N° d'ordre : .....

N° de série : .....

République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Mémoire de Fin d'Étude

Présenté à

L'Université Echahid Hamma Lakhdar d'El Oued

Faculté de Technologie

Département de Génie Electrique

En vue de l'obtention du diplôme de

MASTER ACADEMIQUE

En Télécommunications

Présenté par

Chennouf Sara

Chaoubi Amina

Thème

Etude comparative des techniques de débruitage de signal parole (pour le cas mono-canal)

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

Mr. Nacereddine Lakhdar.	Maitre de conférences	Président
Mr. Ajgou Riadh.	Maitre de conférences	Rapporteur
Mr. Chemsa Ali.	Maitre de conférences	Examinateur

Année Universitaire 2015/2016

Remerciements

En premier lieu, nous remercions
Dieu qui nous a procuré ce succès a
Travers ce modeste travail, nous
tenons à remercier vivement notre encadreur
«Mr. AJGOU Riad», *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 à messieurs le président et les membres de jury d'avoir accepté d'examiner etd'évaluer notre travail.
Nous exprimons également notre gratitude à tous les professeurs qui ontcollaboré à notre formation jusqu'à la fin de
Notre cycle universitaire.
Sans omettre bien sûr de depuis notre premier cycle d'étude jusqu'à la fin de Notre cycle universitaire.
Sans omettre bien sûr de remercier profondément à la réalisation du présent travail.
Tous ceux qui ont contribué de près ou de loin à la réalisation du présent travail.
Enfin ce travail est dédié en hommage de notre .
collègue, notre amie et notre frère

Dédicace

Je dédie cette thèse à . . .

A ma très chère mère Aicha
Tu es l'exemple de dévouement qui n'a pas cessé de m'encourager. Je te souhaite lasanté et une longue vie avec plein de bonheur.
A mon cher père Azzedine
Rien au monde ne vaut les efforts fournis jour et nuit pour mon éducation et mon bien être. Ce travail est le fruit de tes sacrifices que tu as consentis pour mon éducation et ma formation.
A mon cher frère Souhaib
A mes très chères sœurs Soumaia ,Hadjer , Zineb et Bouthaina .
A mes chèresgrands-parents.
A toute ma famille ,mes oncles ,mes tantes ,mes cousins et mes cousines .
A ma camarade chaoubi amina qui a partagé avec moi les moment difficiles de ce travail et à sa famille.
A tout mes amis que j'ai connu dans ma vie .
A mes fleurs
Chibani Meriem,Trad Ouafa,Bouaoun Hayat, Ganadra imane,Hniyat Nour El Houda,Agab Souhir,Bn arouba Zineb.

A mes enseignants de l'école primaire jusqu'à l'université dont leurs conseils précieuxm'ont guidée, qu'ils trouvent ici l'expression de ma reconnaissance.

CHENNOUF SARA

TABLE DES MATIERES

CHAPITRE II: Techniques de débruitage de signal parole (Speech enhancement)

CHAPITRE III : Techniques d'évaluation des algorithmes d'amélioration de signal parole

CHAPITRE IV : Résultats et Simulations

Liste des figures

Liste des Tableaux

Liste des Symboles et abréviations

F_0	La fréquence fondamentale
LPC	Linear Predicting Coding
TFD	Transformée de Fourier Discrète
DCT	Transformation en Cosinus Discret
FFT	Fast Fourier Transform
DSP	Densité Spectrale de Puissance
IFFT	Inverse Fast Fourier Transform
RSB	Le Rapport Signal sur Bruit
MEQM	Minimum de l'erreur Quadratique Moyenne
TOP	Le traitement Optimisé de la Parole
MMSE	Minimum Mean Square Error
MMSE-STSA	Minimum Mean Square Error Short-Term Spectral Amplitude
EVD	La décomposition en Valeurs Propres
SVD	La décomposition en Valeurs Singulières
MAN	Maskee to Audible Noise
CM	La courbe de Masquage
AMPF	Anti-MAN Perceptual Filter
DRT	Diagnostic Rhyme Test
STI	Speech Transmission Index
SII	Speech Intelligibility Index
AI	Articulation Index
PESQ	Perceptual Evaluation of Speech Quality
PSQM	Perceptual Speech Quality Measure
MBSD	Modified Bark Spectral Distortion
BSD	Bark Spectral Distortion
CD	Le Cepstral Distance
WSS	Weighted Spectral Slope Measures
LLR	Likelihood Linear Regression

Liste des Symboles et abréviations

DMOS	Degradation Mean Opinion Score
MOS	Mean Opinion Score
CMOS	Comparison Mean Opinion Score
ACR	Absolute Category Rating
CCR	Comparison Category Rating
DCR	Degradation Category Rating
AR	Auto Regressive
dB	Decibel
Fe	Fréquence d'échantillonnage

Résumé

L'amélioration de la parole vise à améliorer la qualité de parole, en utilisant divers algorithmes. Le débruitage de signal parole est l'amélioration de l'intelligibilité et la qualité de perception globale du signal dégradé et cela on utilisant différentes techniques. L'objectif de ce travail est de faire une étude comparative des performances de plusieurs approches de débruitage (amélioration de signal) de signal parole dans lesquelles un canal d'acquisition unique est disponible (le cas monocanal). Dont, on étudie les différentes méthodes bayésiennes de débruitage du signal de parole. Cette étude peut être imposée par le système utilisé comme les applications basées sur le téléphone mobile ou par la disponibilité du signal désiré comme les applications pré enregistrées. Le problème de débruitage consiste à enlever le bruit à partir du signal corrompu sans l'altérer. Par conséquent, l'étude comparative se fait en termes de qualité de la parole. La qualité de la parole reconstruite est mesurée avec la technique de mesurer de l'évaluation perceptuel de la qualité de parole PESQ (Perceptual Evaluation of Speech Quality), segSNR (segmentaire SNR), WSS (Weighted Spectral Slope Measures) et LLR (Likelihood Linear Regression). Ainsi l'évaluation peut être effectuée en termes du taux d'améliorations du SNR. Toutes méthode a été évaluée en présence de différents types de bruit (les bruits inclus : bavardage, blanc gaussiene, réel, voiture, restaurant, salle d'exposition) en utilisant la base de données TIMIT et le corpus de NOIZEUS.

Mot clés: Amélioration de signal parole, PESQ, segSNR, LLR, WSS, NOIZEUS, TIMIT, Matlab.

Abstract

Improvement of the speech aims to improve the quality of the saound by using various algorithms. The objective of débruitage of signal speech is the improvement of the intelligibility and quality of total perception of the degraded signal by using various techniques. The objective of this work is to make a comparative study on performances of several approaches of débruitage (improvement of signal) of signal speech in which a single channel of acquisition is available (the case monocanal). Although we study the various methods bayésiennes of débruitage of the signal of saound. This study can be imposed by the system used like the applications based on the mobile phone or by the availability of the signal wished like the preregistered applications. The problem of débruitage consists in removing the noise starting from the signal corrupted without deteriorating it. Consequently, the comparative study is done in terms of quality of the speech. The quality of the rebuilt saound is measured with the measurement technique of the perceptuel evaluation of the quality of word PESQ (Perceptual Evaluation of Speech Quality), segSNR (segmentary SNR), WSS (Spectral Weighted Measures Slope) and LLR (Likelihood Linear Regression). Thus the evaluation can be carried out in terms of the rate of improvements of the SNR. All method was evaluated in the presence of various types of noise (the noise included: chattering, white Gaussian, real, car, restaurant, showroom) by using database TIMIT and the corpus of NOIZEUS.

Keyword: Improvement of signal speech, PESQ, segSNR, LLR, WSS, NOIZEUS, TIMIT, Matlab.

ملخص

يهدف تحسين الصوت إلى تحسين جودته و ذلك باستعمال خوارزميات مختلفة (متعددة)، و بذلك يتم الحد أو التقليل من التشويش (الضوضاء). إن الهدف من تقليل التشويش في الإشارة الصوتية هو تحسين الوضوح و جودة الفهم للإشارة بشكل عام و ذلك باستخدام عدة تقنيات. إن هدف هذا البحث هو إجراء دراسة مقارنة للعديد من أداء الأصوات للحد من التشويش و تعزيز جودة الإشارة الصوتية عبر قناة استقبال واحدة متاحة (حالة القناة الواحدة).

حيث يدرس أيضا الأساليب النظرية و الافتراضية للتشويش أثناء الكلام. قد يفرض النظام المستعمل مثل التطبيقات على الهاتف النقال أو تلك المثبتة مسبقا على الجهاز على هذه الدراسة أو عن طريق إمكانية اختيار الإشارة المتاحة المفضلة. تتطلب إزالة التشويش من الإشارة الصوتية دون إحداث أي تغيير. لذلك تتم دراسة مقارنة من حيث نوعية الصوت. إن جودة الصوت معادة التشكيل تقاس عن طريق تقنية قياس جودة الصوت و هي:

PESQ (Perceptual Evaluation of Speech Quality), segSNR (segmentaire), SNR WSS (Weighted Spectral Slope Measures) et LLR (Likelihood Linear Regression)

تقويم كافة الطرق ومناهج التشويش في عدة مواضع في الإشارة الصوتية من بينها التالي: الثرثرة (التكلم)، التشويش الأبيض و الحقيقي، السيارة، المطعم، قاعة العرض. نستعمل في هذه الدراسة قاعدة البيانات TIMIT و مجموعة قوانين NOIZEUS.

الكلمات المفتاحية : amélioration de signal parole (تحسين الإشارة الصوتية) ، PESQ, segSNR, LLR, WSS, NOIZEUS, TIMIT, Matlab

Introduction générale

La parole est un système structuré qui permet aux êtres humains de communiquer entre eux L'information d'un message parlé est transmise par les fluctuations de la pression de l'air qui sont émises par l'appareil phonatoire, c'est le signal vocal. Ce signal est analysé par l'oreille et les informations résultantes sont transmises au cerveau qui ils interprètent. Au sens strict, le contenu d'un signal vocal est représenté uniquement par son intelligibilité. Mais la qualité des signaux paroles souffrent des bruits ambiants ou réverbérations. Dans ce travail on discute différentes techniques de rehaussement (débruitage).

Le problème de débruitage de la parole n'est pas récent. Cependant, il constitue toujours un champ d'étude vaste et encore riche des 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 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. Ce modèle omet tout bruit convolutif, multiplicatif….

Les méthodes classiques, comme la soustraction spectrale ou le filtrage de Wiener, réussissent à réduire le bruit additif, mais en contrepartie, ils introduisent un bruit résiduel (bruit musical) gênant pour la perception humaine. Le besoin de réduire ce type de bruit tout en préservant l'intelligibilité de la parole a poussé les chercheurs proposé d'autres solutions à ce problème, mais aussi pour réduire certaines limitations des systèmes mono-capteur de débruitage de la parole.

Ces premières tentatives ont apporté des améliorations sur la procédure classique de soustraction spectrale afin d'éviter ses effets indésirables et d'améliorer ainsi l'intelligibilité de la parole. Mais par la suite, en vue des progrès du traitement du signal avec nouvelles solutions ont été proposées, par exemple, l'emploi des ondelettes et les méthodes à sous-espace signal.

L'intérêt s'est porté aussi sur l'amélioration des mesures de qualité de la parole en vue d'une évaluation objective s'approchant au mieux du jugement de l'auditeur.

Bien que les tests subjectifs soient plus décisifs et traduisent l'opinion des sujets humains, leur coûteuse mise en œuvre a nécessité de développement d'autres critères. Les plus usuels sont ceux évaluant la qualité de la parole débruité en termes de distorsion de forme en comparaison avec le signal de parole de référence. Certes, ce type de mesure délivre une information sur les performances du débruiteur, mais ne garantit pas d'obtenir une qualité perçue qui peut satisfaire l'auditeur. d'où la proposition de mesures objectives de qualité se basant sur des notions de psychoacoustique pour simuler la perception humaine sans avoir besoin d'effectuer des tests subjectifs.

Notre travail se divise en quatre chapitres :

- ✓ Le premier chapitre : nous discutons la description et les caractéristiques du signal de parole comme la fréquence fondamentale F_0 , formants.
- ✓ Le deuxième chapitre : ce chapitre est consacré à l'étude des différentes techniques de débruitage de signal parole : Les techniques d'atténuation spectrales, soustraction spectrale, la soustraction spectrale au sens d'Ephraim et Malah, filtrage de Wiener, réduction du bruit musical, détection de la parole par le TOP, MMSE et MMSE-LSA (Minimum Mean Square Error-Short-Term Spectral Amplitude), méthodes à sous-espace signal, le phénomène MAN (Maskee to Audible Noise).
- ✓ Le troisième chapitre : nous avons exploré les techniques d'évaluation de rehaussement de signal parole avec une étude de quelque méthode de débruitage.
- ✓ Le quatrième chapitre : ce chapitre est consacré à la partie de simulation en mené une série de simulations à l'aide du logiciel " Matlab 2011 ". On a simulé quelques techniques étudiées dans les chapitres précédents à savoir Wiener-Scalart96, MMSE-Cohen2004, MMSE-STSA84 et MMSE-STSA85.

A la fin et avec une conclusion générale et perspectives on termine notre travail.

Chapitre 1

Généralité sur le signal parole

I.1 Introduction

La parole est le principal moyen de communication dans toute société humaine. Son apparition peut être considérée comme concomitante à l'apparition des outils, l'homme ayant alors besoin de raisonner et de communiquer pour les façonner. D'un point de vue humain, la parole permet de se dégager de toute obligation de contact physique avec la machine, libérant ainsi l'utilisateur qui peut alors effectuer d'autres tâches.

Avant de voir les différents techniques de débruitage de signal parole, il est important de commencer par comprendre ce qu'est la parole ?Quels sont les caractéristiques de son continues?

I.2 Qu'est-ce que la parole ?

La parole est la manière naturelle et, en conséquence,la forme la plus commune de communication humaine et n'est autre qu'un ensemble de voix articulées qui expriment ses pensées, ses désirs, et même ses sentiments.

I.3 Fonctionnement de l'appareil vocal

L'ensemble du système vocal se compose des poumons et du conduit trachéobronchique, du larynx et du conduit vocal, formé par le pharynx et les cavités nasales et orales [1].

La figure I.1 représente l'appareil phonatoire et modèle mécanique de production de la parole.

- L'ensemble poumons et conduit trachéobronchique se comporte comme un générateur d'air qui alimente le larynx.
- Le larynx est l'ensemble de cartilages articulés, ligaments muscles et muqueuses, qui grâce à son action sur les cordes vocales (muscles élastiques), permet de déterminer la nature du flux d'air qui va exciter le conduit vocal.
- Le conduit vocal par des organes articulateur imprime au son émis les caractéristiques spécifiques permettant de distinguer les différents phonèmes et ceci en tant que :
  - résonateur de l'onde glottique pour la production des voyelles
  - générateur de bruit pour la production des consonnes.

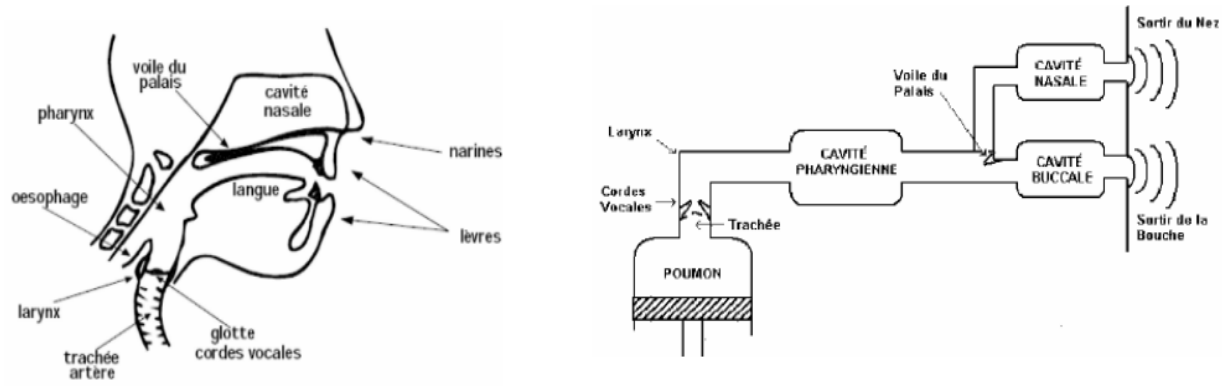


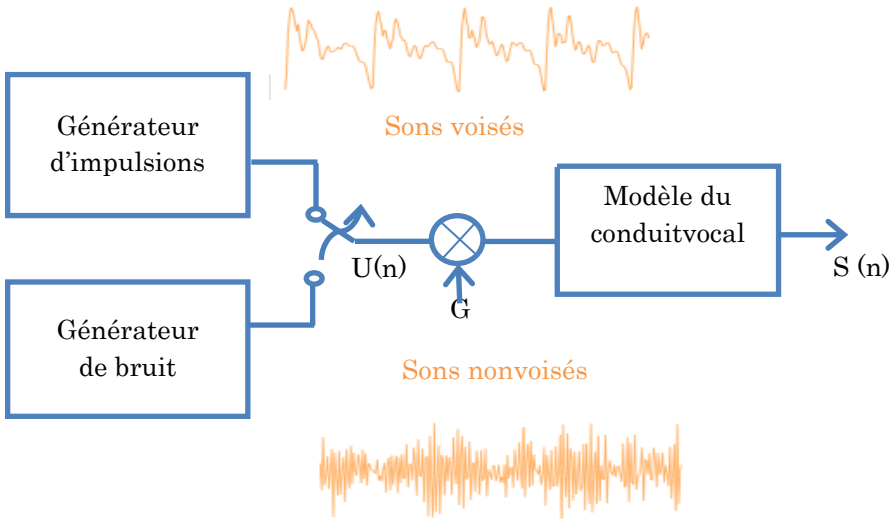
Figure I.1 : Appareil phonatoire et Modèle mécanique de production de la parole [2].

I.4 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 figureI-2 [3].

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 F_0.

Quand les 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.



FigureI-2 : Modèle simple de production de la parole

La Figure I-3 représente l'évolution temporelle, ou audiogramme, du signal vocal pour les mots « parenthèse », et « effacer ». On y constate une alternance de zones assez périodiques et de zones bruitées, appelées zones **voiséeset** **non-voisées**. La figure I-4 (donne une représentation plus fine de tranches de signaux voisés et non voisés. L'évolution temporelle ne fournit cependant pas directement les traits acoustiques du signal [4].

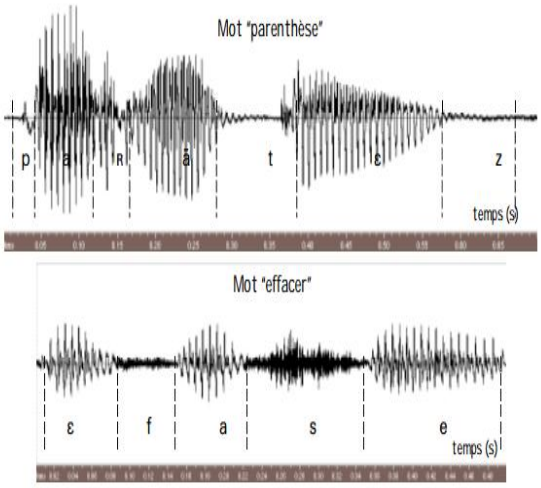


Figure I-3 : Audiogramme de signaux de parole

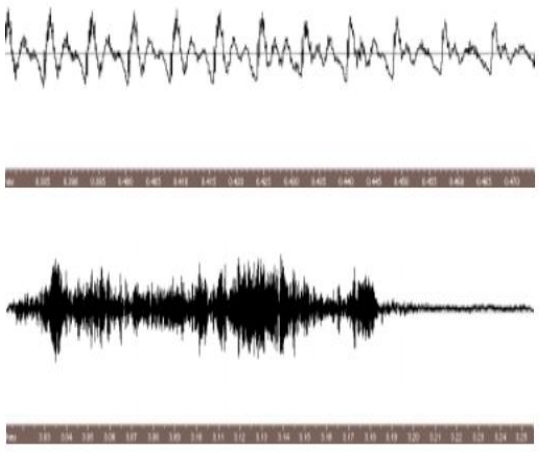


Figure I-4 : Exemples de son voisé (haut) et Non-voisé (bas).

I.5 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 (ou 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 [5].

I.6 Caractéristiques de la parole

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 :

- **Le domaine temporel**

- L'enveloppe, qui contient l'énergie du signal
- La structure fine, qui contient les différentes variations du signal

- **Le domaine fréquentiel**

- La fréquence fondamentale,F_0 et ses harmoniques
- Les formants, F_1, F_2, F_3,.....

Des méthodes existent pour trouver ces différentes informations.L'enveloppe temporelle peut être obtenue en calculant le module de la transformée de Hilbert. La F_0 se calcul par une autocorrélation du signal [6]. Et les formants par la méthode de la LPC (Linear Predicting Coding [6]).

La figure I-5 représente le spectre d'un signal parole présentant trois résonances formantiques.

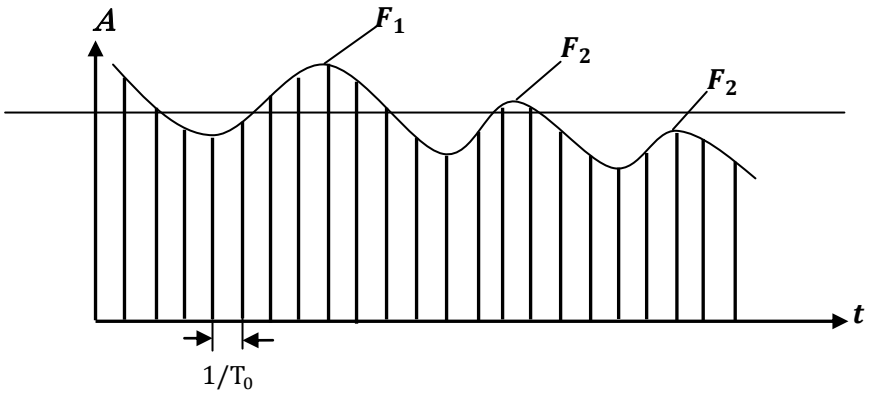


Figure I-5 : Spectre présentent trois résonances formantiques [1].

I.6.1 Fréquence fondamentale F_0

La fréquence fondamentale dans le signal de parole « représente » la vibration des cordes vocales (le signal glottique). Ce signal est composé de la fondamentale et de ses harmoniques. Elle caractérise la personne qui parle [7] :

– Pour un homme :≈ $100Hz$

– Pour une femme:≈ $200Hz$

– Pour un enfant :≈ 300 à 400 Hz

La figure I-6 représente un sont voisé avec sonfréquence fondamentaleF_0.

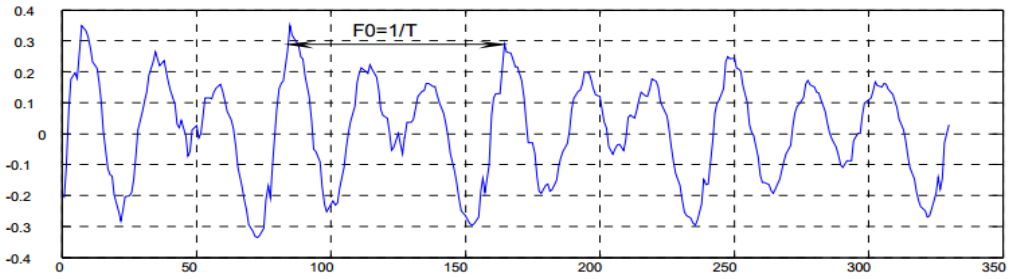


Figure I-6 : Un son voisé avec son fréquence fondamentale « F_0 » (Amplitude V_s échantillon).

