrograme d'un signal de parole propre* …… 45

Figure III.4. *Specrtrograme d'un exemple de débruitage par la méthode de soustraction spectrale classique.……..* 46

Figure.III.5. *Spectrograme d'un exemple de débruitage par la méthode de soustraction spectrale multi-bande.……* 46

Figure.III.6. *Spectrograme d'un exemple de débruitage par la méthode filtrage de wienner (bruit airport).……* 46

Figure III.7 .*Comparaison en terme de moyenne de PESQ.……* 47

Figure III.8 .*Comparaison en terme de moyenne de PESQ.……..... 49*

Figure III.9 .*Comparaison en terme de moyenne de RSBseg.…….* 50

Figure III.10 .*Comparaison en terme de moyenne de WSS*…….. 51

Tableau. I.1. Echelle catégorielle de mesure pour le test ACR (UIT-T P.800) 9

Tableau. I.2. Echelle catégorielle de mesure pour le test DCR (UIT-T P.800). 10

Tableau. I.3. Echelle catégorielle de mesure pour le test CCR (UIT-T P.800) 10

Tableau. I.4. Echelle catégorielle de mesure pour le test MUSHRA (UIT-R BS.1534) 11

Tableau I.5. Différentes classes de bruit 19

Tableau III.1. *Mesure du LLR ,SNRseg ,WSS et PESQ pour la méthode soustraction spectrale classique où RSB=0dB* 40

Tableau III.2. *Mesure du LLR ,SNRseg ,WSS et PESQ pour la méthode soustraction spectrale classique où RSB=5dB* 40

Tableau III.3. *Mesure du LLR ,SNRseg ,WSS et PESQ pour la méthode soustraction spectrale classique où RSB=10dB* 41

Tableau III.4. *Mesure du LLR ,SNRseg ,WSS et PESQ pour la méthode soustraction spectrale classique où RSB=15dB* 41

Tableau III.5. *Mesure du LLR ,SNRseg ,WSS et PESQ pour la méthode soustraction spectrale multi-bande où RSB=0dB* 42

Tableau III.6. *Mesure du LLR ,SNRseg ,WSS et PESQ pour la méthode soustraction spectrale multi-bande où RSB=5dB* 42

Tableau III.7. *Mesure du LLR ,SNRseg ,WSS et PESQ pour la méthode soustraction spectrale multi-bande où RSB=10dB* 43

Tableau III.8. *Mesure du LLR ,SNRseg ,WSS et PESQ pour la méthode soustraction spectrale multi-bande où RSB=15dB* 43

Tableau III.9. *Mesure du LLR ,SNRseg ,WSS et PESQ pour la méthode filtrage de wienner où RSB=0dB* 44

Tableau III.10. *Mesure du LLR ,SNRseg ,WSS et PESQ pour la méthode filtrage de Wienner où RSB=5dB* 44

Tableau III.11. *Mesure du LLR ,SNRseg ,WSS et PESQ pour la méthode filtrage de wienner où RSB=10dB* 45

Tableau III.12. *Mesure du LLR ,SNRseg ,WSS et PESQ pour la méthode filtrage de wienner où RSB=150dB* 45

ACR	Absolute Category Rating
AFNOR	Association française de normalisation
CCR	Comparison Category Rating
CMOS	Comparison Mean Opinion Score
DCR	Degradation Category Rating
DMOS	Degradation Mean Opinion Score
DRT	Diagnostic Rhyme Test
DSP	Densité Spectrale de Puissance
FFT	Fast Fourier Transform
IFFT	Inverse Fast Fourier Transform
KLT	Transformée de Karhunen-Loeve
LLR	Log-Likelihood Ratio
LPC	Linear Predictive Coding
MEQ M	Minimum de l'Erreur Quadratique Moyenne
MUSHRA	Multi Stimulus test with Hidden Reference and Anchor
PESQ	Perceptual Evaluation of Speech Quality
RAP	Reconnaissance Automatique de la Parole
RSB	Rapport Signal sur Bruit
RSBseg	Rapport Signal à Bruit segmentaire
SSC	Soustraction Spectrale Classique
SSM	Soustraction Spectrale Multi-bande
TFD	Transformée de Fourier Discrète
WSS	Weighted Spectral Slope

Table des Matières

Dédicaces

Remerciements

Résumé

Mots clés

Liste des figures

Liste des tableaux

Abréviations

Tables des matières

Introduction générale ……... 1

Chapitre 1 : Généralités sur la parole et le bruit

1. Introduction……..……..……..... 2

2. Définition de parole……..…..... 2

3. Production du signal de parole ……..… 2

4. Perception de la parole ……..……..…… 3

5. Caractéristiques ……..……... 3

5.1 Le domaine temporel ……...……...……. 3

5.2 Le domaine fréquentiel ……..……..….... 4

5.2.1 La fréquence fondamentale ……..……..... 4

5.2.2 Les formants ……..……...……..... 4

6.Visualisation des sons ……..…...……... 4

6.1 L'oscilloscope ……...……... 4

6.2 Spectrogramme ……. 5

7. Les fenêtres temporelles ……...……... 6

8. Notions de psychoacoustique……..…… 7

9. Qualité et intelligibilité de la parole ……...……... 7

10.Evaluation de la qualité vocale ……...…….... 8

10.1 critères subjectifs ……...…… 8

10.1.1.Test ACR (Absolute Category Rating) …..…..……... 8

10.1.2.Test DCR (Degradation Category Rating) ….....…..…….... 9

10.1.3.Test CCR (Comparison Category Rating) …….…...……. 10

10.1.4. Test MUSHRA …… 11

11. Le bruit ……...……. 12

12. Définition du bruit en acoustique ….……...……..... 12

13. Les différents types de bruit ……...……..……..... 13

13.1. Les bruits additifs ……..……...……... 15

13.2. Les bruits convolutionnels ……..... 15

13.3. Les bruits physiologiques ……...……. 15

14. bruit blanc et bruit coloré ……...……..... 16

14.1. Définition techniques ……...……..…..... 17

14.2. Le bruit blanc ……..……...……..... 17

14.3. Le bruit coloré.……...…… 17

15.Qu'est ce qu'un bruit musical.……..……..... 18

16. Les caractéristiques du bruit ….…….…….……..... 18

17. Problématique de la réduction de bruit ……...…… 19

17.1. La gêne due au bruit ……...…….. 19

17.2. Les applications … ….……...……...……... 20

18. Conclusion …...…….... 21

Chapitre 2 : Etat de l'art des algorithmes de débruitage de la parole

1. Introduction…...……..... 22

2. Soustraction spectrale ……..…..……..... 23

2.1.Principe …… 23

2.2.Soustraction spectrale par filtrage …… 24

2.3.Amélioration de la soustraction spectrale par Berouti ……..……..... 25

2.4.Amélioration de la soustraction spectrale par Boll ..…….…..…….. 25

2.4.1.La moyenne d'amplitude ……..……..……..... 26

2.4.2.Correction de l'estimateur ……..……..…….. 26

2.4.3.Réduction du bruit résiduel ……...…… 27

2.4.4.Atténuation additionnelle durant les silences ……..... 27

2.5.Amélioration de la soustraction spectrale par Virag……...…..... 27

2.6.Amélioration psychoacoustique ……..……..…… 28

2.7.Soustraction spectrale multi-bande ……..…… 28

3.Filtrage de wienner ……...……...……..... 29

3.1.Principe ……...……..... 29

3.2.Amélioration psychoacoustique ……..……... 29

4. Méthode à sous-espace ……..…….. 29

5. Conclusion ……..…….. 31

Chapitre 3 : Evaluation et Comparaison entre les méthodes de débruitage de la parole.

1.Introduction ……...…….. 32

2.Soustraction spectrale multi-bande ……...…… 32

2.1. Mise en œuvre…… 35

3.Evaluation de la qualité de la parole ……...…….. 36

3.1.Les critères objectifs ……...……. 36

3.1.1 RSB segmentaire (RSBseg) ……...…...…… 36

3.1.2 Perceptual Evaluation of Speech Quality (PESQ) ……... 37

3.1.3.Log-Likelihood Ratio (LLR)……...…..…….. 38

3.1.4.Weighted Spectral Slope (WSS)……...…..…...…….. 38

4. Comparaison entre la méthode de soustraction spectrale classique , la méthode de multi-bande et le filtrage de Wienner……... 39

4.1 La base des signaux……..…….. 39

4.2.Les graphes ……. 47

5.Conclusion …… 52

Conclusion générale..... 53

Bibliographie 54

Introduction générale

La dégradation d'un signal vocal par un bruit d'environnement engendre des problèmes perceptifs ennuyants pour l'oreille humaine et détériore les performances des systèmes de communication homme machine. La réduction du bruit en parole continue constitue un des problèmes majeurs, qui a donné naissance à un grand besoin de développer des algorithmes appelés les algorithmes de débruitage de la parole capables d'éliminer les signaux indésirables dégradant le signal vocal et dont l'objectif essentiel est d'améliorer l'aspect perceptuel et l'intelligibilité du signal vocal. Actuellement, les méthodes d'amélioration de la qualité vocale, intègrent des hypothèses sur le signal utile et le bruit. Ces techniques sont associées à un compromis entre une distorsion introduite dans le signal qu'on cherche à minimiser et un effet de réduction du bruit qu'on cherche à maximiser.

Dans le premier chapitre, nous allons présenter des généralités, on invoquera alors les définitions du bruit et la parole, la production et la perception du signal de parole, les différents type de bruit leurs principaux caractéristiques, les principaux notions dans le domaine de traitement de la parole avec la citation des différents critères d'évaluation subjectifs de la qualité vocale.

Dans le deuxième chapitre nous présenterons quelques algorithmes de débruitage de la parole et on termine par le troisième chapitre qui concerne des évaluations comparatives sur des critères objectifs de qualité sont présentées pour plusieurs types de bruit et de rapport signal à bruit. Les résultats ont révélé la supériorité des méthodes proposés dans ce travail.

1. Introduction :

La parole est l'un des principaux moyens de communication entre êtres humains, sa simplicité en fait d'ailleurs le moyen de communication le plus populaire dans la société humaine (il est plus facile de parler à quelqu'un que de lui écrire ou de lui faire un schéma) [1].

Ce chapitre résume les principaux caractéristiques de la parole et du bruit : pour la parole, de sa production à sa modélisation, sa perception et la qualité et intelligibilité de la parole ; pour le bruit, on citera les aspects acoustique du bruit, la définition techniques du bruit blanc et coloré et les différentes classes aux quelle un bruit peut appartenir.

Nous aborderons aussi les critéres subjectives d'évluation de la qualité vocale et quelques notions nécessaire au traitement de la parole telque la visualisation du son, la transformé de Fourier et le fenêtrage .

2. Définition de parole :

Une onde de pression située entre 20 Hz et 20 kHz devient un son car il est situé dans le domaine audible (c'est le spectre de l'audition). Il peut alors être interprété comme un signal sonore par le cerveau.

Le son a plusieurs origines. Il est dépendant du lieu dans lequel il est produit. Il peut être de la parole, de la musique, du bruit et la plupart du temps, un mixage de ces trois informations. Notre oreille capte le son, qui est ensuite transféré au cerveau pour le comprendre et l'interpréter.[2]

3. Production du signal de parole :

Le signal de parole est le résultat de l'excitation du conduit vocal par un train d'impulsions ou un bruit donnant lieu respectivement aux sons voisés et non voisés figure I.1 Dans le cas des sons voisés, l'excitation est une vibration périodique des cordes vocales suite à la pression exercée par l'air provenant de l'appareil respiratoire.

Ce mouvement vibratoire correspond à une succession de cycles d'ouverture et de fermeture de la glotte. Le nombre de ces cycles par seconde correspond à la fréquence fondamentale F0. Quant au signaux non-voisés, l'air passe librement à travers la glotte(du moins pas dans tout le conduit vocal) sans provoquer de vibration des cordes vocales.

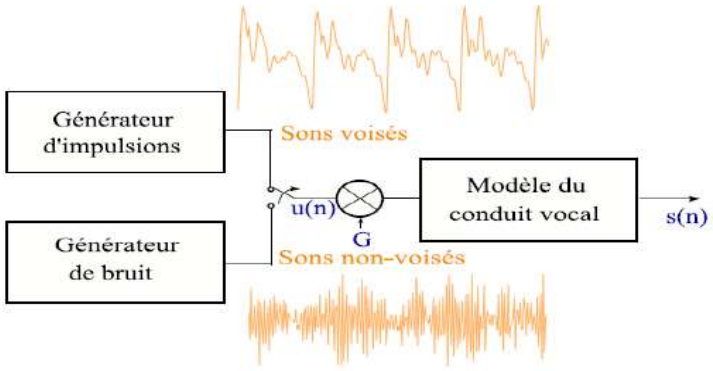


Figure I.1. *.modèle simple de production de la prole[3]*

4. Perception de la parole:

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

Le premier trait est la fréquence fondamentale, fréquence de vibration des cordes vocales. Ses variations définissent le pitch qui constitue la perception de la hauteur (où les sons s'ordonnent de grave à aigu). Seuls les sons quasi-périodiques (voisés) engendrent une sensation de hauteur tonale bien définie.

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

Le dernier trait acoustique est l'énergie correspondant à l'intensité sonore. Elle est habituellement plus forte pour les segments voisés de la parole que pour les segments non voisés [3].

5. Caractéristiques :

Comme tous les autres sons arrivant aux oreilles, c'est une onde de pression. Elle varie suivant les personnes (hommes, femmes ou enfants), suivant l'intensité (énervement, chuchotement)... La parole est décrite dans plusieurs domaines :

5.1.Le domaine temporel :

– L'enveloppe, qui contient l'énergie du signal

– La structure fine, qui contient les différentes variations du signal [3].

5.2.Le domaine fréquentiel :

5.2.1. La fréquence fondamentale :

La fréquence fondamentale F0 (ou pitch [période des sons voisés]) joue un rôle important dans la parole. C'est elle qui véhicule une grande partie de l'information .

L'intensité de la voix et les durées successives des syllabes complètent ces informations.

D'une manière générale, la prosodie, qui peut être considérée comme l'effet des différentes variations de la fréquence fondamentale F0, de l'intensité et de la durée, peut faire ressortir bien des caractéristiques du locuteur, comme son genre, ses origines géographiques et culturelles, ses émotions, etc. mais participe aussi à la caractérisation de la langue elle-même, par la manière dont elle est utilisée pour différencier les divers éléments syntaxiques comme les énoncés (interrogatifs, exclamatifs ou déclaratifs), l'importance de certains mots, ou bien même pour caractériser les différences lexicales entre les mots. F0 moyen-homme 100 à 150 Hz, F0 moyen-femme 200 Hz à 300 Hz et F0 moyen-enfant 350 à 400 Hz.

5.2.2.Les formants :

Le voisement que nous percevons est différent de celui produit à la source par les cordes vocales. Ce que nous entendons est le fruit d'un phénomène de filtrage sur une onde complexe. La capacité d'amplifier ou, au contraire, d'atténuer certaines fréquences est la propriété de tout résonateur. Dans le cas précis de la parole, le stimulus est fourni par l'onde périodique complexe provenant du mouvement des cordes vocales et ce sont les cavités supra-glottiques qui assurent la fonction de résonateur. La forme, la dimension ainsi que la matière qui compose ces cavités sont autant de particularités qui détermineront les fréquences qui seront mises en évidences et celles qui seront atténuées. Les cavités supra-glottiques ont la capacité de neutraliser certaines harmoniques et d'en mettre d'autres en évidence par un simple changement de configuration. Donc le conduit vocal possède une fonction de transfert (filtre) appliquée à une source produit un phonème, on appelle formant les maxima locaux (pics) de cette fonction de transfert notés F1, F2, F3… et correspondent aux zones de renforcement maximal. Néanmoins, du point de vue perceptif, seuls quelques formants jouent un rôle central au niveau de la parole [1].

6. Visualisation des sons :

6.1. L'oscilloscope :

L'oscilloscope, l'une des plus anciennes de ces représentations, montre l'évolution temporelle de l'amplitude du signal. C'est une simple fonction du temps qui ne dévoile pas la structure interne du son (sa composition fréquentielle) et qui se révèle peu intéressante pour des objets sonores complexes et notamment pour l'étude de la parole. Un objet sonore est généralement défini à l'aide de trois grandeurs principales (trivariance acoustique) : son intensité, sa composition fréquentielle et sa durée. Mettre en image un son consiste à trouver

une représentation qui soit liée à ces trois grandeurs ce qui nécessite, donc, un espace à trois dimensions (amplitude, fréquence, temps).

6.2 Spectrogramme :

Le spectrogramme est un outil de visualisation utilisant la technique de la transformée de Fourier et donc du calcul de spectres. Il a commencé à être largement utilisé en 1947, à l'apparition du sonagraphe, et est devenu l'outil incontournable des études en phonétique pendant de nombreuses années. L'apparition de l'informatique puis d'écrans graphiques de bonne qualité a permis d'abandonner tout matériel comme le sonagraphe mais la technique du spectrogramme est encore aujourd'hui largement utilisée du fait de sa simplicité de mise en oeuvre et du grand nombre d'études qui ont déjà été réalisées.

Le spectrogramme permet de mettre en évidence les différentes composantes fréquentielles du signal à un instant donné, une transformée de Fourier rapide étant régulièrement calculée à des intervalles de temps rapprochés. Avant le calcul des transformées successives, le signal doit d'abord être préaccentué par un filtre du premier ordre pour égaliser les hautes fréquences dont l'énergie est toujours plus faible que celle des basses fréquences.

Cette phase de préaccentuation du signal est suivie par une phase de fenêtrage, nécessaire du fait de la théorie qui sous-tend la transformée de Fourier. Dans cette méthode d'analyse, le signal est considéré comme indéfiniment stable et constitué d'une somme invariable de fonctions sinusoïdales de fréquences différentes. Pour contourner cette contrainte théorique d'invariabilité du signal, il faut convoluer le signal avec une fenêtre temporelle qualifiée de glissante puisque chaque calcul de spectre nécessite de convoluer le signal avec la fenêtre temporelle à un instant particulier.

Différentes fenêtres temporelles existent mais chacune introduit une erreur résiduelle plus ou moins importante dans le spectre obtenu du fait de la forme choisie qui peut être, dans le pire des cas, triangulaire ou carrée.[1]

$$X[f] = \sum_{n=0}^{N-1} x(n)\, e^{-j2\pi nf/N} \qquad \textbf{avec} \quad 0 \leq f < N \qquad \textbf{(I.1)}$$

Formule de la transformée de Fourier discrète

Où :
x(n) est la composante temporelle et
X(f) est la composante fréquentielle

7. Les fenêtres temporelles :

En pratique, la transformée de Fourier discrète (TFD) n'est appliquée que sur une partie du signal et non sur le total. Cette partie est appelée « trame ». Une fenêtre de pondération est ensuite appliquée pour éviter les discontinuités au début et à la fin de la trame. Introduisant dans le spectre du signal des informations non existantes et donc l'introduction d'erreur. Il existe plusieurs types de fenêtre :

– Rectangulaire : 0, lorsque ce n'est pas la trame du signal et 1, pendant la durée de la trame. Si le signal commence et finit très proche de 0 (type signaux impulsionnels) alors il n'y a pas beaucoup de déformation au niveau du signal fréquentiel car une telle fenêtre engendre un sinus cardinal au niveau du spectre.

