

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



M/ELE.SY.COM
2012/09

جامعة جيجل
مكتبة كلية العلوم والتكنولوجيا
رقم المذكرة: M 1163
02/02

Université de Jijel
Faculté des Sciences et de la Technologie
Département d'Electronique

Projet de fin d'études pour l'obtention du Diplôme de
Master II en Electronique

Option : Electronique et Systèmes de Communication

Thème :

**Application du filtre de Wiener
au rehaussement de la parole**

Présenté par :

M^{elle} KISSOUM Asma

Encadré par :

Mr. BOUBAKIR Chaâbane

Promotion : Juin 2012.

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



Université de Jijel
Faculté des Sciences et de la Technologie
Département d'Electronique

Projet de fin d'études pour l'obtention du Diplôme de
Master II en Electronique

Option : Electronique et Systèmes de Communication

Thème :

**Application du filtre de Wiener
au rehaussement de la parole**

Présenté par :

M^{elle} KISSOUM Asma

Encadré par :

Mr. BOUBAKIR Chaâbane

Promotion : Juin 2012.

Remerciements

La louange avant tout à Dieu seul qui m'a donné la force, le courage et la volonté pour accomplir ce modeste travail.

J'adresse mes plus sincères remerciements à mon promoteur Mr. BOUBAKIR Chaâbane d'avoir proposé et dirigé ce projet, je tiens également à lui exprimer ma profonde gratitude pour son aide, sa disponibilité et ses conseils attentifs.

Mes vifs remerciements à tous les membres de jury de m'avoir honoré en acceptant de juger et évaluer ce modeste travail.

Je veux remercier tous les enseignants qui ont participé à ma formation de près ou de loin.

Enfin, je remercie toute personne ayant participé de près ou de loin à l'élaboration de ce travail.

Dédicaces

Je dédie ce travail à mes chers parents pour leurs encouragements et leur soutien moral et matériel durant toutes mes années d'études que Dieu le tout puissant me les gardes.

Ma chère mère et mon cher père

A mes chers frères : Younes et Mohammed Akrame

A mes chères sœurs : Sara, Safia et Selma

Je dédie également ce travail à ma grande famille « KISSOUM »,

A mes chères amies : Houria, Khadija, Souad, Besma et Houda

A mes collègues de la promotion d'électronique 2012 et à toutes mes amies de l'université de jijel.

A tous ceux qui ont participé de près ou de loin à la réalisation de ce travail.

ASMA

Liste des figures

Figure 1.1 : L'appareil phonatoire humain	4
Figure 1.2 : Synoptique du PESQ	16
Figure 2.1 : Principe de débruitage d'un signal de parole	19
Figure 2.2 : Schéma général du problème de filtrage statistique	20
Figure 2.3 : Courbe de l'atténuation du filtre de Wiener en fonction du SNR a priori.....	28
Figure 2.4 : Courbes d'atténuation du filtre de Wiener paramétrique ($\alpha = 1, \beta$ variable)..	34
Figure 2.5 : Courbes d'atténuation du filtre de Wiener paramétriques(α variable, $\beta = 1$)..	35
Figure 2.6 .a) : Formes d'ondes, cas d'un bruit blanc avec $\alpha = 1, \beta = 0.5$ et un SNR=0 dB.....	40
Figure 2.6 .b): Spectrogrammes, cas d'un bruit blanc avec $\alpha = 1, \beta = 0.5$ et un SNR=0....	40
Figure 2.7 .a) : Formes d'ondes, cas d'un bruit blanc avec $\alpha = 1, \beta = 0.5$ et un SNR=5 dB.....	41
Figure 2.7 .b): Spectrogrammes, cas d'un bruit blanc avec $\alpha = 1, \beta = 0.5$ et un SNR=5 dB	41
Figure 2.8 .a) : Formes d'ondes, cas d'un bruit blanc avec $\alpha = 2, \beta = 1, \text{SNR}=0 \text{ dB}$	42
Figure 2.8 .b): Spectrogrammes , cas d'un bruit blanc avec $\alpha = 2, \beta = 1$ et un SNR=0 dB	42
Figure 2.9 .a) : Formes d'ondes , cas d'un bruit blanc avec $\alpha = 2, \beta = 1, \text{SNR}=5 \text{ dB}$	43
Figure 2.9 .b): Spectrogrammes , cas d'un bruit blanc avec $\alpha = 2, \beta = 1$ et un SNR=5 dB	43
Figure 3.1 : Organigramme du filtre de Wiener itératif.....	46
Figure 4.1 : schéma bloc du filtre de Wiener itératif avec contrainte	61

Liste des tableaux

Tableau 1.1 : Classification du bruit	10
Tableau 1.2 : Test subjectif de qualité de type MOS.....	12
Tableau 2.1 : Mesures objectives pour un bruit blanc, ($\alpha = 1$ et β variable).....	37
Tableau 2.2 : Mesures objectives pour un bruit babble, ($\alpha = 1$ et β variable).....	37
Tableau 2.3 : Mesures objectives pour un bruit blanc, ($\beta = 1$ et α variable).....	38
Tableau 2.4 : Mesures objectives pour un bruit babble, ($\beta = 1$ et α variable).....	38
Tableau 3.1 : Mesures objectives, cas d'un bruit blanc pour différentes itérations.....	54
Tableau 3.2 : Evaluation objective de la qualité pour un bruit blanc.....	55
Tableau 3.3 : Evaluation objective de la qualité pour un bruit de voiture.....	55
Tableau 3.4 : Evaluation objective de la qualité pour un bruit de parole.....	55
Tableau 4.1 : Valeurs de la pondération triangulaire.....	61
Tableau 4.2 : Evaluation objective de la qualité pour un bruit blanc.....	62
Tableau 4.3 : Evaluation objective de la qualité pour un bruit de parole.....	62

Liste des abréviations

ACR	Absolute Category Rating
AR	Auto Regressif
ARMA	Auto Regressif à Moyenne Ajustée
BSD	Bark Spectral Distortion
DFT	Discrete Fourier Transform
DMOS	Degradation Mean Opinion Score
DRT	Diagnostic Rhyme Test
DSP	Densité Spectrale de Puissance
EQM	Erreur Quadratique Moyenne
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
IFFT	Inverse Fast Fourier Transform
IIR	Infinite Impulse Response
IS	Itakura-Saito
LAR	Log Area Ratio
LLR	Log –Likelihood-Ratio
LP	Linear Prediction
LPC	Linear predictive Coding
LSF	Line Spectrum Frequency
LSP	Line Spectrum Position
MAP	maximum a posteriori
ML	Maximum Likelihood
MM- LSA	Multiplicatively Modified – Log Spectral Amplitude
MMSE	Minimum Mean Square Error
MOS	Mean Opinion Score
MRT	Modified Ryme Test
MS	Minimum Statistics method
MSE	Mean Square Error
MBSD	Modified Bark Spectral Distortion

PESQ	Perceptual Evaluation of Speech Quality
PSQM	Perceptual Speech Quality Measure
REL P	Risidual Excited Linear Prediction
RSB	Rapport Signal sur Bruit
STSA	Short Time Spectral Amplitude
SVQ	Split Vector Quantization
WLD	Wiener Levinson Durban
WSS	Weighted Slope Spectral distance

Sommaire

Liste des figures.....	i
Liste des tableaux.....	ii
Liste des abréviations.....	iii
Introduction générale.....	01

Chapitre 1 : Généraliés

1.1 Introduction.....	3
1.2 Généralités sur le signal de la parole.....	3
1.2.1 Le mécanisme de phonation.....	3
1.2.2 Classification des sons de la parole	4
1.2.3 Caractéristiques du signal de la parole.	4
1.3 Nature et caractéristiques du bruit.	6
1.3.1 Classification du bruit.....	6
1.3.2 Physiologie du bruit.....	10
1.4 Mesures de qualité.....	11
1.4.1 Mesures subjectives.....	12
1.4.2 Mesures objectives	13
1.5 Conclusion	17

Chapitre 2 : Etude de la théorie du filtre de Wiener

2.1 Introduction.....	18
2.2 Généralités sur le rehaussement de la parole.....	18

2.2.1 Divers types de dégradation de la parole.....	18
2.2.2 Intérêt du rehaussement de la parole	19
2.2.3 Classification des techniques de rehaussement	19
2.2.4 Systèmes monovoie.....	20
2.3 Introduction à la théorie du filtre de Wiener	20
2.3.1 Filtre de Wiener dans le domaine temporel.....	21
2.3.2 Filtre de Wiener dans le domaine fréquentiel.....	24
2.3.3 Filtres de Wiener pour la réduction du bruit.....	26
2.4 Estimation du SNR a priori	29
2.4.1 Approche d'estimation du maximum de vraisemblance	29
2.4.2 Approche d'estimation de décision dirigée (decision directed)	31
2.5 Filtre de Wiener racine carrée (Square-Root Wiener Filter)	32
2.6 Filtre de Wiener paramétrique	33
2.7 Simulations et résultats.....	35
2.7.1 Conditions d'expérimentations	35
2.7.2 Evaluation des performances	36
2.7.3 Interprétations	38
2.6 Conclusion	44

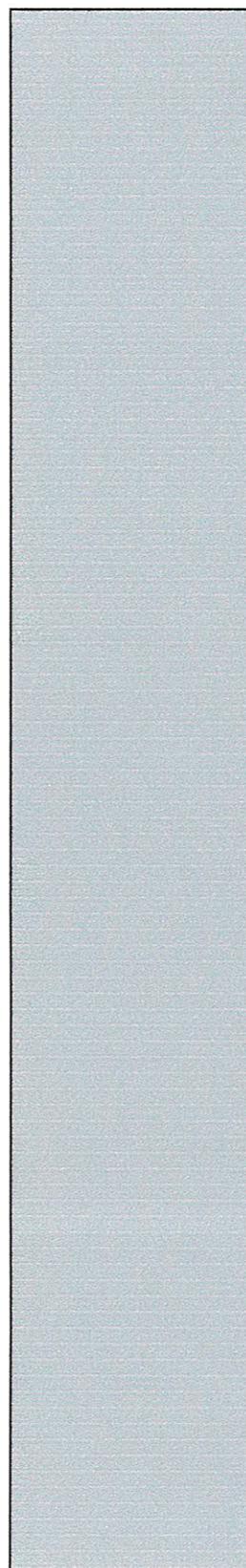
Chapitre 3 : Filtre de Wiener itératif

3.1 Introduction.....	45
3.2 Filtre de Wiener itératif.....	45
3.3 Calcul itératif basé sur le modèle LPC.....	46
3.3.1 Modèle mathématique de la production de la parole.....	46
3.3.2 Estimation statistique des paramètres du modèle tout-pôles dans le bruit.....	48
3.4 Test et résultats.....	54
3.5 Conclusion	56

Chapitre 4 :Filtre de Wiener itératif avec contraintes

4.1 Introduction.....	57
4.2 Filtre de Wiener itératif avec contraintes.....	57
4.2.1 Contraintes spectrales au cours du temps (interframe)	57
4.2.2 Contraintes au cours des itérations.....	60
4.3 Filtre de Wiener itératif avec contraintes.	60
4.4 Tests et résultats	61
4.5 Conclusion	63
Conclusion générale.....	64
Bibliographie et Webographie.....	65

Introduction générale



Introduction générale

Les systèmes de télécommunications sont en constante évolution depuis plusieurs années et les applications de téléphonies numériques et mobiles se multipliant, le codage et la reconnaissance de la parole ont bénéficié d'un grand nombre d'études. A cause des perturbations et des bruits apparaissant dans ces applications, le signal reconstruit ne pouvait être une réplique du signal propre.

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

L'investigation des méthodes de rehaussement de la parole a permis d'atteindre des résultats spectaculaires. Cependant, certaines problématiques et questions restent ouvertes. Une de ces problématiques est de parvenir à un compromis entre la réduction du bruit, la distorsion du signal et le bruit résiduel et musical. Le besoin de réduire ce type de bruit tout en préservant la qualité et l'intelligibilité de la parole a poussé les chercheurs à proposer d'autres solutions à ce problème, mais aussi à réduire certaines limitations des systèmes mono-capteur de débruitage de la parole.

Les méthodes basées sur le filtre de Wiener ont suscité beaucoup d'intérêt ces dernières décennies. Des améliorations importantes ont été apportées sur ce type de filtrage, dans le but d'atteindre une réduction efficace du bruit en limitant la présence du bruit musical et la distorsion du signal de parole rehaussé.

Ce travail présente l'application du filtre de Wiener et ses variantes au rehaussement de la parole en situation monovoie et contient quatre (04) chapitres :

Le premier chapitre présentera des notions élémentaires et les termes relatifs à la description de la parole et du bruit, ainsi que les mesures de qualité pour bien juger la qualité et l'intelligibilité de la parole.

Le deuxième chapitre, abordera le rehaussement de la parole. On présentera une description détaillée sur la théorie du filtre de Wiener classique et ses deux variantes paramétrique et racine carrée, en plus de ses résultats d'implémentation.

Au cours du troisième chapitre, on étudiera des améliorations apportées au filtre de Wiener de type itératif et leur impact sur les performances du rehaussement de la parole. Ainsi que les résultats de simulation seront présentée.

Dans le dernier chapitre, on appliquera des contraintes spectrales sur le filtre de Wiener itératif, pour augmenter d'avantage la qualité du signal rehaussé.

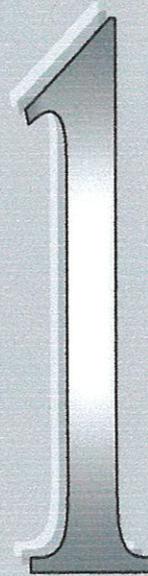
Enfin, la conclusion générale donnera les interprétations appropriées aux différents résultats et une vue sur les perspectives du travail.

Chapitre

Généralités

Dans ce chapitre :

- **Introduction.**
- **Généralités sur le signal de la parole.**
- **Nature et caractéristiques du bruit.**
- **Mesures de qualité.**
- **Conclusion.**

A large, stylized number '1' is centered vertically within a light blue rectangular background. The number has a 3D effect with a white highlight on its left side and a dark shadow on its right side.

1.1 Introduction

Si l'on tente une définition, la parole est la capacité de l'être humain de communiquer la pensée par l'intermédiaire de sons articulés, c'est le moyen de communication privilégié entre les humains qui sont les seules êtres vivants à utiliser un tel système structuré. Dû à son importance, la parole a préoccupé depuis toujours les scientifiques.

Le but de ce chapitre est de présenter des notions sur la parole et le bruit. Pour cela on présentera tout d'abord le principe de production de la parole, ses caractéristiques, les techniques d'analyse les plus utilisées, ainsi que les sources et les types de bruit, dernière chose il faut définir les mesures de qualité pour pouvoir comparer les méthodes utilisées.

1.2 Généralités sur le signal de la parole

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 les interprète. Au sens strict, le contenu d'un signal vocal est représenté uniquement par son intelligibilité. Dans un sens plus large, il faut aussi tenir compte de tout ce qui représente la qualité du signal vocal, c'est-à-dire les intonations, la prosodie et les perturbations du milieu ambiant. On va ci-après décrire certaines propriétés de l'appareil phonatoire humain [1].

1.2.1 Le mécanisme de phonation

Le signal vocal est le résultat de plusieurs actions conjuguées. En effet, la parole résulte de l'action coordonnée des appareils respiratoires et masticatoires contrôlés par le système nerveux central.

L'appareil respiratoire fournit l'air qui est expiré par la trachée artère. L'air passe ensuite par le larynx où sa pression est modulée grâce aux cordes vocales qui déterminent la taille de l'ouverture (la glotte) par laquelle il peut passer. Finalement, l'air transite par le conduit vocal qui s'étend du pharynx aux lèvres pour devenir le signal vocal qui est émis par le locuteur [2]. La figure (1.1) représente un schéma général de l'appareil phonatoire humain.

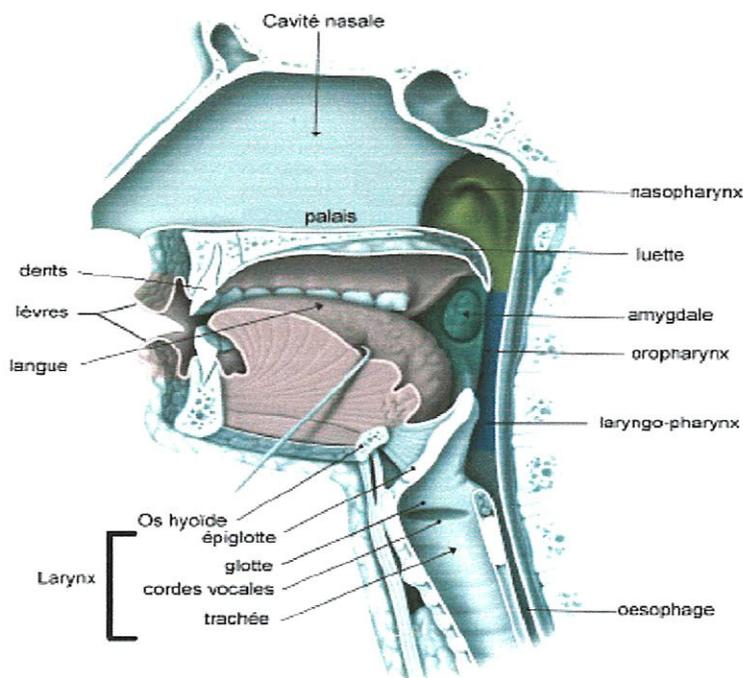


Figure 1.1 : L'appareil phonatoire humain [3].