I.6.2 Formants

On peut les caractériser comme étant le filtre analogique du signal glottique. En effet les formants représentent les résonances entre le signal glottique et la sortie de la bouche. Ils caractérisent les voyelles. Ce sont les maximums d'énergie dans le signal. On peut représenter les deux premiers formants (F_1et F_2) dans le triangle de Delattre (table et figure) [7].

Voyelle (API)	F_1(Hz)	F_2(Hz)
[u]	320	800
[o]	500	1000
[α]	700	1150
[a]	1000	1400
[ø]	500	1500
[y]	320	1650
[$\in$]	700	1800
[e]	500	2300
[i]	320	3200

Table I-1 : Exemple de valeursdes formants pour différentes voyelles.

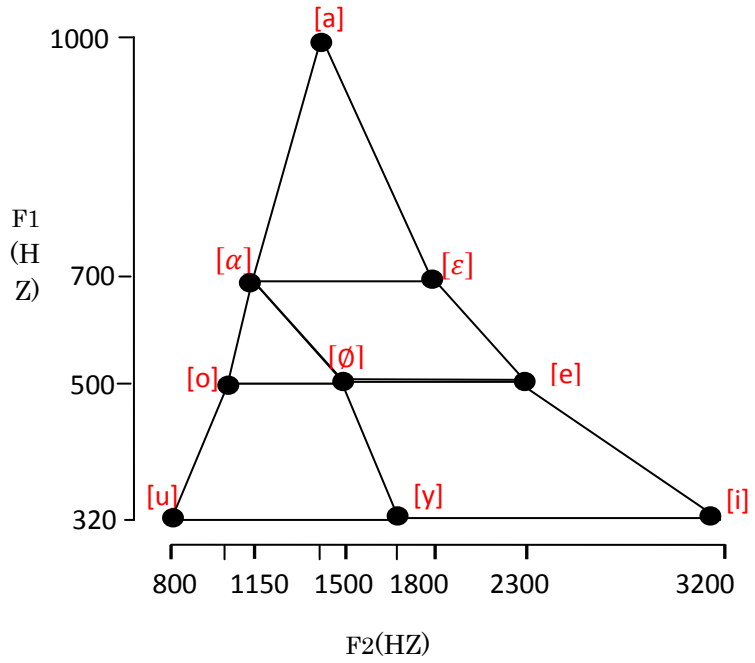


Figure I-7 : Triangle formantiques de Delattre

I.7 Numérisation de signale parole

Le parole est produite par l'articulation des membres phonatoires de être humain et prend une forme analogique apériodique,ce qui est impossible pour que la machine puisse l'interpréter ou le prédire car elle ne comprend que du numérique. Pour cela on doit faire un traitement de numérisation sur ce signal. L'une des méthodes les plus utilisés dans la numérisation est la méthode Delta ou MIC qui consiste en trois étapes : l'échantillonnage, la quantification et le codage [8].

I.7.1 L'échantillonnage

L'échantillonnage consiste à transformer une fonction $a(t)$ à valeurs continues en une fonction â(t) discrète constituée par la suite des valeurs $a(t)$ aux instants d'échantillonnage $t = kT$ avec k un entier naturel (figureI-8). Le choix de la fréquence d'échantillonnage n'est pas aléatoire car une petite fréquence nous donne une présentation pauvre du signal. Par contre une très grande fréquence nous donne des mêmes valeurs, redondance, de certains échantillons voisins donc il faut prélever suffisamment de valeurs pour ne pas perdre l'information contenue dans $a(t)$. Le théorème suivant traite cette problématique :

Théorème (de Shannon):La fréquence d'échantillonnage assurant un non repliementdu spectre doit être supérieur à 2 fois la fréquence haute du spectre du signal analogique.

$$F_{ech} \geq 2 \times F_{max}$$

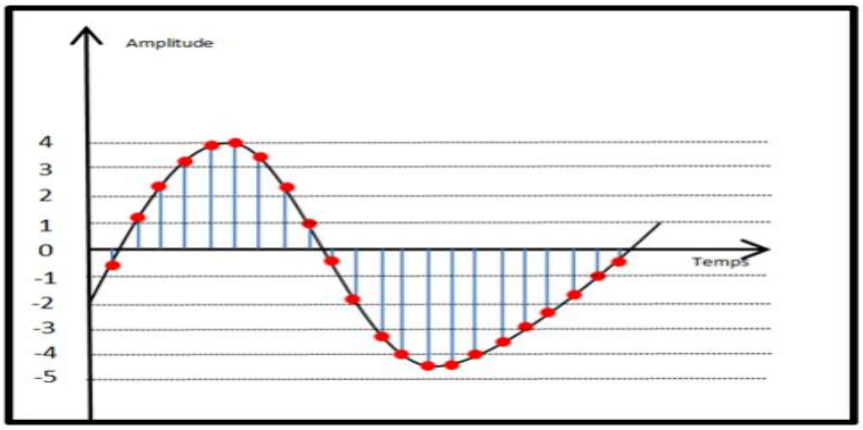


Figure I-8 : Un signal échantillonné

Pour la téléphonie, on estime que le signal garde une qualité suffisante lorsque son spectre est limité à 3400 Hz et l'on choisit f_e = 8000 Hz. Pour les techniques d'analyse, de synthèse ou de reconnaissance de la parole, la fréquence peut varier de 6000 à 16000 Hz.Par contre pour le signal audio (parole et musique), on exige une bonne représentationdu signal jusque 20 kHz et l'on utilise des fréquences d'échantillonnage de 44.1 ou 48kHz. Pour les applications multimédia, les fréquences sous-multiples de 44.1 kHz sont de plus en plus utilisées : 22.5 kHz, 11.25 kHz [9].

I.7.2 Quantification

Cette étape consiste à approximer les valeurs réelles des échantillons selon une échelle de n niveaux appelée échelle de quantification. Il y a donc 2n valeurs possibles comprises entre $-2n-1$ et $2n-1$ pour les échantillons quantifiés (figureI-9).L'erreur systématique que l'on commet en assimilant les valeurs réelles de l'écart au niveau du quantifiant le plus proche est appelé bruit de quantification [8].

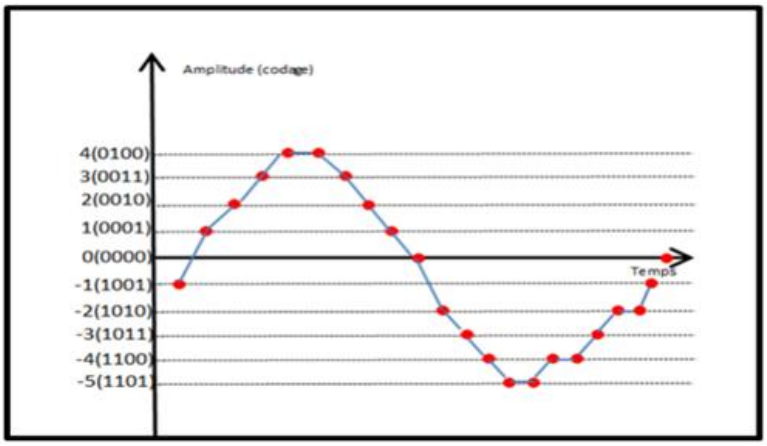


Figure I-9 : Un signal quantifié

I.7.3 Codage

C'est la représentation binaire des valeurs quantifiées qui permet le traitement du signal sur machine (figureI-9)[8].

I.8 Analyse et modélisation de la parole

Le signal de parole est un processus aléatoire non-stationnaire à long terme, mais il est considéré comme stationnaire dans des fenêtres temporelles d'analyse de l'ordre de 20 à 30ms. Cette propriété de stationnaritéà court terme permet donc une analyse et modélisation progressive du signal de parole accompagnée, bien sûr, d'un chevauchement de fenêtres pour permettre une continuité temporelle des caractéristiques de l'analyse et du modèle [5].

I.8.1 Méthode d'autocorrélation basée sur analyse LPC

Dans l'analyse par prédiction linéaire LPC, le conduit vocal est modélisé par une fonction de transfert qui suit un modèle autorégressif. Cette analyse est fort utilisée

dans le codage de parole dans le but de réduire la redondance du signal vocal, ou pour extraire des paramètres pertinents pour la reconnaissance de parole [10].

L'estimation des coefficients de la fonction de transfert du conduit vocal est faite en supposant connaitre le signal d'excitation. Pour les sons non voisés, le signald'excitation est un bruit blanc de moyenne nulle et de variance unité. Pour les sons voisés, cette excitation est une suite d'impulsions d'amplitude unité. La fonction de transfert du conduit vocal dans le domaine Z est donnée par :

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1-A(z)} \quad \text{(I.1)}$$

Ou $A(z) = \sum_{k=1}^{p} a_k z^{-k}$ est le prédicteur linéaire, a_k sont les coefficients de prédiction,S(z) est le signal de parole produit en sortie, U(z) est le signal d'excitation et G est un gain. Le signal de parole s(n) à la sortie du modèle est donc représenté par la somme d'une combinaison linéaire des échantillons précédents et de la fonction d'excitation, tel que :

$$S(n) = \sum_{k-1}^{p} a_k\, S(n-k) + Gu(n) \quad \text{(I.2)}$$

Le modèle de prédiction exploite le fait que les échantillons successifs du signal de parole sont corrélés, d'où l'intérêt de ce modèle dans le codage de la parole dans le sens où il permet de représenter la parole juste par ses paramètres pertinents, sans redondance. Signalons également que les coefficients sont choisis de façon à minimiser l'erreur quadratique de prédiction sur chaque segment de la fenêtre d'analyse.

I.8.2 Méthode Cepstre

Le cepstre est basé sur une connaissance du modèle de production de la parole. Comme nous l'avons vu dans la section précédente, une modélisation du signal de parole consiste à définir ce signal comme le résultat de la convolution de la fonction de transfert du conduit vocal (filtre) par un signal d'excitation (source). Le but du cepstre est de séparer ces deux contributions (source et filtre) par application de la convolution à travers une transformée en cosinus discret.

Le processus de calcul du cepstre est le suivant ou s,u et h le signal de parole, le signal d'excitation (source) et la fonction de transfert du conduit vocal (filtre).

$$s = u * h \quad \text{(I.3)}$$

$$TFD(s) = UH \quad \text{(I.4)}$$

Le logarithme de l'amplitude transforme le produit de la TFD en somme, on obtient alors :

$$log\ |S(v)|\ =\ log\ |U|\ +\ log\ |H| \qquad (I.5)$$

Par transformation en cosinus discret (DCT), on obtient le cepstre. L'expression du cepstre réel est donc :

$$c\ =\ DCT(\log(TFD(s))) \qquad (I.6)$$

L'espace fréquentiel de représentation du cepstre est équivalent à un espace temporel. A partir du cepstre (FigureI-10), il est possible de définir la fréquence fondamentale de la source u en détectant les pics périodiques (harmoniques) au-delà d'un certain nombre N de coefficients.

En effet, les N premiers points du cepstre contiennent l'information la plus pertinente sur le spectre et permettent d'obtenir un spectre lissé, débarrassé des harmoniques dus à la contribution de la source. Cependant, déterminer la fréquence fondamentale d'un signal de parole reste encore un problème difficile.

En effet, les algorithmes classiques manquent de robustesse quand le bruit est présent, quand la fréquence fondamentale change rapidement ou quand la valeur de celle-ci n'est pas assez élevée [5].

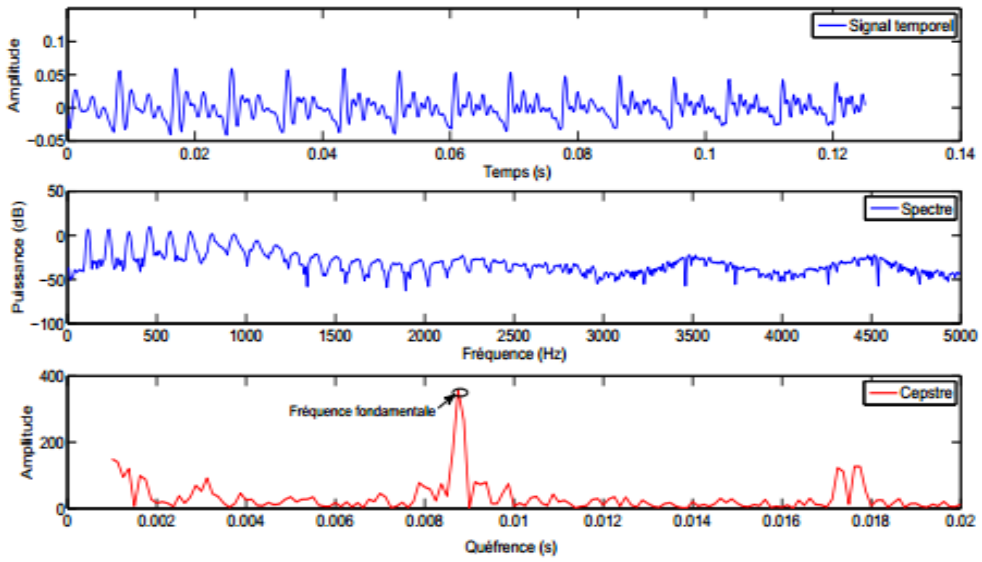


Figure I-10 : Détermination de la fréquence fondamentale Parcepstre.

I.9 Conclusion

Un signal de parole est une séquence de sons correspondant à une suite d'états de l'appareil phonatoire. Le signal de parole est un processus aléatoire non stationnaire à long terme.

Nous avons pu voir au cours de ce chapitre, le phénomène de la production de la parole, les différentes sources permettant la génération des sons d'une langue donnée.

Nous avons aussi présenté quelques méthodes de base d'analyse et de modélisation du signal de parole. Ces méthodes peuvent être utilisées pour mettre en évidence les caractéristiques fréquentielles de ce signal et pour mettre en œuvre, en exploitant ces caractéristiques.

Chapitre 2

Techniques de débruitage de signal parole (speech enhancement)

II.1 Introduction

L'oreille humaine a des capacités impressionnantes pour reconnaitre 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).

Dans ce chapitre, la base de la technique de réduction de bruit dans un système monovoie va être présentée. La plupart des méthodes de réduction de bruit sont issues de deux principales techniques. Le premier est la « Soustraction Spectrale » et le second est le « filtrage de Wiener», il y a aussi Plusieurs méthodes de débruitage de la parole.

Avant l'étude, nous nous pencherons sur ces techniques pour identifier les différents types de bruit et caractéristiques.

II.2 Nature et 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 II.1, extrait de [11], représente les différentes classes auxquelles un bruit peut appartenir.

Propriétés	Types
Structure	Continue/Iimpulsif/périodique
Type d'interaction	Additif/multiplicatif/convolutif
Comportment temporal	Stationnaire/non-stationnaire
Bande de fréquence	Etroit/large
Dépendence	Corrélé/décorrélé
Propriétés statistiques	Dépendant/independent
Propriétés spatiales	Cohérent/incoherent

Tableau II.1 : Différentes classes du bruit

Comme notre but est essentiellement le débruitage de parole, on se limite dans notre étude aux bruits additifs, stationnaires ou non stationnaires, et décorrélé avec la parole (indépendance au sens statistique), tels que le bruit de conversation appelé bavardage (babble), le bruit de voiture appelé (car) et le bruit blanc gaussien (ce dernier est souvent utilisé mais peu réaliste).

II.3 Réduction du bruit pour les signaux monovoies

La réduction de bruit monovoie est effectuée à partir d'un seul microphone, il n'y aura alors qu'un seul signal à traiter. Les démonstrations qui vont suivre seront décrites de la même façon pour ces méthodes. Dans tous les cas, une amélioration de la compréhension de la parole est nécessaire, ces algorithmes sont faits pour améliorer la parole par rapport au reste du signal [12].

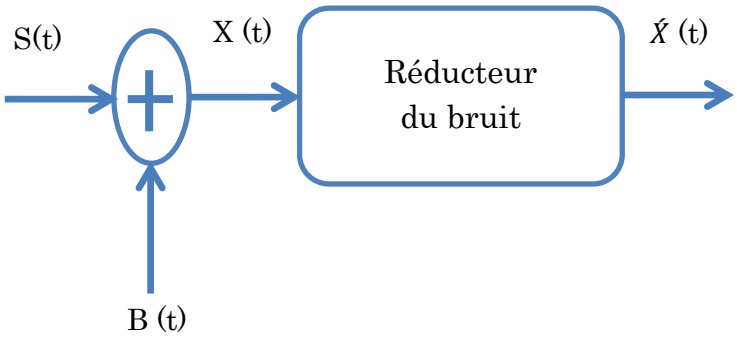


Figure II.1: Modèle de débruitage monovoie

Au départ, on peut considérer que le signal qui arrive au microphone est compose 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 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 considéré le signal arrivant au microphone est un mixage entre un signal utile et un signal bruité :

$$x(t) = s(t) + b(t) \quad \text{(II.1)}$$

Où:

– t : représente le temps

– x : représente le signal capté par le microphone

– s : représente le signal utile

– b : représente le bruit

Cette équation (EqII.1) doit *ê*tre modifiée pour traiter le signal.Il doit *ê*tre échantillonné. De ce fait, l'équation devient :

$$x(n) = s(n) + b(n) \quad \text{(II.2)}$$

Où:

– n : représente le numéro de l'échantillon

– x : 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 x(n) pourrait alors *ê*tre faite pour enlever le bruit dans x(n). Cependant, le bruit n'est jamais connu par avance. Il faut poser des hypothèses. Le bruit b(n) doit *ê*tre 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 connait est x(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 m*ê*me que celui du bruit. Il est actuellement très difficile de travailler dans le domaine temporel par échantillon car le filtrage doit *ê*tre un produit de convolution.

II.4 Estimation du bruit

La première étape consiste à transformer le signal temporel par l'intermédiaire de la FFT en nombre complexe et plus particulièrement le module. Il est très rare d'utiliser la phase du signal car l'amplitude suffit habituellement. L'équation (II.2) devient alors [12] :

$$X(f) = S(f) + B(f) \quad \text{(II.3)}$$

Et en module :

$$|X(f)|^2 = |S(f)|^2 + |B(f)|^2 \quad \text{(II.4)}$$

Où :

– f : représente les fréquences

– X : représente le spectre du signal capté par le microphone

– S : représente le spectre du signal utile

– B : représente le bruit

Le but de cette sous-section est d'estimer de bruit B(f) afin d'appliquer en sortie, un algorithme de réduction de bruit (figureII.1). Pour trouver le spectre du bruit, la première hypothèse consiste à dire qu'il est stationnaire (hypothèse qui n'est pas toujours valable) et indépendant sur la trame d'analyse. La seconde hypothèse est que la parole varie beaucoup plus rapidement que le bruit. On peut dire par contre que sur des

trames d'analyse courte (durée de la trame < 25 ms) la parole est stationnaire. C'est pourquoi une trame d'analyse longue renseigne sur le bruit alors qu'une trame d'analyse courte donne des informations sur la parole.

Le calcul du bruit se fait généralement sur la densité spectrale de puissance (DSP) qui est obtenue en calculant l'énergie par fréquence du signal (théorème de Parseval).

$$PX(f) = |X(f)|^2 \quad \text{(II.5)}$$

Où :

– PX(f) : représente la DSP du signal.

L'équation (II.4) devient alors :

$$PX(f) = PS(f) + PB(f) \quad \text{(II.6)}$$

Il existe plusieurs façons pour estimer cette DSP au niveau du bruit (PB(f)). Le plus simple est de prendre une trame d'acquisition longue de 200ms pour connaitre le spectre moyen assimilé au bruit et des trames courtes pour le signal de parole d'une durée de 8 ms.

La comparaison ensuite de ces deux aspects permet d'estimer le bruit des trames à l'entrée du signal. Une fois le module du bruit considéré, il faut essayer de l'atténuer dans la trame correspondante, il existe plusieurs méthodes (classiquement utilisées) pour exécuter ce débruitage. Les trois principaux filtres en monovoie seront présentés dans les sous sections suivantes.

II.5 Les techniques d'atténuation spectrales

Ces techniques constituent une famille d'algorithmes de référence pour la réduction du bruit en parole continue, opérant dans le domaine fréquentiel au moyen de modifications spectrales. L'idée de base est d'atténuer plus ou moins fortement les composantes spectrales du signal dégradé en fonction de l'estimation du niveau du bruit. L'ensemble des caractéristiques de telles méthodes est géré, en premier temps, par la nature de la transformation elle-même qui permet une interprétation spectrale du signal analysé, et dans un deuxième temps, par une règle de suppression qui repose sur un mécanisme permettant de décider de l'atténuation à apporter à chaque canal fréquentiel qui s'effectue à partir d'une estimation de la densité spectrale du bruit et une estimation locale du signal.

II.5.1 Soustraction spectrale

La soustraction spectrale est le débruiteur le plus ancien. Elle a été introduite par Boll [13]. Comme son nom l'indique, elle effectue son travail dans le domaine spectral et a pour principe de soustraire le bruit estimé au signal. L'estimation du bruit se fait sur plusieurs trames d'acquisition (≃ 300ms).

$$PX'(f) = PS(f) + PB(f) - EPB(f) \quad \text{(II.7)}$$

Où :

– EPB(f) : représente l'estimation du bruit

– PX′(f) : représente la DSP après la soustraction spectrale

Principe

Il existe deux versions de base pour la soustraction spectrale se différenciant par l'amplitude ou la puissance.

$$|X'(f)| = |X(f)| - |EB(f)| \quad \text{(II.8)}$$

Dans ce cas, il s'agit de la « soustraction spectrale d'amplitude ». Mais le plus souvent comme indiqué dans le paragraphe précédent, la soustraction se fait au niveau de la puissance.

$$|X'(f)|^2 = |X(f)|^2 - |EB(f)|^2 \quad \text{(II.9)}$$

Le problème de ces deux équations II.8 et II.9, est que le second terme peut être négatif. On peut le rendre positif en changeant de signe ou bien en l'annulant comme dans l'équation II.10. C'est la première amélioration que l'on peut proposer.

$$|X'(f)|^2 = \begin{cases} |X(f)|^2 - |EB(f)|^2 \; si \; |X(f)|^2 > |EB(f)|^2 \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad sinon \end{cases} \quad \text{(II.10)}$$

Le passage dans le domaine temporel se fait par une IFFT. La phase du signal d'entrée est gardée, une estimation du bruit de la phase s'avère être une tâche compliquée.

II.5.1.1 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.

$$|X'(f)| = G(f).|X(f)| \; 0 \leq G(f) \leq 1 \tag{II.11}$$

Et pour la soustraction spectrale de puissance, on obtient :

$$G(f) = \begin{cases} \sqrt{1 - \frac{|EB(f)|^2}{|X(f)|^2}} & si \; |X(f)|^2 > |EB(f)|^2 \\ 0 & sinon \end{cases} \tag{II.12}$$

Dans la littérature, la soustraction spectrale est très utilisée car elle est très simple à mettre en œuvre. Néanmoins, elle génère des artéfacts après la réduction du bruit ainsi qu'une distorsion du signal et le bruit musical [12].

II.5.1.2 Amélioration de la soustraction spectrale par Berouti

Berouti [14], 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 ß |EB(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 :

$$|X'(f)|^2 = \begin{cases} |X(f)|^2 - \alpha|EB(f)|^2 \; si \; (|X(f)|^2 - \alpha|EB(f)|^2) > \beta|EB(f)|^2 \\ \beta|EB(f)|^2 \qquad\qquad sinon \end{cases} \tag{II.13}$$

L'introduction de la quantité $\beta|EB(f)|^2$, au lieu d'un zéro (comme dans l'équation II.12), permet d'ajouter un bruit large bande qui, selon Berouti, va masquer les composantes tonales voisines de même amplitude (ou d'amplitudes comparables). Les paramètres α et β ont pour objectif de trouver un compromis entre la quantité du bruit résiduel, celle du bruit musical et finalement la distorsion du signal. Ajuster convenablement ces deux paramètres est une tâche qui influe beaucoup sur la qualité des résultats.