– Hanning : C'est de loin la fenêtre la plus utilisée. Elle est particulièrement utile pour les signaux pseudo stationnaires. Partant de 0 et finissant à0, elle rend le signal périodique sur la durée de la trame. C'est elle qui introduit le meilleur compromis entre résolution fréquentielle et temporelle en amplitude.

– Hamming : Elle est similaire à Hanning, c'est en fait la dérivée de la fenêtre de Hanning, mais conduit une meilleure réponse en fréquence.

– Kaiser-Bessel : C'est l'inverse de la fenêtre de Hamming, elle donne une meilleure réponse sur les amplitudes.

– Flat Top : c'est la seule fenêtre qui donne l'amplitude exacte pour les raies spectrales. Elle est utilisée pour l'étalonnage des signaux.

– Blackman : Elle permet de réduire fortement les lobes secondaires. C'est la fenêtre la plus étroite.

La multiplication d'une fenêtre avec le signal de la trame entraine automatiquement un sinus cardinal (figure I.2). Celui-ci est plus ou moins atténué selon les caractéristiques de la fenêtre.

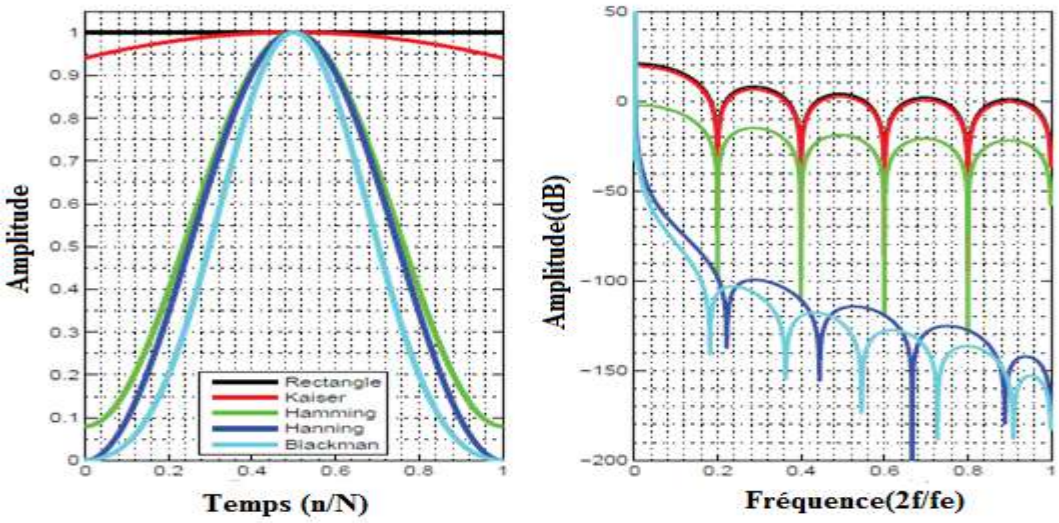


Figure I.2. *Différence entre les différents types de fenetres [2]*

8. Notions de psychoacoustique:

La psychoacoustique est l'étude de la perception des sons. Elle rassemble et décrit les relations qui existent entre le phénomène acoustique physique, sa perception par notre oreille et la description que l'on en fait. L'oreille humaine perçoit un son pur dans l'intervalle de 20 Hz à 20 kHz ; cet intervalle est variable selon les individus et décroit avec l'âge. Il existe un seuil d'audition absolu en dessous duquel l'oreille ne perçoit pas de son (figure I.3). Ce seuil caractérise l'énergie dont a besoin un son pur pour être perçu per par l'oreille en silence absolu. Il existe également un seuil traduisant la limite supérieur de la perception de l'oreille. Il est connu sous le nom de seuil de douleur car des douleurs aiguës de l'oreille apparaissent à ce niveau. Il se situe à environ 130 dB. Expérimentalement, on s'arrête à 90 dB, car là d'ores et déja les dégradation de l'audition apparaissent[3].

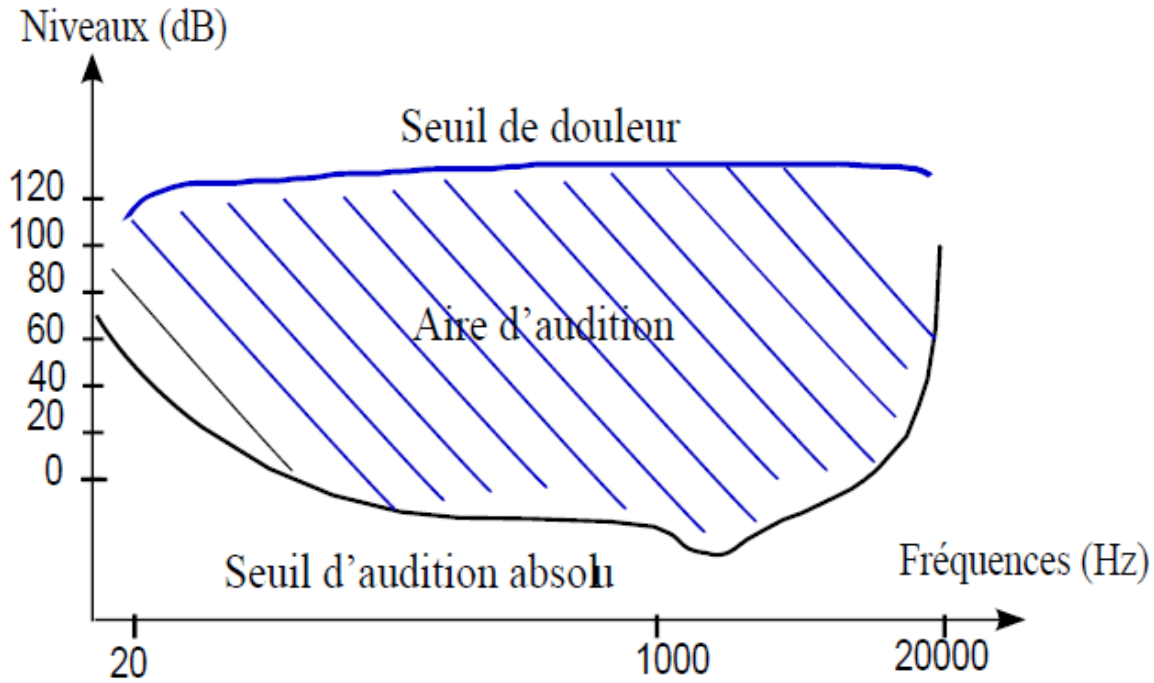


Figure I.3. *Seuil d'audition absolu[3]*

9. Qualité et intelligibilité de la parole:

L'intelligibilité de la parole correspond à la capacité de comprendre un message linguistique contenu dans un signal de parole. L'intelligibilité est donc une mesure objective définie par le nombre de mots prononcés correctement identifiés par l'auditeur. Chaque mesure d'intelligibilité est une interaction entre le locuteur, l'environnement de transmission et l'auditeur. Le meilleur moyen de juger l'intelligibilité est d'effectuer des tests d'écoute avec des sujets, dont la capacité d'écoute est normale, en utilisant par exemple la méthode du test de rime DRT (Diagnostic Rhyme Test); celui-ci permet d'évaluer la transparence du

message reçu à travers une mesure du degré de dégradation des caractéristiques élémentaires des consonnes lorsque celles-ci se trouvent au début de mots . Une version plus générale du test DRT a permis de tester tout aussi bien les voyelles que les consonnes et ce quelle que soit leur position dans un mot .

La qualité d'un signal de parole permet de prendre en compte la présence d'agents extérieurs "perturbateurs" (environnement bruyant, distorsions,. . .). La clarté du message peut en effet être affectée par ce bruit environnemental, ce qui nuit au confort d'écoute. C'est donc une mesure subjective liée à l'aspect agréable de l'écoute du signal de parole par l'auditeur. Cependant, même après le débruitage, la qualité de la parole n'est pas totalement restituée ; elle est même parfois encore plus dégradée. Les éléments fondamentaux qui influent sur la qualité de la parole après débruitage sont les distorsions du signal et le bruit résiduel communément appelé bruit musical. Les tests de jugement de la qualité par des auditeurs sont les seuls moyens d'évaluation valables et sûrs d'un système de débruitage de la parole. Mais comme pour l'intelligibilité, il existe des critères objectifs d'évaluation de la qualité tels que le PESQ, LLR,WSS ,RSBseg etc. Ces critères ont un caractère perceptuel justement parce qu'ils sont fondés sur des notions psychoacoustiques pour simuler notre perception vis-à-vis du signal de parole. Plus de détails sur ces différents critères feront partie de chapitre 3.

Pour conclure, l'intelligibilité est donc une notion à ne pas confondre avec la qualité de la parole. Une amélioration de la qualité de la parole n'implique pas une amélioration en terme d'intelligibilité. Dans les environnements bruyants, améliorer l'intelligibilité de la parole s'avère une tâche plus difficile qu'améliorer la qualité de la parole [3].

10.Evaluation de la qualité vocale :

10.1.critères subjectifs :

Les tests subjectifs représentent le meilleur moyen d'évaluer la qualité vocale en demandant directement l'opinion des sujets sur les échantillons vocaux testés.

La recommandation de l'UIT intitulée "Méthodes d'évaluation subjective de la qualité de transmission" décrit trois tests subjectifs d'évaluation de la qualité vocale : le test ACR, le test DCR et le test CCR.

10.1.1 Test ACR (Absolute Category Rating) :

Le test subjectif le plus courant et le plus utilisé est le test appelé "Absolute Category Rating" (ACR)

Les sujets expriment leurs jugements sur la qualité du signal vocal à travers une échelle catégorielle à cinq niveaux présentée dans le Tab. III.1. Pour forcer l'utilisation de toute l'échelle, il est demandé aux sujets d'utiliser au maximum les cinq catégories par rapport à l'ensemble de la base sonore. Cette consigne a pour but d'obtenir des résultats présentant le plus de variations significatives possibles entre les conditions, au détriment dans certains cas de la spontanéité et du réalisme des jugements de préférence.

La mesure de la qualité du signal vocal se fait dans l'absolu, sans le signal de référence. L'auditeur se crée sa propre référence interne de qualité vocale qui est de ce fait très variable suivant les sujets interrogés . C'est pourquoi un apprentissage doit être réalisé préalablement, afin de présenter aux sujets les variations maximales des dégradations des sons proposés, afin que les sujets jugent de la qualité vocale par rapport à l'ensemble des stimuli proposés. La base sonore devra contenir des voix de sexes différents, afin de s'affranchir des effets du timbre de la voix. De plus, il est conseillé que chaque échantillon sonore dure entre 6 et 10 secondes et contienne deux ou trois phrases séparées entre elles par un intervalle d'au moins une seconde, pour permettre d'entendre le bruit de fond de la liaison. Le bruit d'ambiance à l'écoute doit être inférieur à 40 dB(A)[4].

Tableau. I.1. *Echelle catégorielle de mesure pour le test ACR (UIT-T P.800)*

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

L'échelle proposée comprend cinq catégories principalement pour simplifier la tâche demandée aux sujets et parce que ces catégories sont estimées suffisantes à l'évaluation de la qualité vocale. La moyenne des résultats de tous les sujets permet d'obtenir la note de qualité représentative de l'ensemble des individus d'un groupe, avec une précision variable suivant le nombre de sujets interrogés et la difficulté du test proposé. Cette précision peut être estimée à travers l'intervalle de confiance entre les différents sujets. La recommandation préconise de faire participer au moins 12 sujets, et si possible 32 sujets, pour obtenir des résultats fiables.

Ce type de test est assez fatigant et peut provoquer une baisse de concentration de la part des sujets. C'est pourquoi la durée du test ne doit pas dépasser 1h30[4].

10.1.2.Test DCR (Degradation Category Rating) :

La méthode d'évaluation par catégories de dégradation (test DCR) permet d'avoir une meilleure précision par rapport au test ACR, lors de l'évaluation de communication présentant de faibles dégradations. Le jugement se fait par comparaison entre le signal débruité et le signal de référence non-dégradé. De ce fait, la tâche demandée aux sujets est plus simple que dans l'absolu (test ACR) et les dégradations fines de la parole sont plus facilement détectables. Par exemple, ce test est utilisé pour faire la différence entre certains codecs.

Il est demandé aux sujets d'évaluer le niveau de dégradation par rapport à un certain attribut perceptif, sur une échelle de catégories à cinq niveaux présentée sur le Tab. I.2. Cette

consigne a l'avantage de pouvoir orienter le test suivant différents attributs comme par exemple le naturel de la voix, le bruit de fond ou encore l'évaluation de la qualité vocale[4].

Tableau. I.2 *Echelle catégorielle de mesure pour le test DCR (UIT-T P.800)*

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

En moyennant les résultats de tous les sujets, nous obtenons la note d'appréciation moyenne de la dégradation représentée par la note DMOS (Degradation Mean Opinion Score).

10.1.3.Test CCR (Comparison Category Rating) :

Le principe du test d'évaluation par catégories de comparaison (CCR (UIT-T P.800)) est semblable au test DCR avec une différence par rapport à l'ordre de présentation des paires des échantillons sonores. Contrairement au test DCR, l'échantillon de référence et l'échantillon dégradé sont présentés dans un ordre aléatoire. L'échelle de mesure est adaptée à cette présentation aléatoire. Elle utilise une échelle catégorielle à sept niveaux, afin d'évaluer la qualité du deuxième son par rapport au premier :

Tableau.I.3 *Echelle catégorielle de mesure pour le test CCR (UIT-T P.800)*

Note	Qualité du 2éme son /au 1er son
3	Bien meilleure
2	Meilleure
1	Légèrement meilleure
0	A peu près équivalente
-1	Un peu moins bonne
-2	Moins bonne
-3	Beaucoup moins bonne

Grâce à cette méthode, les auditeurs formulent deux jugements avec une même réponse :

- "Quel est l'échantillon de meilleure qualité ?"
- "Quelle est la différence de qualité entre les deux échantillons ?"

Les méthodes DCR et CCR sont efficaces pour prédire des différences de qualité vocale entre les stimuli. Elles sont plus précises que lors d'une évaluation dans l'absolue (test ACR). L'avantage de la méthode CCR par rapport à la procédure DCR est qu'elle permet d'évaluer le traitement de la parole qui dégrade ou améliore la qualité de la parole.

En moyennant les résultats de tous les sujets, nous obtenons la note moyenne d'opinion par comparaison, représentée par le score CMOS (Comparison Mean Opinion Score).

10.1.4. Test MUSHRA :

Le test MUSHRA "*MUlti Stimulus test with Hidden Reference and Anchor"* UIT-R BS.1534 est utilisé pour évaluer la qualité de la parole avec une grande précision, en utilisant des signaux d'ancrage (aussi appelés signaux de référence). Ce test comporte deux étapes. La première consiste à identifier le signal d'ancrage (la référence cachée) proposé parmi les signaux dégradés. La deuxième consiste à demander aux sujets de donner leurs jugements de qualité vocale sur une échelle continue représentée sur le Tab. I.4, par rapport au signal d'ancrage. La première étape est évidente pour des signaux comportant de fortes dégradations, mais elle est aussi très utile dans le cas de faibles dégradations afin de vérifier les capacités des sujets à bien détecter les défauts.

Il est possible de tester un maximum de 15 signaux sur une même interface avec au moins un signal de référence caché. Le test MUSHRA se comporte alors comme un test de comparaison par paire entre chaque signal testé[4].

Tab. I.4 *Echelle catégorielle de mesure pour le test MUSHRA (UIT-R BS.1534)[4]*

Note	Qualité de la parole
80-100	Excellente
60-79	Bonne
40-59	Passable
20-39	Médiocore
0-19	Mauvaise

11. Le bruit :

Le bruit est le phénomène sonore le plus représenté dans le monde. Il est à la fois présent dans les sons de l'environnement (naturels, urbains, domestiques, humains), dans les différentes formes de communication sonore (humaine et animale) et dans les sons destinés à la musique (sons vocaux, instrumentaux, synthétiques). Le bruit se propage dans tous les milieux (gazeux, liquide, solide) et sur toute l'étendue des fréquences, des infrasons aux ultrasons en passant par toute la zone de fréquences audibles par l'oreille humaine (20 Hz-20 kHz). Même l'océan – le soi-disant « monde du silence » – recèle encore plus de bruits que la surface de la terre. Seul l'espace est réfractaire aux vibrations acoustiques quelles qu'elles soient.

La permanence du « bruit » dans notre quotidien a donné lieu à de nombreuses métaphores. Dans son usage courant, il est synonyme de nuisance sonore (tapage, tintamarre, vacarme). Mais sa polysémie s'étend à des significations telles que rumeur (bruit qui court), invasion (bruits de botte), retentissement (faire grand bruit), etc. Synonyme d'aléatoire, de parasitage, perturbation ou dégradation de l'information, le bruit devient instrument de mesure en physique, en astronomie, en électronique ou en cardiologie.

L'Association française de normalisation (AFNOR) a homologué deux définitions du bruit. L'une est physique : « Vibration acoustique erratique, intermittente ou statistiquement aléatoire », l'autre se place du point de vue des effets que le bruit produit sur l'homme : « Toute sensation auditive désagréable ou gênante ». Chacune de ces définitions se fonde sur une dimension spécifique du son, respectivement le contenu spectral et l'intensité. La notion de bruit sera approchée complémentairement des points de vue de l'acoustique (caractérisation d'un phénomène physique) et de la psychoacoustique (caractérisation d'une sensation auditive)[5].

12. Définition du bruit en acoustique:

Avant de définir les attributs physiques du bruit, nous allons brièvement rappeler quelques notions d'acoustique. Le son est la conséquence du mouvement des particules d'un milieu élastique (gazeux, liquide ou solide). Dans l'air, un objet en vibration régulière, comme une corde pincée, crée une succession de compressions et de dilatations des particules. Les couches d'air successives sont comprimées et décomprimées à mesure que l'onde s'éloigne de la source perturbatrice. La perturbation se propage de proche en proche sous forme d'une onde de pression : une onde sonore. Lorsqu'un mouvement se reproduit identique à lui-même, on dit qu'il est périodique. On représente ce mouvement en fonction du temps par une courbe sinusoïdale. Une oscillation complète (un cycle) est définie par sa période (la durée en seconde d'un cycle), son amplitude (l'élongation maximale du cycle à partir de sa position de repos, ce qui correspond à l'intensité du son), sa fréquence (le nombre de cycles par seconde, ce qui correspond à la hauteur du son) et sa phase. Un mouvement périodique simple, représenté par une courbe sinusoïdale, est qualifié de son « pur ».