1.2.2 Classification des sons de la parole

On peut distinguer deux grandes classes de sons émis par l'appareil phonatoire : les sons voisés et les sons non voisés.

– Les sons voisés, de forte énergie, résultent de l'excitation du conduit vocal par des impulsions périodiques de pression liées aux oscillations des cordes vocales. À chaque impulsion, la glotte s'ouvre brusquement et libère la pression accumulée en amont puis elle se referme plus graduellement.

– Les sons non voisés, beaucoup moins énergétiques que les sons voisés, résultent quant à eux de l'écoulement libre de l'air par la glotte qui reste ouverte. Le chuchotement est un mode de phonation particulier car la glotte reste ouverte en permanence, il n'y a donc pas de production de sons voisés pendant le chuchotement.

1.2.3 Caractéristiques du signal de la parole

Le signal de parole est un signal non-stationnaire : ses propriétés statistiques changent au cours du temps. La non stationnarité résulte des changements au cours du temps de la source ainsi que la forme et les dimensions du conduit vocal. Pour avoir des signaux localement stationnaires, l'utilisation d'une fenêtre glissante dans le temps est donc indispensable.

Cette stationnarité est assurée pour des fenêtres de durée compatible avec les ordres de grandeurs typiques de variations des paramètres acoustiques. Le découpage temporel consiste alors à multiplier le signal original par la fenêtre choisie, décalée dans le temps.

Comme les paramètres acoustiques évoluent rapidement, il est nécessaire d'utiliser un pas d'analyse court (en général de 5 ms à 20ms) [4].

➤ Fenêtrage

Le but du fenêtrage est de décomposer le signal de parole en petites tranches (chacune de 20ms environ) où il peut être considéré comme localement quasi-stationnaire. De plus, et pour profiter de l'évolution lente du signal vocal, le fenêtrage permet le traitement en temps réel et facilite aussi l'analyse des signaux.

Il existe plusieurs types de fenêtres d'analyse, comme la fenêtre rectangulaire, la fenêtre de Hanning et la fenêtre de Hamming. Les deux dernières sont les plus convenables à la parole, car elles entraînent un minimum de distorsion spectral du signal de la parole, par rapport aux autres fenêtres. La fenêtre de Hamming est donnée par :

$$w(n) = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi n}{N}\right), & 0 \leq n \leq N-1 \\ 0, & \text{ailleurs} \end{cases} \quad (1.1)$$

Et celle de Hanning par :

$$w(n) = \begin{cases} 0.5 + 0.5 \cos\left(\frac{2\pi n}{N}\right), & 0 \leq n \leq N-1 \\ 0, & \text{ailleurs} \end{cases} \quad (1.2)$$

➤ Variabilité

Le signal de la parole est un signal très complexe, par ce qu'il englobe des informations linguistiques, des informations caractéristiques du locuteur et des informations relatives au matériel utilisé pour la transmission ou l'enregistrement du signal, etc. De ce fait, le signal de parole présente une très grande variabilité; variabilité due au locuteur, et variabilité due au matériel.

➤ **Fréquence fondamentale ou F0**

La fréquence fondamentale est la fréquence de vibration des cordes vocales. Ses variations définissent le pitch qui constitue la perception de la hauteur (où les sons s'ordonnent de grave à aigu). Seuls les sons quasi-périodiques (voisés) engendrent une sensation de hauteur tonale bien définie. La fréquence fondamentale peut varier selon le genre (masculin ou féminin) et l'âge du locuteur [5] :

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

➤ **Le spectre fréquentiel**

Le spectre fréquentiel dépend principalement du 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.

➤ **L'énergie**

L'énergie correspond à l'intensité sonore, elle est habituellement plus forte pour les segments voisés de la parole que pour les segments non voisés.

1.3 Nature et caractéristiques du bruit

Le bruit est tout signal nuisible qui se superpose au signal utile en un point quelconque d'une chaîne 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.

1.3.1 Classification du bruit

Plusieurs méthodes de classification des sources de bruits peuvent être opérées (il n'existe pas de classification typique). Ces classifications dépendent le plus souvent de l'application et de l'objectif poursuivis. Ainsi, selon l'importance accordée à tel ou tel paramètre du signal, une classification peut apparaître comme étant plus intéressante qu'une autre. On donne dans ce qui suit, celles qui sont les plus communément rencontrées [6].

- **Classification basée sur la source du bruit**

Les bruits peuvent être endogènes ou exogènes au système.

Les sources de bruits externes : agissent sur le système par influence. Leur origine peut être naturelle ou artificielle. Les perturbations artificielles (*man-made noise*) sont principalement générées par les équipements et les parasites créés par les activités humaines : utilisation des appareils électroménagers, imprimantes, etc... Les perturbations naturelles, quant à elles, sont associées aux conditions et au milieu de propagation (influence des conditions atmosphériques, Fading , etc.). Leur intensité dépend le plus souvent de facteurs spécifiques tels que la localisation, l'environnement électromagnétique, l'architecture et le mode d'utilisation du système, etc.... Dans ce cas, une évaluation des performances (du système de reconnaissance par exemple) n'est possible que lors d'un fonctionnement in situ.

Les sources de bruits internes : sont très souvent difficiles à éliminer. Le cas du bruit thermique est le plus marquant. Généralement, le bon fonctionnement du système n'est assuré que si le niveau du signal utile dépasse de plusieurs ordres de grandeurs le bruit (RSB de quelques dizaines de décibels). Toutefois, dans la réalité, on est souvent confronté au traitement de signaux à très faible RSB, en exploitant au mieux les connaissances *a priori* dont on dispose à la fois sur ces signaux et sur les bruits les entachant.

- **Classification basée sur la stationnarité**

Un bruit peut être stationnaire ou non stationnaire.

Un processus (bruit) est dit non stationnaire si ses propriétés statistiques (moyenne, écart type, etc.) varient en fonction du temps. Généralement travailler avec une hypothèse de non stationnarité est très difficile. Même si le concept de processus stationnaire n'est qu'un modèle simplifié, il est cependant commode d'emploi et très utile en pratique surtout lorsqu'on se met dans un contexte d'observation limité (fenêtres temporelles, ce qui est le cas en parole) pendant lequel le phénomène (bruit) peut être considéré comme présentant un caractère permanent.

- **Classification basée sur la densité de probabilité**

Ne pouvant disposer ni d'une formulation analytique et ni évidemment d'une représentation mathématique déterministe du processus aléatoire, on le caractérise alors par ses propriétés statistiques. La théorie des probabilités est alors utilisée pour la description

statistique des bruits et autres signaux aléatoires. On peut alors distinguer au moins trois types de bruits.

Le bruit blanc : est un processus dont la densité spectrale de puissance est constante pour toute valeur de la fréquence. Il est ainsi appelé par analogie à la lumière blanche qui est composée des radiations de toutes les longueurs d'onde.

Un tel bruit est bien sûr théorique car sa puissance moyenne est infinie. Cependant ce concept est très utilisé dans les cas où le spectre réel peut être, en première approximation, remplacé par un spectre constant.

Le bruit Gaussien : se dit d'un processus qui possède une densité de probabilité Gaussienne. L'importance des processus Gaussiens résulte du fait qu'ils sont le modèle asymptotique d'un grand nombre de phénomènes naturels. D'autre part, ils sont entièrement définis par leurs valeurs moyennes et covariances. L'autre propriété fondamentale réside dans le fait qu'un bruit Gaussien, lorsqu'il transite par un système linéaire, reste toujours Gaussien.

Le bruit Poissonien : permet l'étude de nombreux phénomènes résultant de la réalisation d'événements aléatoires distribués dans le temps. Ceux-ci peuvent être représentés par une séquence aléatoire d'événements qui sont susceptibles de survenir à n'importe quel instant avec la même probabilité.

• Classification fréquentielle

Une classification basée sur la répartition de l'énergie du bruit en fréquence peut également être utilisée.

Le bruit à large bande fait référence à l'étendue du spectre du bruit qui est définie entre deux fréquences de coupures F_1 et F_2 est dont la différence $F_2 - F_1$ qui constitue cette étendue est relativement élevée.

Le bruit à bande étroite est un processus aléatoire dont le spectre constant est circonscrit dans une bande de fréquences B finie relativement étroite, définie entre deux fréquences F_1 et F_2 avec $B = F_2 - F_1$. Le spectre est nul en dehors de cette bande. Cette classification est générale et les différentes catégories se recouvrent.

- **Classification selon le type d'interaction**

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

- **Les bruits additifs**

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

Les bruits additifs peuvent être subdivisés en trois groupes en fonction des lieux où ils peuvent être rencontrés :

- **Bruits des systèmes industriels :** Ils correspondent aux bruits émis par des machines possédant une faible isolation phonique. Ils peuvent être très intenses et sont, par nature, non stationnaires. Ils sont très souvent des bruits rythmiques, ou périodiques, correspondant à la répétition d'une tâche de nature productive. L'automatisation totale des sites de production n'étant pas encore atteinte, il faut également considérer les bruits produits directement ou indirectement par l'homme.
- **Bruit des moyens de transport :** ils correspondent aux bruits qui peuvent être observés dans divers véhicules tels que les voitures, les trains ou les avions. Ils se caractérisent généralement par une très forte stationnarité qui correspond à la vitesse d'un ou plusieurs harmoniques et ne comporte que des micro-fluctuations.
- **Bruits des milieux administratifs et urbains :** se sont les bruits présents dans les bureaux, les domiciles ou dans les concentrations urbaines. Ces bruits peuvent être très variés (climatisation ou bruit de parole) mais sont peu intenses et sont toujours momentanés ou contraire des bruits de moyens de transport où l'auditeur est passager.

➤ Les bruits convolutionnels

Les bruits convolutionnels (ou multiplicatifs) sont dus à la distorsion induite par la voie de communication. Ils résultent de mauvaise qualité d'un ou de plusieurs éléments de support du message ou, tout simplement, de son étroitesse en bande passante.

Les sociétés modernes utilisent de plus en plus des moyens de communication à longue distance tels que le téléphone, les moyens radiophoniques et radiotéléphoniques. Ces moyens de communication à longue distances ont été élaborés à partir d'un compromis coût / efficacité. La parole, lorsqu'elle est transmise par un tel moyen, est forcément dégradée tout en gardant une intelligibilité acceptable.

Le tableau 1.1 résume les classes les plus importantes auxquelles un bruit peut appartenir.

Propriétés	Types
Structure	Continu/Impulsif/Périodique
Type d'interaction	Additif/Multiplicatif(Convolutif)
Comportement temporel	Stationnaire/Non-stationnaire
Bande de fréquence	étroite/large
Dépendance	Corrélé/ Décorrélé
Propriétés statistiques	Dépendant/Indépendant
Propriétés spatiales	Cohérent/Incohérent

Tableau 1.1 : Classification du bruit [6].

1.3.2 Physiologie du bruit

Selon sa fréquence, son intensité, le temps d'exposition et la distance par rapport à la source, le bruit est plus ou moins nuisible.

La fréquence : elle s'exprime en Hertz (Hz). L'oreille entend les fréquences de 20 Hz à 20 kHz. La fréquence de la parole se situe entre 500 Hz et 2 000 Hz. Plus le son est aigu, plus sa fréquence est grande et les ondes sont rapprochées et plus il est dangereux pour l'oreille. Plus le son est grave, plus sa fréquence est basse et les ondes sont espacées.

Cependant il n'y a pas de différence de nature entre les vibrations audibles et les vibrations inaudibles, aussi celles-ci ont-elles reçu le nom d'infrasons (vibrations de fréquence plus

faible que les fréquences audibles) et ultrasons (fréquences plus élevées que les fréquences audibles).

L'intensité ou niveau sonore : s'exprime en décibels (dB) et se mesure avec un sonomètre. Les niveaux sonores audibles vont de 0 dB à 120 dB environ.

Le bruit est d'autant plus nocif que son niveau sonore est grand : le niveau 85 dB constitue la cote d'alerte et 90 dB la cote de danger.

Le temps d'exposition au bruit : la fatigue auditive s'estompe avec le repos et l'oreille récupère son potentiel, mais si l'exposition dure trop longtemps ou se répète trop souvent une surdité peut apparaître.

La distance par rapport à la source du bruit : s'éloigner diminue le niveau sonore et donc le risque du bruit.

La répétition : il existe certains bruits qui n'ont pas besoin d'être à forte puissance pour être gênants, c'est le cas de certains événements sonores répétitifs selon une fréquence précise qui peuvent entrer en résonance avec notre émotivité, comme par exemple, le bruit de la goutte d'eau, après un temps important d'exposition au phénomène.

1.4 Mesures de qualité

La qualité vocale est un phénomène multidimensionnel : le terme « qualité » est une manière de combiner des attributs différents, et il y a beaucoup de façons de les pondérer. Mais trois facteurs semblent les plus importants :

- Intelligibilité (une mesure de compréhensibilité).
- Marque (score) d'articulation (une mesure de reconnaissance des phonèmes).
- Identification du locuteur.

Le meilleur moyen de juger l'intelligibilité est d'effectuer des tests d'écoute. La qualité vocale dépend de la personne qui la juge, elle est donc une notion complexe à définir du fait de sa forte subjectivité, elle peut être évaluée objectivement par une série de calcul appliquée aux signaux sonores testés [7].

1.4.1 Mesures subjectives

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.

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) [8].

Par exemple, 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 1.2).

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

Tableau 1.2 : Test subjectif de qualité de type MOS.

D'autres mesures subjectives orientées vers l'intelligibilité des signaux de mauvaise qualité peuvent être utilisées comme le DRT (Diagnostic Rhyme Test) et le MRT (Modified Rhyme Test), différents mots sont présentés et on doit choisir la plus proche au son écouté.

D'une manière générale, lors de ces types de tests, 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.

Les tests subjectifs sont indispensables pour l'évaluation de la qualité vocale, puisqu'ils représentent le jugement humain de la qualité vocale. Cependant, les tests subjectifs nécessitent de mobiliser beaucoup de moyens (temps, personnes et argent). Ces tests sont difficiles à mettre en œuvre et par conséquent ils ne seront pas implémentés dans le cadre de ce travail. Les méthodes objectives se présentent comme une alternative aux méthodes subjectives et permettent d'automatiser l'évaluation de la qualité vocale. Néanmoins, elles doivent présenter une forte corrélation avec les résultats des tests subjectifs, qui représentent le jugement des utilisateurs [7].

1.4.2 Mesures objectives

Les mesures objectives de qualité des signaux vocaux les plus communément utilisées sont citées et classées dans le tableau 1.3 :

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

Tableau 1.3 : Les différents types de mesures objectives.

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 [9], d'où l'intérêt d'introduire le facteur psychoacoustique pour tout système ayant pour objectif de comparer 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 ([10], [11], [12],. . .). Par exemple, la mesure de la qualité de la parole perçue (PSQM) (Perceptual Speech Quality Measure) [13] et sa version améliorée PESQ(Perceptual Evaluation of Speech Quality) [14], le BSD (Bark Spectral Distortion)[9] et sa version améliorée, MBSD (Modified Bark Spectral Distortion) [15].

Les mesures qui sont généralement utilisées et qui seront implémentées dans ce mémoire sont :

- **Rapport signal sur bruit**

Le rapport signal sur bruit, RSB (en anglais *signal to noise ratio SNR*), comme son nom l'indique, fournit le rapport entre la puissance moyenne du signal et celle du bruit, c'est le critère le plus couramment utilisé pour désigner la qualité d'une transmission d'information par rapport aux parasites. Il est défini –en décibel (dB)- par :

$$SNR \text{ (dB)} = 10 \log_{10} \left[\frac{\sum_{n=-\infty}^{\infty} s_{\Phi}^2(n)}{\sum_{n=-\infty}^{\infty} (s_d(n) - s_{\Phi}(n))^2} \right] \quad (1.3)$$

Où $s_{\Phi}(n)$ est le signal propre et $s_d(n)$ le signal bruité ou rehaussé, cependant la mesure SNR n'est pas une très bonne estimation de la qualité de la parole.

Le signal de parole étant par nature non stationnaire, certains segments du signal peuvent avoir une énergie plus au moins grande. En supposant, que l'énergie du bruit soit à peu près constante, le SNR pourra être soit très faible dans certains segments, soit très élevé dans d'autres segments. Pour avoir une idée de la qualité de la parole synthétique on utilise plutôt le SNR segmental.

- **Rapport signal sur bruit segmental (SNR_{seg})**

Le SNR_{seg} est la moyenne des mesures SNR calculées sur différentes trames. La mesure SNR_{seg} en dB sur M segments de parole est définie comme suit :

$$SNR_{seg} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \frac{\sum_{n=Nm}^{Nm+N-1} s_{\Phi}^2(n)}{\sum_{n=Nm}^{Nm+N-1} [s_d(n) - s_{\Phi}(n)]^2} \quad (1.4)$$

Où M est le nombre de segments dans le fichier audio et chaque segment “ m ” est de longueur N . Pour un signal de parole avec une fréquence de 8 KHz les valeurs de N varient typiquement entre 160 et 240 échantillons (20 ms à 30 ms).

Cette mesure présente l’avantage de tenir compte de l’évolution du SNR au cours du temps et en particulier de bien prendre en compte les segments de faible énergie. Les trames qui ont un SNR qui dépasse 35 dB peuvent affecter la corrélation du SNR segmental avec les tests subjectifs, généralement une limite supérieure à 35 dB est utilisée. Egalement, durant les périodes de silence, les valeurs du SNR peuvent atteindre des valeurs négatives très faibles, une limite inférieure est utilisée égale -10 dB.