Les expériences [13] ont montré que le paramètre α dépend du SNR segmental, notéseg_{SNR}, selon l'équation:

$$\alpha = \alpha_0 - \frac{segSNR}{s} \quad \text{(II.14)}$$

Où :

- α : représente le facteur de soustraction
- α_0 : réprésente la valeur désirée du α dans SNR = 0 dB
- 1/s : représente la pente de la droite.
- SNR : est le rapport d'environ segmentaire signal sur bruit.

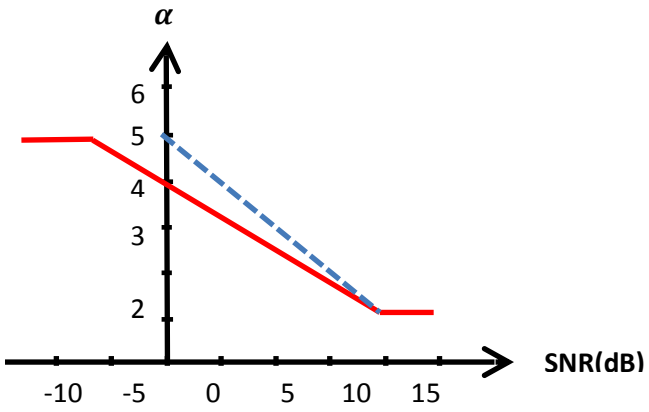


Figure II.2 : Valeur des facteurs de soustraction α par rapport au SNR.

Dans la Figure II.2 par exemple pour (α_0= 4, s = 20 / 3), Pour une plage de seg_{SNR} variant de −5 dB à 5 dB, $\alpha_0$est compris entre 3 et 6. Le paramètre β est très sensible au niveau de bruit. Pour de très hauts niveaux de bruit (−5 dB), le paramètre β doit être compris dans l'intervalle $0.02 \leq \beta \leq 0.06$. Pour des niveaux bas du bruit (0 dB ou 5 dB), il vaut mieux choisir β tel que $0.005 \leq \beta \leq 0.02$.

II.5.1.3 Amélioration de la soustraction spectrale par Boll

Boll [13] 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.

1) La moyenne d'amplitude

$$X'(f) = [|X(f)| - \mu(f)]\, e^{i\theta\acute{X}(f)} \quad \text{(II.15)}$$

Où : μ(f) = EB(f) est la moyenne du bruit calculé pendant les silences. Le filtre est alors de la forme suivante :

$$H(f) = 1 - \frac{\mu(f)}{|X(f)|} \quad \text{(II.16)}$$

On en déduit l'erreur de l'estimation :

$$e(f) = X(f) - \acute{X}(f) = B(f) - \mu(f)\, e^{i\theta \acute{X}(f)} \quad \text{(II.17)}$$

Ou e(f) dépend à la fois de B(f) et de la moyenne de μ(f). Il faut réduire au minimum e(f) pour avoir la meilleure estimation du bruit.

Pour cela, il faut que : B(f) ≃ μ(f). L'introduction de la moyenne de l'amplitude dans le signal bruité$\overline{|X(f)|}=\frac{1}{N}\sum_{f=0}^{N-1}$X(f) permet d'obtenir une nouvelle équation(II.18).

$$X'(f) = \left[\overline{|X(f)|} - \mu(f)\right] e^{i\theta \acute{X}(f)} \quad \text{(II.18)}$$

L'erreur devient alors :

$$e(f) = X(f) - \acute{X}(f) \simeq \mu(f) - \overline{|B(f)|} \quad \text{(II.19)}$$

Où,$\overline{|B(f)|}=\frac{1}{N}\sum_{f=0}^{N-1}$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.

2) La réduction du bruit résiduel

La méthode consiste à remplacer les valeurs négative dans X'(f) par des zéros. C'est une méthode de rectification demi-onde. La nouvelle expression de H(f) devient :

$$H^R(f) = \frac{H(f) + |H(f)|}{2} \quad \text{(II.20)}$$

Et l'équation devient :

$$X'(f) = H^R(f)X(f) \quad \text{(II.21)}$$

3) La réduction du bruit résiduel

On peut résumer cette partie par l'équation suivante :

$$|X'_k(f)| = \begin{cases} |X'_k(f)| & si\ X'(f) \geq max|B^R(f)| \\ min|X'_j(f)| & ,j = k-1, k, k+1 \quad si\ X'(f) < max|B^R(f)| \end{cases} \quad \text{(II.22)}$$

Où:

– k : représente le numéro de la trame

– max$|B^R(f)|$: représente le maximum du bruit résiduel mesuré pendant les silences.

La réduction du bruit résiduel s'effectue ainsi en sélectionnant le minimum de l'amplitude estimée durant trois trames j.

4) Atténuation du signal pendant les périodes de silences

Boll propose de manière empirique un seuil de détection d'activité vocale.

$$\tau = 20.\log_{10}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{X'(f)}{\mu(f)}\right| \qquad \text{(II.23)}$$

Si $\tau < -12$dB, Boll considéré qu'il n'y a pas d'activité vocale. Dans ce cas, au lieu d'enlever le signal totalement, il propose de l'atténuer.

II.5.1.4 Amélioration de la soustraction spectrale par Virag

Virag [11] a mixé les deux précédentes méthodes : Boll et Berouti avec la généralisation de Lim [15] 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 $G_k(f)$ est donné par :

$$G_k(f) = \begin{cases} \left(1 - \left(\frac{B(f)}{X(f)}\right)^{\eta 1}\right)^{\eta 2} & si\ \left(\frac{B(f)}{X(f)}\right)^{\eta 1} < \frac{1}{\alpha+\beta} \\ \left(\beta\left(\left|\frac{B(f)}{X(f)}\right|\right)^{\eta 1}\right)^{\eta 2} & sinon \end{cases} \qquad \text{(II.24)}$$

Où:

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

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

On peut remarquer que le choix de α et ß est beaucoup plus critique que η1,2. Si η1 = η2 = 1, il s'agit d'une soustraction spectrale d'amplitude. Si η1 = η2 = 0.5, il s'agit d'une soustraction spectrale de puissance et si η1 = 2, η2 = 1, il s'agit d'un filtrage de Wiener.

II.5.2 La soustraction spectrale au sens d'Ephraim et Malah

Cette technique [16], proposée par Ephraim et Malah, est basée sur une estimation bayésienne du spectre d'amplitude du signal débruité au sens des moindres carrés, elle évalue l'atténuation spectrale courante en se basant sur la précédente. L'estimation du spectre d'amplitude du signal débruité est donnée par la relation suivante :

$$|X(f)| = G(f)|Y(f)| \quad \text{(II.25)}$$

Avec G(f) est le gain donné par :

$$G(f) = \frac{\sqrt{\pi}}{2}\sqrt{\left(\frac{1}{1+R_{post}}\right)}\left(\frac{R_{priori}}{1+R_{priori}}\right)M\left[\left(1+R_{post}\right)\left(\frac{R_{priori}}{1+R_{priori}}\right)\right] \quad \text{(II.26)}$$

Le niveau relatif local à posteriori définit la valeur mesurée à la $m^{iéme}$ fenêtre :

$$R_{post}(f, m-1) = \frac{|Y(f,m)|^2}{\hat{B}(f,m)} \quad \text{(II.27)}$$

Le rapport signal sur bruit à priori est donné par :

$$R_{priori\,(f,m-1)} = (1-\boldsymbol{\alpha})\left[R_{post}(f, m-1) - 1\right] + \boldsymbol{\alpha}\frac{|Y(f,m-1)|^2}{\hat{B}(f,m)} \quad \text{(II.28)}$$

Avec α un paramètre compris entre 0 et 1, $Y|(f, m-1)|^2$est le spectre de puissance de la fenêtre précédente et R_{post}(f, m) est le niveau relatif local à posteriori.

La règle de suppression d'Ephraim et Malah fournit un moyen simple de contrôle du niveau du bruit résiduel, malgré que pour les signaux fortement bruités cette réduction est limitée par la distorsion du signal.

II.5.3 Filtrage de Wiener

Le filtrage de Wiener a été introduit à la fin des années 60 [17] pour essayer d'améliorer la qualité de la trace recueillie dans les potentiels évoqués. Le problème du type de filtrage proposé est qu'il n'est pas applicable sur une moyenne d'acquisition mais pour chaque trace. Doyle [18], 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 [18].

Tout comme la soustraction spectrale, les calculs sont effectués dans le domaine fréquentiel. La DSP est calculée comme précédemment dans l'équation (II.5). Le système peut être isolé entre le bruit et le signal utile. Ce qui revient à dire que le rapport $\frac{PS(f)}{PB(f)}$

doit être maximisé pour obtenir le signal utile. Le filtre de Wiener est défini de la façon suivante.

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.

$$W(f) = \frac{P_S(f)}{P_S(f)+P_B(f)} = \frac{1}{1+\frac{P_B(f)}{P_S(f)}} \quad \text{(II.29)}$$

En utilisant les définitions du RSBprio, le filtre de Wiener peut s'exprimer sous cette forme :

$$W(f)\frac{RSB_{prio}(f)}{1+RSB_{prio}(f)} \quad \text{(II.30)}$$

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 trame 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.

II.5.4 Réduction du bruit musical

Le bruit musical est par définition un bruit résiduel perpétuellement gênant suite à un débruitage d'un signal par des algorithmes. Il est généralement induit par ce sont des réducteurs de bruit à court terme tel que les deux algorithmes cités précédemment, la soustraction spectrale et le filtrage de Wiener. Le spectre de ce bruit est tonal d'où le caractère musical.

Sa valeur moyenne est plus faible que le bruit du signal d'entrée mais la dispersion fréquentielle est beaucoup plus grande. Le bruit est étalé sur les différentes bandes de fréquences.

D'un point de vu perceptif, le bruit musical est beaucoup plus gênant que le bruit de base [19]. Les principales raisons de l'apparition de ce type de bruit sont :

- Le traitement non linéaire des composantes négatives du signal bruité.
- L'évaluation non précise de la densité spectrale de bruit.

- L'estimation basée sur les spectrogrammes.
- La variabilité de la fonction de gain appliquée au signal bruité.
- La variance des estimateurs locaux de la DSP du signal.

Estimation et réduction du bruit

Si l'on prend un problème de débruitage classique linéaire ou le but est de trouver H(f) qui est l'estimateur, l'erreur dûe à ce filtrage est :

$$e_k(f) = S_k(f) - \tilde{S}_k(f)$$
$$= (H_k(f) - 1).S_k(f) + H_k(f).B_k(f) \quad \text{(II.31)}$$

Où :

- $(H_k(f) - 1).S_k(f)$: représente la distorsion du signal.
- $H_k(f).B_k(f)$: représente le bruit résiduel qui contient le bruit musical.
- k : représente le numéro de la trame.
- $\tilde{S}_k(f)$: représente la fonction estimée.

Dans le cas ou $0 \leq H(f) \leq 1$, le bruit musical est très difficile à réduire sans augmenter et apporter de la distorsion au signal débruité. Comme dans beaucoup de cas, un compromis entre la distorsion et le bruit musical doit être trouvé pour que l'intelligibilité de la parole soit bonne et que le bruit musical soit moins gênant.

II.5.5 Détection de la parole par le TOP

Principe

Le but de la détection d'activité de parole par le traitement optimisé de la parole (TOP) est que pendant les périodes de silence, il n'y a aucune activité et pendant les périodes de parole, le signal est transmis. Sohn [20], propose une méthode statistique. La DSP du bruit est considéré constante pendant tout le traitement jusqu'à une nouvelle période de silence.

Cette hypothèse ne fonctionne qu'avec des trames à court terme. Le TOP peut être amélioré par la méthode de Martin [21], qui n'est autre qu'une soustraction spectrale avec un suivi du minimum statistique de la DSP du bruit. Il est basé sur chaque composante spectrale supposé comme du bruit.

II.5.6 MMSE et MMSE-STSA (Minimum Mean Square Error-Short-Term Spectral Amplitude)

Principe

La particularité de cette règle provient du fait que la valeur de l'atténuation spectrale dépend essentiellement des valeurs du spectre à court terme mesurées dans les trames précédant la trame courante.

La règle de suppression d'Ephraim et Malah [22] est fondée sur une estimation bayésienne du spectre à court terme dans le sens des moindres carrés, d'où l'appellation d'estimateur de l'amplitude spectrale à court terme au sens de l'erreur quadratique moyenne MMSE-STSA (Minimum Mean Square Error-Short-Terme Spectral Amplitude). Elle est une des méthodes les plus populaires donnant des résultats satisfaisants aussi bien du point de vue réduction de bruit que vis-à-vis du bruit musical.

La fonction du gain de cette règle (tel que$\hat{S}(f) = G(f) \cdot Y(f)$), dans la trame k et à la fréquence f, est donnée par :

$$G_k(f) = \frac{\sqrt{\pi}}{2}\sqrt{\frac{1}{X_k(f)}\frac{\hat{\xi}_k(f)}{1+\hat{\xi}_k(f)}}\; V\left[X_k(f)\left(\frac{\hat{\xi}_k(f)}{1+\hat{\xi}_k(f)}\right)\right] \qquad \text{(II.32)}$$

Où $\xi_k(f)$ est l'estimée du rapport signal à bruit a priori dans la trame k donnée par l'Eq. (II.34) et V est une fonction définie par :

$$V(x) = exp\left(\frac{-x}{2}\right)\left[(1+x)I_0\left(\frac{x}{2}\right) + xI_1\left(\frac{x}{2}\right)\right] \qquad \text{(II.33)}$$

Où I_0 et I_1 sont respectivement les fonctions de Bessel modifiées d'ordre 0 et 1 et l'expression de $\xi_k(f)$ est la suivante :

$$\hat{\xi}_k(f) = (1-\alpha)h\left(\underbrace{X_k(f)-1}_{\text{RSB}_{\text{instantané}}}\right) + \alpha\underbrace{|G_{k-1}(f)Y_{k-1}(f)|^2/\gamma_k(f)}_{\text{RSB}_{\text{passé}}} \qquad \text{(II.34)}$$

Dans l'expression (II.34), $X_k(f)$est l'estimée de $E\,[|Y_k(f)|^2]/\gamma_k(f)$, le Rapport Signal sur Bruit a posteriori. Afin d'éviter d'éventuelles valeurs négatives de $X_k(f)$, la

fonction h permet de considérer seulement la partie positive: $h(x) = x$, si $x \geq 0$ et $h(x) = 0$ ailleurs.

Cet estimateur (II.34) est récursif et s'avère performant du fait qu'il apporte des améliorations sur la qualité du signal débruité. Il permet de réduire le bruit musical et les distorsions du signal de par ses propriétés de lissage fréquentiel. Cet estimateur est connu sous le nom de Directed-Decision.

On s'aperçoit, à partir de (II.34) et (II.32), que l'estimateur $G_k(f)$ dépend essentiellement des valeurs du spectre à court terme mesurées dans les trames précédentes. Effectivement, l'estimée $\xi_k(f)$ prend en compte la trame bruitée courante avec un poids de $(1 - \alpha)$ et la trame débruité précédente avec un poids de α (sachant que $0 \leq \alpha \leq 1$).

Dans [23], une analyse asymptotique du gain $G_k(f)$ en fonction de $\xi_k(f)$ montre que, pour des valeurs de $\xi_k(f)$ très petites, on applique une forte atténuation. Dans ce cas de figure, le comportement de $G_k(f)$ en fonction de $(X_k(f) - 1)$, en fixant la valeur de $\xi_k(f)$, montre que pour des valeurs petites de $\xi_k(f)$ l'influence de $(X_k(f) - 1)$ devient importante. Cette influence est même contre intuitive puisque des fortes atténuations sont appliquées quand $(X_k(f) - 1)$ est grand, alors que la logique veut plutôt qu'on débruité plus quand le rapport signal à bruit est faible. Dans [23], l'auteur indique que cette contre intuition est utile pour le traitement de

Segments de parole de faible énergie. Dans une comparaison entre le filtre de Wiener et l'estimateur MMSE-STSA [13], les auteurs constatent que :

- L'erreur quadratique moyenne de l'estimateur MMSE-STSA ne peut pas dépasser 1 alors que pour le filtre de Wiener, elle peut même atteindre la valeur $\frac{1}{1-\frac{\pi}{4}}$
- L'estimateur MMSE-STSA et le filtre de Wiener sont peu sensibles à des petites variations dans l'estimation de $\xi_k(f)$. Ils tolérent en l'occurrence une sur-estimation de cette grandeur plutôt qu'une sous-estimation. Une sur-estimation de $\xi_k(f)$ implique même une atténuation de l'erreur quadratique moyenne dans le cas du filtre de Wiener. Ceci est dû au fait que le filtre de Wiener n'est pas optimal au sens du MMSE quand il emploie l'expression (II.34).
- En utilisant l'expression (II.34) avec une valeur de α égale à 0.98, le filtre de Wiener introduit moins de bruit résiduel que l'estimateur MMSE-STSA, sachant que le bruit résiduel est de nature moins colorée et moins gênant que

le bruit musical pour les deux estimateurs. L'estimateur MMSE-STSA introduit moins de distorsion que le filtre Wiener.

La réduction du bruit musical est fortement liée à l'expression du RSB a priori (II.34), qui constitue d'ailleurs l'originalité du travail présenté dans [13]. Analysant cette expression :

- Si $X_k(f) - 1 \leq 0$, alors $\xi_k(f)$ correspond à une version lissée du rapport signal à bruit a posteriori. Ceci implique que la variance du RSB a priori est plus petite que celle du RSB a posteriori. Puisque $G_k(f)$ dépend essentiellement de $\xi_k(f)$, l'atténuation appliquée au signal bruité ne changera pas brusquement d'une trame à l'autre, d'où la réduction de l'apparition du bruit musical.
- Si $X_k(f) - 1 > 0$, alors $\xi_k(f)$ est une version lissée et retardée d'une trame du RSB a posteriori.
- Quand α diminue, les distorsions diminuent et le bruit musical augmente et vice-versa. Sachant que si α diminue, le poids de h $(X_k(f) - 1)$ augmente, on peut donc conclure que le bruit musical est très sensible à ce terme.

L'inconvénient de cette expression est la sur-atténuation au moment des transitoires dans le cas de l'apparition d'une composante de parole à faible niveau [23].

II.5.7 Méthodes à sous-espace signal

Principe

L'une des approches de débruitage de la parole qui a suscité beaucoup d'intérêt est le filtrage à sous-espace signal. Dans cette approche, on développe un estimateur linéaire non paramétrique, du signal de parole propre, obtenu par décomposition du signal observé en deux sous-espaces orthogonaux : **le sous-espace signal** et **le sous-espace bruit**.

La décomposition est achevée soit par valeurs singulières SVD ou par valeurs propres EVD. Le principe des méthodes à sous-espace signal, décrit dans cette section, se fera premièrement en supposant que le bruit est additif, blanc et décorrélé de la parole, deuxièmement, en raisonnant par rapport à une décomposition en valeurs propres. Pour plus de détails sur l'utilisation des valeurs singulières dans ce genre d'application le lecteur peut se référer à [24]. La réduction du bruit par cette approche est obtenue par

annulation des composantes du sous-espace bruit en premier lieu et en supprimant la contribution du bruit dans le sous-espace signal en second (figure II.3).

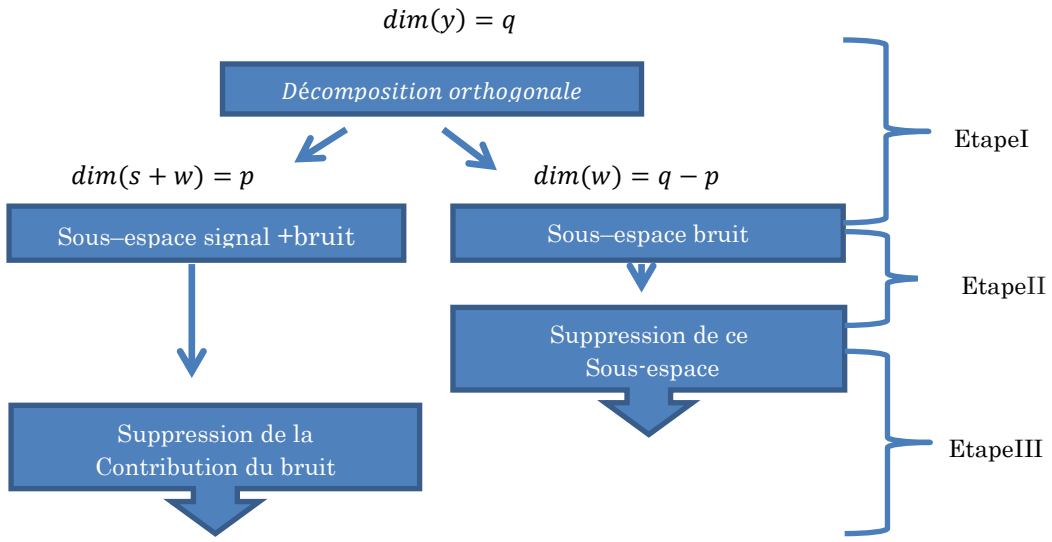


Figure II.3: Débruitage à sous-espace signal

La première étape est simple dans le cas où le bruit est blanc (on verra par la suite ce qui se passe dans le cas où le bruit est coloré). La deuxième étape est indispensable contrairement à la troisième qui est souvent omise pour éviter les distorsions puisque, dans l'espace signal, le bruit et le signal interférent.

Comment peut-on décomposer un vecteur R^nde en deux composantes orthogonales Soient **y**, **s** et **b** les vecteurs correspondant respectivement au signal bruité, au signal propre et au bruit, tels que :

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_q \end{pmatrix}, \boldsymbol{S} = \begin{pmatrix} s_1 \\ s_2 \\ \cdot \\ \cdot \\ s_q \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ b_q \end{pmatrix}$$

On a :

$$\boldsymbol{y} = \boldsymbol{s} + \boldsymbol{b} \qquad \text{(II.35)}$$

SoitR_y, $R_s$etR_b, les matrices d'autocorrélation de y, s et b. Sous l'hypothèse que la parole est décorrélé du bruit, on écrit :

$$R_y = R_s + R_b \qquad \text{(II.36)}$$

La **décomposition en valeurs propres** EVD de ces matrices d'autocorrélation donne lieu aux équations suivantes :

$$R_s = U\Lambda_s U^T \quad \text{(II.37)}$$

$$R_s = U(\delta^2 I)U^T \quad \text{(II.38)}$$

$$R_b = U\Lambda(\Lambda_s + \delta^2 I)U^T \quad \text{(II.39)}$$

Λ_sEst la matrice diagonale contenant les valeurs propres λ_s deR_s, U est une matrice ortho normale en colonnes ; δ^2 est la variance du bruit et I est la matrice identité.

D'après les équations (II.37), (II.38), et (II.39),on remarque que les vecteurs propres du bruit sont identiques aux vecteurs propres du signal de parole grâce à l'hypothèse de bruit blanc. Ces vecteurs propres peuvent donc être calculés à partir de R_y (c-`a-d à partir du signal observé).

En supposant le sous-espace signal de dimension p avec p <q, la matrice d'autocorrélation$R_y$possède ainsi p valeurs propres λ_s non nulles siλ_s>δ^2. Dans ce cas, le bruit peut être séparé de la parole et R_y peut-être réécrite en supposant que les vecteurs propres sont en ordre décroissant :

$$R_y = [U_p U_{p-q}]\left[\begin{pmatrix}\Lambda_s & 0\\ 0 & 0\end{pmatrix} + \delta^2\begin{pmatrix}I_p & 0\\ 0 & I_{p-q}\end{pmatrix}\right][U_p U_{p-q}]^T \quad \text{(II.40)}$$

Indifféremment du critère d'optimisation, le débruitage de la parole est obtenu :

- En annulant les composantes du signal bruité dans le sous-espace bruit (de dimension q − p).
- En atténuant les valeurs propres du sous-espace signal (de dimension p).