Cependant, les sons sinusoïdaux n'existent pas dans la nature, on les produit avec des générateurs de sons ou des logiciels de synthèse. Les sons de l'environnement (naturel, domestique ou urbain), les sons produits par le corps humain, les sons instrumentaux, sont toujours des mouvements périodiques complexes. Selon le théorème de Fourier, toute onde complexe peut se décomposer en une somme de mouvements périodiques simples. Lorsque les fréquences des composantes sont des multiples entiers du son fondamental, le son complexe est périodique. La forme de l'onde complexe se répète périodiquement. À l'inverse, lorsque les fréquences des sinusoïdes ne sont pas des multiples entiers du son fondamental, le son complexe est qualifié de non périodique. Les voyelles du langage parlé ou la plupart des instruments de musique produisent des sons complexes périodiques. Les bruits appartiennent en revanche à la catégorie des sons complexes non périodiques. D'un point de vue acoustique, on définit donc le bruit comme un son complexe non périodique.

Un son complexe est composé d'une somme de mouvements simples (sinusoïdes) dont la décomposition en composantes est appelée spectre. Lorsque ces mouvements simples sont des multiples entiers du fondamental, on les nomme des harmoniques, dans le cas inverse, on les nomme des partiels. Tous les harmoniques d'une fréquence fondamentale ne possèdent pas la même énergie (amplitude). C'est ce qui produit les différences de timbre entre les sons. Un son complexe peut être composé d'un très grand nombre d'harmoniques, mais l'oreille humaine étant limitée à environ 20 000 Hz, seules les premières composantes du spectre ont un rôle dans la détermination du timbre. On peut produire une infinité de formes d'onde en modulant le nombre et l'amplitude des harmoniques. Il existe des formes d'onde « typiques » comme l'onde carrée, l'onde triangulaire et l'onde en dents-de-scie (Figures I.4 et I.5) qui sont à la base de nombreux sons instrumentaux et que l'on retrouve dans les synthétiseurs [5].

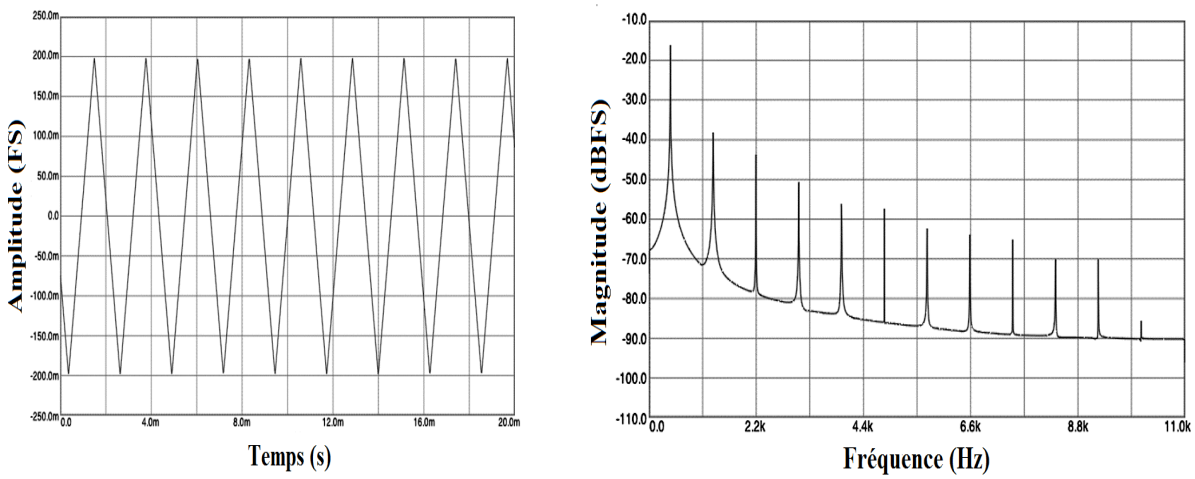


Figure I.4. *Onde triangulaire 440 Hz (forme d'onde à gauche, FFT à droite).*

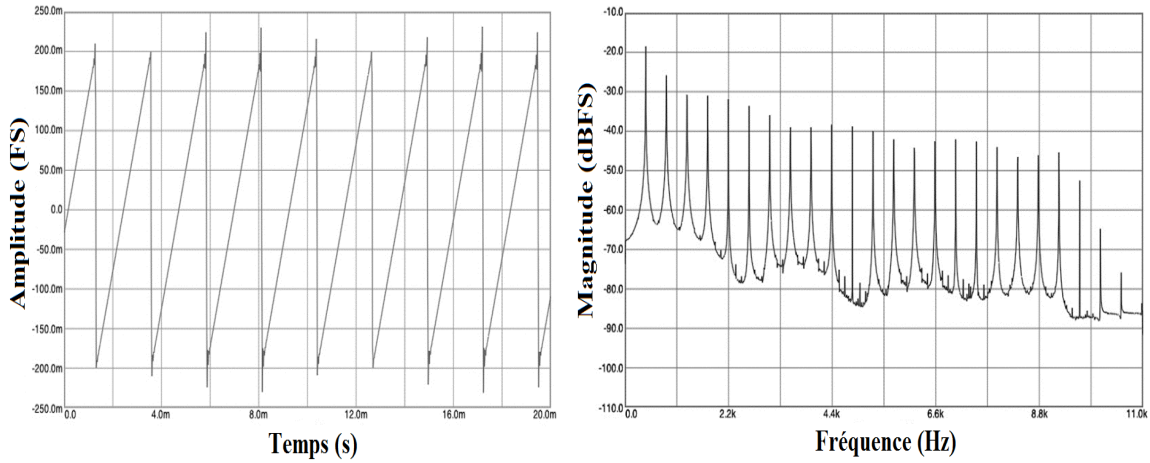


Figure I.5. *Onde en dents-de-scie 440 Hz (forme d'onde à gauche, FFT à droite).*

13. Les différents types de bruit :

Les différents bruits pouvant influer sur un message peuvent être divisés en deux grandes catégories : les bruits additifs et les bruits convolutionnels. La distinction entre les deux peut être faite par le nombre d'agents agresseurs extérieurs à la transmission du message. Les bruits additifs sont causés par des agents extérieurs au trinôme source-voie-destinataire alors que les bruits convolutionnels sont causés par la moindre qualité de la voie de communication, celle-ci ayant alors un rôle ambigu, du point de vue du message, de médium et d'agresseur, l'effet est plus dévastateur si le bruit n'est pas stationnaire .

13.1. Les bruits additifs :

Les bruits additifs sont dus à la multiplicité des systèmes de communication dans un même environnement. Plusieurs émetteurs et plusieurs receveurs pouvant être confinés dans u même espace, les messages de tous les émetteurs peuvent donc se trouver en concurrence sur une même voie sans que les récepteurs possèdent un mécanisme infaillible pour isoler le message qui leur est destiné. L'émetteur et le récepteur peuvent aussi se trouver en présence d'un ou de plusieurs équipements générant un bruit de fond de force variable.

13.2 Les bruits convolutionnels :

Les bruits convolutionnels (ou multiplicatifs) sont dus à la distorsion induite par la voie de communication. Ils résultent de la mauvaise qualité d'un ou de plusieurs éléments de support du message ou, tout simplement, de son étroitesse en bande passante. Les sociétés modernes utilisent de plus en plus de moyens de communication à longue distance tels que le téléphone.

Ces moyens de communication à longue distance ont été élaborés à partir d'un compromis coût/efficacité. La parole, lorsqu'elle est transmise par un tel moyen, est forcément dégradée tout en gardant une grande intelligibilité. Un des champs possibles

d'application de la Reconnaissance Automatique de la Parole (RAP) sont les serveurs vocaux accessibles par les lignes téléphoniques. Mais la parole transmise par téléphone souffre de déformations variables induites par la qualité de la connexion. Une transmission peut ainsi souffrir de l'étroitesse de la bande passante, de la mauvaise qualité des microphones de certains terminaux téléphoniques, de bruits additifs stationnaires et de porteuses basse fréquence.

La qualité de la transmission varie cependant très peu au cours d'une même communication. De manière plus générale, le bruit convolutionnel est présent dans toute application de RAP par l'intermédiaire du microphone utilisé pour la saisie de la voix. Un système de RAP mis au point avec certain microphone pourra voir ses capacités diminuer de manière conséquente lorsqu'un autre microphone sera employé. La parole enregistrée dans tous les corpus utilisés pour la recherche est en effet toujours bruitée puisque le microphone utilisé effectue toujours un filtrage linéaire.

Enfin, certains milieux d'enregistrement sont de mauvaise qualité et peuvent provoquer des phénomènes de réverbération. C'est notamment le cas des pièces possédant de grandes surfaces faites d'un matériau dur ou lorsque le microphone utilisé pour l'enregistrement est placé assez loin du locuteur.

13.3 Les bruits physiologiques :

D'autres bruits peuvent également être considérés dans le domaine de la RAP mais ils n'ont pas la généralité des bruits de type additif ou convolutionnel car ils sont spécifiques à l'être humain lors de sa phase de production de parole. Les plupart des systèmes de RAP fonctionnent mal en milieu bruité car les contraintes posées par de tels environnements n'ont pas été prises en compte dès le départ.

L'homme essaie, lui, de s'adapter aux conditions sonores rencontrées en modifiant sa méthode de production de parole. Un des phénomènes les plus remarquables de modification de production de la parole par l'homme est l'effet Lombard. Lorsqu'un locuteur est placé dans un environnement bruité, il modifie sa voix, et son effort vocal, en "haussant le ton" de manière à ce que la parole produite conserve un bon RSB par rapport à l'environnement.

Cette accentuation de la voix pose cependant un problème majeur aux systèmes de RAP car les spectres de tous les phonèmes peuvent être modifiés ce qui a pour effet de nettement amoindrir les taux de reconnaissance. Certaines études montrent que l'homme arrive à avoir de meilleures capacités de compréhension dans le cas de la parole Lombard que pour de la parole normale lorsqu'il lui est demandé de reconnaître des mots isolés ou de la parole continue masqués par du bruit[1].

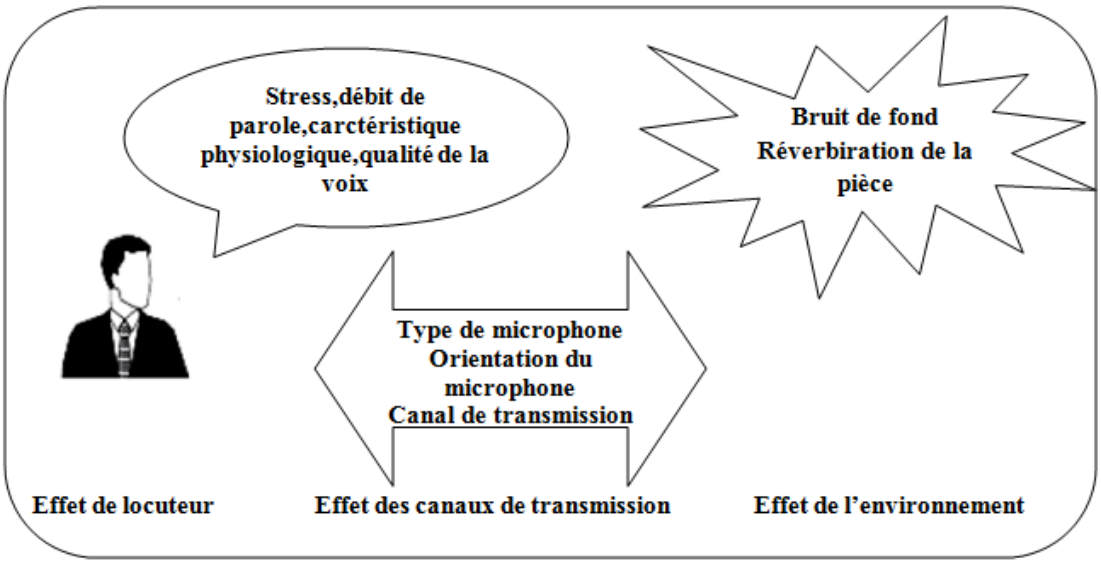


Figure I.6. *Représentation schématique de différentes sources de variabilité qui peuvent dégrader les performances d'un système de reconnaissance de la parole[1]*

14. bruit blanc et bruit coloré :

14.1. définition techniques :

L'idée d'associer chaque type de son à une couleur provient d'une grossière analogie entre les ondes sonores et lumineuses. On peut représenter les spectres de fréquence des ondes sonores sous forme de diagramme , de même, pour les spectres des ondes lumineuses. De ce fait, si le dessin du diagramme de l'onde sonore du « bruit bleu » était transposé dans un diagramme d'ondes lumineuses, cela correspondrait à une lumière bleue, et ainsi de suite.

14.2. Le bruit blanc :

De même que la lumière blanche, le bruit blanc est un signal (ou processus) avec une énergie équivalente par cycle (en hertz). Cela se traduit par un spectre « plat » lorsqu'on en trace le diagramme. Autrement dit, le signal a une puissance constante sur n'importe quelle bande du spectre. Par exemple, l'intervalle de fréquence allant de 40 à 60 Hz contient la même puissance que l'intervalle allant de 4000 à 4020 Hz.

Un signal de bruit blanc qui serait constant sur une bande passante infinie est purement théorique. En cumulant la puissance de chaque fréquence, la puissance totale d'un tel signal serait infinie.

En pratique, un signal est dit « blanc » s'il a un spectre uniforme (plat) dans une bande de fréquence définie (comme l'intervalle de fréquence perçu par l'oreille humaine, ou l'intervalle d'émission d'équipements audio/vidéo).

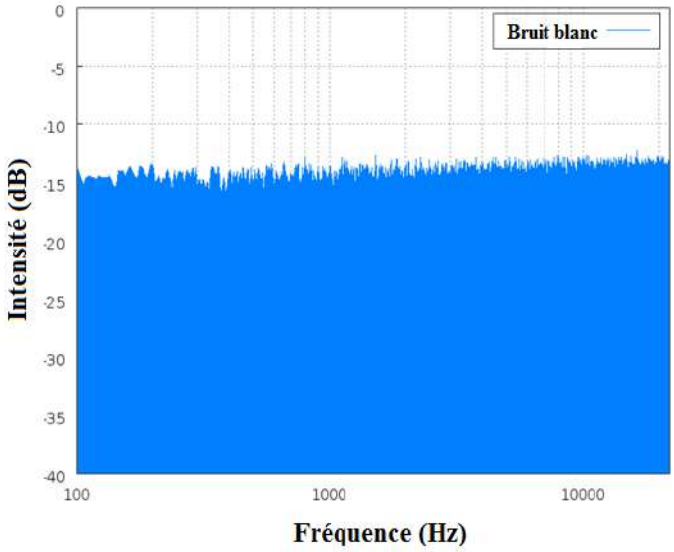


Figure I.7. *Spectre d'un bruit blanc*

14.3. Le bruit coloré :

La lumière blanche est composée de radiations de toutes longueur d'onde entre 400 et 800 nm

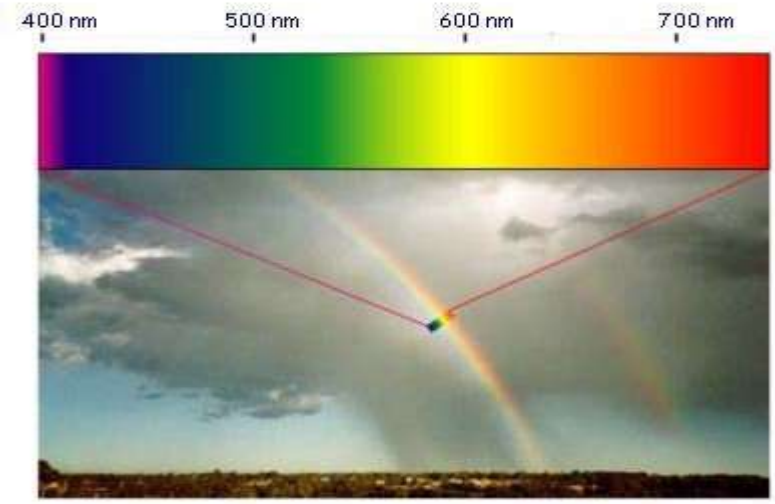


Figure I.8 *Les radiations qui compose la lumière blanche*

Lorsque la lumière contient davantage de radiations d'une certaine longueur d'onde, elle est colorée ,devient rouge si elle contient de grandes longueurs d'onde (fréquences basses) ou bleue si elle contient des radiations de longueur d'onde basse (fréquences élevées).

De la même façon et par analogie avec la lumière, lorsqu'un bruit traverse un filtre, certaines composantes sont atténuées. Le spectre n'est donc plus constant et on dit que le bruit est coloré.

- si les basses fréquences prédominent, on a un bruit rose
- si les fréquences élevées prédominent, on a un bruit bleu[6]

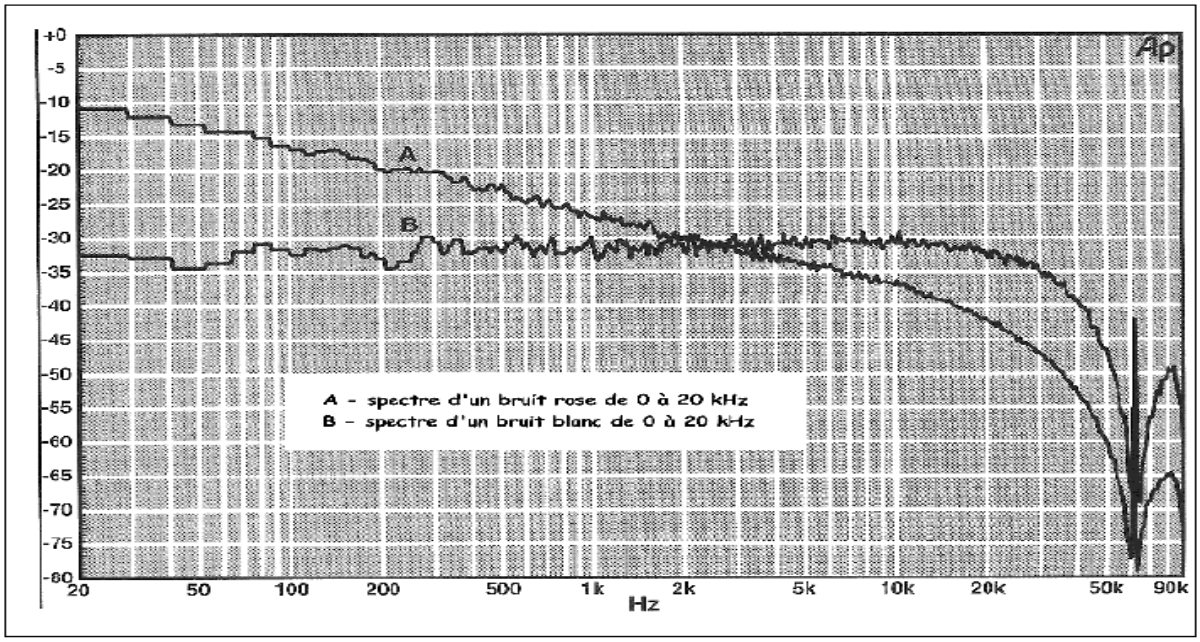


Figure I.9.*Spectre d'un bruit blanc et d'un bruit rose dans la bande audio[6]*

15. Qu'est ce qu'un bruit musical :

Le bruit musical est un bruit résiduel perceptuellement gênant qui apparait suite au débruitage de la parole par des algorithmes d'atténuation spectrale à court terme tels que la soustraction spectrale ou le filtrage de Wiener. Le spectre du bruit musical est particulièrment tonal, d'où le caractère musical. Son énergie moyenne est plus faible que celle du bruit initial, mais sa dispersion en fréquence est plus grande, ce qui le rend plus gênant que le bruit de départ du point de vue perception. Parmi les raisons d'apparition de ce type de bruit , on cite :

- Le traitement non linéaire des composantes négatives du signal débruité.
- L'estimé non précise de la densité spéctrale de bruit.
- La variabilité de la fonction de gain appliqué au signal bruité
- La variance des estimateurs locaux de la densité spéctrale des signaux.[3]

16. Les caractéristiques du bruit:

On appelle bruit tout signal nuisible qui se superpose au signal utile en un point quelconque d'une chaine de mesure ou d'un système de transmission. Il constitue donc une gêne dans la compréhension du signal utile, qui est dans notre cas, la parole. En physique, en acoustique et en traitement du signal, bien que le bruit soit, par nature, aléatoire, il possède certaines caractéristiques statistiques, spectrales ou spatiales. Le tableau suivant, représente des exemples de différentes classes auxquelles un bruit peut appartenir :

Tableau I.5. *Différentes classes de bruit [3]*

Propriété	Type
Structure	Continu/Impulsif/Périodique
Type d'interaction	Additif/Multiplicatif/Convolutif
Comportement temporel	Stationnaire/Non-stationnaire
Bande de fréquence	Etroit/large
Dépendance	Corrélé / Décorrélé
Propriété statistique	Dépendant/Indépendant
Proptirété spatiale	Cohérent/Incohérent

Comme notre but est essentiellement le débruitage de la parole pour des applications simples, on se limite dans notre étude aux bruits additifs, stationnaires, et décorrélés avec la parole (indépendance au sens statistique).