- **Mesure d’Itakura Saito**

La mesure d’Itakura Saito repose sur l’analyse LPC, son expression fait intervenir le modèle tout pôle du signal de référence $s_\phi(n)$ et celui du signal testé $s_d(n)$.

Pour une trame de la parole propre avec le vecteur des coefficients de la prédiction linéaire, \vec{a}_ϕ , et un vecteur des coefficients de la parole rehaussée, \vec{a}_d , la mesure de distorsion d’Itakura-saito est donnée par :

$$d_{IS}(\vec{a}_d, \vec{a}_\phi) = \left[\frac{\sigma_\phi^2}{\sigma_d^2} \right] \cdot \left[\frac{\vec{a}_d R_\phi \vec{a}_d^T}{\vec{a}_\phi R_\phi \vec{a}_\phi^T} \right] + \log \left(\frac{\sigma_d^2}{\sigma_\phi^2} \right) - 1 \quad (1.5)$$

Où : σ_d^2 et σ_ϕ^2 représentent les gains de la prédiction linéaire pour le signal rehaussé et le signal propre respectivement, et R_ϕ la fonction d’autocorrélation du signal propre.

- **Mesure LLR (Log Likelihood Ratio)**

La mesure LLR est presque identique à la mesure d’Itakura-saito (IS), elle s’intéresse seulement à la différence entre la forme spectrale des deux modèles LPC des signaux propre et rehaussé, elle est donnée par :

$$d_{LLR}(\vec{a}_d, \vec{a}_\phi) = \log \left(\frac{\vec{a}_d R_\phi \vec{a}_d^T}{\vec{a}_\phi R_\phi \vec{a}_\phi^T} \right) \quad (1.6)$$

➤ **Mesure PESQ (Perceptual Evaluation of Speech Quality)**

La mesure PESQ est une méthode objective de prévision de la qualité subjective de la parole perçue. Lors d'un test subjectif d'écoute, le sujet juge le signal de sortie dégradé par le système à tester, sans avoir accès au signal de référence correspondant à la phrase prononcée.

Le modèle PESQ permet d'évaluer la qualité d'écoute dans de nombreuses conditions de dégradation (pertes de paquets, distorsion due au codage, bruit ambiant du côté émission,...), aboutissant à une corrélation élevée avec les tests subjectifs.

L'évaluation PESQ compare un signal original $s_o(n)$ à un signal dégradé $s_d(n)$, signal qui n'est autre que, dans notre application, l'estimé du signal utile à partir du signal bruité. Elle permet de prévoir les notes de qualité de perception qu'attribuerait un sujet participant à un test d'écoute subjectif au signal $s_d(n)$.

Le schéma est représenté sur la figure (1.2). Cette note PESQ est appliquée à une échelle de type MOS, sous forme d'un scalaire compris entre -0.5 et 4.5 [16].

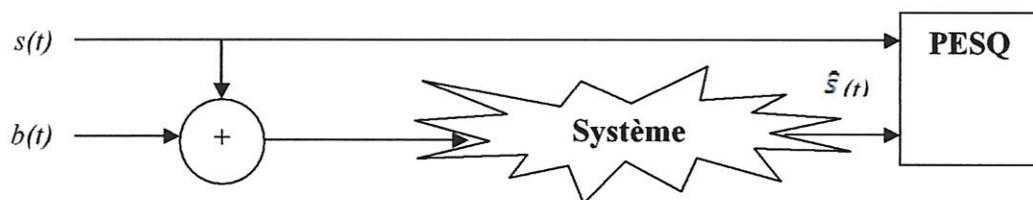


Figure 1.2: Synoptique du PESQ [16].

➤ **Mesure WSS (Weighted Slope Spectral distance) [17]**

Le critère WSS se base sur l'hypothèse que la perception d'un son dépend de sa dynamique spectrale. Le principe de ce critère est le suivant :

- Le spectre de puissance à court terme est subdivisé en 36 bandes. On s'intéresse à la fréquence centrale de chaque bande en déterminant sa puissance en dB. Elle est notée $dB(k)$, où k indique le numéro de la bande.

- On calcule la pente spectrale entre deux bandes adjacentes pour le signal propre et pour le signal traité. On les note $s_\phi(k)$ et $s_d(k)$ respectivement :

$$s_\phi(k) = dB_\phi(k+1) - dB_\phi(k) \quad (1.7)$$

$$s_d(k) = dB_d(k+1) - dB_d(k) \quad (1.8)$$

- Pour tenir compte de la différence d'intensité de deux spectres, on introduit, dans la métrique, un paramètre qui reflète la différence en dB des énergies totales du signal propre et du signal traité. Ces paramètres sont notés respectivement k_ϕ et k_d . La distance de type pente spectrale pondérée (ou WSS) d'une trame est obtenue comme suit :

$$WSS = k_{spl}(k_\phi - k_d) + \sum_{k=1}^{36} W(k)[s_\phi(k) - s_d(k)]^2 \quad (1.9)$$

Où k_{spl} est un paramètre variable et $W(k)$ est un poids de pondération.

La dégradation globale WSS est obtenue en moyennant les dégradations des trames.

1.5 Conclusion

La connaissance du signal de la parole, ses propriétés et les outils utilisés pour son traitement. En plus des caractéristiques du bruit qui dégrade la qualité et l'intelligibilité du signal de la parole seront très utiles pour entamer les méthodes de rehaussement dans les prochains chapitres. L'efficacité de ces méthodes sera évaluée en se basant sur les mesures objectives introduites dans ce chapitre.

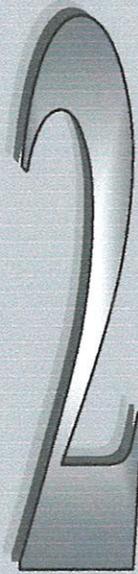


Chapitre

Etude de la théorie du filtre de Wiener

Dans ce chapitre :

- **Introduction.**
- **Généralités sur le rehaussement de la parole.**
- **Introduction à la théorie du filtre de Wiener.**
- **Estimation du SNR a priori.**
- **Filtre de Wiener racine carrée (Square-Root Wiener Filter).**
- **Filtre de Wiener paramétrique.**
- **Simulations et résultats.**
- **Conclusion.**



2.1 Introduction

Le débruitage de la parole est un problème de traitement du signal assez courant et largement utilisé. Dans ce chapitre, des algorithmes de débruitage seront étudiés dans le cadre du bruit additif. Le but de ces algorithmes est de réduire le bruit et par conséquent d'améliorer la qualité et l'intelligibilité de la parole.

On va présenter d'une manière générale, une description détaillée du filtre de Wiener classique avec ses deux types : le filtre de Wiener racine carrée et paramétrique qui nécessitent de passer par le principe et les techniques de rehaussement de la parole. Les résultats de simulation de ces méthodes seront aussi présentés.

2.2 Généralités sur le rehaussement de la parole

2.2.1 Divers sources de dégradation de la parole

La parole peut être corrompue par un bruit à n'importe quelle étape de la chaîne de communication. Par exemple :

- **A la source de la parole**

Quand la source elle-même est dans un environnement bruité, le bruit s'ajoute au signal de la parole prononcé par un locuteur qui se trouve dans cet environnement.

- **Durant la transmission**

Le signal de la parole est généralement transmis à un récepteur lointain à travers des canaux de transmission. Durant cette transmission, le bruit additionnel s'ajoute au signal de parole à cause du comportement non idéal du canal. Le bruit peut également être ajouté pendant la conversion de données faite avant la transmission ou durant la reproduction de la parole aux auditeurs.

- **Le bruit à la réception**

Parfois, bien que l'émetteur soit dans un environnement silencieux, le récepteur peut se trouver dans un environnement fortement bruité. Par conséquent, le rehaussement de la parole est nécessaire dans ce cas aussi [18].

2.2.2 Intérêt du rehaussement de la parole

Le bruit dégrade généralement la qualité et l'intelligibilité de la parole, cette dégradation crée des problèmes non seulement pour la communication entre les personnes, mais des problèmes plus sérieux dans les applications dans lesquelles la décision ou la commande est faite sur la base du signal vocal.

Ainsi, l'objectif principal du rehaussement de la parole est d'améliorer un ou plusieurs aspects perceptuels de la parole tels que la qualité et l'intelligibilité globale. La figure (2.1) montre l'allure d'un signal bruité traité par un processus de débruitage.

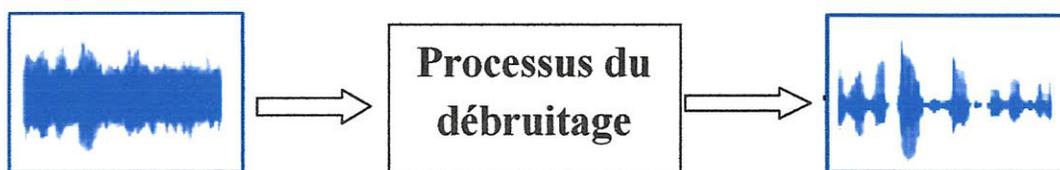


Figure 2.1 : Principe de débruitage d'un signal de parole.

2.2.3 Classification des techniques de rehaussement

Les systèmes de rehaussement de la parole peuvent être classés selon le nombre de canaux d'entrée (un ou plusieurs), le domaine de traitement (temps/fréquence), le type d'algorithme (adaptatif/non adaptatif). Typiquement, la classification la plus utilisée est celle qui divise les techniques de rehaussement en : techniques monovoies (un seul microphone) et techniques multivoies (plusieurs microphones).

- **Techniques monovoies**

Quand seulement un seul microphone d'acquisition est disponible, des techniques de rehaussement monovoie sont employées. Ces systèmes sont simples à mettre en œuvre, mais la contrainte principale de ces méthodes est qu'il n'y a aucun signal de référence pour le bruit disponible, par conséquent la densité spectrale de puissance du bruit doit être estimée [18].

Cette catégorie aura une attention particulière par la suite, car on s'intéressera dans le cadre de ce projet que des techniques monovoies.

- **Techniques multivoies**

S'appliquent aux situations dans lesquelles plusieurs microphones d'acquisition sont disponibles. Ces techniques ont l'avantage de l'existence de plusieurs signaux à l'entrée, en

plus d'un signal de référence pour le bruit. Elles sont plus complexes que les systèmes monovoies.

2.2.4 Systèmes monovoies

Dans toutes les techniques monovoies, on suppose que le signal de parole disponible est : $y(n) = x(n) + d(n)$, où $x(n)$ est le signal de parole propre, $d(n)$ est le bruit additif et $y(n)$ représente le signal de parole bruité.

Chaque technique de rehaussement monovoie exploite une ou plusieurs propriétés du signal de parole pour séparer le bruit.

Il n'existe aucune classification standard pour les méthodes disponibles, la plupart d'entre elles relèveront d'une des catégories présentées ci-dessous :

- Les méthodes basées sur la périodicité de la parole.
- Les méthodes basées sur un modèle de la parole.
- Les méthodes basées sur l'estimation de l'amplitude spectrale à court terme (STSA).
- Les méthodes basées sur les critères perceptuelles.

On s'intéressera dans ce qui suit aux différentes variantes du filtre de Wiener qui peuvent être classées dans les méthodes (STSA) et perceptuelles.

2.3 Introduction à la théorie du filtre de Wiener :

Considérant le problème du filtrage statistique illustré par la figure suivante :

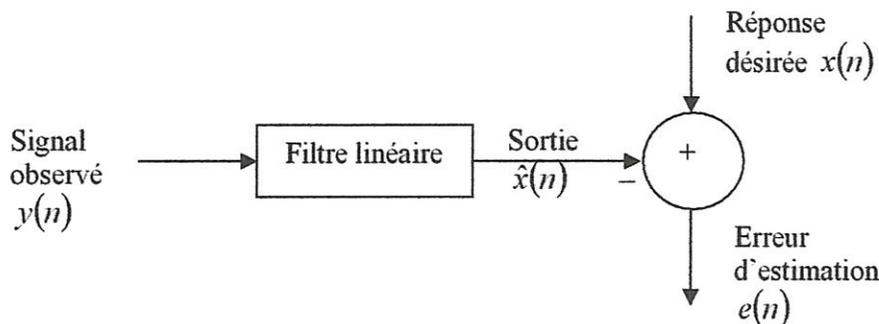


Figure 2.2 : Schéma général du problème de filtrage statistique.

Le signal d'entrée passe à travers un système linéaire et invariant dans le temps pour produire le signal de sortie $\hat{x}(n)$. On doit concevoir un système de telle manière que le signal de sortie $\hat{x}(n)$ soit le plus proche possible du signal désiré $x(n)$. Ceci peut se faire en calculant l'erreur d'estimation $e(n)$, et la rendre aussi faible que possible. Le filtre optimal qui minimise l'erreur d'estimation s'appelle le filtre de Wiener, baptisé du nom du mathématicien Wiener Norbert [19] qui a formulé et résolu pour la première fois ce problème de filtrage dans le domaine continu.

Ce filtre est linéaire et il peut être à réponse impulsionnelle finie (FIR) ou à réponse impulsionnelle infinie (IIR). En général, les filtres de type FIR sont les plus utilisés car : ils sont stables et la solution résultante est linéaire et facile à évaluer.

Considérant un système FIR pour le filtre linéaire de la figure 2.2, on a :

$$\hat{x}(n) = \sum_{i=0}^{N-1} h_i y(n-i) \quad n = 0, 1, 2, \dots \quad (2.1)$$

Où $\{h_i\}$ sont les coefficients du filtre FIR et N est le nombre de coefficients. L'objectif est de calculer les coefficients $\{h_i\}$ du filtre de telle sorte que l'erreur d'estimation $x(n) - \hat{x}(n)$ soit minimale.

L'erreur quadratique moyenne (Mean Square Error : MSE) est généralement utilisée comme critère de minimisation, et les coefficients du filtre optimal peuvent être obtenus dans le domaine temporel ou fréquentiel.

2.3.1 Filtre de Wiener dans le domaine temporel

Généralement, le signal d'entrée $y(n)$ et le signal désiré $x(n)$ sont deux processus aléatoires stationnaires au sens large. L'erreur d'estimation $e(n)$ est donnée par :

$$e(n) = x(n) - \hat{x}(n) = x(n) - \mathbf{h}^T \mathbf{y} \quad (2.2)$$

Où \mathbf{h} est le filtre qu'on cherche et N la longueur de sa réponse impulsionnelle donnée avec une notation matricielle par: $\mathbf{h} = [h_0 \ h_1 \ \dots \ h_{N-1}]^T$, avec $\mathbf{y} = [y(n) \ y(n-1) \ \dots \ y(n-(N-1))]^T$.

Pour trouver les coefficients du filtre optimal, on minimise l'erreur quadratique moyenne $E[e^2(n)]$, qui a pour expression :

$$\begin{aligned} J &= E[e^2(n)] = E[(x(n) - \mathbf{h}^T \mathbf{y})^2] \\ &= E[x^2(n) - 2\mathbf{h}^T \mathbf{y}x(n) + \mathbf{h}^T \mathbf{y} \mathbf{y}^T \mathbf{h}] \\ &= E[x^2(n)] - 2\mathbf{h}^T \mathbf{r}_{yx} + \mathbf{h}^T \mathbf{R}_{yy} \mathbf{h} \end{aligned} \quad (2.3)$$

Où \mathbf{R}_{yy} est la matrice d'autocorrélation ($N \times N$) du signal d'entrée définie par :

$$\mathbf{R}_{yy} = E[\mathbf{y} \mathbf{y}^T]$$

Et \mathbf{r}_{yx} est le vecteur d'intercorrélacion ($N \times 1$) entre le signal d'entrée et la sortie désirée, définie par : $\mathbf{r}_{yx} = E[\mathbf{y}x(n)] = E[(y(n) y(n-1) \dots y(n-N+1))x(n)]$

L'équation (2.3) montre que la fonction J dépend de la réponse impulsionnelle \mathbf{h} . Pour obtenir le minimum, il suffit de chercher les conditions d'annulation de la dérivée de la fonction J par rapport à la réponse impulsionnelle du filtre.

La dérivée de la fonction J par rapport à la réponse impulsionnelle est donnée par :

$$\frac{\partial J}{\partial h_i} = 0 = 2E\left[e(n) \frac{\partial e(n)}{\partial h_i}\right], \quad i = 0, 1, \dots, N-1 \quad (2.4)$$

A partir de l'équation (2.2) on a : $\frac{\partial e(n)}{\partial h_i} = -y(n-i)$, alors l'équation précédente

devient :

$$\frac{\partial J}{\partial h_i} = -2E[e(n)y(n-i)] = 0, \quad i = 0, 1, \dots, N-1 \quad (2.5)$$

L'équation précédente donne la condition nécessaire et suffisante pour que la fonction du coût J atteigne son minimum.

L'erreur d'estimation $e(n)$ doit être orthogonal au signal d'entrée $y(n)$ (deux variables aléatoires X et Z sont orthogonaux si $E[XZ] = 0$). Cet énoncé constitue ce qu'on appelle le principe d'orthogonalité du filtrage linéaire optimal.