Mathématiquement, le débruitage se ramène à un filtrage **F** tel que $\hat{\boldsymbol{s}}$ **=Fy**, ou $F = U_p G_p U_P^T$, avec G_p une matrice (p × p) diagonale contenant les facteurs de pondération g_i appliqués aux p premières valeurs propres de R_y, tel que :

$$F = \sum_{i=1}^{p} g_i u_i u_i^T \quad \text{(II.41)}$$

Est une sommation de filtrages intermédiaires appliqués sur chaque vecteur propre oug_i est le $i^{éme}$élément diagonal de G. La suppression de la contribution du bruit dans le sous-espace signal se fait selon un critère dont l'objectif est de trouver les éléments de la matrice G.

Dans la littérature, plusieurs critères ont fait l'objet de travaux. Ils sont de trois classes temporels, fréquentiels et perceptuels [25], [26],[27], [28]. Ils sont tous basés sur la minimisation de la distorsion du signal en contraignant le bruit résiduel à être au-

dessous d'un certain seuil (la courbe de masquage dans le cas des estimateurs perceptuels).

1. Estimateur dans le domaine temporel

Le critère d'estimation dans le domaine temporel s'écrit sous forme d'un problème d'optimisation (minimisation de l'énergie de la distorsion du signal) sous contrainte (seuil maximal du bruit résiduel).

$$\min_{G} \varepsilon_s^2$$

$$\text{Sous contrainte que} \varepsilon_b^2 \leq q\delta^2$$

Oùε_s^2désigne la distorsion du signal, δ^2 la variance du bruit et q contrôle le niveau admissible du bruit résiduel (0 < q < 1). Ce problème est résolu par la méthode du Lagrangien en résolvant l'équation:

$$\frac{dL(G,\mu)}{dG} = \frac{d\left(\varepsilon_s^2 + \mu(\varepsilon_b^2 - q\delta^2)\right)}{dG} = 0 \qquad \text{(II.42)}$$

Où μ est le multiplicateur de Lagrange. Tout calcul fait, on aboutit à l'expression du filtre optimal G suivante :

$$G_{opt} = \frac{R_s}{R_s + \mu R_b} = \frac{R_s}{R_s + \mu\delta_b^2 I} \qquad \text{(II.43)}$$

En utilisant la décomposition en valeurs propres des matrices d'autocorrélation $R_s \; et \; R_b$ le filtre G (Eq II.43) peut être simplifié par :

$$G_{opt} = U \begin{pmatrix} G_\mu & 0 \\ 0 & 0 \end{pmatrix} U^T \qquad \text{(II.44)}$$

Où :

$$G_\mu = \Lambda_s(\Lambda_s + \mu\delta_b^2 I)^{-1} \qquad \text{(II.45)}$$

2. Estimateur dans le domaine spectral

Cet estimateur est une généralisation de celui du domaine temporel, de telle façon à minimiser l'énergie de la distorsion du signal en gardant un certain niveau du bruit résiduel cette fois-ci pour chaque composante spectrale. Soit $U_k^T \in_b^2$ la$k^{\frac{1}{2}e}$composante spectrale du bruit résiduel.

L'estimateur H qu'on cherche peut accepter cette fois-ci des valeurs d'entrée qui sont complexes. Le critère, dans le domaine spectral, s'écrit ainsi :

$$\min_{H} \epsilon_{\mathbf{b}}^2$$

$$\text{Sous contrainte que } \begin{cases} E\left|U_k^T \epsilon_{\mathbf{b}}^2\right| \le \alpha_k \delta^2 & k = 1,2,\dots\dots\dots p \\ E\left|U_k^T \epsilon_{\mathbf{b}}^2\right| = 0 & k = p+1,\dots\dots q \end{cases} \quad \text{(II.46)}$$

L'énergie du signal dans le sous-espace bruit est nulle pour tout composante spectrale k, tel que p+1 < k < q. La solution de ce problème est aussi donnée par le multiplicateur de Lagrange qui débouche sur l'estimateur optimal H satisfaisant l'équation suivante :

$$HR_s + \delta^2\left(U\Lambda_\mu U^T\right)H - R_s = 0 \quad \text{(II.47)}$$

Tel que Λ_μ = diag $(\mu_1, \mu_2, \dots \mu_p)$ est la matrice diagonale des multiplicateurs de Lagrange.En utilisant la décomposition en valeurs propres de R_s (Eq II.37) et en l'injectant dans (II.47), on obtient :

$$(I - U^T HU)\Lambda - \delta_{\boldsymbol{w}}^2 \Lambda_\mu U^T HU = 0 \quad \text{(II.48)}$$

En posant Q = $U^T HU$ et en supposant que cette matrice est diagonale, le filtre optimal à la même expression que celui du domaine temporel H = UQU^T. Les éléments diagonaux de Q ont alors la forme :

$$q_k = \begin{cases} \frac{\lambda_y(k)}{\lambda_s(k) + \delta_{\boldsymbol{b}}^2 \mu_k} & k = 1, \dots \dots p \\ 0 & k = p+q, \dots q \end{cases} \quad \text{(II.49)}$$

L'hypothèse de départ pour le développement des méthodes à sous-espace signal est de supposer que le bruit est blanc. Dans ce cas, la matrice de variance du bruit est diagonale de forme $\delta_{\boldsymbol{b}}^2$I et les vecteurs propres du signal bruité sont identiques à ceux du signal propre et du bruit. La relation reliant ensuite les valeurs propres de ces signaux est :

$$\begin{aligned} \Lambda_y(k) &= \Lambda_{\boldsymbol{s}}(k) + \Lambda_{\boldsymbol{b}}(k) \quad k = 1,2,\dots\dots q \\ \Lambda_y(k) &= \Lambda_{\boldsymbol{s}}(k) + \delta_{\boldsymbol{b}}(k) \quad k = 1,2,\dots\dots q \end{aligned} \quad \text{(II.50)}$$

II.5.8 Le phénomène MAN (Maskee to Audible Noise)

Le phénomène de masquage est inhérent à notre système d'audition. Il est fortement dépendant de la nature du son et ainsi du signal. Lors du débruitage

perceptuel, on conçoit les filtres en se basant sur la courbe de masquage du signal propre dont on ne dispose pas en réalité.

Mais, le vrai problème n'est pas là. Lorsqu'on évalue la qualité du signal débruité, que ce soit par des critères objectifs ou subjectifs, ce qui compte en dernier c'est le signal débruité lui même.

Du point de vue perceptif, ce signal possède sa propre courbe de masquage. En effet, quand on débruité le signal, on atténue forcément le signal de parole. L'atténuation, certes, dépend du gain du filtre linéaire, mais elle reste non négligeable en général. Qui dit atténuation du signal de parole dit atténuation de sa courbe de masquage (voir figure II.6)

Le fait de percevoir le bruit résiduel (bruit musical en particulier), après filtrage, prouve que le signal débruité n'a pas pu le masquer. Ce bruit est donc situé au-dessus de la courbe de masquage du signal débruité.

Par cette analyse, en mettant en évidence le problème d'atténuation de la courbe de masquage, nous introduisons un phénomène qui est une conséquence immédiate de cette atténuation. Il s'agit du bruit qui, masqué au départ, peut ne plus l'êtreaprès débruitage, d'une part, parce qu'il est situé au-dessus du seuil d'audition absolu et d'autre part parce que le niveau d'atténuation de la courbe de masquage lui permet de se dégager du spectre atténué et de devenir audible, engendrant ainsi une partie du bruit musical. Ce phénomène, nous l'avons baptisé, dans cette thèse, le phénomène MAN (Maskee to Audible Noise) [5].

1. Illustration du phénomène MAN

Afin d'illustrer expérimentalement le phénomène MAN (figure II.4), on suppose qu'un masquant est présent à la fréquence $f_0$de sorte que la courbe de masquage résultante est au-dessus du seuil d'audition absolu au voisinage de cette fréquence.

Elle masque ainsi le bruit adjacent et moins puissant qui est présent à la fréquencef_2, tandis que la deuxième composante de bruit à la fréquence f_1 reste audible du fait que son niveau acoustique est suffisamment élevé pour être perçue en présence du signal masquant.

Supposant maintenant qu'on procède à un filtrage perceptuel classique consistant à traiter uniquement le bruit audible, ce filtrage va réduire le bruit à la fréquencef_1, cequi va entraîner une atténuation du signal à cette fréquence et ainsi une atténuation

desa courbe de masquage dans ce voisinage. Il s'ensuit que le bruit masquéà la fréquence$f_2$devient audible.

Ce phénomène peut se produire quand l'énergie d'une composantede bruit masquée est comprise entre la courbe de masquage et le seuil d'auditionabsoluT^*, à condition que son masquant soit atténué. C'est donc un phénomène qui [5].

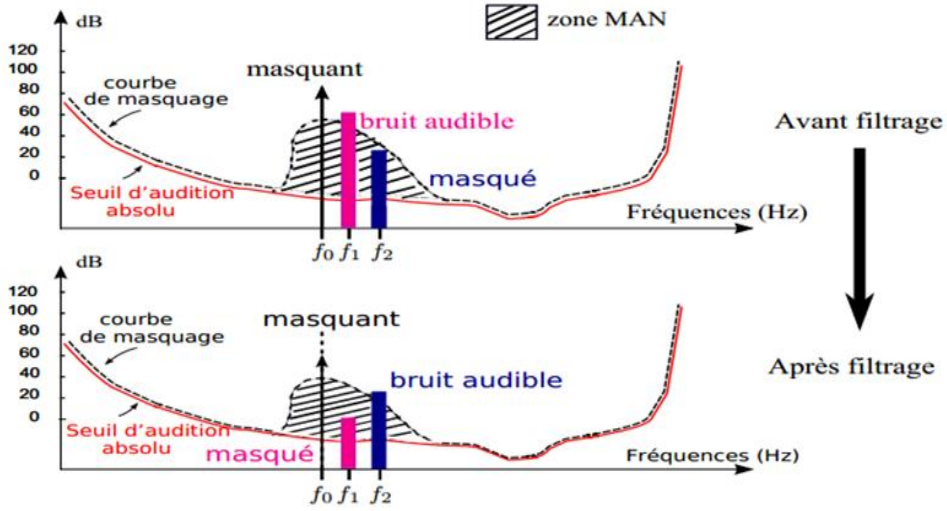


Figure II.4: Maskee to audible noise phénomène on description

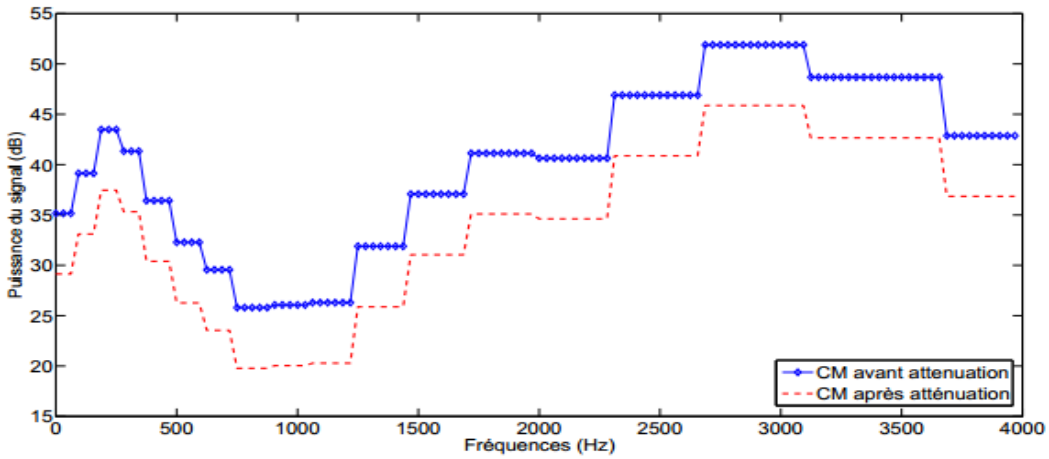


Figure II.5: Atténuation spectrale du signal implique une atténuation de sa courl de masquage.

Peut se produire fréquemment et dont la conséquence immédiate est l'apparition de tonales isolées et dispersées accentuant la perception du bruit musical. L'autre effet en découlant est le masquage de certaines composantes du signal de parole dont

l'énergie est faible à cause de certaines de ces tonales de bruit plus puissantes, ce qui peut induire des distorsions du signal.

Pour illustrer l'effet d'une atténuation spectrale sur la courbe de masquage d'un signal, la figure II.5 présente une comparaison entre la courbe de masquage (CM) d'un signal avant atténuation et celle du même signal, mais après avoir subi une atténuation,dans le domaine fréquentiel, par un facteur ß = 1/2 pour toutes les fréquences (juste à titre d'exemple).

Cette deuxième courbe est notée « CM après atténuation » dans la figure II.5. Sur cette figure, on constate que la deuxième courbe de masquage est une translation de la première avec un facteur de −6.02 dB (ce qui est normal puisque $20\,log_{10}\left(\frac{X}{2}\right) = -20\,log_{10}\,2 + 20\,log_{10}\,X = -6 + 20\,log_{10}\,X$)

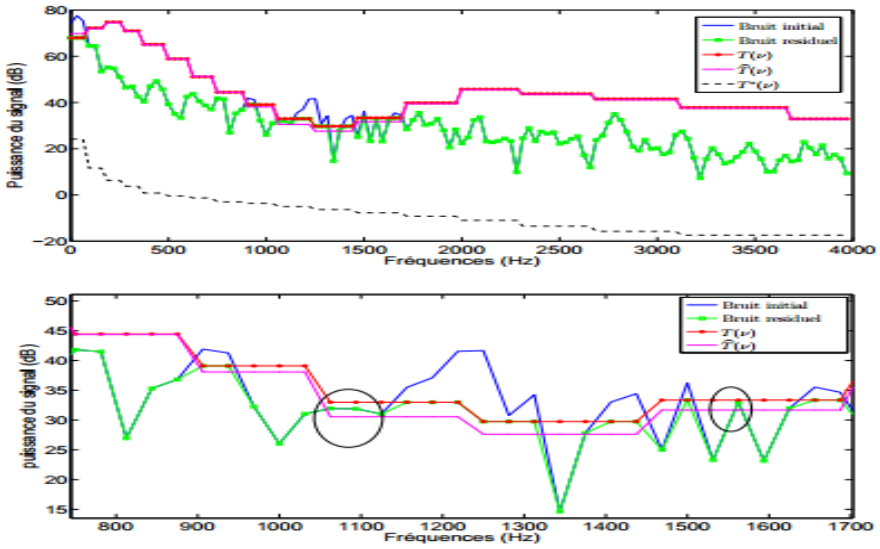


Figure II.6: Apparition du phénomène MAN après filtrage du bruit audible uniquement

On s'attendait peut être à une translation de −3 dB mais cela prouve encore que les transformations qui donnent lieu à la courbe de maquage ne sont pas linéaires. Trouver donc une expression ou une méthode adéquate pour déduire le niveau de la nouvelle courbe de masquage en connaissant celui de l'atténuation du signal n'est, a priori, pas une tâche facile.

Dans l'exemple de la figure II.6, on suppose connaître la courbe de masquage T(v) du signal propre (un échantillon de la base Timis) et la densité spectrale du bruit γ(v)

(un bruit de voiture de la base Noise à 5 dB) et $T^*(v)$ est le seuil d'audition absolu. On effectue le débruitage du signal bruité par un filtrage perceptuel qui traite uniquement le bruit audible.

Sur cette figure, en analysant le bruit résiduel, on constate que certaines composantes du bruit additif, qui n'étaient pas audibles au départ, se retrouvent maintenant au-dessus de la courbe de masquage $\hat{T}(v)$ du signal débruité. Elles seront ainsi audibles après débruitage. Si ce phénomène se produit répétitivement dans chaque trame, plusieurs tonales de bruit, éparpillées en fréquences, vont ainsi apparaître et contribuerontà la perception du bruit musical [5].

2. Double filtrage pour éviter le phénomène MAN

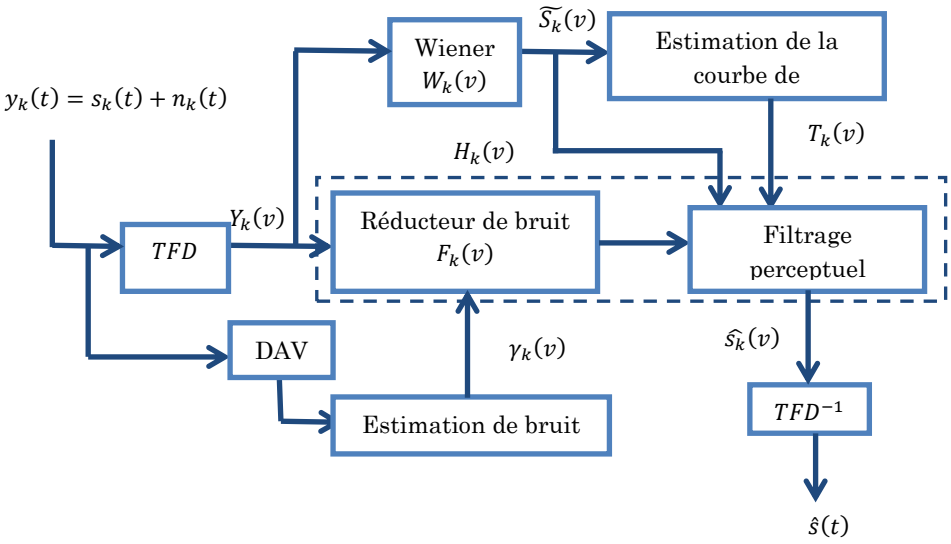


Figure II.7: Principe du double filtrage DF pour une trame k donnée.

Principe

Dans le but d'éviter l'apparition du phénomène MAN, pour les raisons citées précédemment, notre première suggestion [29] consiste à appliquer un double filtrage dont le synoptique est décrit par la figure II.7. Ce synoptique permet d'améliorer le réducteur de bruit F(v) grâce à une pondération perceptuelle à travers un second filtrage G(v).

La figure II.7 pourrait certainement être envisagée pour plusieurs types de réducteurs de bruit non perceptuels (Wiener, soustraction spectrale,...) suivis d'une pondérationde type perceptuel.

Dans notre cas, nous avons considéré le filtre de Wiener comme réducteur de bruit (F_k(v) = W_k(v)) de par ses performances reconnues. Nousavons opté pour l'adaptation du filtre perceptuel de l'équation au domaine deFourier, ce qui donne lieu à l'équation suivante :

$$G_k(v) = \frac{|\tilde{S}_k(v)|^2}{|\tilde{S}_k(v)|^2 + max(\gamma_k(v) - T_k(v), 0)} \quad \text{(II.51)}$$

Ou$|\tilde{S}_k(v)|$est l'amplitude du signal restitué à la sortie du filtrage de Wiener, $T_k(v)$ est la courbe de masquage estimée et $\gamma_k(v)$ est la densité spectrale de puissance du bruit.

L'intérêt de l'approche du double filtrage est d'atténuer d'abord toutes les composantes du bruit, même celles initialement inaudibles, par le biais du réducteur de bruit, d'appliquer ensuite un filtrage perceptuel qui agira en accentuant le débruitagedans les fréquences ou le bruit est perceptuelle ment significatif. En procédant ainsi, onlimite l'apparition du phénomène MAN. Le double filtrage DF résultant a donc pourexpression :

$$H_k^{DF}(v) = W_k(v) G_k(v) \quad \text{(II.52)}$$

II.6 Conclusion

Dans de ce chapitre, nous avons présenté l'ensemble des techniques de réduction de bruit monovoie les plus répandues dans la littérature. Les méthodes découlant de chaque technique ont chacune leur intérêt, et nous avant expliquée le principe des algorithmes de débruitage, chaque algorithme a un principe différent pour obtenir en sortie un signal contenant moins de bruit.

Chapitre 3

Techniques d'évaluation des algorithmes d'amélioration de signal parole

III.1 Introduction

L'évaluation subjective de la qualité de la parole est une étape indispensable dans tout processus de traitement, automatisé ou non. Elle permet de tenir compte du jugement humain à travers des essais d'écoute de laboratoire par plusieurs auditeurs.

Des méthodes statistiques sont ensuite mises en œuvre pour classer les différentes opinions avec un intervalle de confiance de largeur minimale. L'évaluation de la qualité subjective est couteuse en termes de temps et de ressources. Cette difficulté a donné lieu au développement d'autres métriques objectives de qualité qui, bien que moins précises, sont beaucoup plus pratiques et moins couteuses. La corrélation entre les mesures objectives et les mesures subjectives est utilisée comme un critère de performance de ces nouvelles métriques. Plus le critère objectif est corrélé avec les mesures subjectives, plus il constitue une bonne mesure pouvant, plus ou moins, remplacer le jugement humain.

Dans de ce chapitre, on étude les technique d'évaluation de bruit dans le parole et on fera aussi le point sur la caractéristique et les inconvénients des tests subjectifs et objectifs.

III.2 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 [31]. L'intelligibilité est donc une mesure objective définie par le nombre de mots prononcés correctement identifiés [32] 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) [33] , 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 [33], [34]. 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 [35]. Il existe cependant des moyens objectifs qui permettent d'estimer l'intelligibilité de la parole et qui sont largement utilisés dans la littérature, à savoir le

test STI (Speech Transmission Index) [36], le test SII (Speech Intelligibility Index) [37] et le test AI (Articulation Index) [38].

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 surs 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, MBSD, etc. Ces critères ont un caractère perceptuel justement parce qu'ils sont fondés sur des notions psycho acoustiques pour simuler notre perception vis-à-vis du signal de parole. Plus de détails sur ces différents critères feront partie de ce chapitre.

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 termes 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.

III.3 Critères objectifs

Les mesures objectives de qualité des signaux vocaux les plus communément utilisées sont citées et classées dans le tableau (III-01) :

Mesures dans le domaine temporel	Mesures dans le domaine fréquentiel	Mesures dans le domaine perceptuel
SNR segSNR	IS CD WSS LLR	BSD, MBSD PSQM PESQ

Tableau III-01 : Classification des critères d'évaluation objective les plus communément utilisés.

Les critères temporels et fréquentiels se basent essentiellement sur l'évaluation de la qualité en termes de comparaison de distorsion de formes entre signal de référence et signal débruité, sans tenir compte de l'aspect perceptif. Certes, c'est une condition nécessaire mais non suffisante dans la mesure où deux signaux pratiquement de même forme peuvent être perçus différemment [39], d'où l'intérêt d'introduire le facteur psychoacoustique pour tout système ayant pour objectif de conserver la qualité de la parole. Diverses mesures objectives perceptuelles sont élaborées conduisant à de bonnes corrélations avec la perception humaine. Elles sont essentiellement dédiées au codage de la parole, mais trouvent leur application en débruitage de la parole ([40], [41], [32],. . .). A part le fait qu'elles donnent une meilleure corrélation avec la qualité vocale, leur application en débruitage n'a pas été justifiée jusqu'à présent. En guise d'illustration, citons la mesure de la qualité de la parole perçue (PSQM) (Perceptual Speech Quality Measure) [42] et sa version améliorée PESQ (Perceptual Evaluation of Speech Quality) [43], le BSD (Bark Spectral Distortion) [39] et sa version améliorée, MBSD (Modified Bark Spectral Distortion) [44]. Dans la suite, nous donnons, à titre d'exemple, plus de détails sur ces différentes mesures. Il en existe évidemment d'autres, comme le WSS et le LLR qui sont bien décrits dans [24].