17. Problématique de la réduction de bruit :

L'essor fantastique des télécommunications de ces 20 dernières années a permis au grand public de bénéficier d'outils de communication mobiles. Il est désormais devenu possible et courant de téléphoner de partout (ou presque) dans des environnements aussi divers et variés que la rue, une gare ou bien encore une voiture. Cependant, tous ces lieux ne bénéficient pas du calme du salon où le téléphone fixe était autrefois cantonné. La gêne occasionnée par la perturbation qualifiée de bruit est généralement source d'inconfort et de fatigue pour les correspondants quand ce n'est pas l'intelligibilité même du message qui est remise en cause. De plus, la volonté de dématérialiser la prise de son (système mains-libres) va favoriser l'émergence du bruit du fait de l'augmentation de la distance entre la bouche et le microphone. Dans ces conditions, la conversation téléphonique peut s'avérer rapidement fastidieuse ce qui justifie le besoin d'un traitement à même de réduire la gêne des utilisateurs.

17.1. La gêne due au bruit :

La gêne due au bruit est de nature différente pour la personne plongée dans l'ambiance bruitée (le locuteur) et pour celle qui subit les perturbations par l'intermédiaire de son téléphone (l'auditeur) . L'auditeur est le plus gêné car, contrairement au locuteur, il n'a aucun contrôle sur le milieu acoustique. Le locuteur garde un certain contrôle sur le bruit ambiant, il a la possibilité de hausser la voix (effet Lombard) ou de se réfugier dans un endroit moins bruyant. Dans le cas d'une communication mains-libres, il peut aussi focaliser son oreille sur le signal utile grâce aux capacités de localisation spatiale de l'oreille alors que l'auditeur subit complètement la perturbation sonore. Ce dernier est le plus pénalisé dans la mesure où la prise de son par microphone et sa restitution sont ponctuelles.

L'ensemble du champ sonore (des sons provenant de toutes les directions) est donc intégré et restitué en une somme des perturbations. L'information de spatialisation ayant disparu, l'auditeur ne peut donc pas séparer l'information utile des différentes sources de bruit bien que l'oreille humaine en soit normalement capable. De plus, la superposition du bruit et de la parole réduit l'intelligibilité du message ce qui demande un effort constant de la part de l'auditeur et le fatigue rapidement.

Il faut souligner aussi que, dans tout réseau de communication, il y a une étape de codage du signal de parole qui est largement basée sur ses propriétés. En présence de bruit le signal à coder s'éloigne du modèle utilisé pour l'opération de codage-décodage et il en résulte une dégradation d'autant plus gênante que le niveau de bruit est élevé. Le timbre de la voix est altéré et le bruit est codé de façon peu naturelle ce qui le rend encore plus désagréable .Il y a donc tout intérêt à réduire les perturbations liées au bruit avant la phase de codage afin de faciliter sa tâche.

17.2. Les applications :

Afin d'améliorer la qualité du signal transmis au correspondant distant, d'accroître son intelligibilité et de réduire la fatigue de ce dernier, il s'avère important de développer des systèmes de réduction de bruit dont le but consiste à extraire l'information utile en effectuant un traitement sur le signal d'observation bruité. En plus de ces applications de communication parlée, l'amélioration de la qualité du signal de parole s'avère également nécessaire pour la reconnaissance vocale, dont les performances sont fortement altérées lorsque l'utilisateur est plongé dans un environnement bruyant.

La réduction de bruit s'applique principalement au domaine du traitement du signal sonore (parole ou musique) dont voici une liste non exhaustive :

– téléconférence et visioconférence en milieu bruité (en salle dédiée ou bien à partir d'un ordinateur multimédia, *etc.*),

– téléphonie : traitement au niveau du terminal (terminaux classiques et terminaux mobiles, *etc.*) et au sein du réseau de transport,

– terminaux mains-libres (bureau, terminaux mobiles, terminaux fixes embarqués en véhicules, *etc.*),

– prise de son dans les lieux publics (gare, aéroport, rue, *etc.*),

– prise de son mains-libres dans les véhicules,

– reconnaissance de parole robuste à l'environnement acoustique,

– restauration d'enregistrements anciens,

– prise de son pour le cinéma et les médias (radio, télévision, par exemple pour le journalisme sportif ou les concerts, *etc.*).

Le principe de la réduction de bruit est également applicable à tous les domaines où l'on cherche à extraire une information utile à partir d'une observation bruitée [7].

18. Conclusion :

Du fait que notre environnement est bruyant, la parole et le bruit sont donc deux signaux qui coïncident souvent. Dans ce chapitre, les propriétés et caractéristiques de chacun d'eux ont été rappelées afin de mieux les traiter dans le cadre de ce mémoire. On a aussi défini et décrit le comportement du bruit musical. En parlant de perception de la parole, on a présenté quelques définitions de psychoacoustique indispensables pour mieux comprendre le débruitage . A cet effet, plus de détails seront donnés ultérieurement pour décrire les méthodes de débruitage dont l'objectif est de réduire le bruit coloré.

1. Introduction :

L'oreille humaine a des capacités impressionnantes pour reconnaître et distinguer la parole du bruit. Mais, pour le bien être de l'auditeur et dans le souci de limiter sa fatigue, on cherche à améliorer la qualité de l'écoute à travers le débruitage de la parole (pour des applications telles que la téléphonie mobile et la téléphonie mains-libres). On peut aussi utiliser des systèmes de débruitage pour des applications où la cible n'est pas humaine telles que la reconnaissance et la synthèse vocale, le codage de la parole, etc. Les systèmes dédiés à ces dernières applications sont très a□ectés par la qualité de la parole et leurs performances peuvent être médiocres en présence du bruit notamment lorsqu'il y a eu apprentissage en milieu non bruité. Plusieurs méthodes de débruitage de la parole ont été proposées dans la littérature pour répondre aux besoins des utilisateurs et des applications dédiées à la parole.

On fera une étude succincte de certaines d'entre elles. Tout au long de ce chapitre, on considère le problème de débruitage schématisé par la figure II.1, les notations en découlant auront ainsi la même signification dans toute la suite.

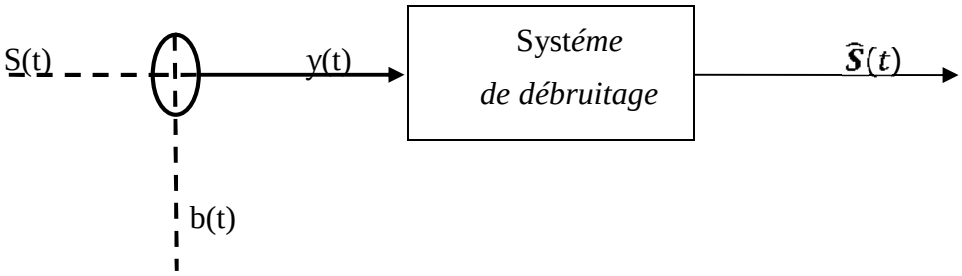


Figure II.1. *Modèle de débruitage utilisé dans ce document*

Au départ, on peut considérer que le signal traité est composé d'un signal utile qui est la parole et d'un bruit qui est ce que l'on doit atténuer. Le but d'un tel algorithme dans l'aide auditive ou bien dans l'implant cochléaire est d'améliorer le rapport signal sur bruit en sachant que le signal est considéré comme étant de la parole et que le reste est un bruit. Ces algorithmes sont implantés dans l'aide auditive et améliorent la parole.

On considère le signal de parole est un mixage entre un signal utile et un signal bruité :

$$\boldsymbol{y(t) = s(t) + b(t)} \qquad \textbf{(II.1)}$$

Où:

– t : représente le temps.

– y : représente le signal capté par le microphone .

– s : représente le signal utile.

– b : représente le bruit .

Une « simple » soustraction du bruit b(n) au signal y(n) pourrait alors être faite pour enlever le bruit dans y(n). Cependant, le bruit n'est jamais connu par avance. Il faut poser des hypothèses. Le bruit b(n) doit être estimé à partir de l'échantillon en cours et des échantillons précédents. Un estimateur de bruit est alors appliqué pour estimer le bruit. Le signal que l'on connaît est y(n), mais il peut aussi avoir des hypothèses sur la parole, par exemple la connaissance du spectre à long terme de la parole, n'est pas le même que celui du bruit. Il est actuellement très di□cile de travailler dans le domaine temporel par échantillon car le filtrage doit être un produit de convolution[2].

2.Soustraction spectrale :

2.1.Principe :

La soustraction spectrale est la méthode de débéruitage la plus ancienne. Elle est introduite par Boll 1979[15]. Elle opère dans le domaine fréquentiel et a pour principe de soustraire une estimé du bruit à partir du signal observé .Le bruit est supposé additif, stationnaire ou légèrement variant ce qui nous permet de l'estimer pendant les périodes de silence[3].

Donc le signal bruité est composé d'un signal de parole s(t) (le signal utile) et d'une composante de bruit b(t),que l'on souhaite éliminer. Le signal observé s'écrit alors :

$$y(t) = s(t) + b(t) \quad \textbf{(II.1)}$$

Il s'agit de passer dans le domaine temps-fréquence en utilisant la Transformé de Fourier Discrète (TFD),d'appliquer un traitement au spectre du signal observé, qui vise a réduit le bruit , puis de repasser dans le domaine temporel par la Transformé de Fourier Discrète inverse (TFDi).[9]

En écrivant le signal observé dans le domaine fréquentiel ,on obtient :

$$Y(f) = S(f) + B(f) \quad \textbf{(II.2)}$$

Il existe deux versions de base de la soustraction spectrale se différenciant l'une de l'autre par l'utilisation soit de la puissance soit de l'amplitude. Si

$$\left|\hat{S}(f)\right| = |Y(f)| - \left|\hat{B}(f)\right| \quad \textbf{(II.3)}$$

Il s'agit de la soustraction spectrale d'amplitude (SSA). Si, par contre, le bruit estimé est donné par son spectre de puissance, on aura la soustraction spectrale de puissance (SSP) :

$$\left|\hat{S}(f)\right|^2 = |Y(f)|^2 - \left|\hat{B}(f)\right|^2 \quad \textbf{(II.4)}$$

Vu que le second terme de l'équation (II.4) peut être négatif, on peut le rendre positif en changeant de signe ou bien en l'annulant comme dans l'équation (II.5). Ceci fait partie des premières améliorations apportées à la soustraction spectrale :

$$|\hat{S}(f)|^2 = \begin{cases} |Y(f)|^2 - |\hat{B}(f)|^2 & si\ |Y(f)|^2 > |\hat{B}(f)|^2 \\ 0 & sinon \end{cases} \quad \textbf{(II.5)}$$

Le passage dans le domaine temporel est réalisé par la transformée de Fourrier inverse en gardant la phase du signal bruité. On se permet de procéder ainsi, d'une part, parce que notre oreille est peu sensible aux variations de la phase et, d'autre part, parce qu'une estimation de la phase est une tâche très compliquée.

$$\hat{S}(t) = IFFT\left[\, |\hat{S}(f)|.e^{i \times arg\, Y(f)} \,\right] \quad \textbf{(II.6)}$$

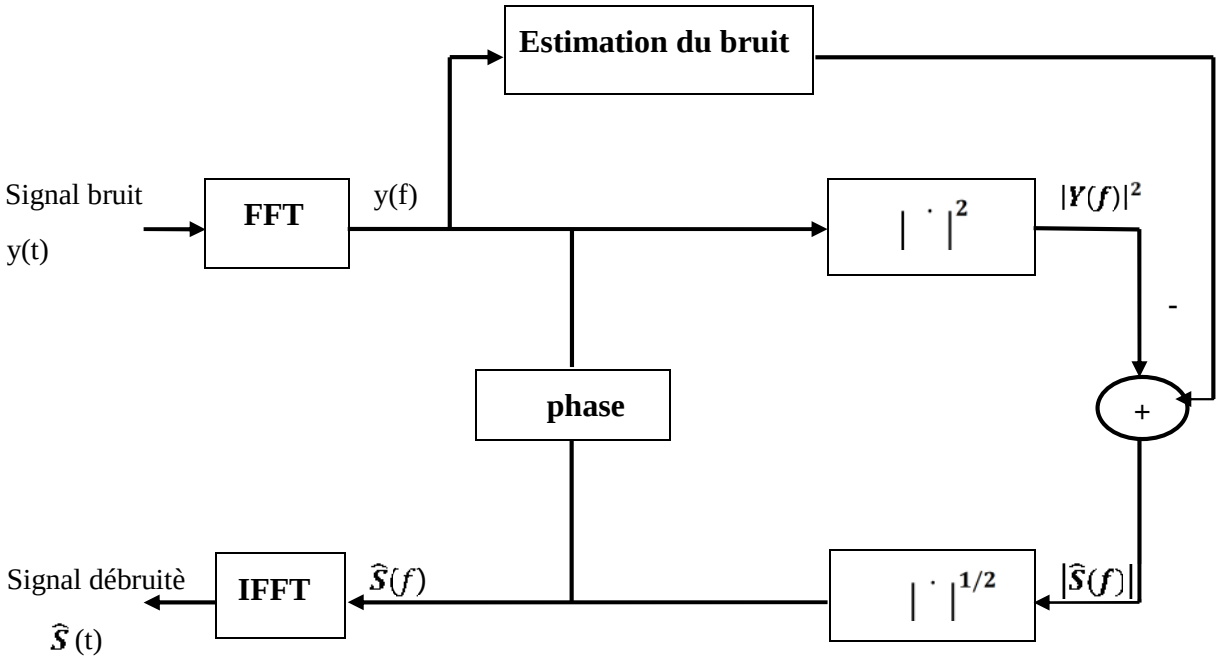


Figure II.2. *Schéma-bloc d'un réducteur de bruit par la soustraction spectrale*

2.2.Soustraction spectrale par filtrage :

En se basant sur un filtre, et en gardant l'estimation du bruit on peut assimiler la soustraction spectrale à un filtrage.

$$|\hat{S}(f)| = G(f).|Y(f)|\,, \qquad 0 \le G(f) \le 1 \quad \textbf{(II.7)}$$

Et pour la soustraction spectrale de puissance, on obtient :

$$G(f) = \begin{cases} \sqrt{1 - \frac{|\hat{B}(f)|^2}{|Y(f)|^2}} & si\ |Y(f)|^2 > |\hat{B}(f)|^2 \\ 0 & sinon \end{cases} \quad \textbf{(II.8)}$$

La soustraction spectrale est largement étudiée dans la littérature vu sa simplicité, mais elle reste limitée par les artefacts qu'elle génère en sortie, à savoir la distorsion du signal et le bruit musical. Les tentatives de réduction de celui-ci ont amené les chercheurs à s'investir sur l'expression adéquate qui peut être le compromis entre la quantité du bruit musical et la distorsion du signal. Ainsi, pour donner plus de souplesse à la soustraction spectrale, plusieurs améliorations ont été apportées à cette technique. L'objectif de cette section est de décrire les plus connues[3].

2.3Amélioration de la soustraction spectrale par Berouti :

Berouti [16], a trouvé qu'après une soustraction spectrale, le bruit résiduel contient deux types de pics spectraux. Ce sont des pics larges perçus comme étant un bruit large bande et des pics étroit comme étant des tonales. Il qualifie ce dernier de bruit musical. Il propose alors dans la soustraction spectrale de rajouter à la surestimation du bruit une quantité $\beta|\widehat{B}(f)|$ au lieu de 0 afin d'éviter que le seuil de tolérance du bruit dépasse la puissance du bruit. On obtient alors l'équation suivante :

$$|\widehat{S}(f)|^2 = \begin{cases} |Y(f)|^2 - \alpha|\widehat{B}(f)|^2 & si\ \left(|Y(f)|^2 - \alpha|\widehat{B}(f)|^2\right) > \beta|\widehat{B}(f)|^2 \\ \beta|\widehat{B}(f)|^2 & sinon \end{cases} \quad \textbf{(II.9)}$$

Ces travaux ont montré que le paramètre α dépend du RSB de la trame correspondante.

$$\alpha = \alpha_0 - \frac{RSB}{s} \quad \textbf{(II.10)}$$

Où:

S=20/3

α_0 = 4 [3]

Ces valeurs sont données pour un RSB compris entre ±6dB, α_0 est compris entre 3 et 6. La variable β est très sensible au bruit. Pour une fort niveau de bruit, RSB = −6dB, β doit être compris entre $0.02 \leq \beta \leq 0.06$. Et pour un bruit faible, $0.005 \leq \beta \leq 0.02$.

2.4.Amélioration de la soustraction spectrale par Boll :

Boll a décomposé la soustraction spectrale en quatre parties. La moyenne d'amplitude, la correction de l'estimateur, la réduction du bruit résiduel et l'atténuation du signal pendant les périodes de silences.