En utilisant les propriétés des dérivées des matrices et des vecteurs, on calcule la dérivée de J en fonction du vecteur \mathbf{h} comme suit :

$$\frac{\partial J}{\partial \mathbf{h}} = -2\mathbf{r}_{yx} + 2\mathbf{h}^T \mathbf{R}_{yy} = 0 \quad (2.6)$$

Le vecteur optimum \mathbf{h}^* est celui qui annule le gradient du critère:

$$\mathbf{R}_{yy} \mathbf{h}^* = \mathbf{r}_{yx} \quad (2.7)$$

On peut exprimer l'équation précédente par :

$$\sum_{i=0}^{N-1} h_i r_{yy}(m-i) = r_{yx}(-m), \quad m = 0, 1, \dots, N-1 \quad (2.8)$$

Qui montre clairement que les coefficients du filtre $\{h_i\}$ peuvent être obtenus en résolvant un système de N équations avec N inconnus, la solution est :

$$\mathbf{h}^* = \mathbf{R}_{yy}^{-1} \mathbf{r}_{yx} \quad (2.9)$$

La solution précédente est appelée la solution de Wiener-Hopf qui peut être exprimée sous forme matricielle comme suit :

$$\begin{bmatrix} h_0 \\ h_1 \\ h_2 \\ \vdots \\ h_{N-1} \end{bmatrix} = \begin{bmatrix} r_{yy}(0) & r_{yy}(1) & r_{yy}(2) & \cdots & r_{yy}(N-1) \\ r_{yy}(1) & r_{yy}(0) & r_{yy}(1) & \cdots & r_{yy}(N-2) \\ r_{yy}(2) & r_{yy}(1) & r_{yy}(0) & \cdots & r_{yy}(N-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{yy}(N-1) & r_{yy}(N-2) & \cdots & r_{yy}(1) & r_{yy}(0) \end{bmatrix}^{-1} \begin{bmatrix} r_{yx}(0) \\ r_{yx}(-1) \\ r_{yx}(-2) \\ \vdots \\ r_{yx}(-N+1) \end{bmatrix} \quad (2.10)$$

Comme le signal d'entrée est considéré comme étant un signal stationnaire au sens large, la matrice d'autocorrélation \mathbf{R}_{yy} est symétrique et les éléments situés sur la diagonale sont identiques, cette forme matricielle est appelée forme de TOEPLITZ. Cette propriété peut

être exploitée pour obtenir des algorithmes efficaces pour la résolution de cette équation. L'algorithme Levinson-Durbin développé [20] représente la solution la plus efficace.

Pour trouver le minimum de l'erreur quadratique moyenne J , on remplace le vecteur du filtre optimal de l'équation (2.9) dans l'équation (2.3), on aura :

$$\begin{aligned}
 J_{\min} &= E[x^2(n)] - 2(\mathbf{R}_{yy}^{-1}\mathbf{r}_{yx})^T \mathbf{r}_{yx} + (\mathbf{R}_{yy}^{-1}\mathbf{r}_{yx})^T \mathbf{R}_{yy} \mathbf{R}_{yy}^{-1} \mathbf{r}_{yx} \\
 &= E[x^2(n)] - 2(\mathbf{r}_{yx})^T \mathbf{R}_{yy}^{-T} \mathbf{r}_{yx} + (\mathbf{r}_{yx})^T \mathbf{R}_{yy}^{-T} \mathbf{r}_{yx} \\
 &= E[x^2(n)] - (\mathbf{r}_{yx})^T \mathbf{R}_{yy}^{-T} \mathbf{r}_{yx} \\
 &= E[x^2(n)] - (\mathbf{r}_{yx})^T \mathbf{h}^*
 \end{aligned} \tag{2.11}$$

Dans les développements précédents, on a utilisé le fait que \mathbf{R}_{yy} est symétrique, et donc : $\mathbf{R}_{yy}^{-T} = \mathbf{R}_{yy}^{-1}$.

2.3.2 Filtre de Wiener dans le domaine fréquentiel

Dans l'étude du filtre de Wiener dans le domaine temporel, on a considéré une réponse impulsionnelle de durée finie. Maintenant, on considère un filtre avec une réponse impulsionnelle infinie de la forme :

$$\hat{x}(n) = \sum_{i=-\infty}^{\infty} h_i y(n-i), \quad -\infty < n < \infty \tag{2.12}$$

A la différence du développement dans le domaine temporel, l'estimation de $x(n)$ à l'instant n est fonction des données antérieurs et futurs. Il est appelé le filtre de Wiener lissé (Wiener Smoother filter).

Il est facile de montrer que les équations de Wiener-Hopf pour le filtre lissé sont identiques à celles de l'équation (2.8).

$$\sum_{i=-\infty}^{\infty} h_i r_{yy}(m-i) = r_{yx}(-m), \quad -\infty < m < \infty \tag{2.13}$$

Contrairement à l'équation (2.8), les équations précédentes sont valables pour toutes les valeurs de m .

A noter que le signal de sortie de l'équation (2.12) $x(n)$, peut être obtenu par la convolution de la réponse impulsionnelle du système $\{h_i\}$ avec le signal d'entrée $y(n)$.

$$\hat{x}(n) = h(n) * y(n) \quad (2.14)$$

Donc, dans le domaine fréquentiel, on a :

$$\hat{X}(k) = H(k)Y(k) \quad (2.15)$$

Où $H(k)$ et $Y(k)$ sont les transformées de Fourier discrètes de $h(n)$ et $y(n)$ respectivement. On peut définir l'erreur d'estimation à la fréquence k par :

$$E(k) = X(k) - \hat{X}(k) = X(k) - H(k)Y(k) \quad (2.16)$$

Par la suite, on cherche à calculer $H(k)$ qui minimise l'erreur quadratique moyenne $E\{|E(k)|^2\}$, dans le domaine fréquentiel. L'erreur quadratique moyenne est donnée par :

$$\begin{aligned} E\{|E(k)|^2\} &= E\{|X(k) - H(k)Y(k)|^2\} \\ &= E\{|X(k)|^2\} - H(k)E\{X^*(k)Y(k)\} - H^*(k)E\{Y^*(k)X(k)\} + |H(k)|^2 E\{|Y(k)|^2\} \end{aligned} \quad (2.17)$$

On peut exprimer l'erreur quadratique moyenne par :

$$J_2 = E\{|E(k)|^2\} = E\{|X(k)|^2\} - H(k)P_{yx}(k) - H^*(k)P_{xy}(k) + |H(k)|^2 P_{yy}(k) \quad (2.18)$$

Où : $P_{xy}(k)$ est la densité interspectrale de puissance de x et y .

$P_{yy}(k)$ est la densité spectrale de puissance de y .

L'équation (2.14) montre que la fonction J_2 dépend de la réponse impulsionnelle H .

Pour obtenir le filtre optimal $H(k)$, il suffit de chercher les conditions d'annulation de la dérivée de la fonction J_2 par rapport à la réponse impulsionnelle du filtre.

$$\frac{\partial J_2}{\partial H(k)} = H^*(k)P_{yy}(k) - P_{yx}(k) = [H(k)P_{yy}(k) - P_{xy}(k)]^* = 0 \quad (2.19)$$

Après utilisation de l'égalité $P_{yx}(k) = P_{xy}^*(k)$, on obtient la forme générale du filtre de Wiener dans le domaine fréquentiel :

$$H(k) = \frac{P_{xy}(k)}{P_{yy}(k)} \quad (2.20)$$

$H(k)$ est complexe car la densité interspectrale de puissance $P_{xy}(k)$ est en général une valeur complexe.

Les équations (2.9) et (2.20) constituent les équations générales du filtre de Wiener dans le domaine temporel et fréquentiel, respectivement.

2.3.3 Filtres de Wiener pour la réduction du bruit

Dans les applications du rehaussement de la parole, le signal d'entrée $y(n)$ est le signal de parole bruité :

$$y(n) = x(n) + d(n) \quad (2.21)$$

Où $x(n)$ est le signal de la parole propre et $d(n)$ est le bruit additif. L'objectif du filtre de Wiener est donc de produire une estimation du signal propre $x(n)$. On peut obtenir les filtres de Wiener correspondants soit dans le domaine temporel ou dans le domaine fréquentiel, en utilisant l'équation (2.9) ou l'équation (2.20).

Pour déterminer le filtre de Wiener dans le domaine temporel, on doit tout d'abord calculer \mathbf{R}_{yy} . On a par définition :

$$\begin{aligned} \mathbf{R}_{yy} &= E[\mathbf{y}\mathbf{y}^T] = E[(\mathbf{x} + \mathbf{d})(\mathbf{x} + \mathbf{d})^T] \\ &= E[\mathbf{x}\mathbf{x}^T] + E[\mathbf{d}\mathbf{d}^T] + E[\mathbf{x}\mathbf{d}^T] + E[\mathbf{d}\mathbf{x}^T] \\ &= \mathbf{R}_{xx} + \mathbf{R}_{dd} \end{aligned} \quad (2.22)$$

Les deux espérances $E[\mathbf{x}\mathbf{d}^T]$ et $E[\mathbf{d}\mathbf{x}^T]$ dans l'équation précédente sont égales à zéro, car le signal et le bruit sont supposés non corrélés et de moyennes nulles. De plus, le vecteur

r_{yx} d'intercorrélation dans l'équation (2.9) est égal à r_{xx} car le signal désiré et les signaux de bruit sont supposés non corrélés. Par conséquent, le filtre de Wiener résultant dans le domaine temporel a la forme suivante :

$$\mathbf{h}^* = (\mathbf{R}_{xx} + \mathbf{R}_{dd})^{-1} \mathbf{r}_{xx} \quad (2.23)$$

Alors, le filtre de Wiener \mathbf{h}^* dépend de la fonction d'autocorrélation du signal propre \mathbf{R}_{xx} (qui est inconnu), et donc le filtre n'est pas réalisable [21].

Pour le calcul du filtre de Wiener dans le domaine fréquentiel, on applique la transformée de Fourier sur les signaux dans l'équation (2.21), on obtient:

$$Y(k) = X(k) + D(k) \quad (2.24)$$

Selon l'équation (2.20), on doit calculer $P_{xy}(k)$ et $P_{yy}(k)$:

$$\begin{aligned} P_{yy}(k) &= E[X(k)\{X(k) + D(k)\}^*] \\ &= E[X(k)X^*(k)] + E[X(k)D^*(k)] \\ &= P_{xx}(k) \end{aligned} \quad (2.25)$$

De même,

$$\begin{aligned} P_{yy}(k) &= E[\{X(k) + D(k)\}\{X(k) + D(k)\}^*] \\ &= E[X(k)X^*(k)] + E[D(k)D^*(k)] + E[X(k)D^*(k)] + E[D(k)X^*(k)] \\ &= P_{xx}(k) + P_{dd}(k) \end{aligned} \quad (2.26)$$

Finalement, en remplaçant l'équation (2.25) et l'équation (2.26) dans l'équation (2.20), on obtient la réponse du filtre de Wiener dans le domaine fréquentiel :

$$H(k) = \frac{P_{xx}(k)}{P_{xx}(k) + P_{dd}(k)} \quad (2.27)$$

Notant que $H(k)$ est réelle, positive et paire car (a) les deux quantités $P_{dd}(k) \geq 0$ et $P_{xx}(k) \geq 0$ et (b) les spectres de puissance $P_{dd}(k)$ et $P_{xx}(k)$ sont symétriques.

Le fait que $H(k)$ est réelle et positive, implique que la réponse impulsionnelle h_i doit être paire aussi. Donc, h_i est non causale et le filtre de Wiener n'est pas réalisable.

En définissant ξ_k :

$$\xi_k = \frac{P_{xx}(k)}{P_{dd}(k)} \tag{2.28}$$

Comme étant le rapport signal sur bruit a priori à la fréquence k , on peut également exprimer le filtre de Wiener dans l'équation (2.27) par :

$$H(k) = \frac{\xi_k}{\xi_k + 1} \tag{2.29}$$

Notant que $0 < H(k) < 1$, et $H(k) = 0$, quand $\xi_k \rightarrow 0$ et $H(k) = 1$ quand $\xi_k \rightarrow \infty$

Ainsi, selon l'équation (2.29), le filtre de Wiener n'affecte pas les régions du spectre où le SNR est élevé et atténue les régions du spectre où le SNR est faible. Ceci est illustré par la figure (2.3), qui trace $H(k)$ en fonction du SNR a priori (ξ_k) en dB.

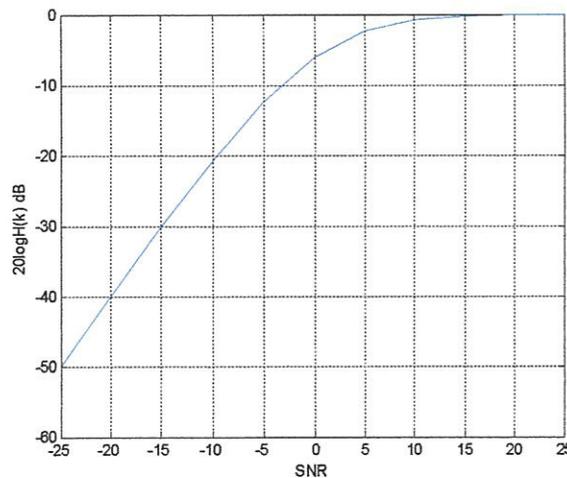


Figure 2.3 : Courbe de l'atténuation du filtre de Wiener en fonction du SNR a priori.

On observe que pour $\xi_k > 10$ dB, aucune atténuation n'est appliquée car $H(k) \approx 1$. Ainsi, le filtre de Wiener atténue chaque composante fréquentielle proportionnellement au SNR a priori estimé (ξ_k) à cette fréquence.

2.4 Estimation du SNR a priori

L'implémentation du filtre de Wiener et ses variantes nécessite le calcul du rapport signal sur bruit a priori ξ_k . Cette quantité doit être estimée pour chaque trame. Deux approches sont considérées ici. Dans la première, un estimateur du maximum de vraisemblance (Maximum Likelihood : ML) de la variance des composantes spectrales de la parole est utilisé. La deuxième approche est basée sur une méthode d'estimation dite de décision dirigée (decision directed), les deux approches nécessitent la connaissance de la densité spectrale du bruit.

2.4.1 Approche d'estimation du maximum de vraisemblance

L'approche d'estimation du maximum de vraisemblance est généralement la plus utilisée pour l'estimation d'un paramètre inconnu, de densité de probabilité donnée, quand aucune information a priori sur ce paramètre n'est disponible. Par exemple, l'estimation de la densité spectrale de puissance du signal rehaussé dans notre cas (la variance $\hat{\lambda}_x(k) = P_{xx}(k)$).

L'estimateur du ML de $\hat{\lambda}_x(k)$ dans la trame d'analyse (l) est basé sur M observations consécutives $Y_k(l) = \{y_k(l), y_k(l-1), \dots, y_k(l-M+1)\}$, qui sont considérés comme statistiquement indépendants. Cette supposition est raisonnable quand l'analyse est faite sur des fenêtres non chevauchées. Cependant, dans le système utilisé ici, le recouvrement est fait.

Néanmoins, on continue cette supposition puisque la dépendance statistique est difficile à être modélisée et manipulée. On considère également que les variances de la $k^{\text{ième}}$ composante spectrale du signal et du bruit $\lambda_x(k)$ et $\lambda_d(k)$ respectivement [$\lambda_d(k) = P_{dd}(k)$], sont deux paramètres qui varient lentement, de sorte qu'ils puissent être considérés comme constants durant les M observations précédentes. Finalement, on suppose que la variance de la $k^{\text{ième}}$ composante spectrale du bruit est connue.

L'estimateur du ML $\hat{\lambda}_x(k)$ de $\lambda_x(k)$, est l'argument non négatif qui maximise la densité de probabilité conjointe conditionnelle de $Y_k(l)$ sachant $\lambda_x(k)$ et $\lambda_d(k)$. En se basant sur le modèle statistique Gaussien et l'indépendance statistique des composantes spectrales, cette densité de probabilité est donnée par :

$$p(Y_k(l)|\lambda_x(k), \lambda_d(k)) = \prod_{m=0}^{M-1} \frac{1}{\pi(\lambda_x(k) + \lambda_d(k))} \cdot \exp\left(-\frac{R_k^2(l-m)}{\lambda_x(k) + \lambda_d(k)}\right) \quad (2.30)$$

Où $R_k(m) = |Y_k(m)|$. La valeur $\hat{\lambda}_x(k)$ est facilement obtenue à partir de (2.30), et égale :

$$\hat{\lambda}_x(k) = \begin{cases} \frac{1}{M} \sum_{m=0}^{M-1} R_k^2(l-m) - \lambda_d(k), & \text{si non négatif} \\ 0 & \text{ailleurs} \end{cases} \quad (2.31)$$

Cet estimateur produit l'estimateur suivant pour le SNR a priori ξ_k :

$$\hat{\xi}_k = \begin{cases} \frac{1}{M} \sum_{m=0}^{M-1} \gamma_k(l-m) - 1, & \text{si non négatif} \\ 0 & \text{ailleurs} \end{cases} \quad (2.32)$$

Où : $\gamma_k(m) = |Y_k(m)|^2 / \lambda_d(k, m)$ est le SNR a posteriori dans la $m^{\text{ième}}$ fenêtre d'analyse.

Dans la pratique, la moyenne courante requise en (2.32) est remplacée par une moyenne récursive avec une constante de temps comparable à la période de corrélation de γ_k .

C'est-à-dire, l'estimateur de ξ_k dans la $l^{\text{ième}}$ fenêtre d'analyse est obtenu par :

$$\bar{\gamma}_k(l) = \alpha \bar{\gamma}_k(l-1) + (1-\alpha) \frac{\gamma_k(l)}{\beta}; \quad 0 \leq \alpha \leq 1; \beta \geq 1 \quad (2.33)$$

$$\hat{\xi}_k(l) = \begin{cases} \bar{\gamma}_k(l) - 1, & \bar{\gamma}_k(l) - 1 \geq 0 \\ 0 & \text{ailleurs} \end{cases} \quad (2.34)$$

β est un facteur de correction.