III.3.1 SNR et SNR segmental (segSNR)

Rapport signal sur bruit (SNR) est l'une des mesures les plus anciennes et largement utilisées objectives. Il est mathématiquement simple à calculer, mais nécessite à la fois déformée et non faussée échantillons (propre) de la parole. SNR peut être calculé comme suit :

$$SNR = 10\,log_{10}\frac{\sum_{n=1}^{N} S^2(n)}{\sum_{n=1}^{N}\left(S(n)-\hat{S}(n)\right)^2}\,[dB] \qquad \text{(III-01)}$$

Où S(n), $\hat{S}$(n), N sont respectivement le signal de référence, le signal débruité, et N le nombre d'échantillons.

Cette définition classique du SNR est connue pour ne pas être bien liés à la parole qualité pour une large gamme de distorsions. Ainsi, plusieurs variantes du SNR classique existent qui montrent une corrélation beaucoup plus élevé avec la qualité subjective.

On a observé que SNR classique ne correspond pas bien avec la qualité de la parole parce que même si la parole est pas un signal stationnaire, moyennes SNR le

rapport sur le signal entier. L'énergie de la parole fluctue au fil du temps, et ainsi de portions où la parole l'énergie est grande, et le bruit est relativement inaudible, ne doit pas être lavé par d'autres portions où l'énergie de la parole est faible et le bruit peuvent être entendus sur la parole. Ainsi SNR a été calculé en trames courtes, puis en moyenne. Cette mesure est appelée segmentaire SNR, et peut être définie comme suit:

$$\mathrm{SNR}_{\mathrm{seg}} = \frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\frac{\sum_{n=Lm}^{Lm+L-1}S^2(n)}{\sum_{n=Lm}^{Lm+L-1}\left(S(n)-\hat{S}(n)\right)^2} \qquad \text{(III-02)}$$

Où L est la longueur de trame (nombre d'échantillons) et M le nombre de trames dans le signal (N = ML). La longueur de trame est normalement réglée entre 15 et 20 ms.

Étant donné que le logarithme du rapport est calculé avant étalement, les trames ayant un exceptionnellement grand rapport est quelque peu pesées moins, alors que les cadres avec un faible taux sont pesés un peu plus élevé.
On peut observer que cela correspond à la qualité perceptive bien, à savoir, trames avec grand discours et pas de bruit audible ne domine pas l'ensemble.

La qualité perçue, mais l'existence de cadres bruyants se distingue et stimuleront la qualité globale inférieure. Cependant, si l'échantillon de parole contient trop de silence, la valeur globales segSNR vont diminuer de façon significative depuis les cadres silencieux montrent généralement segSNR grandes valeurs négatives. Dans ce cas, les parties silencieuses devraient être exclues de la en moyenne en utilisant l'activité de la parole détecteurs. De la même manière, l'exclusion de trames avec des valeurs trop grandes ou petites de la moyenne se traduit généralement par segSNR valeurs qui conviennent bien à la qualité subjective. Une valeur typique pour la partie supérieure et la limite de rapport inférieur est de 35 et -10 dB [45].

Le SNR segmental souffre de deux limitations : d'abord si le signal de parole contient des segments de silence, ce qui est très probable, le s(n) sera nul et n'importe quelle quantité de bruit entraînera un SNR en dB négatif pour ce segment, du coup le SNR total sera faussé par cette quantité. Ce problème peut être résolu partiellement en choisissant un seuil d'énergie au-delà duquel le SNR segmental sera calculé. Ensuite, il faut nécessairement que les deux signaux comparés soient alignés temporellement car ce critère est très sensible aux déphasages.

III.3.2 Mesure d'Itakura Saito

La mesure d'Itakura Saito repose sur l'analyse LPC. Son expression fait intervenir le modèle tout pole du signal de référence S et celui du signal testé Y. Soient P(ω), $\hat{P}(\omega)$ les densités spectrales de puissance du modèle AR du signal de référence et du signal de test. La distance d'Itakura Saito est donnée par :

$$d_{IS}\left(P(w),\widehat{P}(w)\right)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{P(w)}{\widehat{P}(w)}-\log\frac{P(w)}{\widehat{P}(w)}-1\right]dw \qquad \text{(III-03)}$$

III.3.3 Le Cepstral Distance (CD)

Le Cepstral Distance est principalement utile pour représenter la distribution de l'erreur au cours du temps. Les coefficients cepstraux c(i) peuvent être calculés à partir des coefficients de prédiction linéaire a(i) à l'aide de la relation suivante [46] :

$$c(1)=-a(1) \qquad \text{(III-04)}$$

$$c(i)=-a(i)-\sum_{k=1}^{i-1}\left(1-\frac{k}{i}\right)c(i-k)a(k) \quad , \ 1\le i\le p \qquad \text{(III-05)}$$

Considérons les coefficients cepstraux C_t(i) et C_r(i) calculés respectivement sur les trames d'indice i du signal-test à évaluer et de la référence. La distance cepstrale d'ordre 2 entre ces deux signaux est donnée par [47] :

$$d_{cep}=\sum_{i=1}^{p}\left(c_t(i)-c_r(i)\right)^2 \qquad \text{(III-06)}$$

Ou p est l'ordre des coefficients LPC. Suite à cette écriture, la distance cepstrale est tout simplement la distance euclidienne entre les coefficients cepstraux générés récursivement à partir de l'analyse LPC.

III.3.4 Weighted Spectral Slope Measures (WSS)

La pente spectrale (WSS) mesure de distance pondérée est une distance spectrale directe mesure. Il est basé sur la comparaison des spectres lissée de la propre et déformée des échantillons de parole. Le spectre lissé peut être obtenue à partir de chaque analyse LP, Cepstral filtrage (un terme inventé pour le filtrage dans le domaine cepstral) ou de filtres une analyse. Une forme de réalisation de WSS peut être définie comme suit:

$$d_{WSS} = \frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{j=1}^{K} W(j,m)\left(S_c(j,m)-S_d(j,m)\right)^2}{\sum_{j=1}^{K} W(j,m)} \quad \text{(III-07)}$$

Où K est le nombre de bandes, M est le nombre total de trames, et S_c (j, m) et S_d (j, m) sont les pentes spectrales (typiquement les différences spectrales entre bandes voisines) de la bande de j dans le cadre de m pour la parole propre et déformée, respectivement. W (j, m) sont des poids, qui peut être calculée comme indiqué par Klatts [48].WSS a été largement étudiée au cours des dernières années, et a bénéficié d'une large acceptation.

III.3.5 Likelihood Linear Regression(LLR)

Il est bien connu que le processus de production de la parole peut être modélisé de manière efficace avec un modèle de prédiction linéaire (LP). Il y a un certain nombre de mesures objectives qui utilisent la distance entre les deux ensembles de coefficients de prédiction linéaire (LPC) calculée l'original et la parole déformée. Nous ne discuterons que quelques-uns de ceux-ci. Le ratio de vraisemblance (LLR) mesure est une mesure de distance qui peut être directement calculé à partir du vecteur LPC du discours propre et déformée. LLR mesure peut être calculée comme suit:

$$d_{LLR}(a_d, a_c) = log\left(\frac{a_d R_c a_d^T}{a_c R_c a_c^T}\right) \quad \text{(III-08)}$$

Où a_c est le vecteur LPC pour le discours propre, a_d est le vecteur LPC pour la déformée la parole, a^T est la transposée d'un a, et R_c est la matrice d'autocorrélation pour le nettoyage discours.

III.3.6 Bark Spectral Distortion (BSD) et Modified Bark Spectral Distortion (MBSD)

La mesure BSD (Bark Spectral Distortion) [39] est parmi les premiers critères à avoir incorporé des notions en relation avec notre système d'audition dans l'évaluation de la qualité de la parole [49]. Le BSD a pour objectif de mesurer la distorsion entre le signal de référence et celui codé, dans le domaine de Bark. La sensation de force sonore connue sous le nom de sonie est mise en jeu pour calculer cette distorsion. En effet, la distorsion totale est la moyenne de la distance euclidienne entre la sonie du signal de référence et celle du signal débruité.

Le MBSD (Modified Bark Spectral Distortion) [50] introduit le seuil de masquage du bruit pour calculer la distorsion dans le BSD, l'idée est de ne tenir compte que de la distorsion audible. Effectivement, tout ce qui est au-dessous du seuil de masquage du bruit est imperceptible à l'oreille humaine. Par conséquent, la distorsion totale est la moyenne de la différence entre les sonies du signal de référence et du signal débruté pondérée par un paramètre s'annulant lorsque la distorsion est inaudible.

III.3.7 Perceptual Speech Quality Measure (PSQM)

Le PSQM (Perceptual Speech Quality Measure) est une version typique aux signaux de parole d´écrite par la norme P.861 [42]. Elle constitue donc un cas particulier du critère PAQM (Perceptual Audio Quality Measure) [51] dédié aux signaux audio en général. L'intérêt de concevoir une mesure uniquement pour la parole revient aux différences de caractéristiques existant entre la parole et la musique. Le PSQM exploite à son tour les propriétés de la perception auditive humaine pour évaluer la qualité de la parole. La moyenne de la différence en sonie, désignée dans la norme par le terme bruit perturbateur, constitue la note PSQM attribuée à la qualité du signal codé.

III.3.8 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 [43] 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 bel et bien d'autres facteurs techniques et applications [43] 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 (différence 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 dûe 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 [52], [53] et [54], 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, tel que [55], 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.

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 des critères d'écoute et sur d'autres critères objectifs.

III.4 Critères subjectifs

Les mesures de qualité subjective les plus fréquemment utilisées sont le MOS (Mean Opinion Score), le DMOS (Degradation Mean Opinion Score) et le CMOS (Comparison Mean Opinion Score) [56].

Le MOS est le résultat de l'analyse par catégories absolues ACR (Absolute Category Rating) dans laquelle un groupe d'auditeurs écoute un ensemble de fichiers audio et les évalue indépendamment, un à un, selon une échelle de notation sur la qualité perçue (tableau III-02). Le CMOS est le résultat de l'analyse par catégories de comparaisons CCR (Comparison Category Rating) dans laquelle on fournit à un groupe d'auditeurs des signaux par paires. L'auditeur compare les deux signaux de chaque paire en termes de qualité en précisant lequel est le meilleur et évalue la différence selon une échelle de notation bien définie (tableau III-03).

Quant au DMOS, il résulte de l'analyse par catégories de dégradations DCR (Degradation Category Rating) dans laquelle on fournit à un groupe d'auditeurs des paires de signaux pour comparer cette fois-ci la qualité en terme de dégradation. Contrairement au CMOS, les auditeurs savent a priori que la qualité du second signal est moins bonne que celle du premier. Ils doivent donc indiquer à quel point le second est justement moins bon suivant l'échelle de DMOS (tableau III-04).

D'une manière générale, lors de ces trois types de tests, les plus communément utilisés surtout pour évaluer les codeurs de parole, la qualité du signal de parole dépend de la personne qui la juge et l'évalue. Sa façon de percevoir met en jeu l'expérience passée, l'environnement dans lequel elle s'est déroulée, son humeur et ses attentes.

Ainsi, afin de diminuer l'effet subjectif sur l'évaluation de la qualité vocale, les notes des participants pour une condition de test donnée sont moyennées pour obtenir la note moyenne d'opinion.

Dans ce qui suit et comme il est d'usage, on désigne par MOS, comme terme général, les trois tests subjectifs déjà définis sauf précision. Par définition donc, le MOS est un sondage auprès d'un échantillon de personnes représentatives du reste de la population. Lors de ce sondage, les auditeurs sont invités à écouter et à juger. Le jugement se fait à travers l'attribution d'une note sanctionnant la qualité perçue du signal de parole qu'ils ont écouté. La moyenne des notes attribuées constitue donc le MOS.

L'avantage du MOS est qu'il quantifie la qualité perçue par les auditeurs participant aux tests. C'est donc une évaluation réelle, fiable et correcte de la qualité des signaux mis en jeu cependant, ce test est souvent écarté du fait qu'il requiert :

- Un grand nombre d'auditeurs.
- Un équipement audio adapté.

Une formation des auditeurs à la bonne façon d'attribuer des notes pour que celles-ci soient exploitables.

- Une collecte d'informations et des traitements statistiques pour réduire l'aléa.

En outre, le MOS n'est pas standardisé et le processus de test ne peut pas *ê*tre complétement automatisé.

Score MOS	Qualité MOS
5	Excellent
4	Bon
3	Passable
2	Mauvais
1	Médiocre

Tableau III-02 : Echelle MOS.

Score CMOS	Qualité du second comparé au premier
3	Bien meilleure
2	Meilleure
1	Légèrement meilleure
0	A peu près équivalente
-1	Un peu moins bonne
-2	Moins bonne
-3	Nettement mediocre

Tableau III-03 : Echelle CMOS

Score CMOS	Qualité MOS
5	Dépourvu de dégradation
4	Dégradation audible mais pas gênante
3	Dégradation un peu gênante
2	Dégradation gênante
1	Dégradation très gênante

Tableau III-04 : Echelle DMOS

III.5 Conclusion

Ce chapitre décrit brièvement deux aspects de la qualité de la parole, comptes d'opinion qui mesure la qualité globale perçue de la parole, et intelligibilité de la parole qui mesure la l'exactitude du contenu de la parole reçue. Ces deux aspects différents ont été décrits, et deux types de mesures pour ces aspects, la qualité subjective et objective les mesures. Mesures de qualité Subjective emploient des auditeurs humains pour évaluer la qualité dès le discours. Il faut souvent nombre considérable d'auditeurs pour obtenir des résultats stables, et il est souvent fastidieux et coûteux.

D'autre part, la qualité Objectif mesures estime la qualité d'une certaine forme de mesures physiques de la parole. Beaucoup sont basées sur des mesures de distance entre le discours original et dégradé. Rapport signal sur bruit (SNR) ou une extension de cette mesure est une forme courante d'une mesure subjective. Mesures subjectives récentes utilisent une distance plus sophistiquée mesures qui sont connus pour être fortement corrélées avec la perception auditive humaine.

Les mesures subjectives ne nécessitent pas des auditeurs humains, et est donc moins coûteux et moins de temps que les mesures subjectives. Cependant, il existe des exemples où les estimations de l'égalité de mesures objectives et les mesures subjectives ne correspondent pas.

Ainsi, les mesures de qualité subjectives sont toujours la manière la plus concluante pour mesurer la qualité perçue. Toutefois, des mesures objectives récentes sont de bons estimateurs de subjective la qualité et peut être utilisée pour obtenir une estimation de qualité grossière, qui peut être suivie par évaluation de la qualité subjective pour des conditions pour confirmer la perception qualité.

Chapitre 4

Résultats et Simulations

IV.1 Introduction

Dans une communication, le signal est souvent altéré lors de son passage dans le canal. Nous proposons, dans ce chapitre, de traiter le signal original par une technique de rehaussement afin de retrouver le signal de départ. Nous effectuerons cette tâche sur un signal de parole bruité par des bruits blancs gaussiens, bavardage (babbel), restaurant, salle d'exposition, voiture et bruit réel à l'aide du logiciel de calcul Matlab.

Ce processus peut se résumer selon le schéma bloque de la figureIV-01.

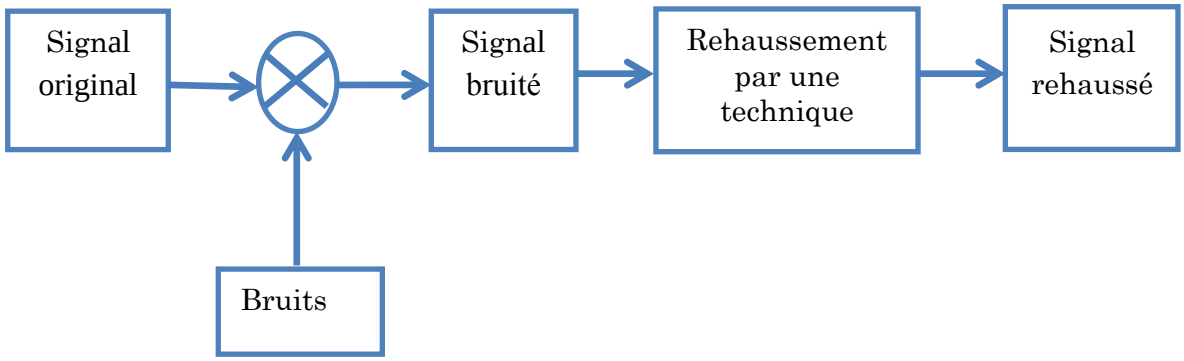


Figure IV-01: Chaine de traitement

IV.2 Résultats et discussion

IV.2.1 Étude comparative des différentes techniques d'élimination de bruit additif au signal parole

Nous essayons de faire l'étude comparative entre quatre méthodes de débruitage pour extraire la partie du signal reçu bruité correspondant au signal original. Nous étudierons, ensuite l'efficacité de ces méthodes en fonction du rapport signal à bruit introduit par le canal et nous chercherons à définir les critères du bon fonctionnement.

En effet on ne peut pas traiter tous les échantillons d'un signal simultanément, on les traite par segment réduit, sélectionné à l'aide d'une fenêtre d'analyse de taille adaptée afin de pouvoir considérer le signal quasi-stationnaire pendant la durée d'analyse.

L'estimation de bruit pour les quatre techniques est faite en tenant compte de la première zone de silence d'une durée de 81.25 ms, voit la figure IV-02.

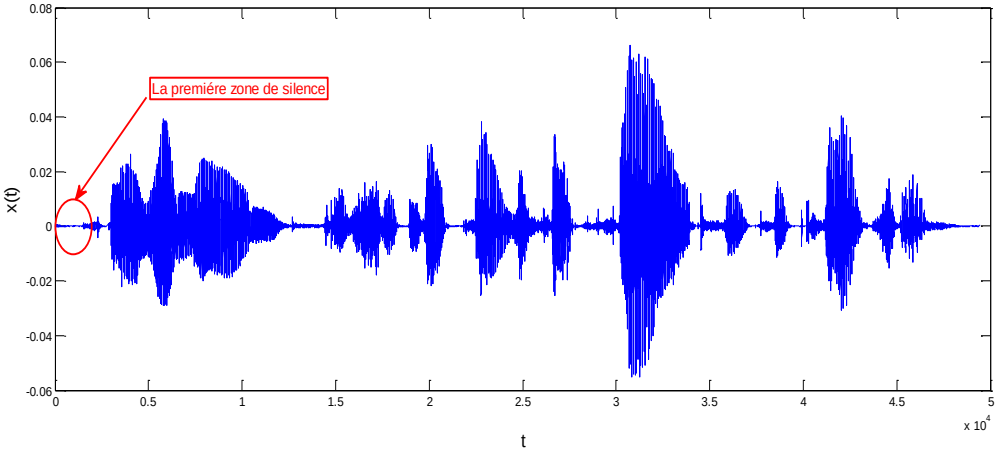


Figure IV-02 : la première zone de silence à considérer pour l'estimation de bruit cette zone est estimée d'une durée de 81.25ms

Ces méthodes sont:

- ✓ **Wiener-Scalart96 [17].**
- ✓ **MMSE-Cohen2004 [22]:** Il est proposé par I.Cohen 2004et utilise Minimum Mean Square Error et Log-Spectral Amplitude Estimator".
- ✓ **MMSE-STSA84 [22]:**Minimum Mean Square Error Short Time Spectral Amplitude (1984)
- ✓ **MMSE-STSA85 [22]:**Minimum Mean Square Error Short Time Spectral Amplitude (1985)

Pour améliorer le signal parole, les quatre techniques d'élimination de bruit (amélioration de signal parole) mentionnées ci-dessus sont évaluées en matière de PESQ, segSNR, WSS et LLR.

IV.2.2 Comparaison des différentes techniques d'élimination des bruits (des bruits blancs gaussiens, bavardage (babbel), restaurant, salle d'exposition, voiture)

a) Bruit réel

La figure IV-03 montre un signal bruité par le bruit réel et le rehaussement de ce bruit par les quatre méthodes: Wiener-Scalart96, MMSE-Cohen2004, MMSE-STSA84 et MMSE-STSA85.

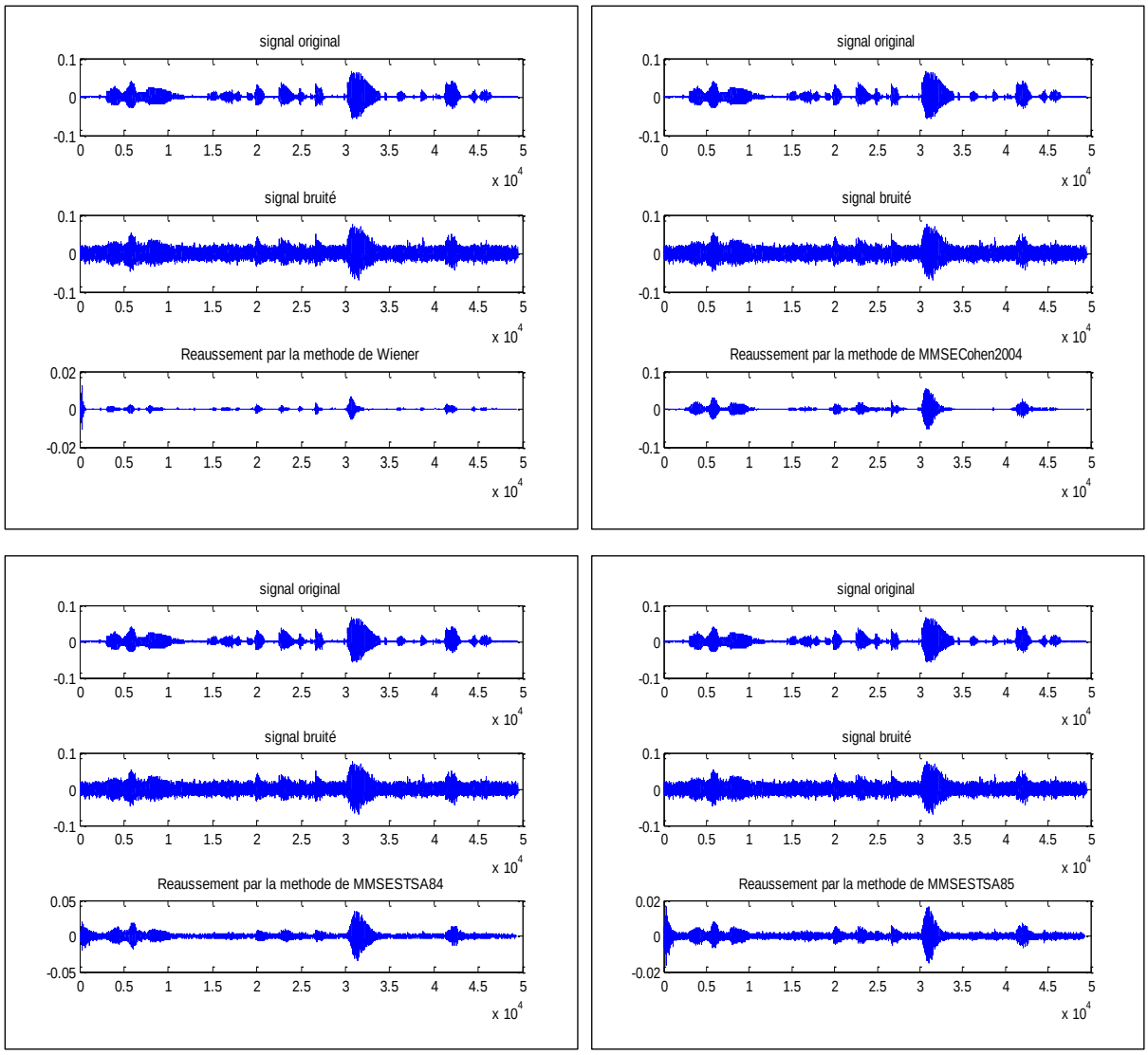


Figure IV-03:rehaussement de signal bruité par bruit réel pour SNR= 0 dB.