2.4.1.La moyenne d'amplitude :

$$\hat{\mathbf{S}}(\mathbf{f}) = [|\mathbf{Y}(\mathbf{f})| - \mu_0]\mathbf{e}^{\mathbf{i}\times\arg \mathbf{Y}(\mathbf{f})} \quad \textbf{(II.11)}$$

Où $\mu(f)$ = E[$|B(f)|$] est la moyenne du bruit calculé pendant les silences. Le filtre équivalent est de la forme :

$$\mathbf{H}(\mathbf{f}) = \mathbf{1} - \frac{\mu(\mathbf{f})}{\mathbf{Y}(\mathbf{f})} \quad \textbf{(II.12)}$$

L'expression de l'erreur devient :

$$\varepsilon(\mathbf{f}) = \mathbf{S}(\mathbf{f}) - \hat{\mathbf{S}}(\mathbf{f}) = \mathbf{B}(\mathbf{f}) - \mu_0\, \mathbf{e}^{\mathbf{i}\times\arg \mathbf{Y}(\mathbf{f})} \quad \textbf{(II.13)}$$

Où ε (f) dépend à la fois de B(f) et de la moyenne de $\mu(f)$. Il faut réduire au minimum $\varepsilon(f)$ pour avoir la meilleure estimation du bruit.
Pour cela, il faut que : B(f) □ $\mu(f)$. L'introduction de la moyenne de l'amplitude dans le signal bruité $\left|\overline{Y(f)}\right| = \frac{1}{N}\sum_{f=0}^{N-1}|Y(f)|$ permet d'obtenir une nouvelle équation:

$$\hat{\mathbf{S}}_{\mathbf{m}}(\mathbf{f}) = \left|\, \left|\overline{\mathbf{Y}(\mathbf{v})}\right| - \mu_0 \mathbf{e}^{\mathbf{i}\times\arg \mathbf{Y}(\mathbf{f})} \right| \quad \textbf{(II.14)}$$

L'erreur devient alors :

$$\varepsilon(\mathbf{f}) = \mathbf{S}(\mathbf{f}) - \hat{\mathbf{S}}_{\mathbf{m}}(\mathbf{f}) \square \mu_0 - \left|\overline{\mathbf{B}(\mathbf{f})}\right| \quad \textbf{(II.15)}$$

O`u, $\left|\overline{B(f)}\right| = \frac{1}{N}\sum_{f=0}^{N-1}|B(f)|$ De cette façon , si l'on moyenne sur une grande période, l'erreur se réduit. Mais dans ce cas, le bruit doit être considéré comme stationnaire. La parole n'étant pas stationnaire, il y a une limite à cette méthode.[3]

2.4.2.Correction de l'estimateur :

La méthode consiste à remplacer les valeurs négative dans $\hat{S}(v)$ par des zéros. C'est une méthode de rectification demi-onde. La nouvelle expression de H(f) devient :

$$\mathbf{H}^{\mathbf{R}}(\mathbf{f}) = \frac{\mathbf{H}(\mathbf{f})+|\mathbf{H}(\mathbf{f})|}{\mathbf{2}} \quad \textbf{(II.16)}$$

Et l'équation devient $\hat{S}(f) = H^R(f)Y(f)$ [2]

2.4.3.Réduction du bruit résiduel :

Cette étape est résumée par l'équation suivante :

$$|\hat{S}_k(f)| = \begin{cases} |\hat{S}_k(f)| & \text{si } \hat{S}_k(f) \geq \max |B^R(f)| \\ \min |\hat{S}_j(f)|, j = k-1, k, k+1 & \text{si} \hat{S}_k(f) < max |B^R(f)| \end{cases} \quad \textbf{(II.17)}$$

Avec k l'indice de la trame, $\hat{S}k(v)=H_k^R(v).Y_k(v)$ oùH^R est donné par l'équation (II.16), max$|B^R(f)|$ est le maximum du bruit résiduel ($B^R(f) = \overline{|B(f)|} - \mu(f)$) mesuré pendant les instants de silence et l'indice j désigne le numéro de la trame. La réduction du bruit résiduel s'effectue ainsi en sélectionnant le minimum de l'amplitude estimée durant 3 trames adjacentes si l'amplitude dans la trame courante j est inférieure au niveau du maximum du bruit résiduel $B^R(\nu)$[3].

2.4.4.Atténuation additionnelle durant les silences :

De manière empirique, Boll propose un seuil de détection d'activité vocale dont la formule est donnée ci-dessous :

$$\zeta = 20\log_{10} \left[\frac{1}{2\pi} \int_{f_1}^{f_2} \frac{\hat{S}(f)}{\mu(f)} \, d(f)\right| \quad \textbf{(II.18)}$$

Si τ < −12dB, il n'y a pas d'activité vocale. A ce moment-là, au lieu de supprimer complètement le bruit, il propose de l'atténuer. Il justifie cette alternative par l'effet désirable de ce bruit comme moyen de rendre moins perceptible le bruit non masqué pendant l'activité vocale. La formule d'atténuation est donnée par :

$$|\hat{S}(f)| = \begin{cases} |\hat{S}(f)| & \text{si } \zeta \geq -12dB \\ cB(f) & \text{si } \zeta < -12dB \end{cases} \quad \textbf{(II.19)}$$

Où 20 log_{10}c = −30 dB [3]

2.5.Amélioration de la soustraction spectrale par Virag :

Virag [17]a mixé les deux précédentes méthodes : Boll et Berouti avec la généralisation de Lin pour obtenir une généralisation de la soustraction spectrale. L'intérêt revient à trouver un compromis entre la réduction du bruit et la distorsion du signal. Le gain G(f) est donné par :

$$G(f) = \begin{cases} \left(1 - \alpha \left(\frac{|B(f)|}{|Y(f)|}\right)^{n_1}\right)^{n_2} & \text{si } \left(\frac{|B(f)|}{|Y(f)|}\right)^{n_1} < \frac{1}{\alpha+\beta} \\ \left(\beta \left(\frac{B(f)}{Y(f)}\right)^{n_1}\right)^{n_2} & \text{sinon} \end{cases} \quad \textbf{(II.20)}$$

Où:

– α est le facteur de sur-soustraction (α > 1).

– β avec 0 ≃ β << 1, est un facteur qui permet d'introduire un léger bruit de fond.

On peut remarquer que le choix de α et β est beaucoup plus critique que n 1,2.

Si $n_1 = n_2 = 1$, il s'agit d'une soustraction spectrale d'amplitude.

Si $n_1 = n_2 = 0.5$, il s'agit d'une soustraction spectrale de puissance et

Si $n_1 = 2$, $n_2 = 1$, il s'agit d'un filtrage de Wiener.

2.6. Amélioration psychoacoustique :

La soustraction spectrale est très simple à mettre en œuvre. Cependant, elle reste une méthode limitée par l'apparition du bruit musical. Des améliorations ont été étudiées sur les aspects psychoacoustiques . Le but est de rendre le bruit musical moins gênant en exploitant un phénomène de masquage. Ce dernier est obtenu par le calcul d'un seuil et que tout bruit au-dessous de ce seuil est imperceptible[2].

2.7. Soustraction spectrale multi-bande :

Dans la méthode de soustraction spectrale multi-bande proposée par Kamath et Loizou, le spectre de la parole est divisé en N bandes et la soustraction spectrale est exécuté indépendamment dans chaque bande. Dans cette méthode, premièrement, le signal fenêtré et le spectre est estimé en utilisant la FFT. Le bruit et la parole sont alors divisés en bande de fréquence différente . À la prochaine étape, les différentes bandes de fréquence est traitées en soustrayant le spectre correspondant de bruit du spectre du signal de parole bruité et à la fin , les bandes de fréquence modifiées sont recombinées et le signal temporelle est obtenu à l'aide de la phase du signal bruité et la Transformé de Fourier Inverse IFFT[10]. On va voir plus de détaille sur cette méthode dans le 3éme chapitre ,où nous la comparons avec la méthode de soustraction spectrale classique.

3.Filtrage de wienner :

Le filtrage de Wiener a été introduit à la fin des années 60 pour essayer d'améliorer la qualité de la trace recueillie dans les potentiels évoqués[18].

Le problème du type de filtrage proposé est qu'il n'est pas applicable sur une moyenne d'acquisition mais pour chaque trace. Doyle , propose une modification pour pouvoir l'adapter à la moyenne (le calcul dans ce cas est beaucoup plus rapide), il n'y a pas besoin de filtrer chaque trace. Cependant, il faut considérer que le bruit est stationnaire dans ce cas .

Tout comme la soustraction spectrale, les calculs sont effectués dans le domaine fréquentiel. La densité spectrale de puissance (DSP) est calculée comme précédemment dans l'équation II.4. Le système peut être isolé entre le bruit et le signal utile. Ce qui revient à dire que le rapport $\frac{|Y(f)|^2}{|\hat{B}(f)|^2}$doit être maximisé pour obtenir le signal utile. Le filtre de Wiener est défini de la façon suivante :

3.1.Principe :

La DSP du bruit est prise dans les périodes de silence. La DSP du signal utile est quant à elle calculée sur chaque trame d'acquisition.

$$\mathbf{w(f)} = \frac{|\mathbf{Y(f)}|^2}{|\mathbf{Y(f)}|^2+\left|\hat{\mathbf{B}}(\mathbf{f})\right|^2} = \frac{\mathbf{1}}{\mathbf{1}+\frac{\left|\hat{\mathbf{B}}(\mathbf{f})\right|^2}{|\mathbf{Y(f)}|^2}} \qquad \textbf{(II.21)}$$

La remarque de ce filtre, est que si le bruit est très bien estimé alors le signal transmis sera directement le signal utile. Le filtrage de Wiener est le filtrage optimal au sens du minimum de l'erreur quadratique moyenne (MEQM). Il adapte le rapport signal sur bruit pour chaque traitée. Cependant, le régime du bruit doit être stationnaire et non transitoire car l'estimation ne sera pas bonne dans ce second cas. Ce type de filtrage est utilisé en général derrière une soustraction spectrale ou autre réducteur de bruit pour améliorer le rapport signal sur bruit de la trace.

3.2.Amélioration psychoacoustique :

Lin , a proposé de décomposer le signal acoustique suivant les bandes de fréquences de l'audition avec un banc de filtres pour simuler le comportement non linéaire de la cochlée. L'erreur quadratique moyenne de chaque bande est ensuite calculée ce qui revient à faire un filtrage de Wiener. Un autre travail proposé par Beaugeant , propose lui un concept différent. C'est de débruiter les canaux contenant du bruit et de ne pas toucher aux autres. Le filtrage sera alors appliqué uniquement dans les bandes spectrales ou le bruit est audible[2].

4. Méthode à sous-espace :

L'approche par sous-espaces se réfère essentiellement a la transformée de Karhunen-Loeve (KLT) et a pour principe d'extraire une base vectorielle orthogonale d'un signal tout en maximisant la concentration de l'énergie. Afin de comprendre la méthode, il faut partir de l'idée que le signal propre **S** est un modèle linéaire d'ordre *K* et dont chaque vecteur qui compose ce modèle est donne par :

$$\mathbf{s = \sum_{i=1}^{M} a_i V_i \quad , M < K} \qquad \textbf{(II.22)}$$

$\{a_i,..., a_m\}$ correspondent en générale a des variables aléatoires complexes et $\{Vi,..., V_M\}$ a des composantes vectorielles complexes d'ordre *K* et considérées comme linéairement indépendantes. Le bruit B d'ordre *K* qui vient corrompre le signal d'entrée est suppose additif, blanc, et décorrèle du signal de parole. Un prétraitement peut être utilise dans le cas de bruit colore afin de blanchir le bruit par filtrage. Le vecteur du signal de parole bruite d'ordre *K* s'écrit alors de la manière suivante :

$$\mathbf{Y = \sum_{i=1}^{M} a_i V_i + B \quad , M < K} \qquad \textbf{(II.23)}$$

La KLT réalise une décomposition en vecteurs propres et en valeurs propres de la matrice de covariance du signal bruite *Y* sur chaque vecteur. Cette matrice, notée *Ryy*, peut s'exprimer par :

$$\mathbf{R_{yy} = R_{ss} + R_{bb} = R_{ss} + \sigma_n^2 I} \qquad \textbf{(II.24)}$$

Avec R_{ss} et R_{bb} les matrices de covariance du signal de parole propre et du bruit, respectivement. La décomposition en valeurs propres peut s'écrire $R_{ss} = U\Lambda_s U^T$. U correspond a une matrice orthogonale de vecteurs propres $\{u_1 \in R^k\}$ de R_{ss} et les valeurs propres se retrouvent dans la diagonale de Λ_s qui peuvent s'exprimer par :

$$\lambda_{s,i} \begin{cases} \lambda_{s,i} + \sigma_n^2 & \text{si } i = 1, \dots, M \\ \sigma_n^2 & \text{si } i = M+1, \dots, K \end{cases} \qquad \textbf{(II.25)}$$

L'idée est d'estimer *s* a partir de *U*, dont les vecteurs propres sont nommes par les vecteurs propres principaux, et qui peut être partitionne en deux sous espace $U = [U_1, U_2]$ - On obtient alors :

$$\mathbf{y = UU^T y = U_1 U_1^T y + U_2 U_2^T y} \qquad \textbf{(II.26)}$$

$U_1 U_1^T y$ constitue la projection de *y* sur le sous-espace du signal, et $U_2 U_2^T y$ est la projection orthogonal sur le sous-espace du bruit. En fait, [Ephraim et Van Trees, 1995] font l'hypothèse que $U_2^T y$ ne contient aucune information sur le signal propre et on estime alors le signal propre qu'a partir de $(U_1^T y, U_1)$, la matrice U_1 peut être écrite de la façon suivante :

$$\mathbf{U_1 = \{u_i : \lambda_{y,i}(k) > \sigma_n^2\}} \qquad \textbf{(II.27)}$$

On se sert de U_1 pour construire un projeté *H* du signal de parole bruite $\hat{S}$ sur le sousespace du signal. La méthode la plus simple pour estimer $\hat{S}$est la méthode des moindres carres. Cette méthode minimise la distance euclidienne entre s et y et elle est définie par :

$$\mathbf{\hat{S} = H_y = U_1 U_1^T y} \qquad \textbf{(II.28)}$$

La méthode proposée par Ephraim et Van Trees, 1995 [8] est plus efficace que la méthode classique par soustraction spectrale proposée par [Boll, 1979], dans le sens ou elle ne laisse aucun bruit musical. Cependant la KLT souffre d'un cout très important de temps de calcul des valeurs propres de la matrice de covariance du signal, ainsi que du prétraitement de

blanchissage dans le cas de bruit colore. Sa complexité en est alors affectée puisqu'elle est cubique *0(n3),* ce qui est trop important pour une implémentation temps-réel.. Des méthodes moins complexes de rehaussement de la parole comme la soustraction spectrale seront sou vent favorisées dans la plupart des cas[8].

5. Conclusion :

Dans ce chapitre, nous avons présenté quelque différentes méthodes de réduction de bruit les plus répandues dans la littérature. Les méthodes découlant de chaque technique ont chacune leur intérêt et les améliorations qui leur ont été apportées au fil du temps sont justifiées. En résumant les difficultés, on s'aperçoit que toutes les techniques ont un point commun, c'est la recherche d'un compromis entre bruit résiduel et distorsion du signal. L'apport de l'aspect perceptuel dans les algorithmes de débruitage de la parole est très prometteur. Un autre point de vue important est la complexité des algorithmes, Par exemple, les méthodes à sous-espace signal, bien que performantes restent pénalisées par la charge de calcul importante qu'elles demandent. Enfin, compte tenu du niveau de complexité, les méthodes basées sur le soustraction spectrale restent très séduisantes du point de vue de leur simplicité et de leur flexibilité par rapport aux améliorations qui peuvent leur être apportées. Nous aurons l'occasion de revenir sur cette technique au chapitre 3, notamment pour présenter un exemple d'une étude comparative dans le but d'établir le meilleur choix de l'algorithme.

1.Introduction :

Le but des outils objectifs d'évaluation de la qualité de la parole utilisé dans cette partie, est de faire une comparaison entre les algorithmes de débruitage. Ce chapitre est composée de trois parties, la première décrive la méthode de soustraction spectrale multi-bande qu'on va la comparer avec la méthode classique et le filtrage de wienner en utilisant les critères objectifs d'évaluation exprimé en deuxième partie, et finalement l'interprétation des résultats.

2.Soustraction spectrale multi-bande :

La méthode de soustraction spectrale proposé par Boll ,est une technique de réduction de bruit populaire en raison de sa simple concept et son efficacité dans l'amélioration de la parole dégradées par un bruit additif. La technique est basée sur l' évaluation directe de l'amplitude spectrale à court terme. Le principe de base de la méthode de soustraction spectrale est de soustraire le spectre du bruit de celui du signal bruité. Le bruit est supposé être décorrélés et additif pour le signal de parole. La méthode de soustraction spectrale réduit considérablement le bruit dans le signal bruité. Toutefois, elle introduit également une distorsion ennuyeux dans le signal appelé bruit musical. En raison des inexactitudes dans l'estimation à court terme du spectre de bruit, les grandes variations spectrales existent dans le spectre débruité causant ces distorsions. En outre, des estimations négatives du spectre de puissance accrue peut se produire. Dans tels cas, les composantes spectrales négatifs sont parqueté à zéro , causant d'autres déformations dans le signal temporel. Des études récentes ont porté sur une approche non-linéaire à la procédure de soustraction . Cette approche a été justifiée en raison de la variation du rapport signal sur bruit au spectre de signal parole. À la différence du bruit gaussien blanc, qui a un spectre plat, le spectre du bruit réel n'est pas plat. Ainsi, le signal de bruit n'affecte pas le signal de parole uniformément sur l'ensemble du spectre. Certaines fréquences sont touchées plus durement que d'autres. Dans le cas du bruit babble, par exemple, les basses fréquences, où la plupart de l'énergie de la parole réside, sont plus touchés que les hautes fréquences. Par conséquent, il devient impératif d'estimer un facteur approprié permet de soustraire seulement la quantité nécessaire de spectre de bruit à partir de chaque intervalle de fréquence, pour empêcher la soustraction destructrice tout en retirant le bruit résiduel.

Kamath[11] a proposé une approche multi-bande pour la méthode de soustraction spectrale d'infrastructures qui répond à ces besoins, c'est-à-dire, qu'il réduit la distorsion mentionnés ci-dessus dans une large mesure tout en maintenant un niveau élevé de la qualité de la parole.

En supposant que le bruit est additif, stationnaire et décorrélé du signal de parole propre, la parole bruité peut être exprimé comme :

$$\boldsymbol{Y(t) = S(t) + B(t)} \quad \textbf{(III.1)}$$

Où y(t), s(nt) et d(nt) sont le signal de parole bruité, propre et le bruit respectivement . Le spectre de puissance du signal de parole bruité peut être approximativement estimé comme:

$$\boldsymbol{|Y(f)|^2 \square\ |S(f)|^2 + |B(f)|^2} \quad \textbf{(III.2)}$$

Le spectre de bruit ne peut pas être obtenus directement, une estimation $\hat{B}(f)$ est calculée pendant les périodes de silence.

Dans la mise en œuvre proposé par Berouti , l'estimation de spectre du signal propre est obtenu par :

$$\boldsymbol{\left|\hat{S}(f)\right|^2 = |Y(f)|^2 - \alpha\left|\hat{B}(f)\right|^2} \quad \textbf{(III.3)}$$

Où α est un facteur qui est en fonction du RSB segmentaire. Cette mise en œuvre suppose que le bruit affecte le spectre du signal uniformément. le facteur α faire la soustraction de sur-estimation du bruit sur l'ensemble du spectre. la Figure III.1 illustre le RSB segmentaire estimés pour quatre bandes de fréquence de parole corrompu par un bruit sous forme de parole.