Les valeurs de α et β sont déterminées par des tests d'écoute.

2.4.2 Approche d'estimation de décision dirigée (decision directed)

On considère ici l'estimation du SNR a priori d'une composante spectrale par une méthode dite de décision dirigée. Cet estimateur s'avère très utile quand il est combiné avec les estimateurs au sens du MMSE ou l'estimateur d'amplitude de Wiener.

Soient $\xi_k(l)$, $A_k(l)$, $\lambda_d(k,l)$ et $\gamma_k(l)$ le SNR a priori, l'amplitude, la variance du bruit et le SNR a posteriori respectivement, de la composante spectrale pour la trame d'analyse (l) .

La dérivation de l'estimateur du SNR a priori est basée ici sur la définition de $\xi_k(l)$ et sa relation au SNR a posteriori $\gamma_k(l)$ comme donnée dans les deux équations suivantes :

$$\xi_k(l) = \frac{E[A_k^2(l)]}{\lambda_d(k,l)} \quad (2.35)$$

$$\xi_k(l) = E[\gamma_k(l) - 1] \quad (2.36)$$

De (2.35) et (2.36), on peut écrire :

$$\xi_k(l) = E\left\{ \frac{1}{2} \frac{A_k^2(l)}{\lambda_d(k,l)} + \frac{1}{2} [\gamma_k(l) - 1] \right\} \quad (2.37)$$

L'estimateur proposé $\hat{\xi}_k(l)$ de $\xi_k(l)$ est déduit de l'équation (2.37), et est donné par :

$$\hat{\xi}_k(l) = \alpha \frac{\hat{A}_k^2(l)}{\lambda_d(k,l-1)} + (1-\alpha)p[\gamma_k(l) - 1], \quad 0 \leq \alpha < 1 \quad (2.38)$$

Où $\hat{A}_k(l-1)$ est l'estimateur de l'amplitude de la $k^{\text{ième}}$ composante spectrale du signal dans la $(l-1)^{\text{ième}}$ fenêtre d'analyse, et $p[\cdot]$ est un opérateur qui est défini par :

$$p[x] = \begin{cases} x & \text{si } x > 0 \\ 0 & \text{ailleurs} \end{cases} \quad (2.39)$$

$p[\cdot]$ est utilisé pour assurer que l'estimateur $\hat{\xi}_k(l)$ est toujours positif si $\gamma_k(l)-1$ est négatif.

L'estimateur proposé pour $\xi_k(l)$ est un estimateur de type « décision dirigée », parce que la mise à jour de $\hat{\xi}_k(l)$ est basée sur l'amplitude estimée de la trame précédente.

$$\text{En utilisant } \hat{A}_k(l) = G[\hat{\xi}_k(l), \gamma_k(l)] \cdot R_k(l)$$

Où $G[\cdot, \cdot]$ est la fonction du gain qui résulte de l'estimateur au sens du MMSE ou de l'estimateur d'amplitude de Wiener. L'équation (2.38) peut être écrite :

$$\hat{\xi}_k(l) = \alpha \cdot G^2[\hat{\xi}_k(l-1), \gamma_k(l-1)] \cdot \gamma_k(l-1) + (1-\alpha) \cdot P[\gamma_k(l)-1] \quad (2.40)$$

Plusieurs conditions initiales ont été examinées par simulation dans les travaux de recherches [22]. L'utilisation de $\hat{\xi}_k(0) = \alpha + (1-\alpha)P[\gamma_k(0)-1]$ est appropriée, puisqu'elle minimise les effets des transitions initiales dans la parole rehaussée.

2.5 Filtre de Wiener racine carrée (Square-Root Wiener Filter)

Supposant qu'on utilise la racine carrée du filtre de Wiener, la sortie du filtre de Wiener dans le domaine fréquentiel sera :

$$\hat{X}(k) = \sqrt{H(k)}Y(k) \quad (2.41)$$

Qui est l'estimation de la densité spectrale de puissance du signal $\hat{x}(n)$. D'après l'équation (2.41) on obtient :

$$\begin{aligned} E|\hat{X}(k)|^2 &= (\sqrt{H(k)})^2 E|Y(k)|^2 \\ P_{\hat{x}\hat{x}}(k) &= H(k)P_{yy}(k) \\ P_{\hat{x}\hat{x}}(k) &= \frac{P_{xx}(k)}{P_{xx}(k) + P_{dd}(k)} P_{yy}(k) \end{aligned} \quad (2.42)$$

Comme le signal et le bruit sont non corrélés, on a $P_{yy}(k) = P_{xx}(k) + P_{dd}(k)$ et l'équation (2.42) devient :

$$P_{\hat{x}\hat{x}}(k) = \frac{P_{xx}(k)}{P_{xx}(k) + P_{dd}(k)} (P_{xx}(k) + P_{dd}(k)) = P_{xx}(k) \quad (2.43)$$

Ainsi, à la présence du bruit, la densité spectrale de puissance de la sortie du filtre de Wiener racine carré $\hat{P}_{xx}(k)$ est identique à la densité spectrale de puissance du signal propre $P_{xx}(k)$.

2.6 Filtre de Wiener paramétrique

La généralisation de la réponse du filtre de Wiener racine carrée donne la réponse suivante du filtre de Wiener paramétrique [23] :

$$H(k) = \left[\frac{P_{xx}(k)}{P_{xx}(k) + \alpha P_{dd}(k)} \right]^\beta \quad (2.44)$$

Pour quelques paramètres α et β . Si $\alpha = \beta = 1$, alors on obtient le filtre de Wiener traditionnel. Si $\alpha = 1$ et $\beta = 1/2$, on obtient le filtre de Wiener racine carré. En variant les paramètres α et β , on peut obtenir les différents types du filtre de Wiener avec des caractéristiques d'atténuations différentes. Pour vérifier l'influence de ces deux paramètres sur l'atténuation, on exprime tout d'abord l'équation (2.44) en fonction de ξ_k :

$$H(k) = \left(\frac{\xi_k}{\alpha + \xi_k} \right)^\beta \quad (2.45)$$

La figure suivante trace l'atténuation du filtre de Wiener en fonction de ξ_k pour une valeur fixe de α ($\alpha = 1$) et pour différentes valeurs de β .

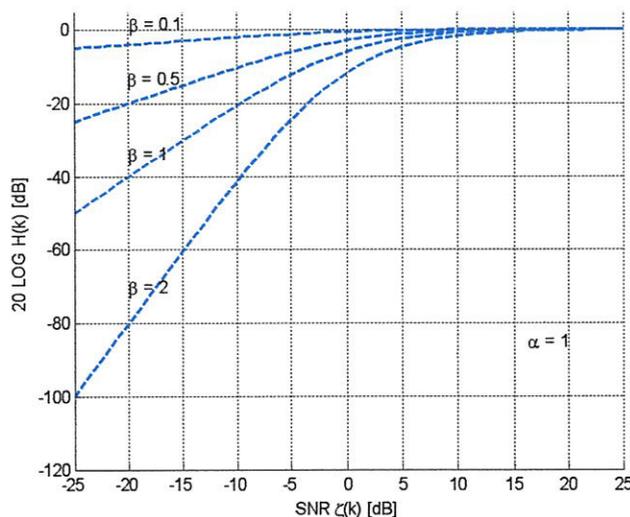


Figure 2.4 : Courbes d'atténuation du filtre de Wiener paramétrique ($\alpha = 1$, β variable).

Comme montré par la figure, le paramètre β affecte l'atténuation aux niveaux des valeurs très faible du SNR. L'atténuation la plus importante est appliquée lorsque β est élevée et $\xi_k \ll 0$ dB. Par exemple quand $\xi_k = -20$ dB et $\beta = 1$, on obtient une atténuation d'environ -40 dB, et quand $\beta = 2$, on obtient une atténuation égale à -80 dB.

Le paramètre β présente un mécanisme qui peut être utilisé pour augmenter artificiellement l'atténuation du signal. Bien que le bruit résiduel puisse être réduit considérablement en augmentant β , une distorsion de la parole sera très probablement introduite en conséquence.

La figure (2.5) illustre l'atténuation du filtre de Wiener en fonction de ξ_k pour une valeur fixe de β ($\beta = 1$) et pour différentes valeurs de α .

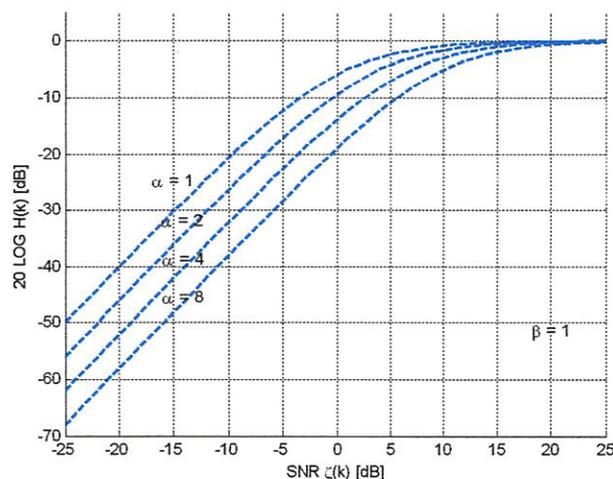


Figure 2.5 : Courbes d'atténuation du filtre de Wiener paramétriques
(α variable, $\beta = 1$).

Comme on peut le voir, le paramètre α décale les courbes d'atténuation vers le bas avec l'augmentation de ses valeurs. Alors, il influe sur l'atténuation aux SNR élevés et faibles à la fois. L'atténuation causée par les valeurs élevées de α dans le cas où les niveaux du SNR sont très faibles n'est pas aussi importante que celle causée par l'augmentation du paramètre β .

Par rapport au paramètre β , le paramètre α offre plus de flexibilité car il peut influencer sur l'atténuation dans le cas où les niveaux du SNR sont faibles ou élevés à la fois.

2.7 Résultats

2.7.1 Conditions d'expérimentations

On applique maintenant les méthodes étudiées précédemment à des signaux pris de la base de données « Noizeus [24] », puis on présente quelques résultats objectifs.

L'utilisation de la base de données (Noizeus) est pour faciliter la comparaison des algorithmes de rehaussement de la parole. Elle est devenue une base standard pour l'évaluation des algorithmes de rehaussement.

Cette base contient 30 phrases phonétiquement équilibrées, ces phrases sont bruitées artificiellement par neuf bruits réels à différent rapport signal sur bruit (0 dB, 5 dB, 10 dB,

15dB). Comme : le bruit de parole, le bruit de voiture, train, restaurant, exhibition, rue, aéroport, bruit blanc et le bruit dans une station de train.

Les trente phrases de la base de données ont été produites par trois locuteurs et trois locutrices (5 phrases/locuteur) et enregistrées dans une salle acoustiquement isolée, avec une fréquence d'échantillonnage de 25KHz, et ils sont rééchantillonnés à 8KHz.

Ainsi, pour chaque méthode, pour un type de bruit bien défini et pour un rapport signal sur bruit désiré, toutes les trente phrases de la parole propre et les trente phrases de la parole bruitée correspondantes, seront utilisées pour chaque mesure objective, où la moyenne est appliquée à la fin.

Les simulations et les tests des méthodes présentées ont été effectués sur un PC Intel® Core™ 2 Duo CPU, 2 GHz, 2 Go de RAM avec des simulations effectuées dans MATLAB Version 7.9.0.529 (R2009b).

Pour les différents tests les paramètres suivants sont utilisés :

- Une fréquence d'échantillonnage de 8 kHz.
- Une fenêtre de Hanning de durée 20 ms (160 échantillons) avec un chevauchement de 50 % est utilisé durant l'étape d'analyse, et une fenêtre rectangulaire pour la synthèse avec la méthode d'addition et chevauchement (overlap and add).
- Une taille de 256 échantillons pour la DFT est utilisée pour le calcul des densités spectrales.
- Une méthode simple d'estimation continue de la densité spectrale de puissance du bruit est utilisée.
- L'approche de décision dirigée a été utilisée pour l'estimation du SNR a priori dans chaque trame, où un facteur de pondération $\alpha = 0.98$ et une valeur minimale de -25 dB du SNR a priori, permettent d'avoir un bon compromis entre la réduction du bruit et la distorsion du signal rehaussé.

2.7.2 Evaluation des performances

Dans cette section, on évalue objectivement les performances des différentes méthodes étudiées précédemment.

Les résultats de test des mesures (LLR, WSS, SNRseg, PESQ) sont présentés dans les tableaux suivants pour les signaux bruités par un bruit blanc et des bruits réels (bruit de

voiture et le bruit de parole (babble)) en appliquant l'algorithme de Wiener classique et ses variantes. Où chaque valeur du tableau correspond à une moyenne effectuée sur 30 phrases différentes de la base de données.

Commençant par le filtre de Wiener paramétrique, on a deux paramètres à choisir dans l'équation (2.44) α et β . Pour une valeur fixe de $\alpha = 1$ et pour différentes valeurs de β , les tableaux (2.1) et (2.2) suivants présentent les mesures objectives obtenues dans le cas d'un bruit blanc et un bruit babble avec un SNR= 0 dB et 5 dB respectivement.

Mesures	SNR = 0 dB				SNR = 5 dB			
	LLR	WSS	SNRseg	PESQ	LLR	WSS	SNRseg	PESQ
Signal bruité	1.802	82.696	-5.081	1.539	1.545	69.972	-2.327	1.799
Wiener paramétrique ($\beta = 0.1$)	1.741	80.062	-4.464	1.592	1.481	67.298	-1.661	1.866
Wiener paramétrique ($\beta = 0.5$) (Wiener racine carrée)	0.910	50.157	0.662	2.333	0.718	40.385	3.077	2.622
Wiener paramétrique ($\beta = 1$) (Wiener classique)	1.237	67.670	1.462	2.255	1.039	55.200	3.827	2.609
Wiener paramétrique ($\beta = 2$)	1.473	78.899	0.149	1.902	1.259	66.136	2.354	2.265

Tableau 2.1 : Mesures objectives pour un bruit blanc, ($\alpha = 1$ et β variable).

Mesures	SNR = 0 dB				SNR = 5 dB			
	LLR	WSS	SNRseg	PESQ	LLR	WSS	SNRseg	PESQ
Signal bruité	0.8950	70.3481	-4.6320	1.7054	0.7153	56.0190	-1.7833	2.0061
Wiener paramétrique ($\beta = 0.1$)	1.741	80.062	-4.464	1.592	1.481	67.298	-1.6608	1.866
Wiener paramétrique ($\beta = 0.5$) (Wiener racine carrée)	0.458	45.419	0.099	2.427	0.344	36.541	2.505	2.713
Wiener paramétrique ($\beta = 1$) (Wiener classique)	0.705	59.936	0.906	2.334	0.556	48.356	3.283	2.696
Wiener paramétrique ($\beta = 2$)	1.473	78.899	0.149	1.902	1.259	66.136	2.354	2.265

Tableau 2.2 : Mesures objectives pour un bruit babble, ($\alpha = 1$ et β variable).

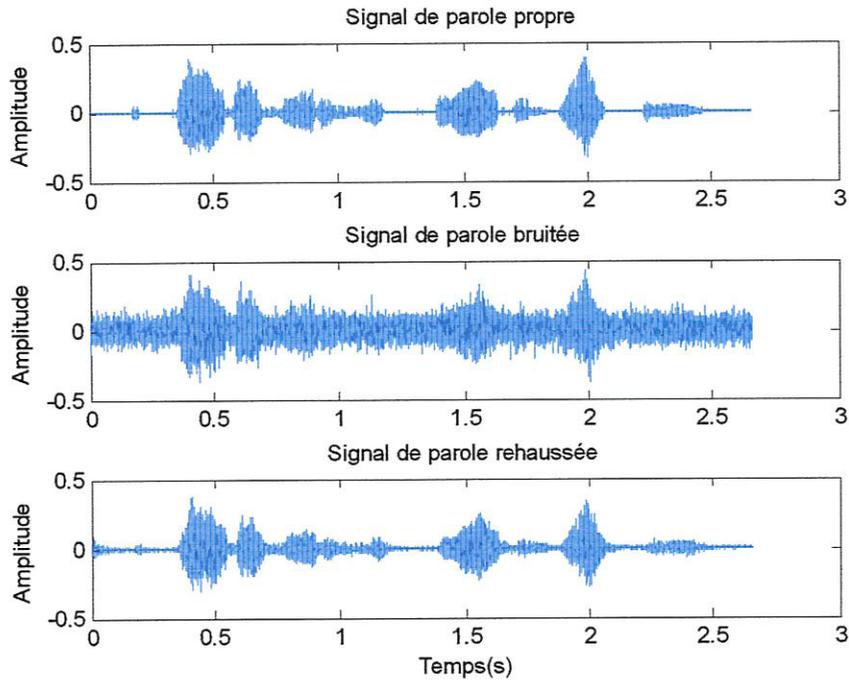


Figure 2.6 .a) : Les formes d'ondes, cas d'un bruit blanc avec $\alpha = 1, \beta = 0.5$ et un SNR= 0 dB.