On observe d'après la figureIV-03 que la technique qui amélioré le signal bruité par un bruit réel est MMSE-Cohen2004.

b) Bruit bavardage

La figure ci-dessous montre un signal bruité par le bruit bavardage et le rehaussement de ce bruit par les quatre méthodes : WienerScalart96, MMSECohen2004, MMSESTSA84et MMSESTSA85.

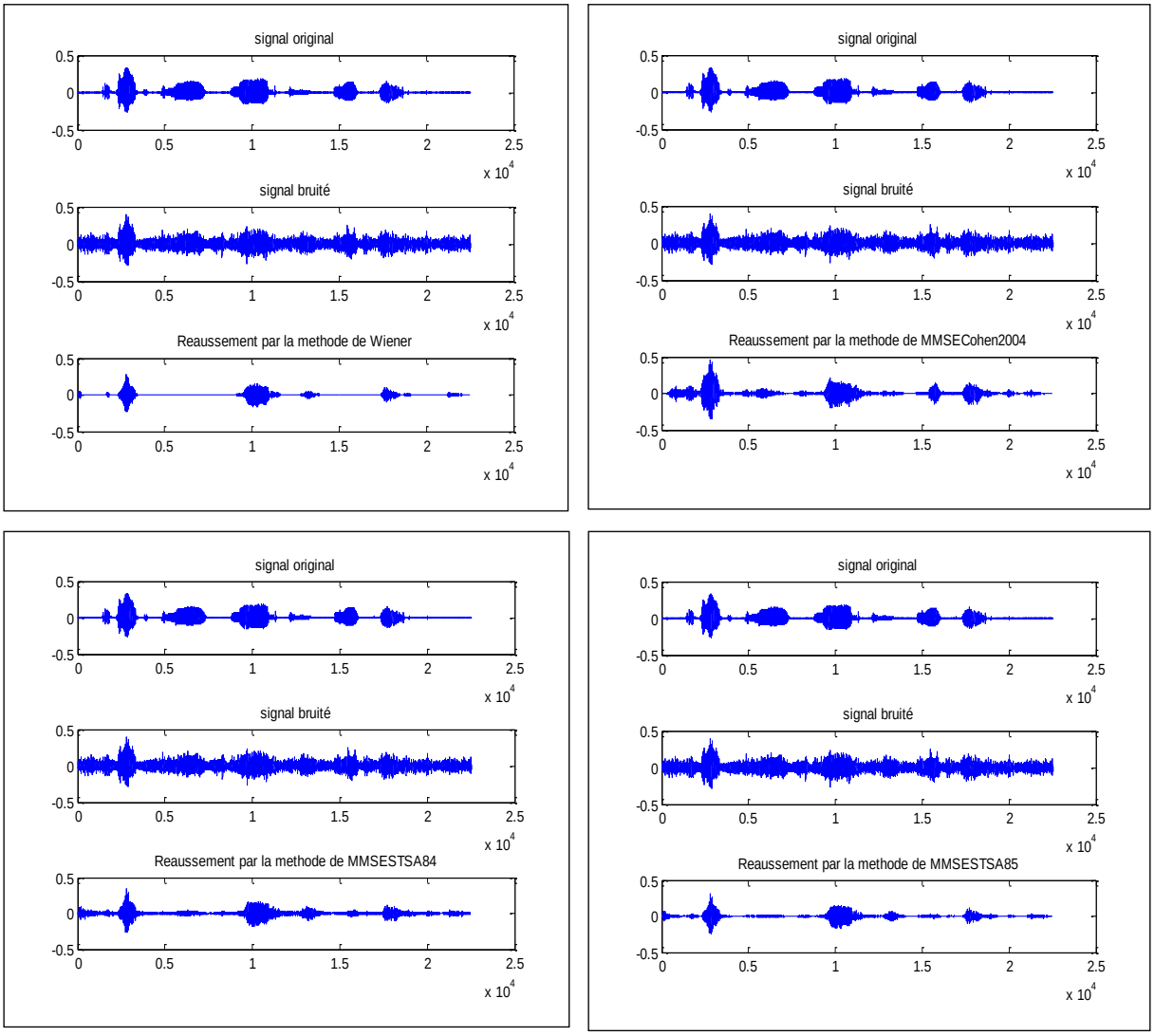


Figure IV-04:rehaussement de signal bruité par bruit bavardage pour SNR= 0 dB.

On observe d'après la figureIV-04 que la technique qui amélioré le signal bruité par un bruit bavardage est Wiener-Scalart96.

c) Bruit blanc gaussien

La figure ci-dessous montre un signal bruité par le bruit blanc gaussien et le rehaussement de ce bruit par les quatre méthodes : WienerScalart96, MMSECohen2004, MMSESTSA84et MMSESTSA85.

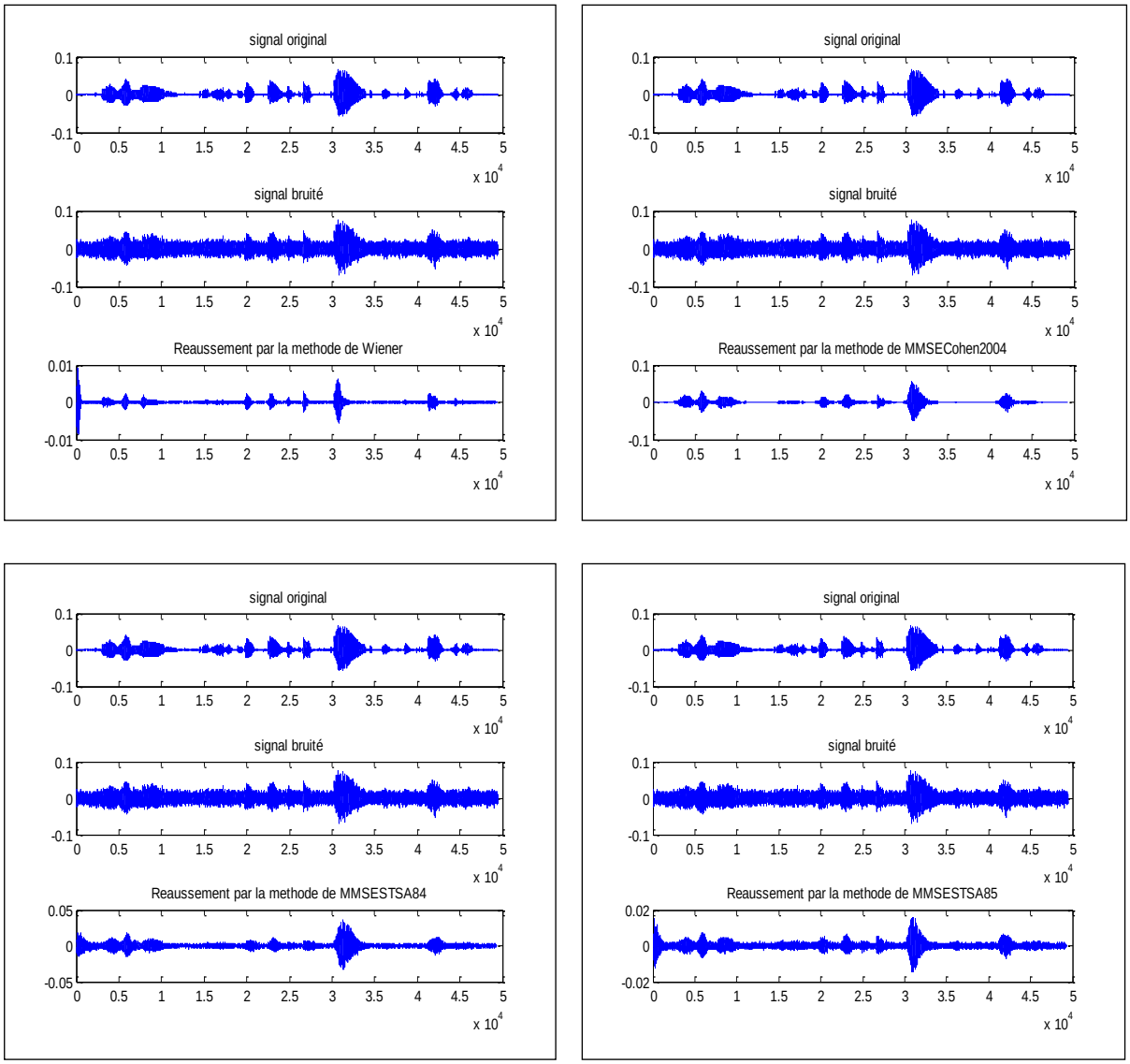


Figure IV-05:rehaussement de signal bruité par bruit blanc gaussienne pour SNR= 0 dB.

On observe d'après la figureIV-05 que la technique qui amélioré le signal bruité par un bruit blanc gaussien est MMSE-Cohen2004.

d) Bruit voiture

La figure ci-dessous montre un signal bruité par le bruit voiture et le rehaussement de ce bruit par les quatre méthodes : Wiener-Scalart96, MMSE-Cohen2004, MMSE-STSA84et MMSE-STSA85.

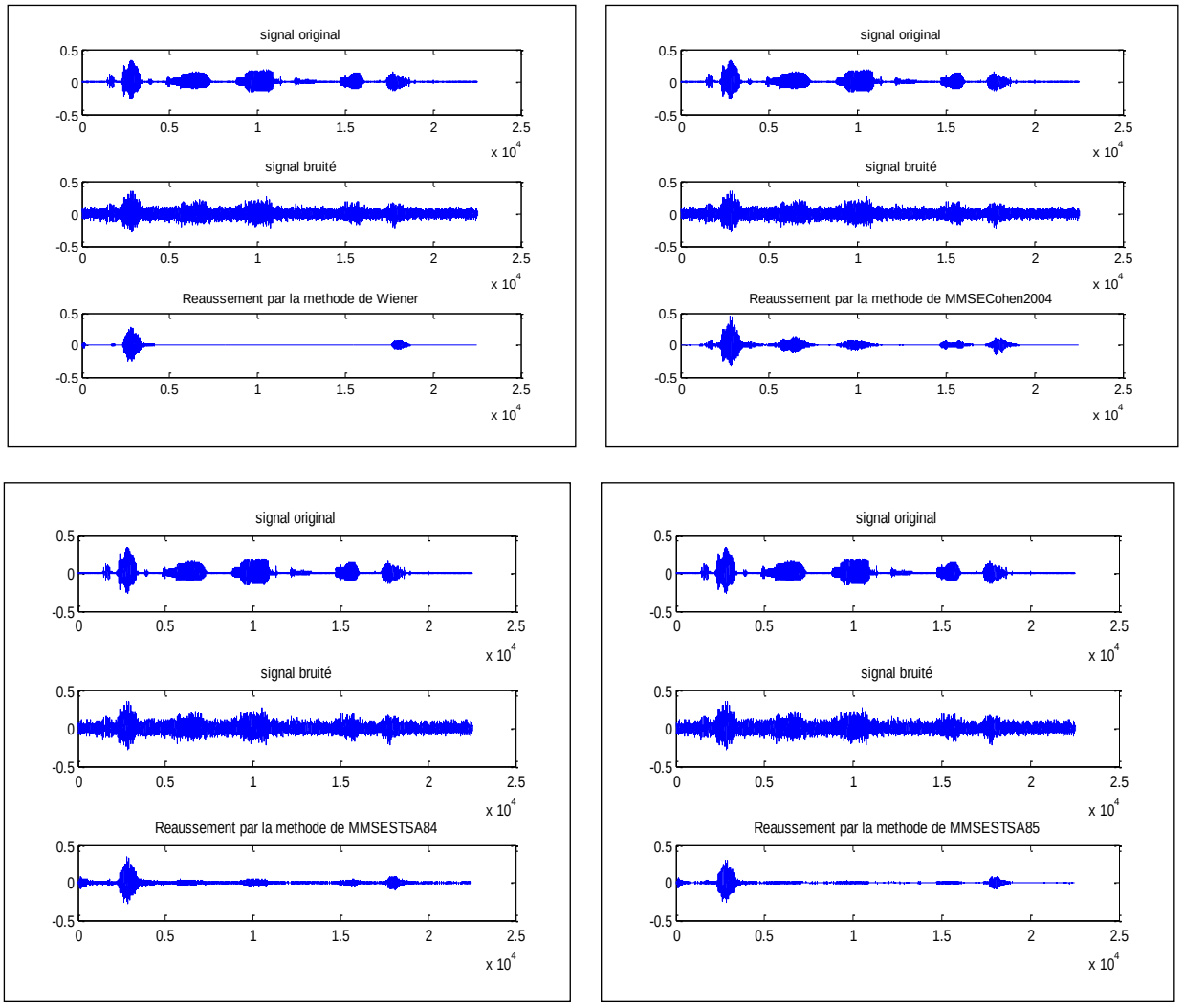


Figure IV-06: rehaussement de signal bruité par bruit Voiture pour SNR= 0 dB

On observe d'après la figure IV-06 que la technique qui amélioré le signal bruité par bruit voiture est MMSE-Cohen2004.

e) Bruit restaurantan

La figure ci-dessous montre un signal bruité par le bruit restaurant et le rehaussement de ce bruit par les quatre méthodes : Wiener-Scalart96, MMSE-Cohen2004, MMSE-STSA84et MMSE-STSA85.

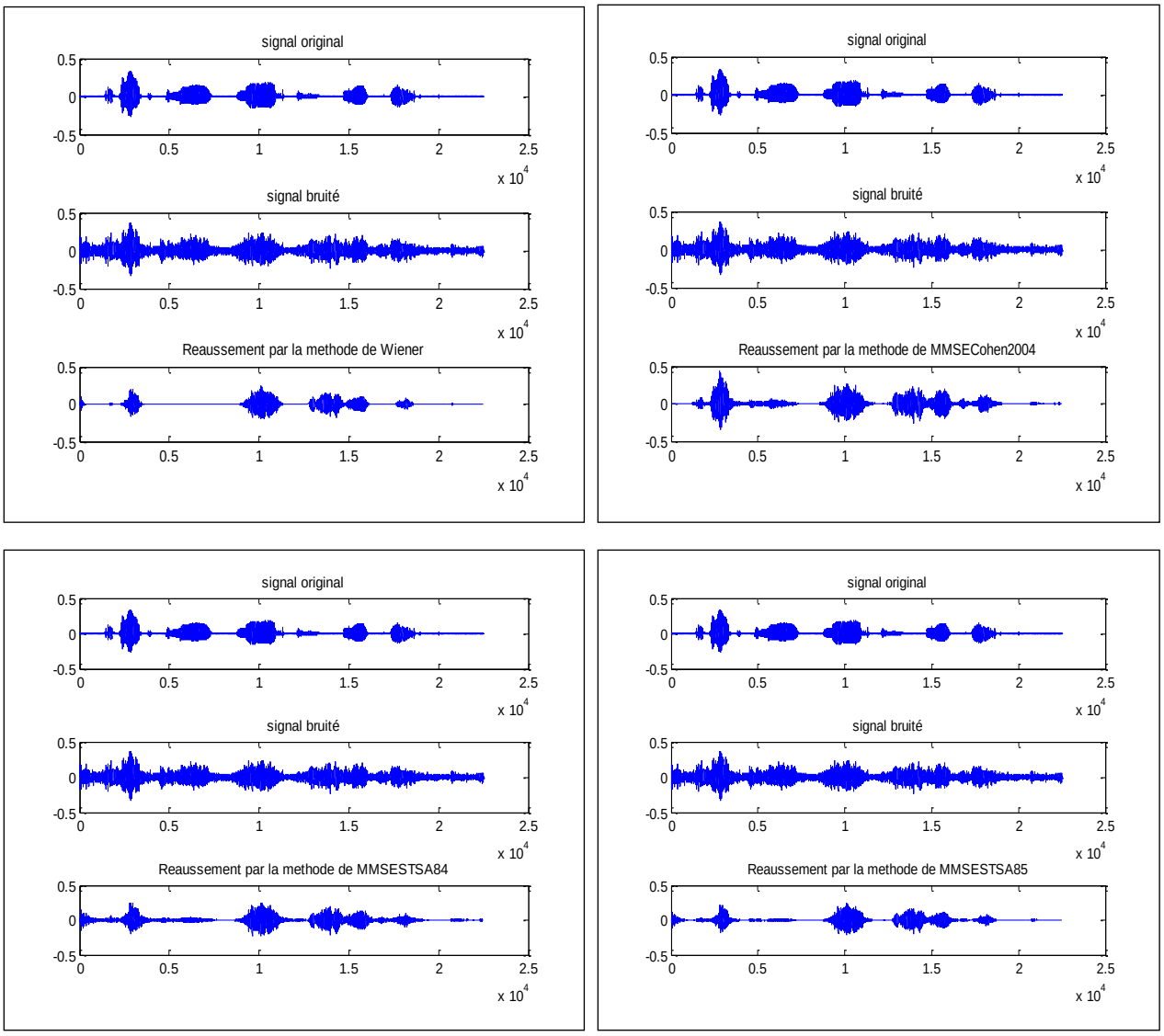


Figure IV-07: rehaussement de signal bruité par bruit restaurant pourSNR= 0 dB

On observe d'après la figureIV-07 que la technique qui amélioré le signal bruité par un bruit restaurant est MMSE-Cohen2004.

f) Bruit salle d'exposition

La figure ci-dessous montre un signal bruité par le bruit salle d'exposition et le rehaussement de ce bruit par les quatre méthodes : Wiener-Scalart96, MMSE-Cohen2004, MMSES-TSA84et MMSE-STSA85

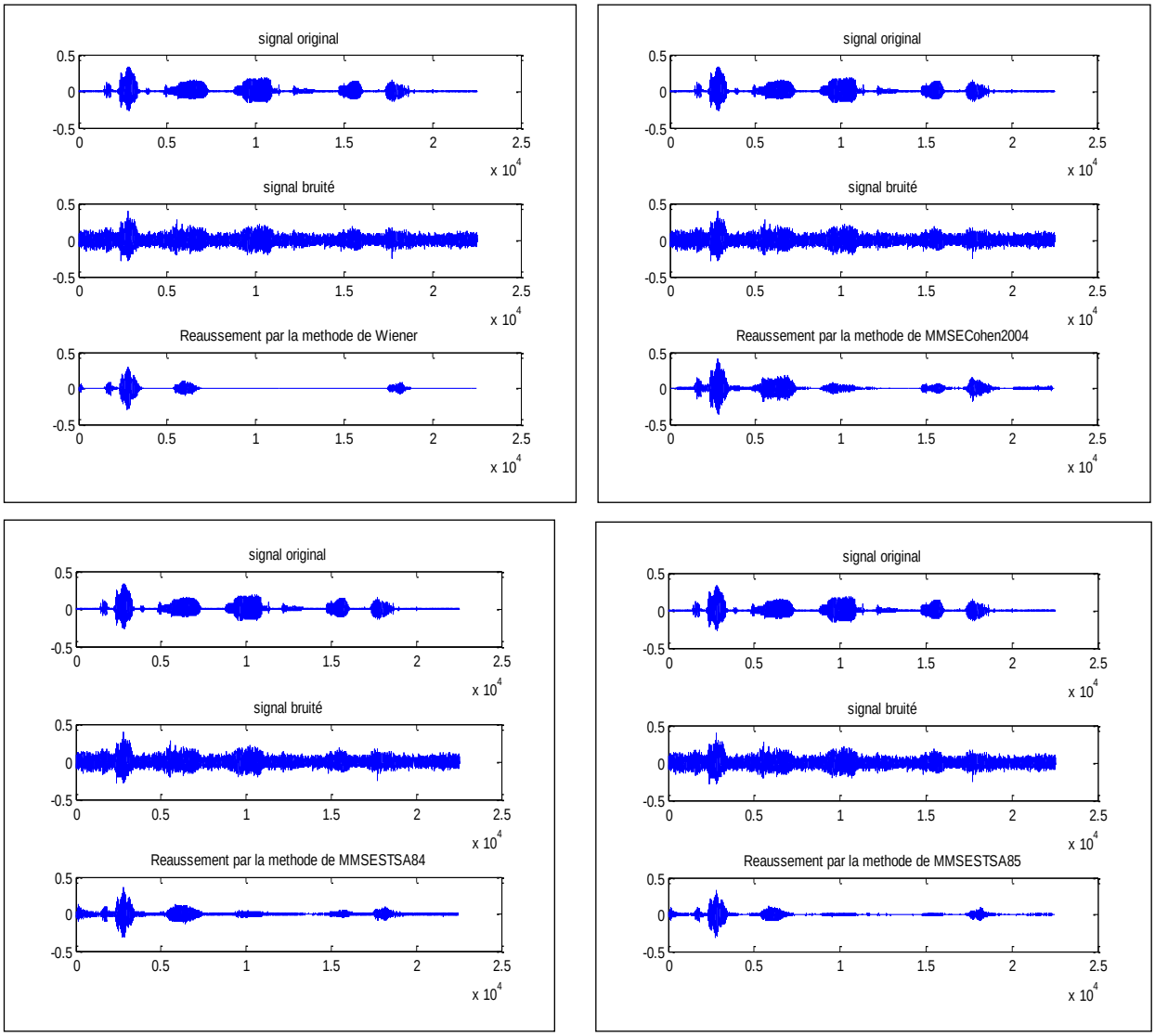


Figure IV-08: rehaussement de signal bruité par bruit salle d'exposition pour SNR= 0 dB

On observe d'après la figureIV-08 que la technique qui amélioré le signal bruité par un bruit salle d'exposition est Wiener-Scalart96.

IV.2.3 Comparaison des méthodes de rehaussement de la parole en termes de PESQ, segSNR, WSS et LLR en présence des bruits (Réel, bavardage, blanc gaussien, voiture et salle d'exposition)

a) Bruit réel

En présence du bruit réel en utilisant la base de données NOIZEUS [57] Les figures ci- dessous représentent la comparaison des performances en termes de ces critères, d' SNR, de -5 à 30 dB par pas de 5 dB.