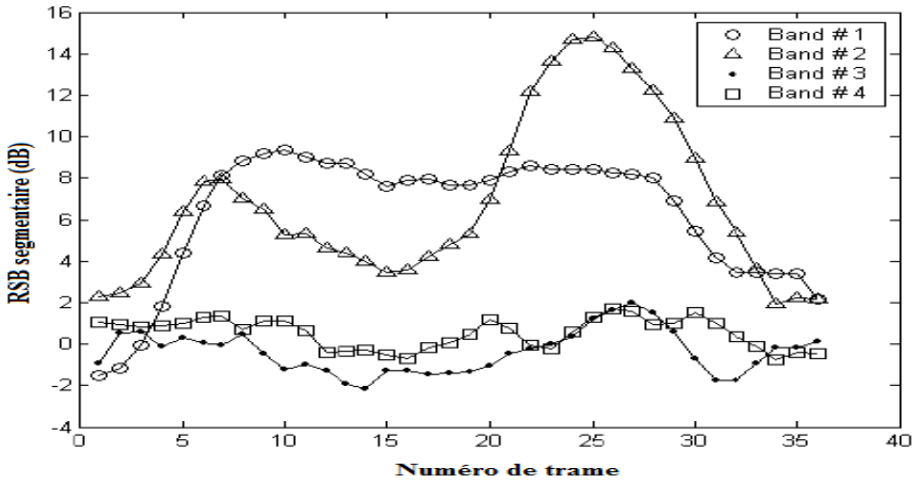


Figure III.1. Le RSB segmentaire de quatre bandes de fréquences de la parole corrompues[11]

Pour prendre en compte que le bruit coloré affecte le spectre de signal parole différemment selon les fréquences, nous proposons une approche multi-bande de soustraction spectrale.

Le spectre du signal est divisée en N bandes et la soustraction spectrale est exécutée indépendamment dans chaque bande. Donc, l'estimation de spectre du signal propre dans la bande iéme est obtenu par :

$$\left|\widehat{S}_i\,(f)\right|^2 = |Y_i(f)|^2 - \alpha_i \delta_i \left|\widehat{B}_i\,(f)\right|^2 \qquad b_i \le f \le e_i \qquad \textbf{(III.4)}$$

Où b_i et e_i sont la fréquence début et la fin de iéme bande, le α_i est le facteur de soustraction de iéme bande et $\boldsymbol{\delta_i}$ est un facteur qui peut être réglées individuellement pour chaque bande de fréquence pour personnaliser les propriétés de suppression du bruit.

Le facteur spécifique α_i est en fonction du RSBi segmentaire de la fréquence d'iéme bande qui est calculé comme :

$$\boldsymbol{RSB_{seg_i}\,(dB) = 10\log_{10}\left(\frac{\sum_{k=b_i}^{e_i}|Y_i(f)|^2}{\sum_{k=b_i}^{e_i}\left|\widehat{B}_i\,(f)\right|^2}\right)} \qquad \textbf{(III.5)}$$

Utilisant la valeur de RSBi calculée dans Eq. (III.5), αi peut être déterminé comme :

$$\alpha_i = \begin{cases} 5 & RSB_{seg_i} \le 5 \\ 4 - \frac{3}{20}\,(RSB_{seg_i}) & -5 \le RSB_{seg_i} \le 20 \\ 1 & RSB_{seg_i} > 20 \end{cases} \qquad \textbf{(III.6)}$$

Alors que l'utilisation de facteur de soustraction αi fournit un degré de contrôle sur le niveau de soustraction de bruit dans chaque bande, l'utilisation de plusieurs bandes de fréquences et l'utilisation de δ_i fournissent un degré supplémentaire de contrôle dans chaque bande.

Les valeurs négatives dans le spectre amélioré dans EQ. III.4 ont été remplacer par le spectre comme :

$$\left|\widehat{S}_i(f)\right|^2 = \begin{cases} \left|\widehat{S}_i(f)\right|^2 & \left|\widehat{S}_i(f)\right|^2 > 0 \\ \beta\left|\widehat{Y}_i(f)\right|^2 & else \end{cases} \quad \textbf{(III.7)}$$

Où β = 0,002.[11]

2.1. Mise en œuvre:

La parole était fenêtré en utilisant la fenêtre de Hamming de 20 ms et un chevauchement de 10 ms entre les trames. La transformée de Fourier rapide (FFT) du signal fenêtré assure les trames précédents et les données qui permettre de bon resultat comme :

$$\overline{Y}_j(f) = \sum_{I=-M}^{M} W_I Y_{j-1}(k) \quad \textbf{(III.8)}$$

Où j est l'indice de la trame , Le nombre de trames M a été limité à 2 pour empêcher l'étalement spectrale. Le filtre W était déterminé et réglé à W =[0,09, 0,25, 0,32, 0,25, 0,09].

Les valeurs de δ_i dans EQ. III.4 ont été déterminé et défini comme :

$$\delta_i = \begin{cases} 1 & f_i \leq 1\,KHZ \\ 2.5 & 1\,KHZ < f_i \leq \frac{F_S}{2} - 2\,KHZ \\ 1.5 & f_i > \frac{F_S}{2} - 2\,KHZ \end{cases} \quad \textbf{(III.9)}$$

Le f_i est la fréquence supérieure de i éme bandes , et le F_S est la fréquence d'échantillonnage. l'utilisation de δ_i pour les bandes basses fréquences est de réduire au minimum la déformation de la parole, puisque la majeure partie de l'énergie de la parole est présenté dans les basses fréquences. Une simple soustraction était aussi utilisé pour les bandes de haute fréquence.

Le spectre amélioré dans chaque bande (Eq. III.4) est combiné, et le signal débruité est obtenu par la transformée inverse de Fourier du spectre amélioré utilisant la phase du spectre bruité original [11].

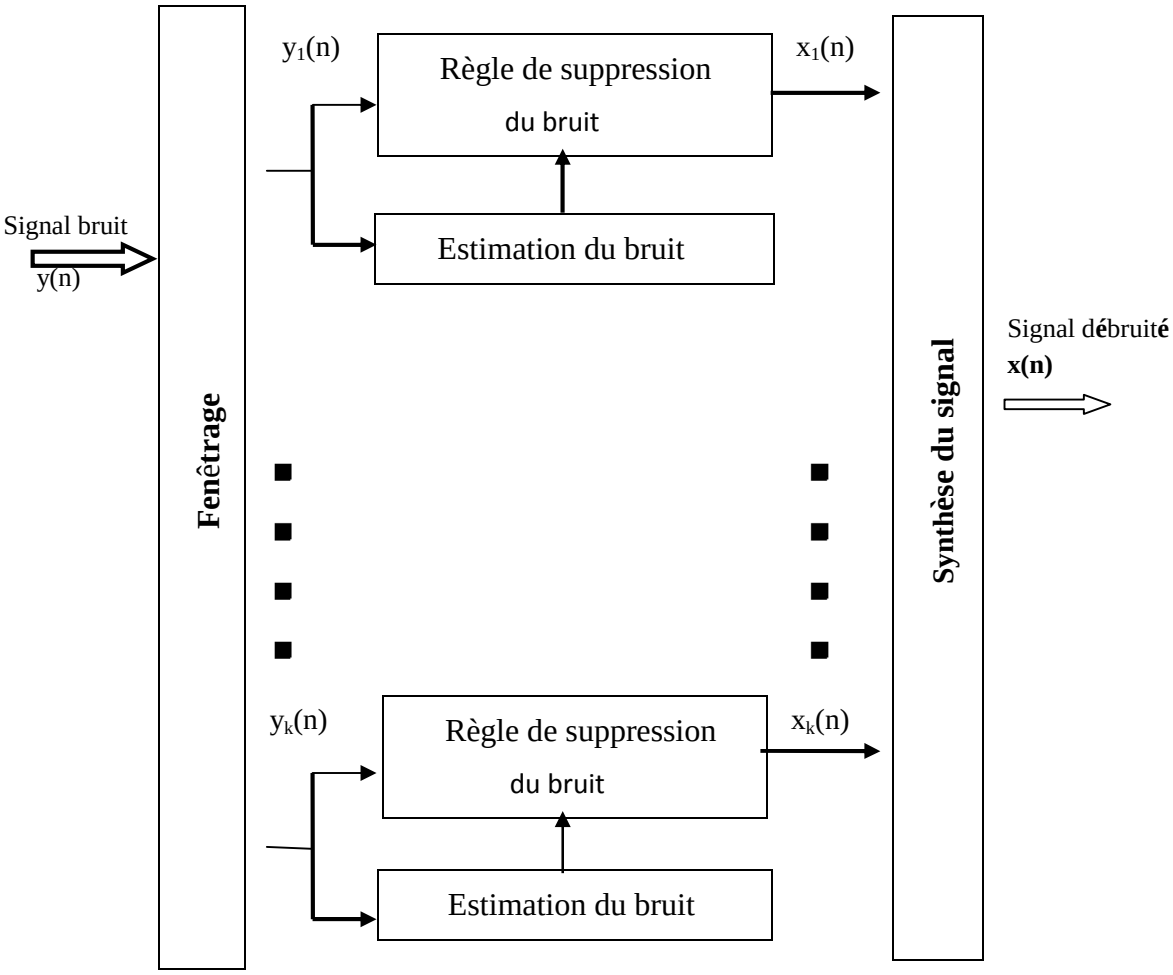


Figure III.2.*Diagramme de la réduction du bruit utilisant un analyse multi-bande.*

3.Evaluation de la qualité de la parole :

Les recherches en psychoacoustique sur la qualité de la parole ont, entre autre, comme but de créer des outils d'évaluation objectifs. Ceux-ci donnent une valeur de la qualité globale de la parole perçue à travers un système de communications. Ainsi, cette étude a comme objectif de contribuer au développement d'un modèle permettant la prédiction de la qualité perçue.

Mais vue que la perception auditive est interne à l'auditeur, la qualité d'un système de téléphonie ne peut pas être mesurée de manière absolue par une méthode objective. Les tests de jugement de la qualité par des auditeurs sont les seules évaluations valables et sûres d'un système de parole. Cependant ces tests sont assez difficiles à mettre en place et assez coûteux.

Des méthodes instrumentales de jugement de la qualité ont donc été développées, essayant de prévoir la qualité de la parole en se basant sur la mesure de signaux caractéristiques[12].

Le prochain paragraphe va donc donner les critères objectifs pour l'évaluation de qualité utilisé dans notre étude comparative .

3.1.Les critères objectifs :

3.1.1 RSB segmentaire (RSBseg) :

Le RSB (Rapport Signal sur Bruit) segmentaire RSBseg est la mesure de qualité objective la plus utilisée dans le domaine temporel. Il définit la moyenne des RSBseg issus de plusieurs segments de courte durée (15 à 20 ms) :

$$\mathbf{RSBseg} = \frac{10}{M}\sum_{m=0}^{M-1}\mathbf{Log}_{10}\frac{\sum_{i=mN}^{mN+N-1}s^2(i)}{\sum_{i=mN}^{mN+N-1}(s(i)-\hat{s}(i))^2} \qquad \textbf{(III.10)}$$

Où s(i), $\hat{s}$(i), N et M sont respectivement le signal de référence, le signal débruité, la longueur d'un segment et le nombre total de segments.

Le RSB segmentaire souffre de deux limitations : d'abord si le signal de parole contient des segments de silence, ce qui est très probable, le s(i) sera nul et n'importe quelle quantité de bruit entrainera un RSB en dB négatif pour ce segment; du coup le RSB total sera biaisé par cette quantité. Ensuite, il faut nécessairement que les deux signaux comparés soient alignés temporellement car ce critère est très sensible aux déphasages[3].

Du fait que ce critère est largement utilisé dans le domaine, nous avons choisi de le conserver comme critère d'évaluation de nos algorithmes bien qu'il donne parfois des résultats incohérents avec ce que nous attendons .en nous basant sur d'autres critères objectifs.

3.1.2 Perceptual Evaluation of Speech Quality (*PESQ) :*

Le PESQ (Perceptual Evaluation of Speech Quality) est l'évaluation de la qualité vocale perçue désignée dans la norme P.862 de UIT comme moyen adapté aux codecs vocaux et aux mesures de bout en bout. De ce fait, d'autres facteurs supplémentaires sont pris en considération pour mieux simuler les conditions réelles, à savoir le temps de propagation, les distorsions dues aux erreurs de transmission, les pertes de paquets...Néanmoins, il existe d'autres facteurs techniques et applications pour lesquels la méthode d'évaluation PESQ n'a pas été encore validée à ce jour, notamment les artefacts causés par les algorithmes de réduction de bruit ainsi que les dégradations liées à l'interaction bidirectionnelle lors de la transmission comme par exemple l'effet d'écho.

Très schématiquement, ce critère se base sur un calcul de distance perceptuelle (di□érence audible entre la représentation perceptuelle du signal de référence et celle du signal de test) suivie d'un modèle cognitif qui permet de prendre en compte le fait qu'une dégradation n'a pas le même impact selon qu'elle est additive ou soustractive, ou selon son contexte (segment de parole ou non) et sa distribution (localisée ou non). La note d'évaluation PESQ finale est une combinaison linéaire de la valeur de perturbation moyenne et de la valeur de perturbation asymétrique moyenne. Le PESQ permet d'évaluer la qualité d'écoute dans de nombreuses conditions de dégradation (perte de paquets, distorsion due au codage et bruit ambiant du coté émission...), aboutissant à une corrélation proche des notes subjectives. Pour les applications de débruitage de la parole, ce critère crée un désaccord au sein de la communauté de recherche bien qu'il soit très utilisé. Dans certains travaux , on dit que la corrélation de ce critère avec la qualité globale n'est importante que dans le cas de la transmission de la parole par le biais de réseaux de communication. D'autres travaux, confirment, par le biais d'études expérimentales et de calculs de corrélation dans le contexte de débruitage de la parole, que ce critère est le plus corrélé parmi six autres mesures objectives, avec un facteur de corrélation de 0.89[3].

3.1.3.Log-Likelihood Ratio (LLR) :

Log-Likelihood Ratio (LLR est une mesure) de distance qui peut être directement calculée à partir du vecteur LPC (Linear Predictive Coding) de parole propre et bruité où le LPC est un codeur particulièrement performant utilise des techniques d'analyse par synthèse LPAS (Linear Prediction Analysis-by-Synthesis). Dans un système de codage LPAS, le signal de parole à transmettre est codé en recherchant la meilleure excitation possible d'un filtre de synthèse, dont les coefficients sont déterminés par une analyse de Prédiction Linéaire. LLR mesure peut être calculée comme suit:

$$d_{LLR}=log\left(\frac{a_d R_c a_d^T}{a_c R_c a_c^T}\right) \quad \textbf{(III.11)}$$

où a_c est le vecteur LPC pour la parole propre, a_d est le vecteur LPC pour le signal bruité, a^T est le transposée de a, et Rc est la matrice d'auto-corrélation pour le signal propre[13].

3.1.4.Weighted Spectral Slope(WSS) :

La mesure de distance WSS calcule la différence entre la pente spectrale de référence et les spectres de parole traité dans chaque bande de fréquences. Les différences de pente spectrales sont pesées selon, le premier, si la bande est près d'une crête ou d'une vallée spectrale et, deuxième, selon que la crête est la plus grande crête dans le spectre.

La mesure de WSS est finalement calculée de chaque cadre de la parole comme :

$$\mathbf{WSS} = \sum_{\mathbf{k=1}}^{\mathbf{L}} \mathbf{W(k)}\,(\mathbf{S_r(k)}\,\bar{\mathbf{s}}_{\mathbf{d}}\mathbf{(k)})^{\mathbf{2}} \qquad \textbf{(III.12)}$$

Où, W (k) est le poids pour la bande k, Sr (k) et $\bar{s}_d(k)$ dénotent les pentes spectrales du référence et le signal bruité du kéme bande et L est le nombre de bandes critiques qui est employé [14].

4. Comparaison entre la méthode de soustraction spectrale classique , la méthode de multi-bande et le filtrage de Wienner :

4.1 La base des signaux :

Les signaux de tests sont échantillonnés à une cadence de 8 KHz. Ces signaux sont corrompus par 8 différents bruits d'environnement réels á différents rapports signal sur bruit variant entre 0 et 15 dB avec un pas de 5 dB. Cette base comprend 30 phrases (en anglais).

Lors de cette étude, nous avons cherché à améliorer les performances des méthodes classique de débruitage par transformée de Fourier. Les mesures objectives sont reliées à des mesures mathématiques entre le signal original et le signal traité afin d'évaluer la technique de rehaussement. L'évaluation porte sur la mesure du rapport signal sur bruit RSB, en particulier sur l'augmentation de cette valeur. nous avons renforcé cette évaluation par des mesures de l'indice de qualité PESQ ,LLR,WSS et RSBseg .Le protocole d'évaluation consiste à comparer la méthode d'atténuation spectrale multi-bandes proposée avec la méthode classique en terme de comparaison entre le signal propre original et le signal débruité .