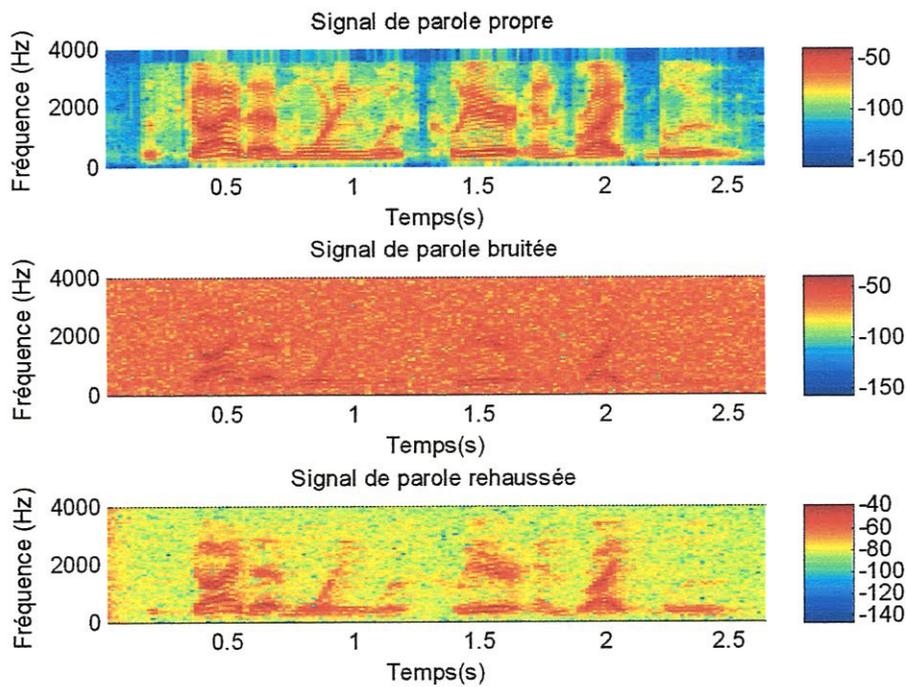


Figure 2.6 .b): Spectrogrammes, cas d'un bruit blanc avec $\alpha = 1, \beta = 0.5$ et un SNR= 0 dB.

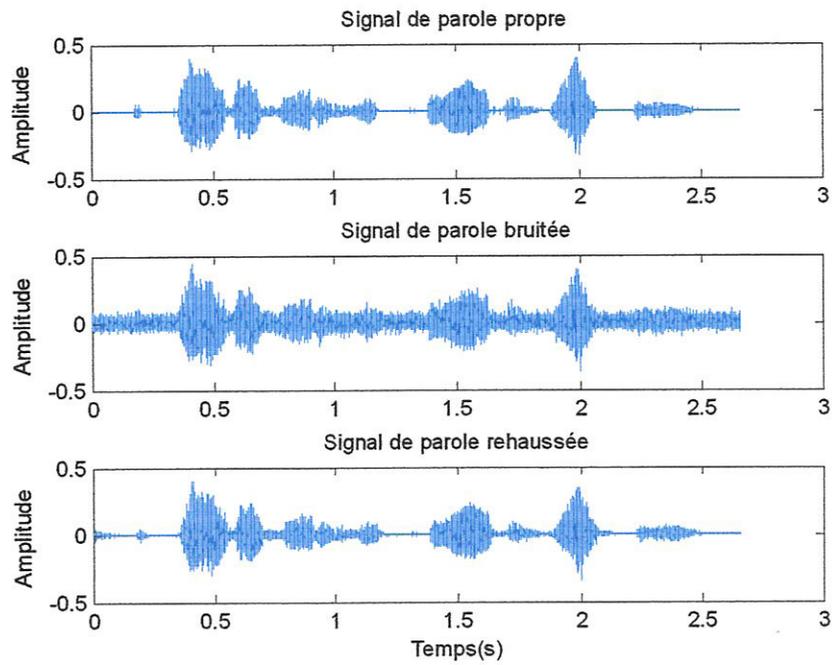


Figure 2.7 .a) : Les formes d’ondes, cas d’un bruit blanc avec $\alpha = 1, \beta = 0.5$ et un SNR= 5 dB.

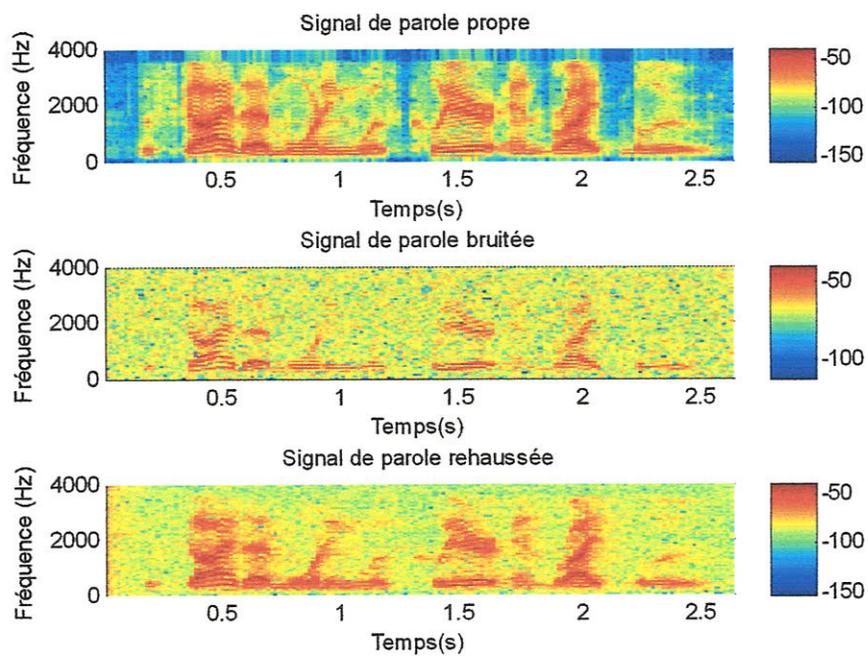


Figure 2.7 .b): Spectrogrammes, cas d’un bruit blanc avec $\alpha = 1, \beta = 0.5$ et un SNR= 5 dB.

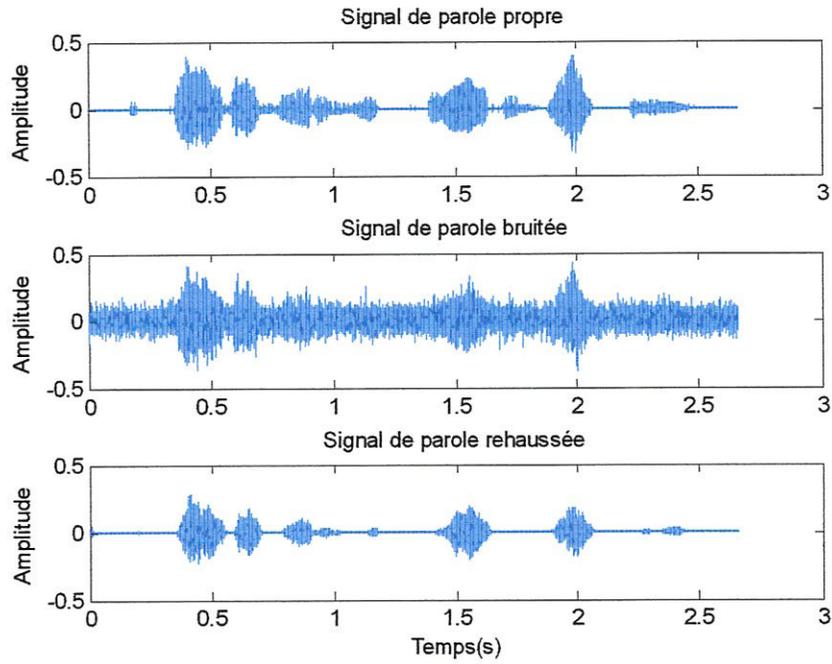


Figure 2.8. a) : Les formes d'ondes , cas d'un bruit blanc avec $\alpha = 2, \beta = 1$ et un SNR=0 dB.

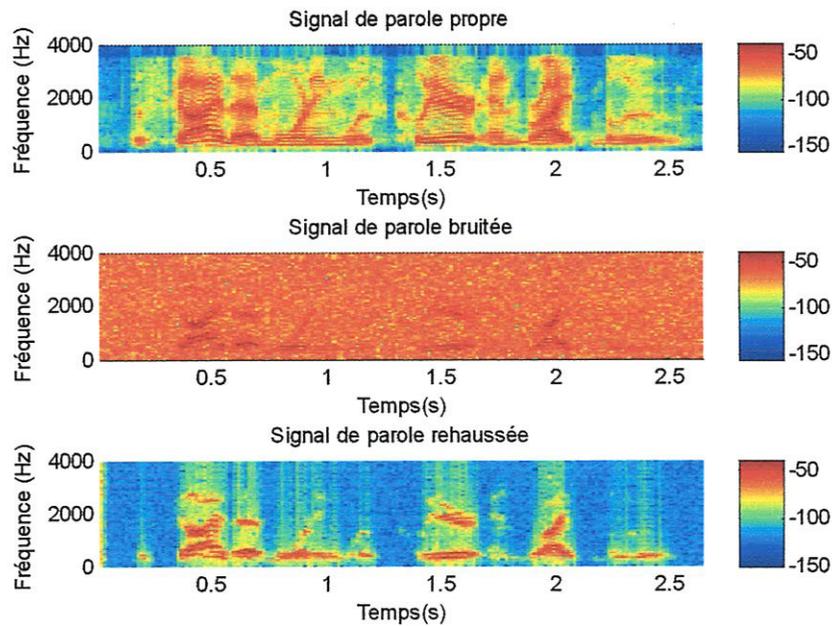


Figure 2.8 .b): Spectrogrammes, cas d'un bruit blanc avec $\alpha = 2, \beta = 1$ et un SNR=0 dB.

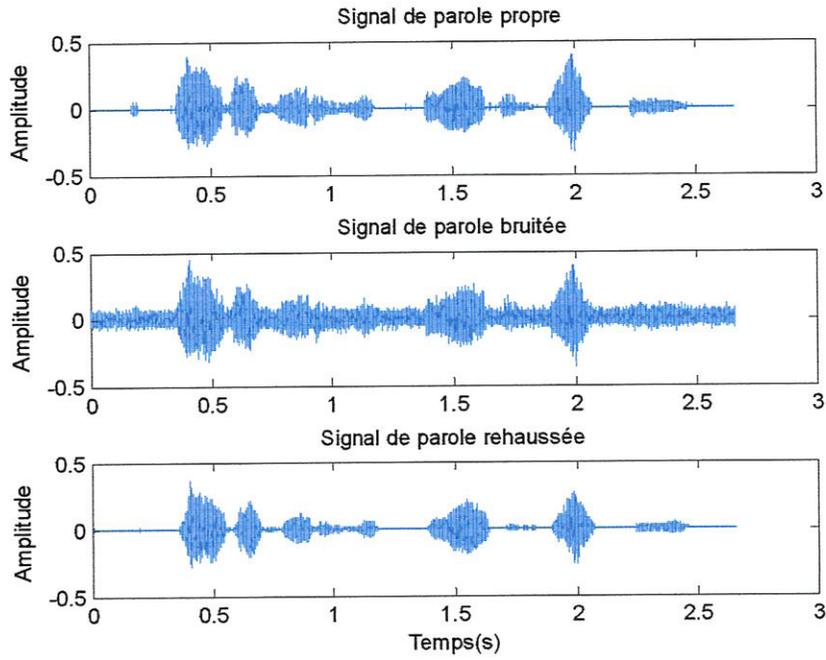


Figure 2.9. a) : Les formes d'ondes, cas d'un bruit blanc avec $\alpha = 2, \beta = 1$ et un SNR=5 dB.

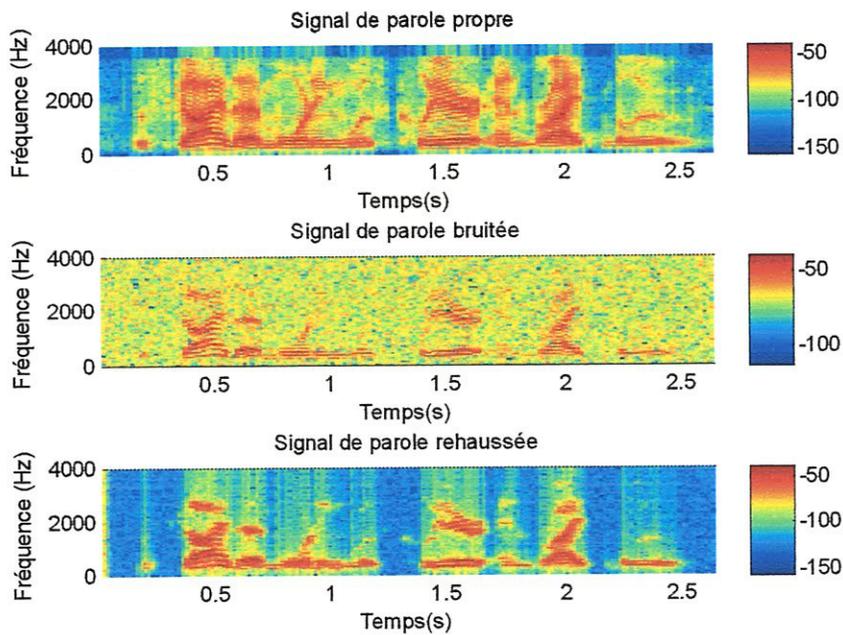


Figure 2.9 .b): Spectrogrammes, cas d'un bruit blanc avec $\alpha = 2, \beta = 1$ et un SNR=5 dB.

2.8 Conclusion

Au vu de ce chapitre, le principe et l'application de l'une des méthodes standards de réduction de bruit, qui est le filtre de Wiener avec sa version paramétrique ont été présentés.

Les résultats obtenus lors de la simulation, les formes d'ondes et les spectrogrammes confirment l'efficacité et la simplicité de ces algorithmes au débruitage de la parole mais, ils ne parviennent pas à éviter complètement le problème du bruit musical et résiduel. On verra dans le chapitre suivant les améliorations apportées sur le filtre de Wiener qui sont de type itératifs et leur impact sur les performances du débruitage de la parole.

Chapitre

Filtre de Wiener itératif

Dans ce chapitre :

- **Introduction.**
- **Calcul itératif du spectre.**
- **Calcul itératif du spectre basé sur le
model LPC.**
- **Tests et résultats.**
- **Conclusion.**



3.1 Introduction

Dans le problème du débruitage de la parole mono-capteur, on ne dispose que du signal observé alors qu'on fait intervenir des quantités a priori. On verra dans ce chapitre d'autres méthodes d'estimations de la densité spectrale de puissance du signal de parole, qui sont basées sur un calcul itératif direct ou après une modélisation LPC du signal de parole.

Ainsi, les différentes étapes des variantes itératives du filtre de Wiener, la détermination de sa fonction du gain et les résultats de simulation seront présentés aussi.

3.2 Calcul itératif du spectre

Dans le deuxième chapitre, on a vu que le filtre de Wiener est non causal et par conséquent non réalisable. De plus, il nécessite la connaissance du spectre de puissance du signal propre, qui n'est pas disponible. La solution est d'estimer la réponse du filtre de Wiener d'une manière itérative. En particulier, on considère un procédé itératif où à l'itération (i+1) le spectre du signal rehaussé est estimé par :

$$\hat{X}_{i+1}(k) = H_i(k)Y(k) \quad (3.1)$$

Où $H_i(k)$ indique la réponse fréquentielle du filtre de Wiener obtenu à l'itération (i).

On doit passer par les deux étapes suivantes à l'itération (i) pour estimer $H_i(k)$:

L'étape 1 : obtenir une estimation du filtre de Wiener $H_i(k)$ basée sur le signal rehaussé $\hat{x}_i(n)$. Dans la première itération, on initialise $\hat{x}_i(n)$ avec le signal de parole bruité $y(n)$.

L'étape 2 : filtrer le signal bruité $y(n)$ à travers le nouveau filtre de Wiener obtenu $H_i(k)$ selon l'équation (3.1), on obtient le signal rehaussé $\hat{x}_{i+1}(n)$.

Utiliser $\hat{x}_{i+1}(n)$ à la place de $\hat{x}_i(n)$ dans **l'étape 1** et répéter le procédé jusqu'à la convergence.

La figure (3.1) illustre les différentes étapes du filtre de Wiener itératif (qu'on désigne par Wiener itératif_A dans les simulations) :

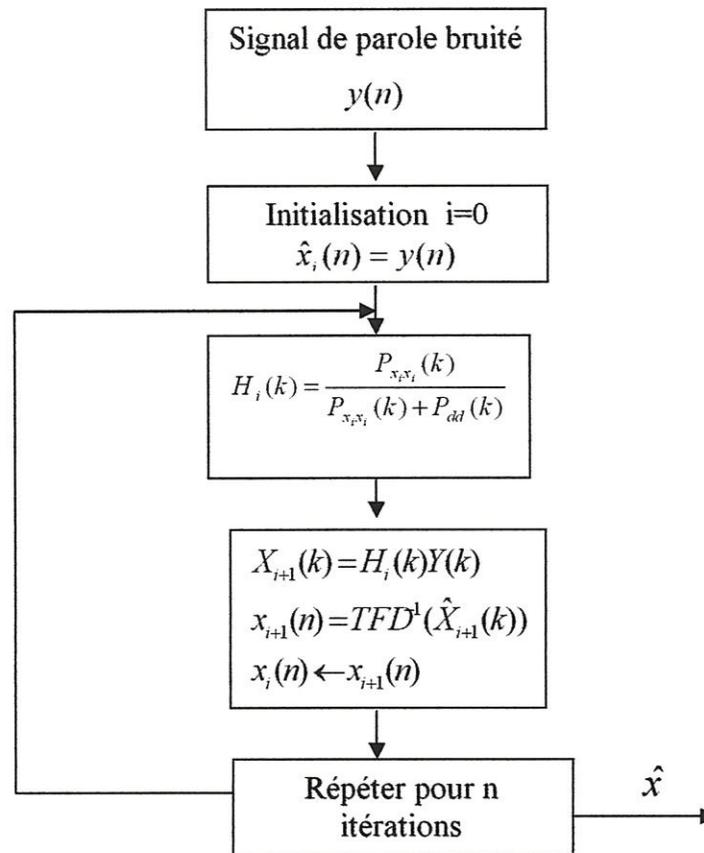


Figure 3.1 : Organigramme du filtre de Wiener itératif.