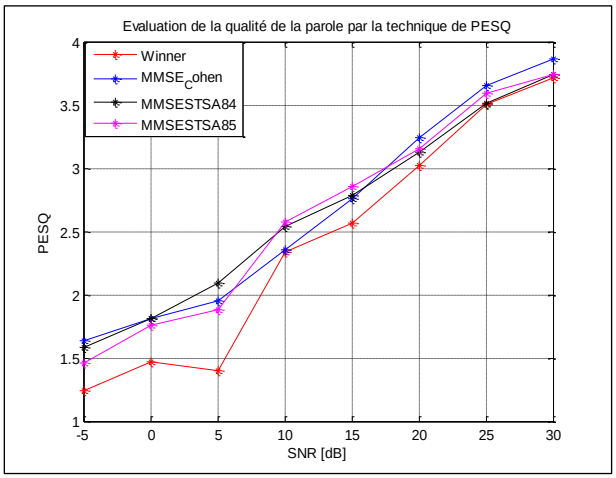


Figure IV-09: Comparaison des Performances en termes de PESQ en présence du bruit Réel (SNR=-5 à 30 dB par pas de 5 dB)

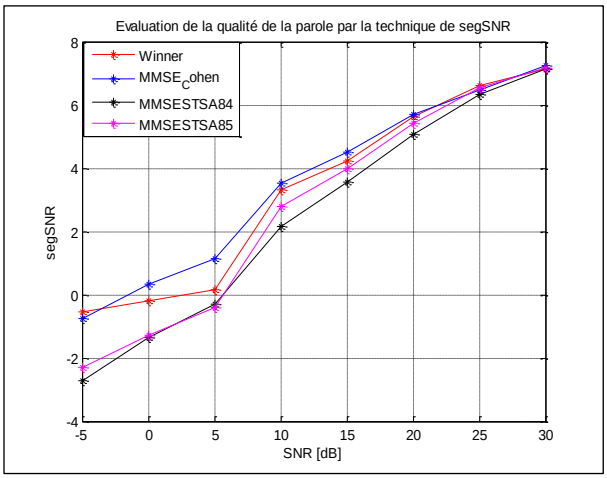


Figure IV-10: Comparaison des performan en termes de segSNR en présence du bruit Réel (SNR=-5 à 30 dB par pas de 5 dB)

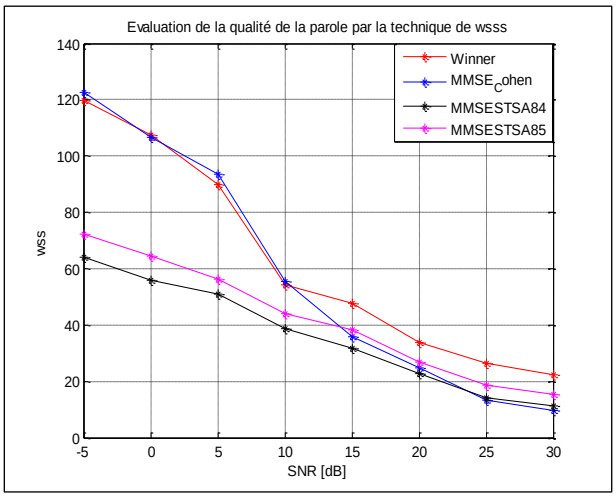


Figure IV-11: Comparaison des performances en termes de WSS en présence du bruit Réel (SNR=-5 à 30 dB par pas de 5 dB)

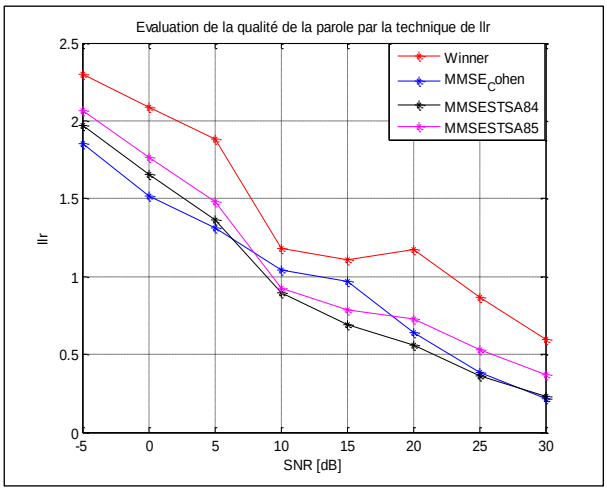


Figure IV-12: Comparaison des performances en termes de LLR en présence du bruit Réel (SNR=-5 à 30 dB par pas de 5 dB)

Nous constatons dans la figure (IV-09) que les méthodes MMSE-Cohen2004, MMSE-STSA84et MMSE-STSA85sont presque proches en terme de PESQ, mais, entre les valeurs des SNR 15 et 30 dB la technique MMSE-Cohen2004 présente un meilleur en terme de PESQ. Par contre la méthode Wiener-Scalart96 est la plus mauvaise.

Dans la figure (IV-10) nous observons que la méthode de MMSE-Cohen2004 présente le meilleur en terme segSNR, tandis que la technique Wiener-Scalart96est comme la seconde en termes de segSNR. Et pour la méthode de MMSE-STSA84, elle est la plus mauvaise.

La figure (IV-11) représente les mesures de WSS. Elle montre que la méthode de MMSE-STSA84est la plus mauvaise, tandis que MMSE-STSA85est comme la seconde en terme de WSS, par contre la méthode MMSE-Cohen2004présente d'une part le meilleur de WSS entre -5 et 10 dB est élevé, et d'autre part la technique de WienerScalart96 est la meilleure entre 10 et 30 dB (voir la figure).

La figure (IV-12) nous indique que la technique Wiener-Scalart96 fournit de meilleur résultat de LLR, tandis que les techniques MMSE-STSA85 et MMSE-Cohen2004 ont le même résultat en termes de LLR. Par contre, la méthode MMSE-STSA84 est la plus mauvaise.

b) Bruit bavardage

En présence du bruit bavardage en utilisant la base de données NOIZEUS [57]. Les figures ci- dessous représentent la comparaison des performances en termes de ces critères d'SNR, de 0 à 15 dB par pas de 5 dB.

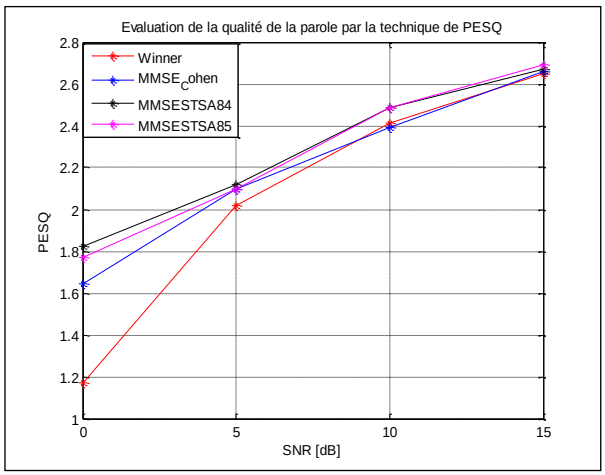


Figure IV-13: Comparaison des performances en termes de PESQ en présence du bruit Babbel (bavardage) (SNR=0 à 15 dB par pas de 5 dB)

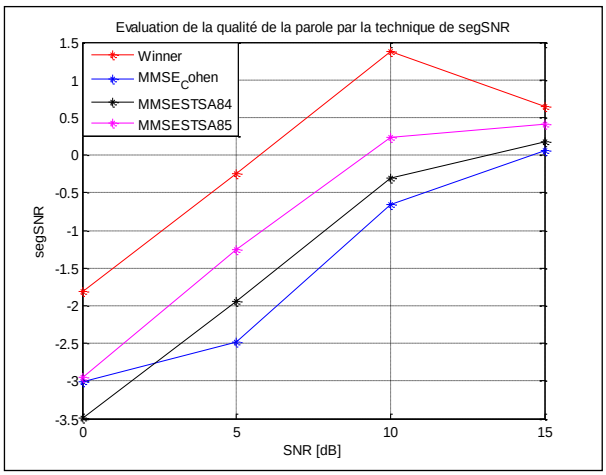


Figure IV-14: Comparaison des performances en termes de segSNR en présence du bruit Babbel (bavardage) (SNR=0 à 15 dB par pas de 5 dB)

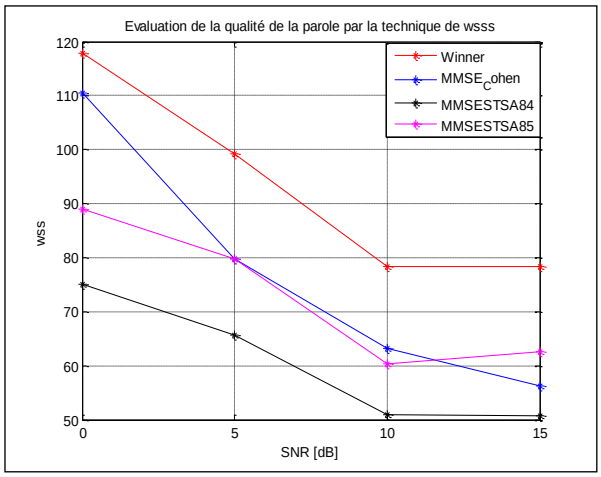


Figure IV-15: Comparaison des performances en termes de WSS en présence du bruit Babbel (bavardage) (SNR=0 à 15 dB par pas de 5 dB)

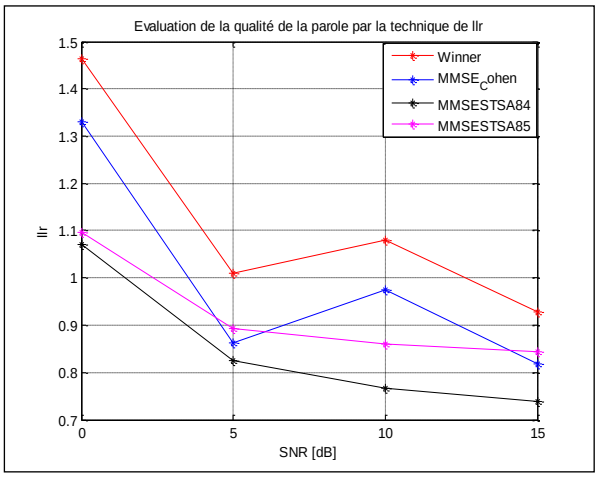


Figure IV-16: Comparaison des performances en termes de LLR en présence du bruit Babbel (bavardage) (SNR=0 à 15 dB par pas de 5 dB)

Depuis la figure (IV-13) nous observons que les méthodes deMMSECohen2004, MMSESTSA84et MMSESTSA85 sont presque proches en terme de PESQ entre 0 et 10 dB, et elles sont présentent les meilleures techniques. Entre 10 et 15 dB, la technique WienerScalart96 est plus faible par rapport aux méthodes MMSESTSA84et MMSESTSA85.

Les figures (IV-14, IV-15 etIV-16) montrent que la technique de WienerScalart96 est la plus haute en terme segSNR, WSS, LLR. Ce qui signifie qu'elle est la meilleure.

Tandis que les techniques MMSESTSA84et MMSESTSA85 comme les secondes en terme de segSNR, WSS, LLR, par contre la méthode de MMSECohen2004 est la plus mauvaise.

c) bruit blanc gaussien

En présence du bruit blanc gaussien en utilisant la base de données TIMIT [58].Les figures ci- dessous représente la comparaison des performances en termes de ces critères d' SNR, de -5 à 30 dB par pas de 5 dB.

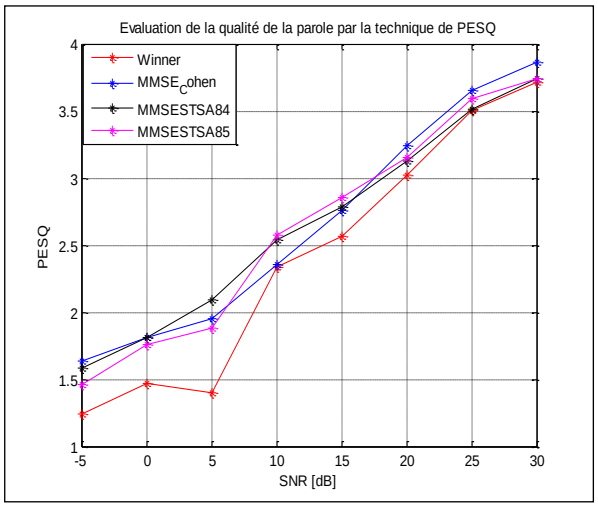


Figure IV-17: Comparaison des performances En termes de PESQ en présence du bruit Blanc Gaussien (SNR=-5 à 30 dB par pas de 5 dB)

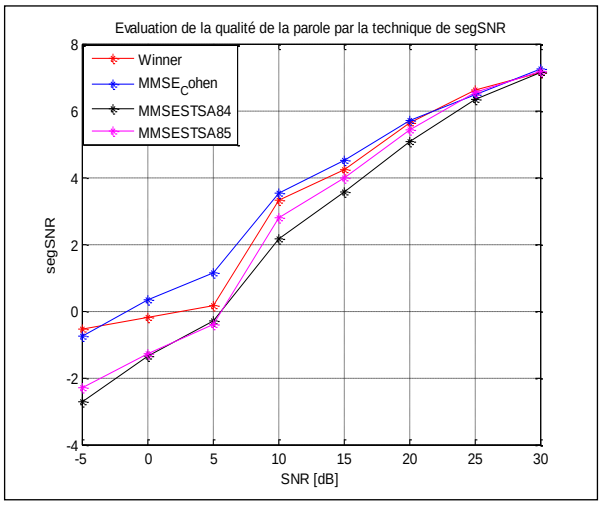


Figure IV-18: Comparaison des performances en termes de segSNR en présence du bruit Blanc Gaussien (SNR=-5 à 30 dB par pas de 5 dB)

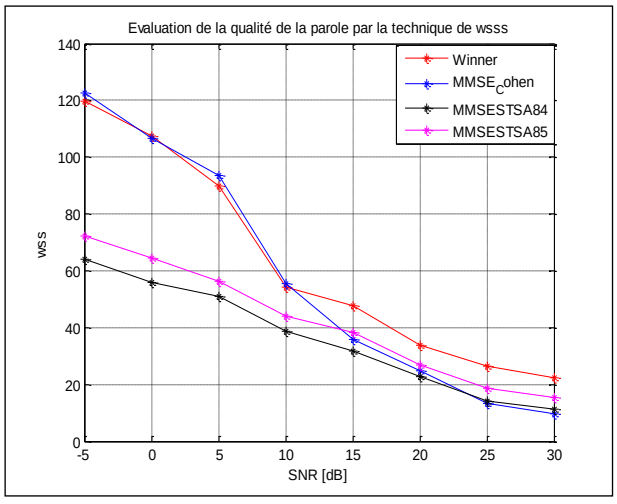


Figure IV-19: Comparaison des performances en termes de WSS en présence du bruit Blanc Gaussien (SNR=-5 à 30 dB par pas de 5 dB)

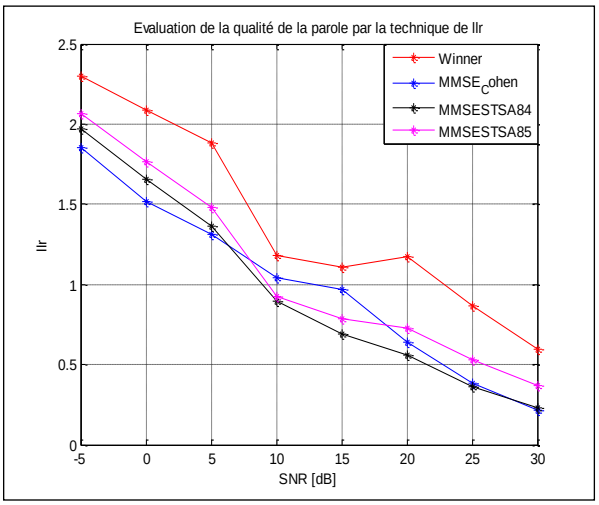


Figure IV-20: Comparaison des performances en termes de LLR en présence du bruit Blanc Gaussien (SNR=-5 à 30 dB par pas de 5 dB)

La figure (IV-17) représente les résultats de mesures de PESQ. A travers cette figure nous constatons que la technique Wiener-Scalart96 est de mauvaise qualité par rapport aux autres méthodes qui sont de qualité plus ou moins proches avec de petites déférences, qui montrent que MMSE-Cohen2004 est dotée d'une meilleure qualité.

Selon la figure (IV-18), les quatre méthodes sont les proches d'améliorées le signal de parole avec une petite différence en terme de segSNR qui montre que la technique la plus performante est MMSE-Cohen2004 et que la technique la moins performante est MMSE-STSA84.

Nous pouvons observer sur les figures (IV-19 et IV-20) que la technique Wiener-Scalart96 est plus performante en WSS et en LLR, pendant que les méthodes MMSE-Cohen2004, MMSE-STSA84et MMSESTSA85 sont de qualité égales. En termes de WSS et de LLR la méthode MMSE-STSA84 est restée la moins performante.

d) Bruit voiture

En présence du bruit voiture en utilisant la base de données NOIZEUS [57] Les figures ci- dessous représente la comparaison des performances en termes de ces critères d' SNR, de 0 à 15 dB par pas de 5 dB.

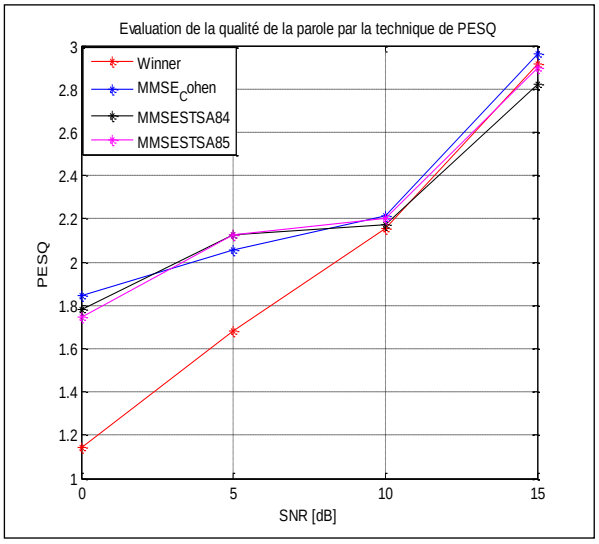


Figure IV-21: Comparaison des performances en termes de PESQ en présence du bruit Voiture (SNR=0 à 15 dB par pas de 5 dB)

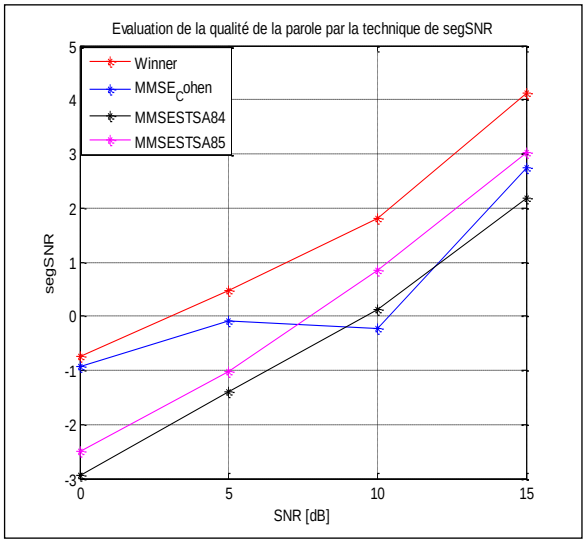


Figure IV-22: Comparaison des performances en termes de segSNR en présence du bruit Voiture (SNR=0 à 15 dB par pas de 5 dB)

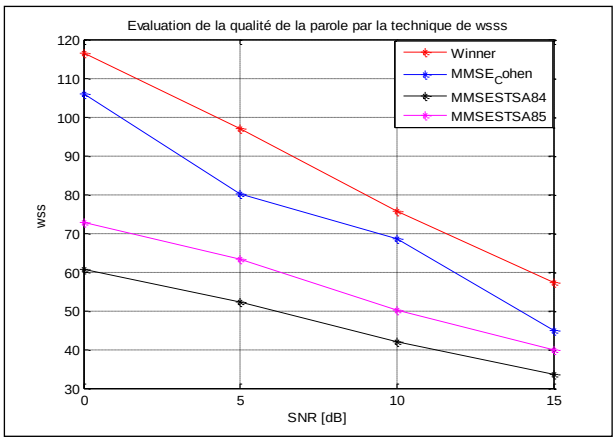


Figure IV-23: Comparaison des performances en termes de WSS en présence du bruit Voiture (SNR=0 à 15 dB par pas de 5 dB)

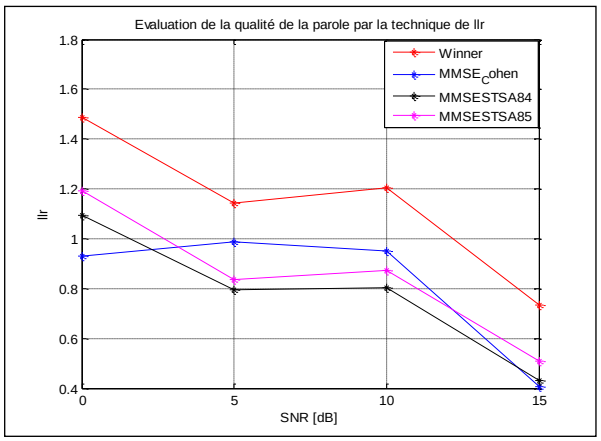


Figure IV-24: Comparaison des performances en termes de LLR en présence du bruit Voiture (SNR=0 à 15 dB par pas de 5 dB)

Dans la figure (IV-21) on observe que les qualités des méthodes MMSE-Cohen2004, MMSE-STSA84et MMSE-STSA85 sont presque semblables en terme de PESQ entre 0 et 10 dB, où elles représentent les meilleurs techniques. Par opposition, nous remarquons que la technique WienerScalart96 constitue la qualité la plus mauvaise entre 10 et 15 dB. Aussi, la méthode MMSE-Cohen2004 est meilleure, tandis que MMSE-STSA84 est la plus mauvaise.

Les figures (IV-22, IV-23 et IV-24) montrent que la technique de Wiener-Scalart96 est la plus haut en termes de segSNR, WSS et LLR. Ce qui signifie qu'elle est de meilleure qualité. Tandis que la technique MMSE-STSA84 constitue la méthode la plus mauvaise.

e) Bruit restaurant

Présence du bruit restaurant en utilisant la base de données NOIZEUS [56]. Les figures ci- dessous représentent la comparaison des performances en termes de ces critères d' SNR, de 0 à 15 dB par pas de 5 dB.

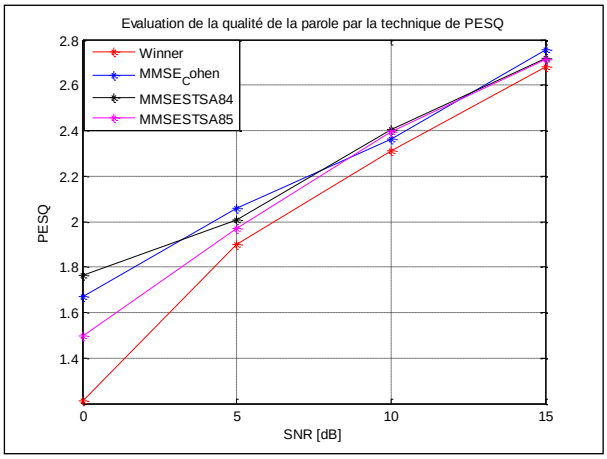


Figure IV-25: Comparaison des performances en termes de PESQ en présence du bruit Restaurant (SNR=0 à 15 dB par pas de 5 dB)

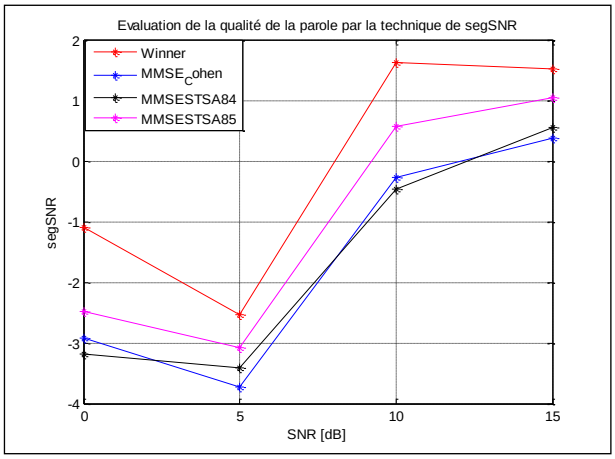


Figure IV-26: Comparaison des performances en termes de segSNR en présence du bruit Restaurant (SNR=0 à 15 dB par pas de 5 dB)

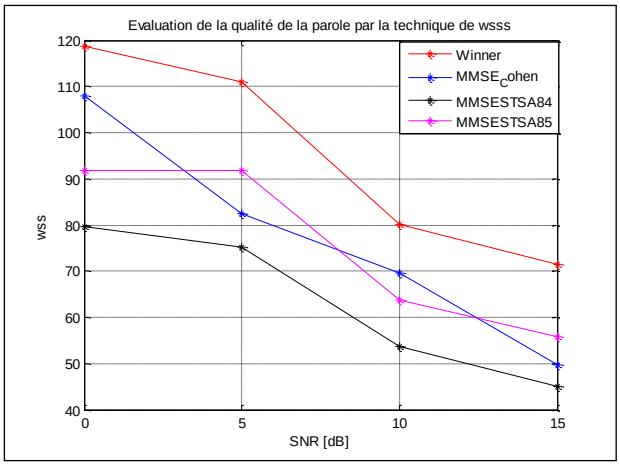


Figure IV-27: Comparaison des performances en termes de WSS en présence du bruit Restaurant (SNR=0 à 15 dB par pas de 5 dB)

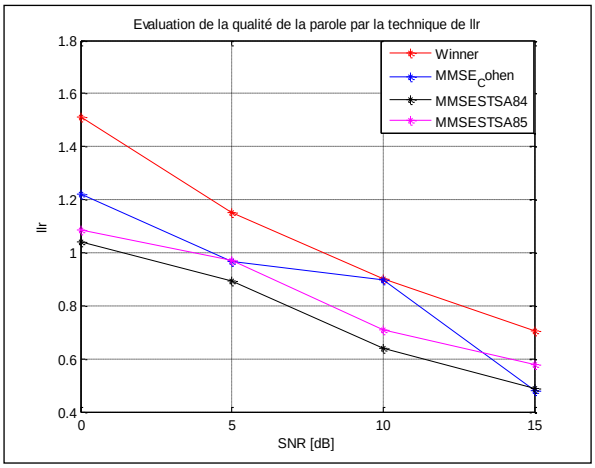


Figure IV-28: Comparaison des performances en termes de LLR en présence du bruit Restaurant (SNR=0 à 15 dB par pas de 5 dB)

La figure (IV-25) illustre les mesures de PESQ. Les méthodesMMSE-Cohen2004, MMSE-STSA84 et MMSE-STSA85 sont presque proches en terme de PESQ entre 5 et 15 dB avec une petite déférence qui en faveur de MMSE-Cohen2004 qui montre que elles sont les meilleur.Par contre la méthode la plus mauvaise est Wiener-Scalart96.