Les résultats des tests sont illustrés dans les tableaux suivants :

Tableau III.1 :*Mesure du LLR ,RSBseg ,WSS et PESQ pour la méthode soustraction spectrale classique où RSB=0dB*

RSB=0dB	PESQ	WSS	RSBseg	LLR
Aéroport	1.7963	80.08063	-3.3695	0.899533
Bavardage	1,734233	79,50663	-3,50093	0,918467
Voiture	1,7679	69,19033	-2,8676	0,970633
Exhibition	1,785367	79,4791	-3,51377	0,905733
Restaurant	1,780567	80,07643	-3,56407	0,926633
Station	1,746433	75,76693	-3,06383	0,921533
Rue	1,7479	69,36987	-2,65737	0,980467
Train	1,658514	71,10473	-3,38664	1,157152

Tableau III.2 :*Mesure du LLR ,RSBseg ,WSS et PESQ pour la méthode soustraction spectrale classique où RSB=5dB*

RSB=5dB	PESQ	WSS	RSBseg	LLR
Aéroport	2,149133	64,74735	-1,7964	0,715833
Bavardage	2,114067	65,86007	-1,73203	0,753767
Voiture	2,1477	56,06653	-1,31257	0,766933
Exhibition	2,057267	65,98837	-1,68249	0,7669
Restaurant	2,0834	66,4466	-2,02603	0,7319
Station	2,146133	59,74323	-1,4692	0,720833
Rue	2,078433	58,82167	-1,3483	0,800033
Train	2,011435	57,69111	-1,53238	0,954165

Tableau III.3 :*Mesure du LLR ,RSBseg ,WSS et PESQ pour la méthode soustraction spectrale classique où RSB=10dB*

RSB=10dB	PESQ	WSS	RSBseg	LLR
Aéroport	2,502367	49,99893	-0,02577	0,538033
Bavardage	2,459233	52,33373	-0,10327	0,576087
Voiture	2,484133	47,01827	0,395553	0,5776
Exhibition	2,450633	52,6084	-0,14627	0,578333
Restaurant	2,505967	51,54633	-0,14917	0,553833
Station	2,481967	47,0749	0,267767	0,5636
Rue	2,482433	48,61313	0,416267	0,596833
Train	2,373779	46,47357	0,007468	0,699599

Tableau III.4 :*Mesure du LLR ,RSBseg ,WSS et PESQ pour la méthode soustraction spectrale classique où RSB=15dB*

RSB=15dB	PESQ	WSS	RSBseg	LLR
Aéroport	2,846133	41,92903	1,525267	0,388933
Bavardage	2,828	41,67027	1,5375	0,400767
Voiture	2,868433	34,70703	2,186033	0,377867
Exhibition	2,821167	40,73453	1,541033	0,409567
Restaurant	2,647567	42,0667	1,350833	0,410367
Station	2,874315	37,26677	1,942611	0,401275
Rue	2,773967	39,45487	1,6579	0,452567
Train	2,781825	36,76559	1,664215	0,501752

Tableau III.5 :*Mesure du LLR ,RSBseg ,WSS et PESQ pour la méthode soustraction spectrale multi-bande où RSB=0dB*

RSB=0dB	PESQ	WSS	RSBseg	LLR
Aéroport	1,734067	97,4326	-1,72715	0,950453
Bavardage	1,727711	100,6035	-2,07671	1,040341
Voiture	1,642741	71,21994	-3,33492	1,141975
Exhibition	1,575779	89,34558	-1,50157	1,223608
Restaurant	1,690751	96,78564	-2,17084	0,99374
Station	1,669148	87,33881	-1,32543	1,030125
Rue	1,680764	85,12023	-1,29538	1,077281
Train	1,505945	80,9799	-1,91834	1,284802

Tableau III.6 :*Mesure du LLR ,RSBseg ,WSS et PESQ pour la méthode soustraction spectrale multi-bande où RSB=5dB*

RSB=5dB	PESQ	WSS	RSBseg	LLR
Aéroport	2,154564	80,03121	0,20315	0,775538
Bavardage	2,09685	85,11543	0,023269	0,842589
Voiture	2,011435	57,69111	-1,53238	0,954165
Exhibition	2,039272	80,74448	0,588452	0,920964
Restaurant	2,108298	79,37931	-0,26681	0,783301
Station	2,122948	71,684	0,763887	0,793102
Rue	2,070224	69,48467	0,272015	0,874122
Train	2,014833	65,4233	0,296152	1,020234

Tableau III.7 :*Mesure du LLR ,RSBseg ,WSS et PESQ pour la méthode soustraction spectrale multi-bande où RSB=10dB*

RSB=10dB	PESQ	WSS	RSBseg	LLR
Aéroport	2,508663	64,71825	2,123183	0,597276
Bavardage	2,497366	64,3615	2,070875	0,625482
Voiture	2,373779	46,47357	0,007468	0,699599
Exhibition	2,444224	66,93205	2,500489	0,710536
Restaurant	2,493047	62,63379	1,879419	0,614466
Station	2,534772	59,88859	2,820581	0,605373
Rue	2,495246	56,56908	2,13925	0,653535
Train	2,390607	52,62984	2,160659	0,751467

Tableau III.5 :*Mesure du LLR ,RSBseg ,WSS et PESQ pour la méthode soustraction spectrale multi-bande où RSB=15dB*

RSB=15dB	PESQ	WSS	RSBseg	LLR
Aéroport	2,877661	52,52162	3,602871	0,416287
Bavardage	2,884177	49,84189	3,567545	0,459894
Voiture	2,781825	36,76559	1,664215	0,501752
Exhibition	2,824334	54,40416	3,838084	0,536069
Restaurant	2,831047	51,72012	3,234545	0,468301
Station	2,921491	48,5494	4,235358	0,44055
Rue	2,793983	49,58077	3,635816	0,523468
Train	2,812693	42,647	3,675318	0,55235

Tableau III.9 :*Mesure du LLR ,RSBseg ,WSS et PESQ pour la méthode filtrage de wienner où RSB=0dB*

RSB=0dB	PESQ	WSS	RSBseg	LLR
Aéroport	1,561139	123,1423	-1,42652	1,29977
Bavardage	1,607849	127,0294	-1,68831	1,376163
Voiture	1,571123	114,697	-0,24738	1,326109
Exhibition	1,378402	124,0925	-1,04719	1,488452
Restaurant	1,486841	122,1279	-2,00721	1,349316
Station	1,534107	120,2955	-1,00905	1,355895
Rue	1,572436	118,6366	-0,90611	1,429463
Train	1,484211	110,8467	-1,23929	1,581553

Tableau III.10 :*Mesure du LLR ,RSBseg ,WSS et PESQ pour la méthode filtrage de wienner où RSB=5dB*

RSB=5dB	PESQ	WSS	RSBseg	LLR
Aéroport	2,08413	100,6925	0,153575	1,034493
Bavardage	2,08413	100,6925	0,153575	1,034493
Voiture	1,990516	92,71356	1,267971	1,105215
Exhibition	1,954853	101,0939	0,713229	1,193792
Restaurant	1,990057	98,10326	-0,40894	1,068538
Station	1,969879	97,74652	0,589757	1,070343
Rue	1,955912	95,00232	0,283431	1,180164
Train	1,974548	87,08131	0,523349	1,296964

Tableau III.11 :*Mesure du LLR ,RSBseg ,WSS et PESQ pour la méthode filtrage de wienner où RSB=10dB*

RSB=10dB	PESQ	WSS	RSBseg	LLR
Aéroport	2,446251	78,29742	1,837782	0,822521
Bavardage	2,413723	81,1343	1,736944	0,869714
Voiture	2,531685	73,98745	2,970929	0,906447
Exhibition	2,422333	81,52601	2,396547	1,002121
Restaurant	2,414653	75,69679	1,502928	0,863067
Station	2,494352	75,53587	2,51709	0,848278
Rue	2,417681	74,02263	1,803592	0,875003
Train	2,426782	69,43949	2,220193	1,013566

Tableau III.12 :*Mesure du LLR ,RSBseg ,WSS et PESQ pour la méthode filtrage de wienner où RSB=15dB*

RSB=15dB	PESQ	WSS	RSBseg	LLR
Aéroport	2,913859	62,09478	3,303295	0,599818
Bavardage	2,873923	60,45031	3,15674	0,644076
Voiture	2,991834	55,3851	4,329935	0,671869
Exhibition	2,817731	65,77016	3,635736	0,818233
Restaurant	2,79769	62,00503	2,807655	0,680461
Station	2,910282	59,03497	3,754803	0,659894
Rue	2,781907	61,87657	3,268185	0,76691
Train	2,878126	55,46978	3,586716	0,808262

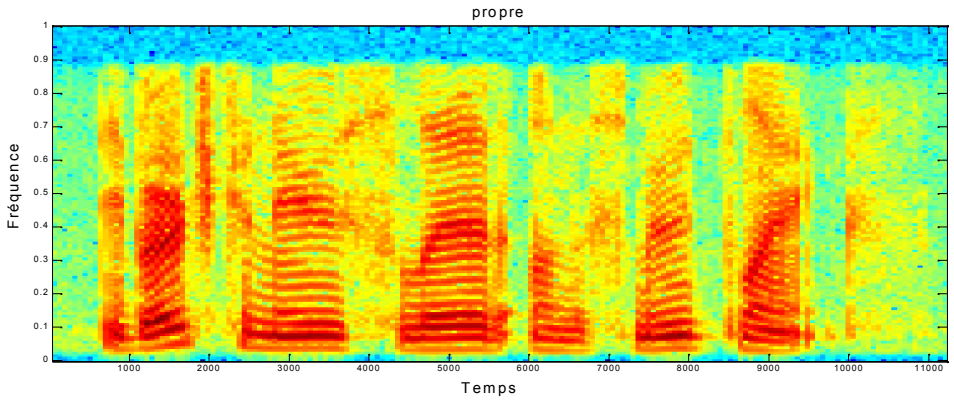


Figure.III.3 *Spectrograme d'un signal de parole propre*

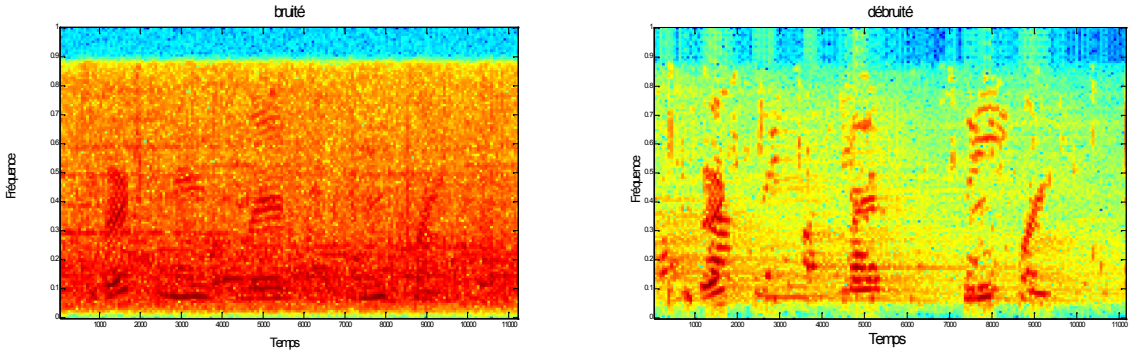


Figure III.4. *Spectrograme d'un exemple de débruitage par la méthode de soustraction spectrale classique (bruit airport)*

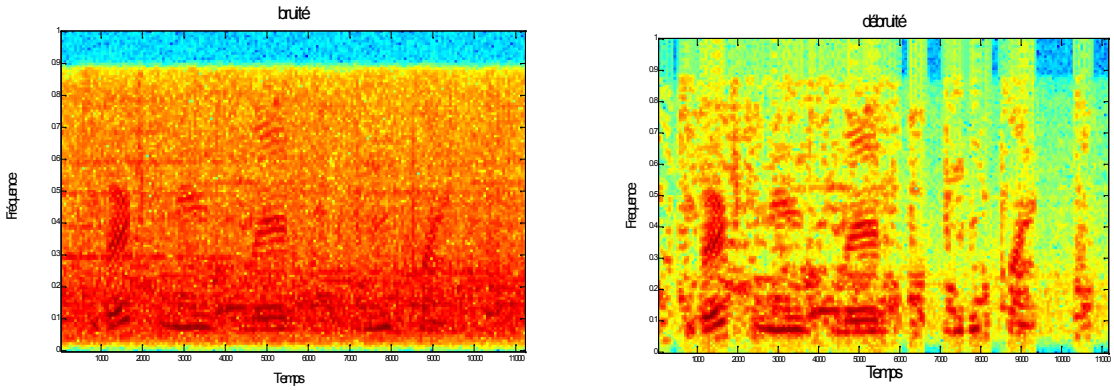


Figure.III.5. *Spectrograme d'un exemple de débruitage par la méthode de soustraction spectrale multi-bande (bruit airport)*

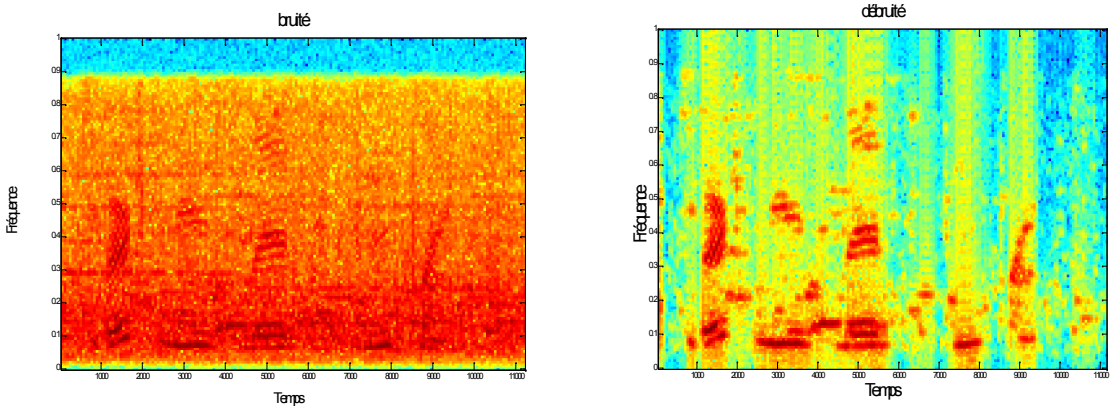


Figure.III.6. *Spectrograme d'un exemple de débruitage par la méthode filtrage de wienner (bruit airport)*

4.2.Les graphes :

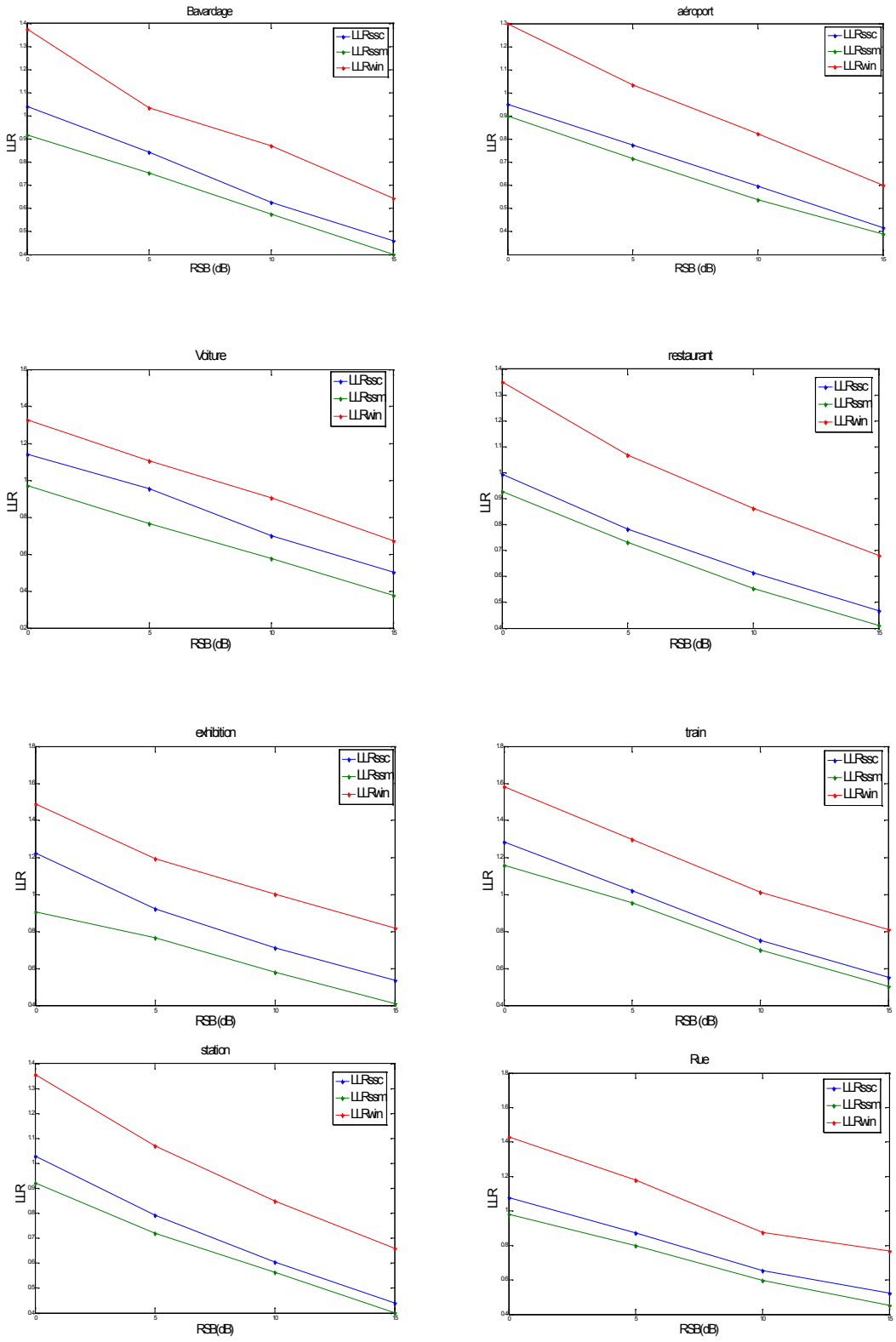
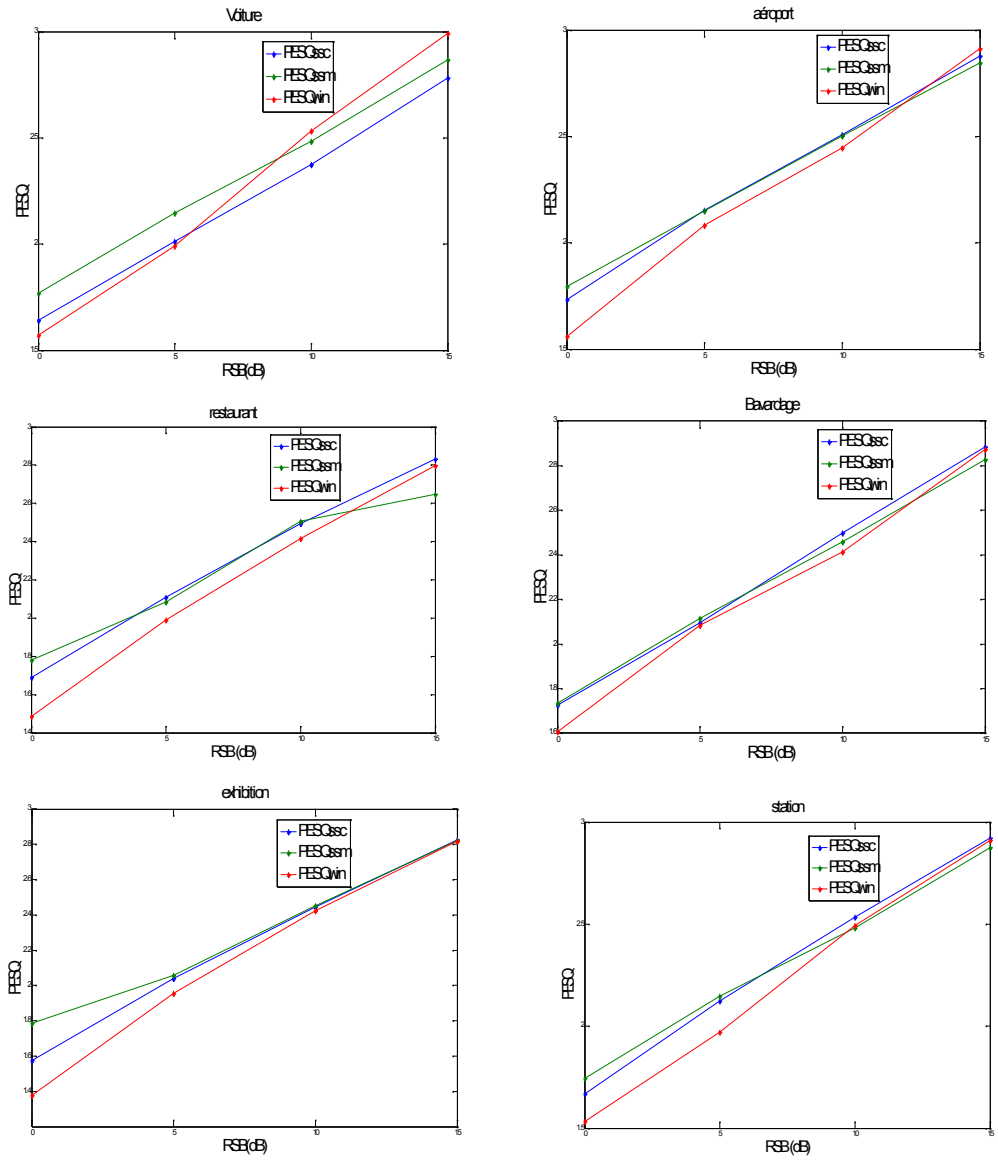


Figure III.7.*Comparaison en terme de moyenne de LLR*

Dans la figure III.7 nous remarquons que le LLR diminue avec l'augmentation du RSB pour les trois méthodes et pour les différent types de bruit, et comme remarque générale, la méthode de filtrage de winner présente le LLR le plus faible puis on trouve la méthode de soustraction spectrale classique et vient au dernier la méthode multi-bande qui présente le LLR le plus élevé quelque soit le type de bruit, on remarque aussi que la méthode de soustraction classique et de multi-bande présente des valeurs proche comparé au filtrage de wienner.