Comme avec tous les algorithmes itératifs, l'étude de la convergence est indispensable. On veut que l'algorithme itératif améliore le signal rehaussé dans chaque itération, en se basant sur certains critères.

3.3 Calcul itératif du spectre basé sur le modèle LPC

Un tel algorithme est proposé par Lim et Oppenheim [25] dans le contexte d'estimation des paramètres du modèle AR d'un signal bruité. Dans leur approche, la parole a été considérée comme la sortie d'un processus AR et le problème devient une estimation des paramètres AR propres. Alors, avec cette approche le problème du rehaussement de la parole a été réduit à l'estimation des paramètres AR supposés avoir été utilisés pour générer le signal propre.

3.3.1 Modèle mathématique de la production de la parole

La parole peut être la réponse d'un système linéaire invariant dans le temps qui est le conduit vocal, à une excitation périodique dans le cas d'un son voisé ou une excitation

aléatoire dans le cas d'un son non voisé. Le système qui modélise le conduit vocal a la forme tout-pôles suivante dans le domaine de Z :

$$V(z) = \frac{g}{A(z)} = \frac{g}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (3.2)$$

Où g est le gain du système, et $\{a_k\}$ sont les coefficients du modèle tout-pôles, qui peuvent être obtenus en utilisant l'analyse par prédiction linéaire et p est le nombre de coefficients (ordre de la prédiction).

Dans le domaine temporel, le signal de parole $x(n)$ est décrit par l'équation de récurrence suivante :

$$x(n) = \sum_{k=1}^p a_k x(n-k) + g.w(n), \quad n = 0, 1, \dots, N-1 \quad (3.3)$$

Où $w(n)$ est l'excitation d'entrée du système. On suppose que $w(n)$ est un bruit blanc Gaussien de moyenne nulle et de variance unité. On peut exprimer l'équation précédente d'une manière plus compacte comme suit :

$$x(n) = \mathbf{a}^T \mathbf{x}_p + g.w(n) \quad (3.4)$$

Où $\mathbf{a}^T = [a_1, a_2, \dots, a_p]$, et $\mathbf{x}_p^T = [x(n-1), x(n-2), \dots, x(n-p)]$.

L'équation (3.3) a besoin des conditions initiales pour $n < p$, on désigne par \mathbf{x}_i le vecteur des conditions initiales, où $\mathbf{x}_i^T = [x(-1), x(-2), \dots, x(-p)]$.

Dans le cas d'une observation bruitée on a :

$$y(n) = x(n) + d(n) = \mathbf{a}^T \mathbf{x}_p + g.w(n) + d(n) \quad (3.5)$$

Où $d(n)$ est le signal du bruit qui suit une distribution Gaussienne de moyenne nulle et de variance σ_d^2 .

A partir de l'équation (3.3), il est clair que le signal d'entrée (propre) $x(n)$ dépend en général de $2p+1$ paramètres : les p valeurs du vecteur de coefficients \mathbf{a} , les p valeurs initiales de \mathbf{x}_1 et le gain g .

Le modèle de production de la parole précédent offre une solution alternative au problème du rehaussement de la parole. Au lieu d'essayer d'estimer le signal propre $x(n)$ lui-même, on peut à la place estimer les paramètres qui sont supposés utilisés dans la génération du signal selon l'équation (3.5).

La valeur du gain g peut être facilement obtenue en utilisant le théorème de Parseval, et les p valeurs initiales de \mathbf{x}_1 n'affectent pas l'estimation de $\mathbf{a}^T = [a_1, a_2, \dots, a_p]$. La technique de la prédiction linéaire peut être utilisée pour l'estimation du vecteur \mathbf{a} dans le cas d'une observation bruitée, mais en général les coefficients estimés ne seront pas précis. Par conséquent, d'autres méthodes sont nécessaires pour estimer les coefficients $\mathbf{a}^T = [a_1, a_2, \dots, a_p]$ du signal propre à partir du signal bruité.

3.3.2 Estimation statistique des paramètres du modèle tout-pôles dans le bruit

Les techniques d'estimation statistiques ont été utilisées par Oppenheim et Lim [25] pour estimer les paramètres tout-pôles \mathbf{a} dans le bruit. Notamment, les techniques du maximum *a posteriori* (MAP). Dans l'approche de base, Oppenheim et Lim [25] considèrent la maximisation de la probabilité du vecteur observé \mathbf{a} sachant le signal de parole bruité :

$$\mathbf{a}_{MAP} = \max_a p(\mathbf{a}/\mathbf{y}) \quad (3.6)$$

Où $\mathbf{y}^T = [y(0) \ y(1) \ \dots \ y(N-1)]$ est le vecteur de parole bruitée.

La maximisation de $p(\mathbf{a}/\mathbf{y})$ conduit à un problème non linéaire difficile à résoudre. Alors, un algorithme itératif sous optimal a été utilisé à la place. Ils ont considéré l'estimation du MAP de $x(n)$ sachant les observations bruitées $y(n)$ et les coefficients \mathbf{a} $[p(x/\mathbf{a}, \mathbf{y})]$. Le gain g et les conditions initiales \mathbf{x}_1 sont supposés connues.

L'algorithme itératif suivant montre les différentes étapes dans ce cas :

Etape 1 : en se basant sur une estimation initiale de \mathbf{a} , qui est notée \mathbf{a}_0 , on estime x par maximisation de $p(x/\mathbf{a}_0, \mathbf{y})$. Cette première estimation de \mathbf{x} est notée \mathbf{x}_1 .

Etape 2 : on utilise \mathbf{x}_1 pour former une nouvelle estimation de \mathbf{a} , qui est noté \mathbf{a}_1 . Les coefficients \mathbf{a} peuvent être obtenus en utilisant la technique de la prédiction linéaire.

Aller à l'étape 1 et utiliser \mathbf{a}_1 à la place de \mathbf{a}_0 .

L'algorithme précédent est répété itérativement jusqu'à l'obtention d'une meilleure estimation des coefficients \mathbf{a} . Il a été prouvé dans [25] que l'algorithme itératif précédent converge vers un maximum local de la densité de probabilité conjointe $p(\mathbf{a}, \mathbf{x}/y)$.

Pour implémenter la première étape de l'algorithme itératif précédent, on doit maximiser $p(x/a_i, y)$ pour toutes les valeurs de $x(n)$. D'après la règle de Bayes $p(\mathbf{x}/a_i, y)$ peut être écrite comme suit :

$$p(\mathbf{x}/\mathbf{a}_i, \mathbf{y}) = \frac{p(\mathbf{y}/\mathbf{a}_i, \mathbf{x})p(\mathbf{x}/\mathbf{a}_i)}{p(\mathbf{y}/\mathbf{a}_i)} \quad (3.7)$$

Où \mathbf{a}_i sont les coefficients obtenus à l'itération i . Le dénominateur est indépendant de $x(n)$ et peut donc être ignoré dans la maximisation. On suppose que le gain g et le vecteur initial \mathbf{x}_i sont connus dans tous les termes de la densité de probabilité dans l'équation (3.7). L'estimation du MAP de \mathbf{x} requise dans l'étape 1 est donc donnée par :

$$\mathbf{x}_{MAP} = \max_x p(\mathbf{x}/\mathbf{a}_i, \mathbf{y}) = \max_x p(\mathbf{y}/\mathbf{a}_i, \mathbf{x})p(\mathbf{x}/\mathbf{a}_i) \quad (3.8)$$

La densité conditionnelle $p(\mathbf{y}/\mathbf{a}_i, \mathbf{x})$ est Gaussienne de moyenne $x(n)$ et de variance σ_d^2 (la même variance que celle du bruit $d(n)$) et a donc la forme suivante :

$$p(\mathbf{y}/\mathbf{a}_i, \mathbf{x}) = \frac{1}{(2\pi\sigma_d^2)^{N/2}} \exp\left[-\frac{1}{2\sigma_d^2} \sum_{n=0}^{N-1} (y(n) - x(n))^2\right] \quad (3.9)$$

La densité conditionnelle $p(\mathbf{x}/\mathbf{a}_i)$ est aussi Gaussienne de moyenne $\mathbf{a}^T \mathbf{x}_p$ et de variance g^2 , elle est donnée par :

$$p(\mathbf{x}/\mathbf{a}_i) = \frac{1}{(2\pi g^2)^{N/2}} \exp\left[-\frac{1}{2g^2} \sum_{n=0}^{N-1} (x(n) - \mathbf{a}^T \mathbf{x}_p)^2\right] \quad (3.10)$$

Substituant l'équation (3.9) et l'équation (3.10) dans l'équation (3.7), on obtient l'expression de $p(\mathbf{x}/\mathbf{a}_i, \mathbf{y})$:

$$p(\mathbf{x}/\mathbf{a}_i, \mathbf{y}) = C \frac{1}{(4\pi^2 \sigma_d^2 g^2)^{N/2}} \exp\left[-\frac{1}{2} \Delta_p\right] \quad (3.11)$$

Où C est une constante, et Δ_p est donné par :

$$\Delta_p = \frac{1}{g^2} \sum_{n=0}^N (x(n) - \mathbf{a}^T \mathbf{x}_p)^2 + \frac{1}{\sigma_d^2} \sum_{n=0}^{N-1} (y(n) - x(n))^2 \quad (3.12)$$

La maximisation de $p(\mathbf{x}/\mathbf{a}_i, \mathbf{y})$ est équivalente à la minimisation de Δ_p , parce que l'exposant dans l'équation (3.11) est négatif.

L'estimation de $x(n)$ (dans l'étape 1) est équivalente au choix des valeurs de \mathbf{x} tels que :

$$\frac{\partial \Delta_p}{\partial x(n)} = 0, \quad n = 0, 1, \dots, N-1 \quad (3.13)$$

L'équation précédente produit un système linéaire de N équations avec N inconnus. Le nombre d'échantillons N utilisés dans l'analyse peut être élevé et donc la solution demande un temps de calcul élevé. Heureusement, la solution de l'équation (3.13) peut être simplifiée en observant que Δ_p peut être exprimé sous la forme $(\mathbf{x} - \mathbf{m}_i)^T \sum_i^{-1} (\mathbf{x} - \mathbf{m}_i)$, et la densité conditionnelle devient :

$$p(\mathbf{x}/\mathbf{a}_i, \mathbf{y}) = C \frac{1}{(4\pi^2 \sigma_d^2 g^2)^{N/2}} \exp\left[-\frac{1}{2} (\mathbf{x} - \mathbf{m}_i)^T \sum_i^{-1} (\mathbf{x} - \mathbf{m}_i)\right] \quad (3.14)$$

Où \sum_i est la matrice de covariance et \mathbf{m}_i est le vecteur de la moyenne.

Selon l'équation précédente, la densité conditionnelle $p(\mathbf{x}/\mathbf{a}_i, \mathbf{y})$ est conjointement Gaussienne, alors l'estimation du MAP de \mathbf{x} est équivalente à l'estimation basée sur la minimisation de l'erreur quadratique moyenne (MMSE) de \mathbf{x} [26]. Car l'estimation de type MMSE de \mathbf{x} est également la moyenne a posteriori de la densité $p(\mathbf{x}/\mathbf{a}_i, \mathbf{y})$. Puisque la densité conditionnelle $p(\mathbf{x}/\mathbf{a}_i, \mathbf{y})$ est conjointement Gaussienne, le maximum et la moyenne de $p(\mathbf{x}/\mathbf{a}_i, \mathbf{y})$ sont identiques. De plus, pour des grandes valeurs de N (c-à-d des trames de longue-durée), on n'a pas besoin d'évaluer la moyenne de $p(\mathbf{x}/\mathbf{a}_i, \mathbf{y})$, mais on peut utiliser le filtre de Wiener non causal pour obtenir l'estimation de type MMSE de \mathbf{x} .

En résumé, l'estimation de type MAP de \mathbf{x} (équation (3.8), ou, d'une manière équivalente, la solution de l'équation (3.13), peut être simplement obtenue en filtrant le signal bruité $y(n)$ par le filtre de Wiener :

$$H(k) = \frac{P_{xx}(k)}{P_{xx}(k) + \sigma_d^2} \quad (3.15)$$

Où $P_{xx}(k)$ est la densité spectrale de puissance de $x(n)$ calculée en utilisant les valeurs de \mathbf{a}_i comme suit :

$$P_{xx}(k) = \frac{g^2}{\left| 1 - \sum_{m=1}^p a_m e^{-jmk} \right|^2} \quad (3.16)$$

Où les coefficients a_m dans l'équation précédente correspondent aux \mathbf{a}_i . Par conséquent, l'étape 1 peut être alternativement implémentée en utilisant le filtre de Wiener.

Le terme g du gain dans l'équation (3.16) peut être estimé en se basant sur le fait que la variance du signal de parole bruité $y(n)$ est égale à la somme des variances du signal propre et du bruit :

$$\sigma_y^2 = \sigma_x^2 + \sigma_d^2 \quad (3.17)$$

Où σ_y^2 est la variance (qui est aussi l'énergie) de $y(n)$, et σ_x^2 est la variance du signal estimé $x(n)$. La variance σ_x^2 du signal estimé peut être calculée dans le domaine fréquentiel en utilisant le théorème de Parseval :

$$\sigma_x^2 = \sum_{n=0}^{N-1} x^2(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} p_{xx}(\omega) d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{g^2}{\left| 1 - \sum_{m=1}^p a_m e^{-jm\omega} \right|^2} d\omega \quad (3.18)$$

Alors, en remplaçant l'équation (3.18) dans l'équation (3.17), on obtient l'équation suivante pour g :

$$g^2 = \frac{\frac{2\pi}{N} \sum_{n=0}^{N-1} y^2(n) - 2\pi\sigma_d^2}{\int_{-\pi}^{\pi} \frac{1}{\left| 1 - \sum_{m=1}^p a_m e^{-jm\omega} \right|^2} d\omega} \quad (3.19)$$

Notant que le dénominateur dans l'intégral peut être évalué en calculant la transformée de Fourier des coefficients tout-pôles.

Enfin, l'algorithme itératif du filtre de Wiener (qu'on désigne par Wiener itératif_B dans les simulations) est le suivant :

Initialisation : Initialiser l'estimation initiale du signal \mathbf{x}_0 par le signal bruité $y(n)$: $\mathbf{x}_0 = \mathbf{y}$.
pour $i = 0, 1, 2, \dots$ faire :

Étape 1 : à partir du signal estimé \mathbf{x}_i à l'itération i , calculer les coefficients \mathbf{a}_i en utilisant la technique de la prédiction linéaire.

Étape 2 : Utilisant \mathbf{a}_i , estimer le gain g^2 selon Equation (3.19).

Étape 3 : Calculer la densité spectrale de puissance à court terme du signal \mathbf{x}_i :

$$p_{x_i x_i}(k) = \frac{g^2}{\left| 1 - \sum_{m=1}^p a_i(m) e^{-jmk} \right|^2} \quad (3.20)$$

Où $\{ a_i(k) \}$ sont les coefficients estimés à l'étape 1.

Étape 4 : Calculer le filtre de Wiener :

$$H_i(k) = \frac{P_{x_i x_i}(k)}{P_{x_i x_i}(k) + \sigma_d^2} \quad (3.21)$$

Étape 5 : Estimer le spectre du signal rehaussé : $X_{i+1}(k) = H_i(k)Y(k)$, où $Y(k)$ est la densité spectrale du signal bruité $y(n)$.

Calculer la transformée de Fourier inverse de $X_{i+1}(k)$ pour obtenir le signal rehaussé $x_{i+1}(n)$ dans le domaine temporel.

Étape 6 : Aller à l'étape 1 en utilisant x_{i+1} pour le signal estimé et répéter jusqu'à la convergence de l'algorithme (selon un critère d'arrêt) ou pour un nombre d'itérations bien défini.

Dans l'algorithme précédent on a supposé que le bruit additif est un bruit blanc Gaussien de moyenne nulle et de variance σ_x^2 . Pour le bruit coloré, l'équation (3.21) prend la forme :

$$H_i(k) = \frac{P_{x_i x_i}(k)}{P_{x_i x_i}(k) + \hat{P}_{dd}(k)} \quad (3.22)$$

Où $\hat{P}_{dd}(k)$ est la densité spectrale de puissance du bruit estimé. La variance du bruit nécessaire pour le calcul du gain dans l'équation (3.19) peut être calculée dans le domaine fréquentiel comme suit :

$$\sigma_d^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{P}_{dd}(\omega) d\omega \quad (3.23)$$

En principe, l'algorithme itératif précédent est exécuté jusqu'à ce qu'un critère de convergence soit satisfait. En pratique, l'algorithme précédent est répété pour un nombre fixe d'itérations. Trois à sept itérations se sont avérés suffisants [27]-[28].