La figure (IV-26) montre que la technique de Wiener-Scalart96 est la plus haut en terme segSNR, qui signifiée qu'elle est dotée d'une meilleure qualité. Tandis que les deux méthodes MMSE-Cohen2004 et MMSE-STSA84 représentent les plus mauvaises.

Les figures (IV-27 et IV-28) montrent que la techniqueWiener-Scalart96 est la plus haut en termes de WSS et de LLR. Ce qui veut dire qu'elle est la meilleure. Alors

que les techniques MMSE-STSA85 et MMSE-Cohen2004 en terme de WSS et de LLR sont les plus mauvaises.

f) Bruit salle d'exposition

En présence du bruit salle d'exposition en utilisant la base de données NOIZEUS [56]. Les figures ci- dessous représente la comparaison des performances en termes des critères PESQ, segSNR, WSS et LLR d' SNR, de 0 à 15 dB par pas de 5 dB.

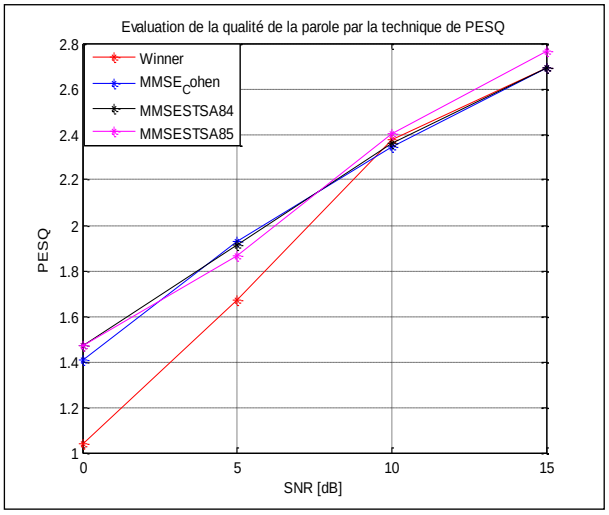


Figure IV-29: Comparaison des performances en termes de PESQ en présence du bruit salle d'exposition (SNR=0 à 15 dB par pas de 5 dB)

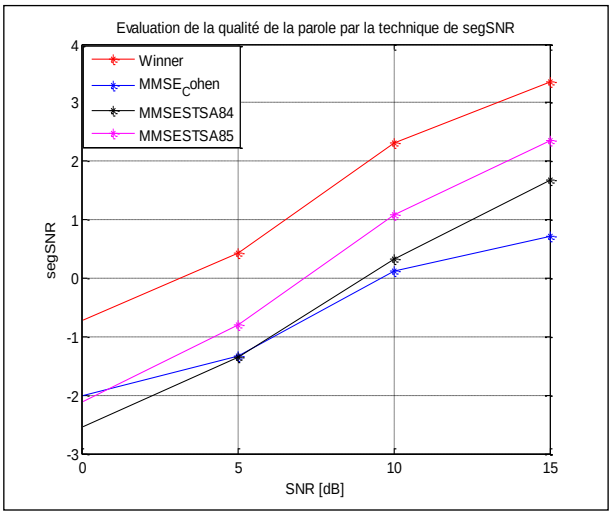


Figure IV-30: Comparaison des performances en termes de segSNR en présence du bruit salle d'exposition (SNR=0 à 15 dB par pas de 5 dB)

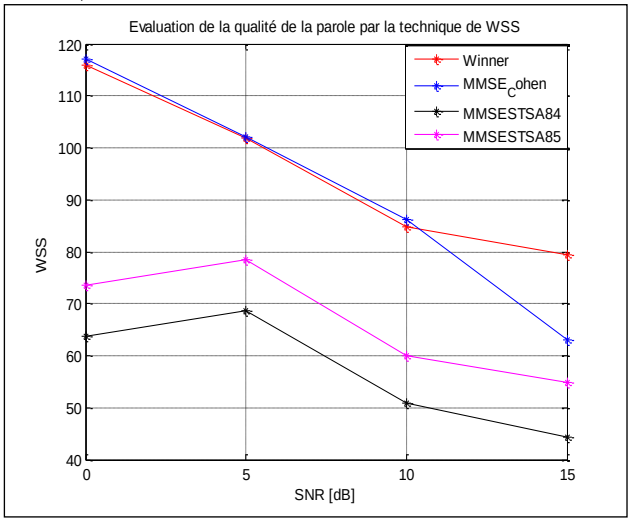


Figure IV-31: Comparaison des performances en termes de WSS en présence du bruit salle d'exposition (SNR=0 à 15 dB par pas de 5 dB)

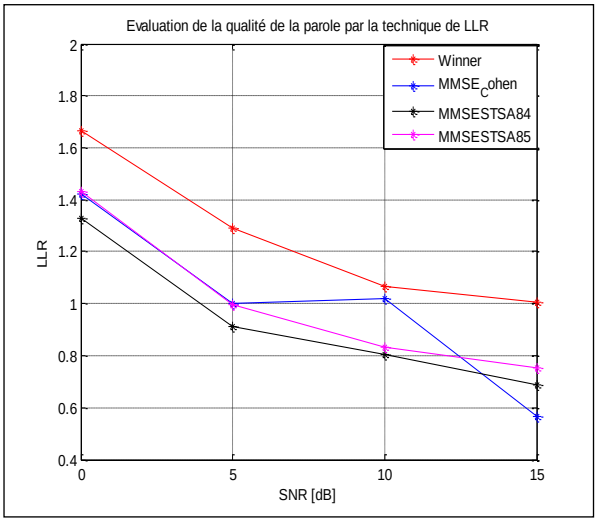


Figure IV-32: Comparaison des performances e termes de LLR en présence du bruit salle d'exposition (SNR=0 à 15 dB par pas de 5 dB)

Nous observons dans la figure (IV-29) que les méthodes deMMSE-Cohen2004, MMSE-STSA84et MMSE-STSA85 sont presque proches en termes de PESQ entre 0 et 10 dB et présentent les meilleurs techniques. Par contre la technique Wiener-Scalart96 est la plus mauvaise entre 10 et 15 dB.

Nous pouvons voir d'après la figure (IV-30) que la technique Wiener-Scalart96 est de meilleure qualité en termes de segSNR.Quandla méthode MMSE-STSA85 en termes de segSNR est un peu moins bonne.Pour les deux autres méthodes MMSE-Cohen2004 et MMSE-STSA84, elles sont les plus mauvaises (surtout pour la technique MMSE-Cohen2004).

Selon la figure (IV-31) les méthodes Wiener-Scalart96 et MMSE-Cohen2004 sont presque proches en termes de WSS entre 0 et 10 dB et représentés les meilleurs techniques. Par contre, la méthode MMSE-Cohen2004 est la plus mauvaise.

La figure (IV-32) montre que la technique de Wiener-Scalart96 est la plus haut en termes LLR, ce qui signifiée qu'elle est de meilleure qualité. Tandis que les techniques MMSE-Cohen2004 et MMSE-STSA85 comme les secondes en termes LLR. Pendant que la technique MMSE-STSA84 est la plus mauvaise.

Remarque : pour les critères objectives PESQ, SegSNR, WSS et LLR, que haut courbe est le meilleur par rapport les autres est ainsi de suite

Tableau IV.1: les calculs des moyennes des PESQ, segSNR, WSS et LLR pour les méthodes mentionnées précédemment pour la parole contaminée par les bruits de : bavardage, réel, voiture, blanc gaussienne, restaurant et salle d'exposition pour (SNR=0 à 15 dB par pas de 5 dB).

Méthodes	critères objectives	Types de bruit					
		Bavardage	Blanc gaussienne	Voiture	Salle d'exposition	Restaurant	Réel
Wiener Scalart96	PESQ	2.062	1.948	1.973	1.946	2.026	1.677
	LLR	1.121	1.564	1.142	1.258	1.066	1.561
	segSNR	-0.016	1.883	1.416	1.346	-0.121	1.637
	WSS	93.845	74.587	86.585	95.574	95.334	72.98
MMSE Cohen2004	PESQ	2.192	2.226	2.268	2.093	2.211	2.127
	LLR	0.996	1.211	0.818	1.001	0.890	1.218
	segSNR	-1.511	2.419	0.375	-0.623	-1.634	2.195
	WSS	77.43	72.4	74.878	92.077	77.422	72.99
MMSE STSA84	PESQ	2.277	2.317	2.223	2.109	2.233	2.244
	LLR	0.850	1.150	0.781	0.932	0.765	1.131
	segSNR	-1.383	1.009	-0.509	-0.472	-1.624	0.899
	WSS	60.617	44.22	47.055	56.912	63.431	43.467
MMSE STSA85	PESQ	2.259	2.269	2.243	2.127	2.144	2.138
	LLR	0.923	1.241	0.852	1.003	0.834	1.229
	segSNR	-0.890	1.283	0.089	0.128	-0.986	1.076
	WSS	72.925	50.635	56.553	66.753	75.822	46.915

IV.3 Comparaison en termes du temps d'exécution

Technique de rehaussement de la parole	Temps exécutions [sec]
WienerScalart96	**3.051**
MMSECohen2004	4.182
MMSESTSA84	3.393
MMSESTSA85	3.491

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

D'après le tableau ci-dessus qui montre les résultats des simulations en matière de temps d'exécution, nous pouvons observer bien que la méthode Wiener-Scalart96 a moins de temps d'exécution que les autres techniques.

IV.4 Conclusion

Dans ce chapitre nous avons fait une comparaison entre les déférentes méthodes d'amélioration du signal parole (éliminé le bruit) en termes de PESQ, segSNR, LLR et WSS et à l'aide de logiciel Matlab. On peut résumer les résultats de ce travail :

- On observe d'après le calcul de critère PESQ que la technique MMSE-STSA84 est la meilleure pour améliorer le signal parole bruité.
- On observe aussi d'après les calculs des critères WSS, LLR, segSNR, que la méthode la plus efficace pour éliminer le bruit est la méthode de WienerScalart96 est ce dernier à moins de temps d'exécution que les autres techniques.

Conclusion générale

Avec l'avènement des télécommunications mobiles grand public, le besoin d'améliorer la prise de son, notamment en réduisant la gêne due au bruit, s'est fait de plus en plus présent. Les techniques de réduction du bruit sont soumises à un compromis entre le niveau effectif de réduction et la distorsion qui affecte le signal de parole. Au vu des performances actuelles, il est souhaitable de supprimer plus de bruit tout en conservant un niveau de dégradation acceptable du signal restauré, ceci en particulier lorsque le niveau de bruit est important. Les techniques qui ont suscité le plus d'intérêt au cours de ces 30 dernières années sont les approches par atténuation spectrale à court terme qui consistent à modifier une transformée à court terme du signal bruité en utilisant une règle de suppression. L'essor de cette famille de techniques s'explique essentiellement par le fait qu'elles permettent de respecter les contraintes de temps réel et de complexité inhérentes aux applications de communication parlée.

Notre travail est devisé en quatre chapitre, le premier sert de présenter en général que ce que un signal parole, comment fonctionner l'appareil vocal et la numérisation de signal parolequi parcoure en trois étapes : l'échantionnage, la quantification et le codage.

Dans le deuxième chapitre nous avons présenté les différents techniques de débruitage de signal parole dans le cas monovoies .Nous intéressons par les méthodes de Wiener-Scalart96, MMSE-Cohen2004, MMSE-STSA84 et MMSE-STSA85, la soustraction spectrale au sens d'Ephraim et Malah.Aussi nous expliquons dans ce chapitre le principe de l'algorithme de débruitag

Nous exposons dans le troisième chapitre les techniques d'évaluation de la qualité de signale parole débruité d' après les mesures des critères objectifs : PESQ, segSNR, WSS, LLR (qui sont l'en basé dans notre étude dans la partie de simulation) et les critères subjectifs : MOS, DMOS et CMOS.

Enfin le dernier chapitre continent les résultats de simulation à l'aide logiciel MATLAB et contient aussi la comparaison entre les différentes méthodes en mesurant les critères objectifs tel que PESQ, segSNR, WSS, LLR, ou ses résultats démontrent à nous que la meilleure technique pour améliorer le signal parole bruité (bavardage, Blanc gaussienne , voiture, salle d'exposition Restaurant, Réel) est le filtrage de Wiener d'après le résultat de

mesure des critères WSS, segSNR, LLR. Et la technique MMSESTS84 est la meilleure d'après la mesure de critère PESQ.

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.
- 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*...).

Bibliographie

[1] Damien Vincent. Thèse« Analyse et contrôle du signal glottique en synthèse de laparole»l'École Nationale Supérieure des Télécommunications de Bretagne 2007.

[2] LE Manh Tuan « Analyse des voyelles spéciale du Vitnamien ». Institut de laFrancophonie pour l'Informatique En collaboration avec le Centre de Recherche MICA, Hanoi.2005.

[3] Lawrence R. Rabiner& Ronald W. Schafer. Introduction to digital speech processing. NowPublishers Inc., Hanover, MA, USA, 2007.

[4] Thierry Dutoit,Introduction auTraitement Automatiquede la Parole,le 20 octobre2000.

[5] Asmaa Amehraye,Débruitage perceptuel de la parole,15 mai 2009.

[6] Octobre 2011, Analyse en ondelette des potentiels évoqués auditifs en réponse à dessons de parole (SABR) : une nouvelle analyse objective de l'encodage sous corticaldes sons de parole voisés et non voisés, SFORL 2010, Paris (75), Présentation orale.

[7] Arnaud Jeanvoine, Intérêt des algorithmes de réduction de bruit dans l'implant cochléaire : Application à la binauralité,Submitted on 22 Oct 2013.

[8] BENAMMAR Ryadh, Traitement Automatique De La Parole Arabe Par Les HMMs: Calculatrice Vocale, 25 Septembre 2012.

[9] Bernard Gosselin. Représentation de l'information et quantification des signauxFaculté Polytechniques de Mons, 2000. Belgique.

[10] Young & all. Thehtk book (for htk version 3.4). Cambridge University Engineering Department, 2006.

[11] N. Virag. Single channel speech enhancement based on masking properties of the human auditory system. IEEE Trans. Speech and AudioProcessing, vol. 7pages 126–137, 1999.

[12] Arnaud Jeanvoine,Intérêt des algorithmes de réduction de bruit dans l'implant cochléaire : Application à la binauralité,Submitted on 22 Oct 2013.

[13] Steven F.Boll.Suppression of acoustic noise in speech using spectral subtraction.IEEE,27(2):113-120,Appril 1979.

[14] M.Berouti, Bolt Beranek, Newman, R. Schwartz, and J. Makhoul.Enhancement of speech corrupted by acoustic noise.Acoustics, Speech,and Signal Processing, IEEE International Conference on ICASSP 79, 4:208–211, Apr 1979.

[15] J.Lim and A. Oppeenheim.Enhancement and bandwidth compression of noisy speech. IEEE, 67:1586–1604, 1979.

[16] Y. Ephraim and D. Malah. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator, IEEE transaction on acoustics speech and signal processing, vol.assp-33,no 2, December 1985.

[17] DO. Walter. A posteriori wiener filtering of average evoked responses. ElectroencephalogrClinNeurophysiol, Suppl27:61- 1968.

[18] DJ. Doyle. Some comments on the use of wiener filtering for the estimation of evoked potentials. ElectroencephalogrClinNeurophysiol, 38(5):533–4,May 1975.

[19] M. Hansen. Effects of multi-channel compression time constants on subjectively perceived sound quality and speech intelligibility. Ear Hear, 23(4):369–80, Aug 2002.

[20] J. Sohn, N. S. Kim, and W. Sung. Statistical model-based voice activity detection. IEEE Signal Process Letts, 6:1–3, 1999.

[21] R. Martin. Spectral subtraction based on minimum statistics. Pages 1182–1185, 1994.

[22] Y. Ephraim & D. Malah. Speech enhancement using a minimum mean square error short-time spectral amplitude estimator. IEEE Trans. Acoustic, Speech, Signal Processing, vol. ASSP-32, pages 1109–1121, Dec 1984.

[23] O. Cappe. Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor. IEEE Trans. on Speech and Audio Processing, vol. 2(2), pages 345–349, Avr 1994.

[24] P. Loizou. Speech enhancement: Theory and practice.CRC; 1 edition, 2007.

[25] Y. Ephraim & H.L. Van Trees. A signal subspace approach for speech enhancement. IEEE Trans. Speech and Audio Processing, vol. 3, pages 251–266, 1995.

[26] F. Jabloun& B. Champagne.Incorporating the human,lebouquin hearing properties in the signal subspace approach for speech enhancement. IEEE Trans. Speech and Audio Processing, vol. 11, pages 700–708, 2003.

[27] Kris Hermus, Patrick Wambacq& Hugo Van hamme. A review of signal subspace speech enhancement and its application to noise robust speech recognition. EURASIP J. Appl. Signal Process., vol. 2007, no 1, pages 195–195, 2007.

[28] Y. Hu & P. Loizou. Evaluation of objective Measures for speech enhancement. in Proc. Interspeech, pages 1447–1450, 2006.

[29] A. Amehraye, D. Pastor & A. Tamtaoui. Perceptual improvement of Wiener filtering. In Proc. of ICASSP, Las Vegas, USA, pages 2081–2084, 2008.

[30] A. Amehraye, L. Fillatre, D. Pastor & D. Aboutajdine.A perceptual filter for unmasked noise prevention. To besubmitted to Speech Communications, 2009.

[31] B. Virole. Psychologie de la surdité. 2éme Edition, De Boeck Université, 2001.

[32] Y. Hu & P. Loizou.A comparative intelligibility study of speech enhancement algorithms.IEEE Signal ProcessingLetters, vol. 4, pages 561–564, Apr 2007.

[33] L. Bunieti. Traitement automatique de la parole en milieu bruité : étude de modèles connexionnistes statiques et dynamiques. Université Henri Poincaré – Nancy 1, 1997.

[34] S. R. Quackenbush, T. P. Barnwell III & M. A. Clements.Objective Measures of Speech Quality. Englewood Cliffs,NJ: Prentice-Hall. 1988.

[35] G. Fairbanks. Test of phonetic differentiation: the rhyme test. Journal of the Acoustical Society of America, vol. 30,pages 596–600, 1958.

[36] IEC-Standard.60268-16. Sound system equipment- Part16 : Objective rating of speech intelligibility by speechtransmission index. 1998.

[37] ANSI. Method for Measuring the Intelligibility of Speechover Communication Systems. 1989.

[38] ANSI. S3.5. American National Standard Methods forCalculation of the Articulation Index. 1969.

[39] S. Wang, A. Sekey& A. Gersho.An objective measure forpredicting subjective quality of speech codes.IEEE J. on Select.Areas in Commun.vol. SAC-10, pages 819–829,Sept 1992.

[40] Y. Hu & P. Loizou.Incorporating a psychoacoustic modelin frequency domain speech enhancement. IEEE SignalProcessing Letters, vol. 11(2), pages 270–273, Feb 2004.

[41] N. Ma, M. Bouchard & R. A. Goubran. Perceptual Kal- man filtering for speech enhancement in colored noise.In Proc. ICASSP'04, Montreal, Canada, volume 4, pages1045–1048, 2004.

[42] UIT-T P.861. Objective quality measurement of telephone-band (300-3400 Hz) speech codes. 1998.

[43] UIT-T P862. Perceptual evaluation of speech quality(PESQ), an objective method for end-to-end speech qualityassessment of narrowband telephone networks and speech codes. 2000.

[44] W. Yang, M. Dixon & R Yantorno. Modified bark spectraldistortion measure which uses noise masking threshold. InIEEE Speech coding Workshop, pages 55–56, 1997.

[45] Hansen, J.H.L., Pellom, B.L.: An effective quality evaluation protocol for speech enhancement algorithms. In: Proceedings of the International Conference on Spoken Language Processing (ICSLP), vol. 7, pp. 2819–2822 (1998)

[46] Y. Tohkura. A Weighted Cepstral Distance Measure for Speech Recognition. IEEE Trans. Acoust., Speech & Signal Process., vol. ASSP-35, pages 1414–1422, 1987.

[47] H. Kobatake&S.Matsunoo. Degraded word recognition based on segmental signal-to-noise ratio weighting. In Proc. ICASSP 04, Adelaide, SA, Australia, volume I, pages 425–428, 1994.

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

[49] S. Wang, A. Sekey&A.Gersho. Modified bark spectral distortion measure which uses noise masking threshold. IEEE Speech Coding Workshop, vol. SAC-10, 1997.

[50] W. Yang, M. Dixon & R Yantorno. Enhanced modified Bark spectral distortion (EMBSD): An objective speech quality measure based on audible distortion and cognition model. Phd thesis, Temple University Graduate Board, May 1999.

[51] J. Beerendes& J. Stemerdink. A perceptual audio quality measurement based on a psychoacoustic sound representation. J. Audio Eng. Soc, vol. 40, pages 963–972, 1992.

[52] Y. Hu & P. Loizou. Evaluation of objective Measures for speech enhancement. in Proc. Interspeech, pages 1447– 1450, 2006.

[53] A. Rix, J. Beerends, M. Hollier& A. Hekstra. Perceptual evaluation of speech quality (pesq) a new method for speech quality assessment of telephone networks and codes. In Proc. ICASSP'04, Adelaide, SA, Australia, volume I, pages 749–752, 2001.

[54] B. Grundlehner, J. Lecoq, R. Balan& J. Rosca. Performance assessment method for speech enhancement systems. In Proc. IEEE BENELUX/DSP Valley signal processing symposium, 2005.

[55] Y. Hu & P. Loizou. Evaluation of objective quality Measures for speech enhancement. Evaluation of objective Measures for speech enhancement, vol. 16, pages 229–238, Jan 2008.

[56] S. Keagy. Integrating voice and data networks: Practical solutions for the new world of packetized voice over data networks.Cisco Press, 2000.

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

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