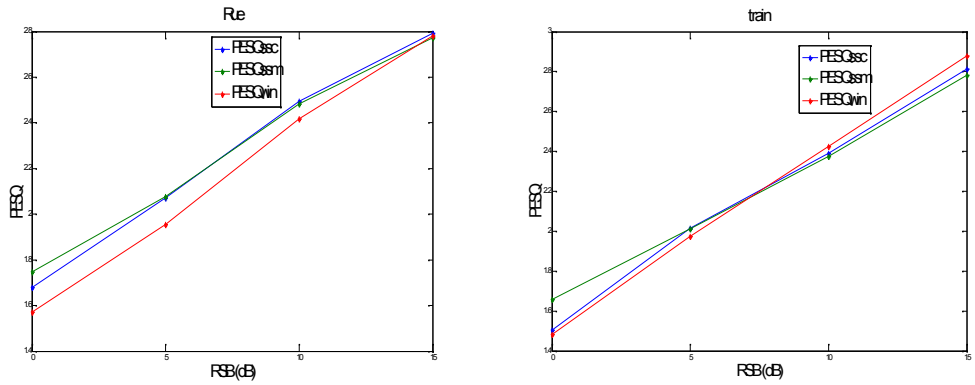
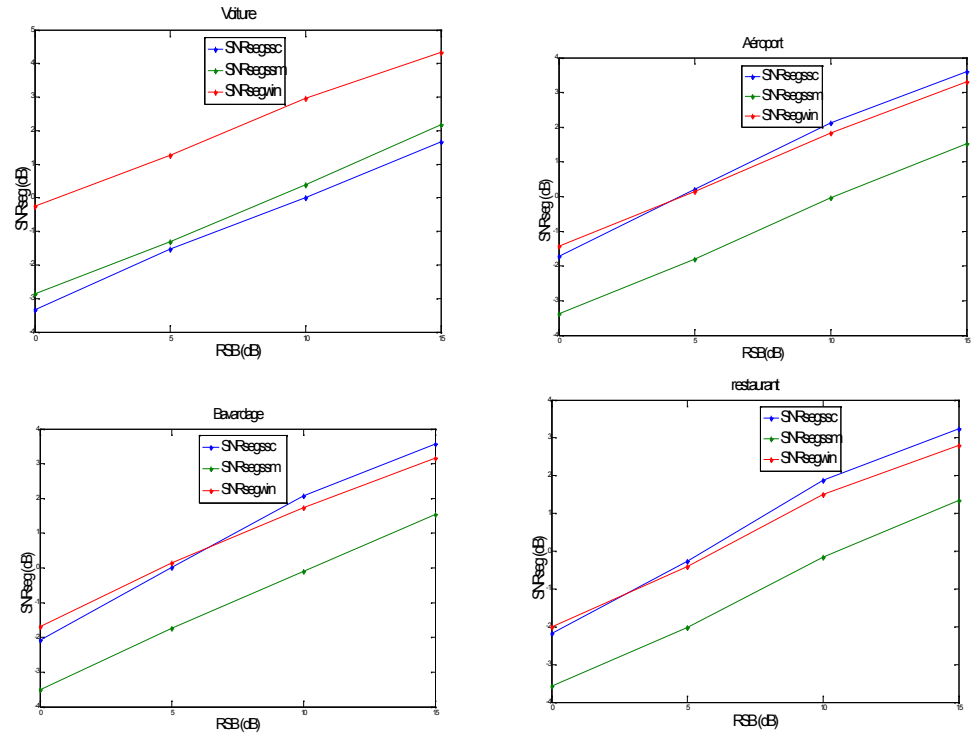


Figure III.8 .*Comparaison en terme de moyenne de PESQ*

Dans la figure III.8 nous remarquons que le PESQ augmente avec l'augmentation de RSB pour les trois méthodes et pour les différents types de bruit, et selon le PESQ la soustraction spectrale multi-bande est bien performé dans la plage (0dB-5dB) où elle devient cohérente avec la méthode classique à partir du 5 dB et elle a mentionnée une chute de performance à partir de 10 dB dans le cas de bruit (restaurant) , contrairement au filtrage de wienner qui est mal fonctionné sur la plage (0dB-5dB) puis il a commencé à améliorer jusqu'à ce qu'il devient le meilleur à 15dB dans le cas de bruit voiture, aéroport et train**.**



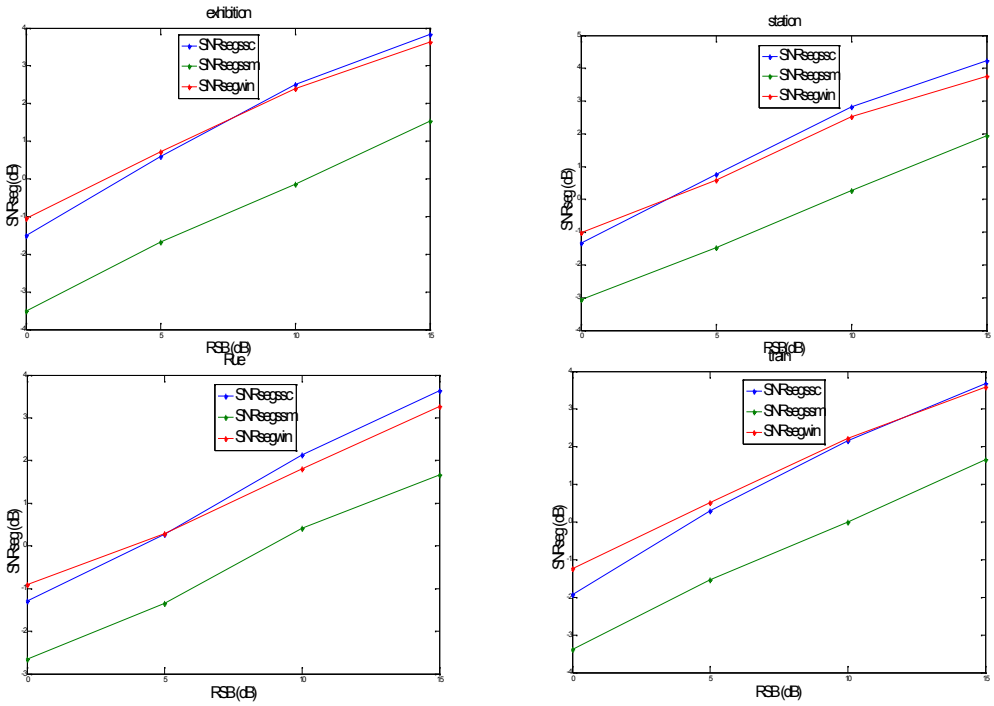


Figure III.9 .*Comparaison en terme de moyenne de RSBseg*

Dans cette figure (III.9) le RSBseg est directement proportionnel à l'augmentation du RSB et cela correspond à la logique parce qu'ils sont presque identique dans la méthodologie de travail sauf que le RSBseg travail sur plusieurs segments de signal , et revenant aux méthodes, cette fois le RSBseg classe la méthode classique dans la première place, puis on trouve la méthode de Wienner , en notons qu'elle est la meilleure dans le cas de bruit (voiture), et vient à la dernière la méthode de multi-bande.

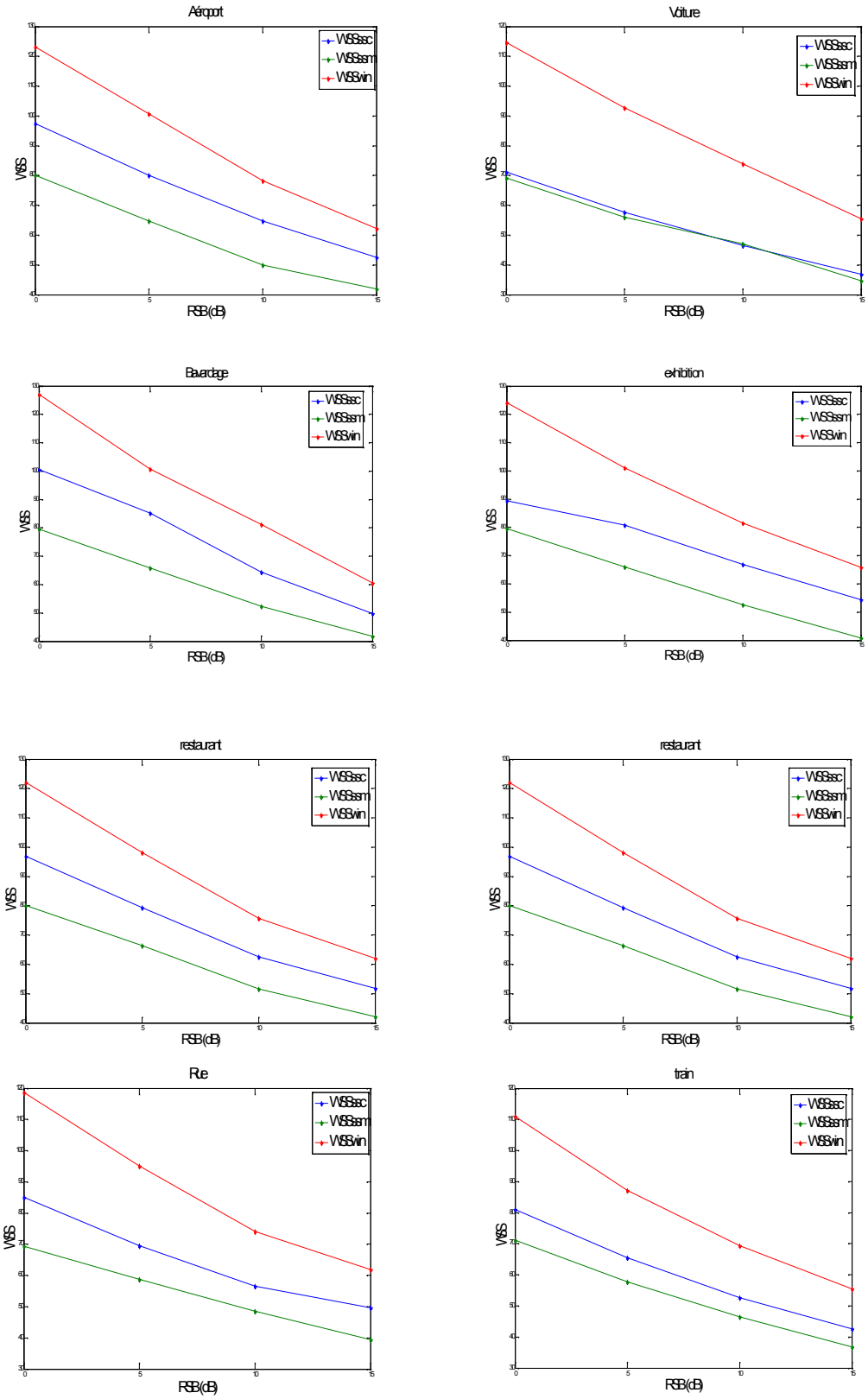


Figure III.10 .*Comparaison en terme de moyenne de WSS*

Dans la figure III.10, le WSS diminue avec l'augmentation du RSB, et soumet l'ordre de la plupart des autres critères telque la méthode multi-bande est plus satisfaire que d'autres, vient à la deuxième la soustraction classique et le filtrage de wienner au dernier, notons que trois méthodes présente des valeurs indépendants sauf dans le cas du bruit voiture où on remarque que les courbes ont des formes de même pente.

5.Conclusion :

Dans ce chapitre, nous avons comparé la technique de réduction du bruit multi-bande avec la méthode classique et le filtrage de Wienner en utilisant les critères objectifes .

D'après les résultats obtenu et les figures, nous remarquons que les valeurs de PESQ et de RSBseg augmentent en temps que le WSS et le LLR diminuent au fur et à mesure que le rapport signal sur bruit augmente pour les trois méthodes et pour les différent types de bruit. Et du point de vue perceptif et de qualité vocale, la méthode multi-bande a permis d'obtenir des résultats meilleurs que ceux fournis par la soustraction spectrale classique et le filtrage de Wienner surtout au faible RSB et pour tous les type de bruits, mais restent les deux autres méthodes utiles pour quelques plages de RSB sur certains types de bruits.

Donc l'approche multi-bande s'avère prometteuse dans le domaine du rehaussement de la parole. Outre les résultats qualitatifs et quantitatifs, la réduction du bruit est nettement considérable.

Conclusion générale

Le problème de débruitage de la parole n'est pas récent. Cependant, il constitue toujours un champ d'étude vaste et encore riche d'idées. L'objectif est de restaurer un signal utile à partir d'observations corrompues par un bruit souvent considéré additif. Cette hypothèse est souvent utilisée, à la fois pour sa simplicité, mais aussi car elle permet de modéliser un grand nombre de situations pratiques. Le signal observé est donc considéré comme la somme du signal de parole et du bruit ambiant.

Donc on a besoin d'un traitement spécifique qui améliore la qualité de la perception de la parole en diminuant le niveau de bruit. Devant ces contraintes, un traitement spécifique du signal de parole est nécessaire afin d'offrir une meilleure qualité de la parole.

En effet, plusieurs techniques de débruitage du signal de parole tels que la méthode de la soustraction spectrale, le filtrage adaptatif de Wiener et la méthode sous espace sont développées et le choix entre ces techniques de rehaussement dépend des données disponibles et des caractéristiques du signal porteurs de l'information.

Lors de cette étude, nous avons cherché à voir les performances de ces méthodes pour le cas d'un bruit coloré. Les mesures objectives sont reliées à des mesures mathématiques entre le signal original et le signal traité afin d'évaluer la technique de réduction de bruit. L'évaluation porte sur la mesure du rapport signal sur bruit RSB, en particulier sur l'augmentation de cette valeur. Or une telle mesure ne renseigne pas nécessairement sur la qualité subjective réelle du rehaussement apporté, nous avons soutenu cette évaluation par des mesures de l'indice de qualité telle que le PESQ , LLR ,SNRseg, et le WSS. Le protocole d'évaluation consiste à comparer la méthode d'atténuation spectrale multi-bande par rapport à la soustraction spectrale classique et le filtrage de Wienner.

Par ailleurs les meilleurs résultats sont fournis par l'analyse multi-bandes. Elle a contribuée d'une manière significative au rehaussement du signal traité et à l'amélioration de la qualité vocale et ce à différents rapport signal sur bruit et pratiquement pour tous les types de bruit.

Les mêmes constatations sont valables, si nous associons la méthode de soustraction spectrale classique avec la méthode de multi -bande, en considérant que cette technique donne comme même des résultats acceptable.

Bibliographie

[1] BENDAHMANE Abderrahmane, « Traitement Automatique de la Parole » ,Cours ,1ère année Master RFIA, Universités Des Sciences Et de la technologie d'Oran– Mohamed BOUDIAF , Année universitaire 2013/2014.

[2] Arnaud Jeanvoine, « Intérêt des algorithmes de réduction de bruit dans l'implant cochléaire : Application à la binauralité », Thèse de doctorat, l'université claude bernard lyon 1´Ecole doctorale E.E.A , 17 Décembre 2012.

[3] Amehraye Asmaa, « Débruitage perceptuel de la parole », Thèse de doctorat, l'Ecole Nationale Supérieure des Télécommunications de Bretagne, 15 mai 2009.

[4] LEMAN Adrien , « Diagnostic et évaluation automatique de la qualité vocle à partir d'indicateurs hybrides (Modéle DESQHI) », thèse de doctorat, L'Institut National des Sciences Appliquées de Lyon 7 juin 2011.

[5] LALITTE Philippe, « ASPECTS ACOUSTIQUE ET SENSORIEL DU BRUIT », l'Université de Bourgogne *Filigrane* n°7, p. 13-31,1er semestre 2008.

[6]MULLER Jean-Philippe, « Le bruit dans les systèmes électroniques », *l'Université* Internationale de Monaco Article Juillet 2002

[7] PLAPOUS Cyril ,« Traitements pour la réduction de bruit. Application à la communication parlée » ,Thése de doctorat, l'université de Rennes 1**,** 12 Décembre 2005.

[8] GIRARD Andre , « Reduction de bruit de signaux de parole mono-capteur basee sur la modelisation par EMD », Memoire de maitrise, UNIVERSITE DE SHERBROOKE Departement de genie electrique et de genie informatique, Fevrier 2010.

[9] FOX Charles , « Débruitage multicapteur appliqué à la téléphonie mains-libres en automobile »,thèse de doctorat ,Institut des sciences et technologie PARIS ,5 décembre 2013.

[10] LALCHHANDAMI and GUPT Rajat , «Different Approaches of Spectral Subtraction Method for Speech Enhancement »,Aricle; Department of Electronics & Communication Engineering Maharishi Markandeshwar University, Mullana (Ambala), INDI,2013

[11] S. Kamath and P. Loizou, «A Multi-Band Spectral Sub- Traction Method for Enhancing Speech Corrupted by Colored Noise », *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, Orlando, 13-17 May 2002, pp. IV-4164.

[12] RAAKE Alexander , «Qualité perçue de parole transmise par voie téléphonique large-bande » , Mémoire pour le Master, Université Pierre et Marie Curie , 12 septembre 2005.

[13] K. Kondo , « Subjective Quality Measurement of Speech » , Article , Signals and Communication Technology, DOI: 10.1007/978-3-642-27506-7_2, 2

[14] NAZANIN Pourmand , «Objective and Subjective Evaluation of Wideband Speech Quality » ,these de doctorat , The School of Graduate and Postdoctoral Studies The University of Western Ontario London, Ontario, Canada ,2012.

[15] S. Boll. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust., Speech, Signal Processing, vol. 27, pages 113–120, 1979.

[16] M. Berouti, R. Schwartz & J. Makhoul. Enhancement of speech corrupted by acoustic noise. In Proc. of ICASSP, volume I, pages 208–211, 1979.

[17] N. Virag. «Single channel speech enhancement based on masking properties of the human auditory system». IEEE, 7 :126–137, 1999.

[18] DO. Walter. «A posteriori wiener filtering of average evoked responses». Electroencephalogr Clin Neurophysiol, Suppl 27 :61+, 1968.

[19] L. Lin, W. H. Holmes, and E. Ambikairajah. «Speech denoising using perceptual midification of winer filtering». IEEE Electronic Letters, 38 :1486–1487, Nov 2002.