Quelques problèmes concernant le filtrage de Wiener itératif doivent être soulevés, comme :

- Le critère de convergence ou le nombre d'itérations qu'on doit utiliser ne sont pas clairs.
- Le spectre tout-pôles peut avoir des pics trop élevés, qui influent sur la qualité de la parole.
- De faux pics spectraux peuvent être produits si l'algorithme est exécuté pour un nombre élevé d'itérations.

Concernant le nombre optimal d'itérations aucune solution n'a été trouvée. De plus, le nombre optimal d'itérations varie suivant les différents types de son. Dans [27], les auteurs ont observé que les voyelles nécessitent en moyenne trois itérations pour une meilleure qualité de la parole, tandis que les consonnes ont besoin de deux itérations. Pour les nasals, les liquides et les fricatives de cinq à six itérations sont nécessaires.

Plusieurs solutions ont été proposées pour remédier aux deuxième et troisième problèmes [28]-[29]. Ces deux problèmes peuvent être traités en imposant des contraintes spectrales sur les caractéristiques du spectre tout-pôles. Ces améliorations seront détaillées dans le chapitre suivant.

3.4 Tests et résultats

Dans cette partie, les mêmes conditions d'expérimentations, la même base de données et les mêmes mesures objectives seront utilisées.

Les deux variantes du filtre de Wiener itératif étudiées dans ce chapitre seront testées et comparées. Ces deux variantes sont désignées par **Wiener itératif_A** et **Wiener itératif_B**.

Comme pour tout algorithme itératif, on commencera par la détermination du nombre optimal d'itérations qui assure un meilleur compromis entre la qualité du signal rehaussé et la distorsion. On appliquera la première version sur les trente (30) phrases de la base de données, dans le cas du bruit blanc, sous les deux valeurs du SNR global et ceci pour un nombre d'itérations qui varie entre 0 et quatre.

	0 dB				5 dB			
	LLR	WSS	SNRseg	PESQ	LLR	WSS	SNRseg	PESQ
Signal bruité	1.802	82.696	-5.081	1.539	1.545	69.972	-2.326	1.799
Wiener itératif_A (itér = 0)	1.535	73.175	-2.415	1.794	1.272	60.572	0.404	2.108
Wiener itératif_A (itér = 1)	1.191	62.207	0.251	2.186	0.945	50.579	2.935	2.538
Wiener itératif_A (itér = 2)	0.783	51.113	3.087	2.724	0.606	41.228	5.467	3.083
Wiener itératif_A (itér = 3)	0.780	51.341	4.978	2.909	0.661	42.824	6.815	3.213
Wiener itératif_A (itér = 4)	3.052	52.705	4.820	2.715	1.001	45.112	6.633	3.052

Tableau 3.1 : Mesures objectives, cas d'un bruit blanc pour différentes itérations.

Le tableau (3.1) montre que deux (02) ou trois (03) itérations sont un bon compromis. D'autres simulations ont été effectuées pour d'autres types de bruits et on a remarqué que trois itérations assurent les meilleures performances.

On cherche maintenant à comparer entre les performances de l'application de la variante du filtre de Wiener itératif_A, qui calcul itérativement le spectre d'une manière directe avec l'application de la version B, qui utilise une modélisation LPC d'ordre $P = 12$ pour le calcul du spectre et du gain. Les tableaux suivants présentent les résultats des mesures objectives obtenues pour trois types de bruits avec un nombre d'itérations égales à trois.

	0 dB				5 dB			
	LLR	WSS	SNRseg	PESQ	LLR	WSS	SNRseg	PESQ
Signal bruité	1.802	82.696	-5.081	1.539	1.545	69.972	-2.326	1.799
Wiener itératif_A	0.780	51.341	4.978	2.909	0.661	42.824	6.815	3.213
Wiener itératif_B	1.728	78.213	1.499	1.736	1.450	65.275	3.370	2.067

Tableau 3.2 : Evaluation objective de la qualité pour un bruit blanc.

	0 dB				5 dB			
	LLR	WSS	SNRseg	PESQ	LLR	WSS	SNRseg	PESQ
Signal bruité	1.014	66.943	-4.960	1.634	0.795	54.083	-2.173	1.891
Wiener itératif_A	0.528	49.887	4.378	2.812	0.451	41.639	6.146	3.109
Wiener itératif_B	2.785	70.689	-0.861	1.758	2.161	56.249	0.685	2.020

Tableau 3.3 : Evaluation objective de la qualité pour un bruit de voiture.

	0 dB				5 dB			
	LLR	WSS	SNRseg	PESQ	LLR	WSS	SNRseg	PESQ
Signal bruité	0.895	70.348	-4.632	1.705	0.715	56.019	-1.783	2.006
Wiener itératif_A	0.492	50.786	4.437	2.823	0.424	41.682	6.241	3.122
Wiener itératif_B	2.798	65.500	-0.551	1.877	2.152	51.942	0.968	2.161

Tableau 3.4 : Evaluation objective de la qualité pour un bruit de parole.

Vu les résultats des mesures objectives des tableaux (3.2) jusqu'à (3.4), la version (A) du filtre de Wiener itératif donne les meilleurs résultats dans toutes les situations et pour les

différents types de bruit. La version (B) demande moins de temps de calcul du spectre par rapport à la version (A), mais ses performances sont inférieures à celles de la version (A). L'utilisation de l'analyse LPC pour l'estimation du spectre du signal rehaussé à partir des coefficients LPC permet de réduire le temps de calcul, mais des contraintes doivent être imposées sur les paramètres dans ce cas.

3.5 Conclusion

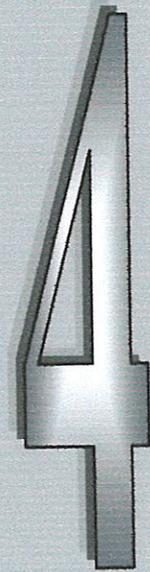
Les variantes de type itératif du filtre de Wiener ont été étudiées et comparées dans ce chapitre. La comparaison de ses variantes avec les variantes étudiées dans le chapitre deux et les améliorations à apporter sur les versions itératives seront présentées dans le chapitre suivant.

Chapitre

Filtre de Wiener itératif avec contraintes

Dans ce chapitre :

- **Introduction.**
- **Contraintes spectrales imposées sur le filtre itératif de Wiener.**
- **Filtre de Wiener itératif avec contraintes.**
- **Tests et résultats.**
- **Conclusion.**



4.1 Introduction

L'application du filtre de Wiener itératif au rehaussement de la parole a soulevée quelques problèmes concernant le critère de convergence et la forme du spectre tout-pôles durant les itérations, qui influent sur la qualité de la parole produite.

Plusieurs solutions ont été proposées [28][29], en imposant des contraintes spectrales sur les caractéristiques du spectre tout-pôles. D'une part, on peut imposer des contraintes de continuité spectrale afin de s'assurer que le spectre obtenu à une trame donnée n'est pas trop différent de celui obtenu au niveau de la trame précédente ou celui de la trame suivante. Ceci est motivé par le fait que les caractéristiques du conduit vocal ne changent pas radicalement d'une trame à l'autre. D'autre part, on peut imposer des contraintes au niveau des itérations pour s'assurer que seulement les spectres tout-pôles valides sont générés.

4.2 Contraintes spectrales imposées sur le filtre itératif de Wiener

Des contraintes spectrales peuvent être imposées au cours du temps (interframe) et au cours des itérations (intraframe) pour améliorer les performances de l'algorithme de filtrage de Wiener itératif [27][29][30].

4.2.1 Contraintes spectrales au cours du temps (interframe)

L'approche la plus simple pour imposer des contraintes au cours du temps est d'appliquer un lissage temporel sur le spectre estimé comme suit :

$$P_{xx}^{(l)}(k) = \tau P_{xx}^{(l-1)}(k) + (1 - \tau) P_{xx}^{(l)}(k) \quad (4.1)$$

Où $P_{xx}^{(l)}(k)$ représente le spectre estimé à la trame l , et τ est une constante de lissage $0 \leq \tau \leq 1$. Encore mieux, plutôt d'utiliser une valeur fixe pour τ , on peut varier τ dans le temps.

Quand le spectre change rapidement, comme dans le cas des transitions voisés/non voisés, par exemple ; on peut appliquer un lissage temporel faible. Par contre, on peut augmenter le lissage quand le spectre est relativement stationnaire, comme dans le cas d'un

segment d'une voyelle. On peut donc varier la constante de lissage τ selon une certaine mesure de stationnarité du spectre de signal. Le spectre lissé résultant est donné par :

$$P_{xx}^{(l)}(k) = \tau(l)P_{xx}^{(l-1)}(k) + (1 - \tau(l))P_{xx}^{(l)}(k) \quad (4.2)$$

Où $\tau(l)$ est dans ce cas, la constante de lissage qui varie dans le temps $0 \leq \tau(l) \leq 1$.

Alternativement, on peut imposer des contraintes directement sur les paramètres tout-pôles $\mathbf{a}^{(l)}$, parce que les pôles du polynôme de la prédiction linéaire $A(z)$ affectent la forme du spectre. Les contraintes spectrales sur les paramètres tout-pôles estimés à la trame l , $\mathbf{a}^{(l)}$, peuvent être imposées en utilisant les paramètres tout-pôles estimés dans les trames passées adjacentes (c.-à-d. $\mathbf{a}^{(l-1)}$, $\mathbf{a}^{(l-2)}$, ...) et les trames suivantes (c.-à-d. $\mathbf{a}^{(l+1)}$, $\mathbf{a}^{(l+2)}$, ...).

Comme le spectre du modèle tout-pôles peut avoir des pics de forme anormale. Ces pics peuvent être générés lorsque les modules des pôles du polynôme $A(z)$ sont très proches du cercle unité.

Une approche simple peut être adoptée dans ce cas, en évaluant le polynôme tout-pôles dans l'équation (3.20) avec un cercle de rayon r , c'est à dire, d'évaluer $A(re^{j\omega})$, où r est un paramètre qui contrôle la largeur des formants. On peut donc remplacer le dénominateur dans l'équation (3.20) avec:

$$\left| A(re^{jk}) \right|^2 = \left| 1 - \sum_{m=1}^p a_i(m) r^m e^{-jmk} \right|^2 \quad (4.3)$$

Pour $r < 1$, l'évaluation précédente déplace les pôles à l'intérieur du cercle unité et augmente la largeur des formants. De même, on peut imposer des contraintes angulaires sur les pôles, pour limiter le mouvement des emplacements des pôles à l'intérieur d'une fenêtre angulaire prédéfinie, ou de façon équivalente en limitant le déplacement des pics des formants dans une petite région. Pour ce faire, on a besoin d'un algorithme de détermination des racines du polynôme $A(z)$.

Comme la détermination des racines nécessite trop de calculs et numériquement moins précise, une procédure plus robuste a été étudiée dans [28] pour mettre en œuvre les contraintes précédentes. Elle est basée sur la représentation des fréquences de raies spectrales (LSF) des coefficients tout-pôles.

La représentation LSF utilise les deux polynômes avant et après de la prédiction linéaire (LPC), $P(z)$ et $Q(z)$ [30]. Les angles des racines de ces deux polynômes sont appelés les fréquences de raies spectrales, que nous désignons par l_i ($i = 1, 2, \dots, M$).

Les fréquences LSF d'indices impairs sont appelées les paramètres de position et correspondent approximativement à des fréquences des formants :

$$p_i = l_{2i-1} \text{ pour } i = 1, 2, \dots, M/2 \quad (4.4)$$

La distance entre les racines du polynôme $P(z)$ et les racines les plus proches du polynôme $Q(z)$ fournit les paramètres de différence, qui indiquent les largeurs des formants. Ceci correspond approximativement à la distance entre les LSF d'indices impairs et celles d'indices paires. Plus la distance est petite, plus la largeur du formant est petite. Les paramètres de différence d_i sont donnés par :

$$|d_i| = \min_{j=-1,1} |l_{2i+j} - l_{2i}|, \quad i = 1, 2, \dots, M/2 \quad (4.5)$$

Ainsi, les paramètres LSF nous fournit un moyen pour imposer des contraintes sur la position des fréquences des formants et la largeur des formants à la fois.

Pour réduire les variations des fréquences des formants d'une trame à l'autre, on peut appliquer des contraintes sur les paramètres de position. Les paramètres de position p_i sont lissés au cours du temps en utilisant les valeurs passées et futures de p_i .

Alors, le paramètre de position estimé à la trame (l) durant l'itération (i), désigné par : $\hat{p}_j^{(i)}(l)$ est lissé en utilisant une fenêtre de pondération triangulaire comme suit :

$$\hat{p}_j^{(i)}(l) = \sum_{m=-L}^L R \left[1 - \frac{|m|}{W} \right] \hat{p}_j^{(i)}(l+m), \quad j=1,2,\dots,M/2 \quad (4.6)$$

Où L est le nombre de trames passées et futures, R et W désignent la taille et la largeur de la fenêtre respectivement.

4.2.2 Contraintes au cours des itérations

En plus des contraintes au cours du temps, des contraintes au cours des itérations (intraframes) peuvent également être imposées [27]-[28]-[30]. Les contraintes ne sont pas appliquées aux coefficients tout-pôles directement, mais elles sont appliquées aux coefficients d'autocorrélation, qui seront utilisés pour le calcul des coefficients tout-pôles. Plus précisément, la valeur de l'autocorrélation estimée à l'itération (i), notée $r_{xx}^{(i)}(n)$, est calculée par une combinaison pondérée des valeurs des autocorrélations des itérations précédentes :

$$r_{xx}^{(i)}(n) = \sum_{m=0}^{i-1} \psi_m r_{xx}^{(i-m)}(n) \quad (4.7)$$

Où ψ_m sont les poids avec : $\sum_m \psi_m = 1$.

4.3 Filtre de Wiener itératif avec contraintes

Les différents types de contraintes spectrales étudiés précédemment combinés avec le filtre de Wiener conduit à la figure (4.1), qui représente le schéma bloc de l'algorithme du filtre de Wiener itératif avec contraintes [28].

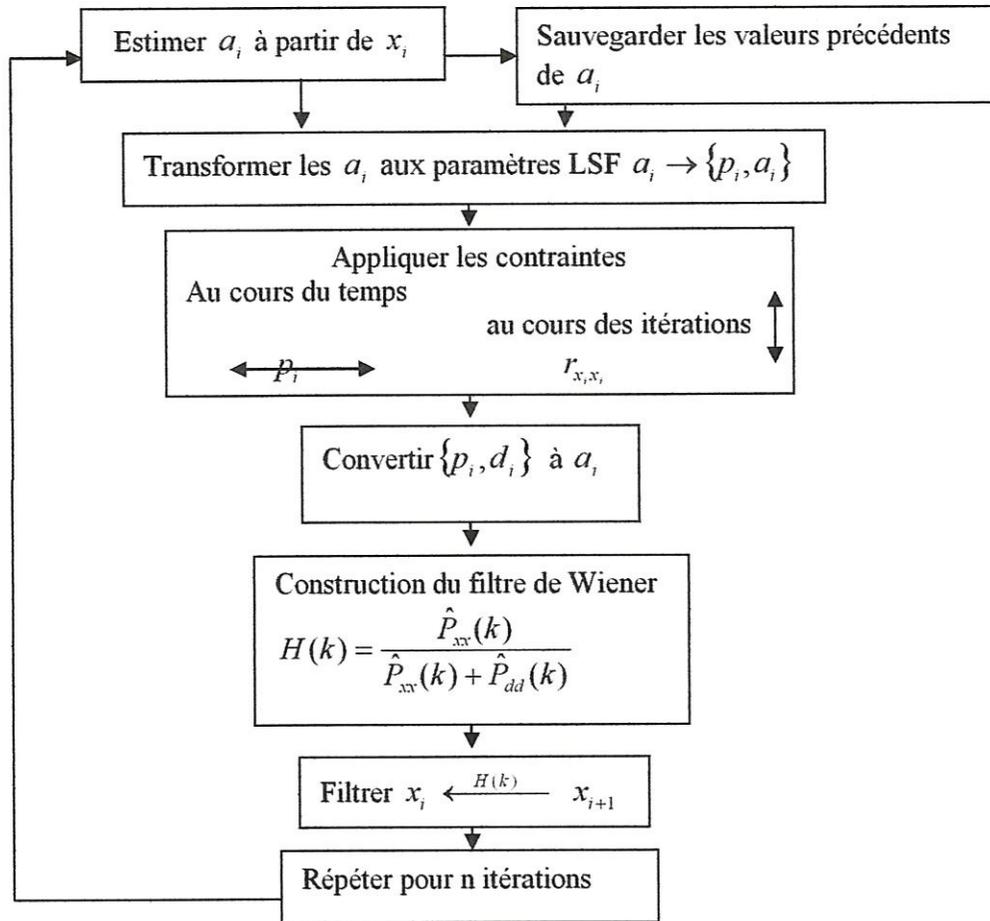


Figure 4.1 : Schéma bloc du filtre de Wiener itératif avec contraintes [28].

4.4 Tests et résultats

Afin d'appliquer les contraintes au cours des itérations de l'équation (4.7) sur les valeurs de la fonction d'autocorrelation de chaque trame, les poids ψ_m doivent satisfaire la condition : $\sum_m \psi_m = 1$. Une pondération de type triangulaire a été utilisée de la forme :

$$\psi_m = R \left(1 - \frac{m}{i} \right) \text{ for } 0 \leq m \leq i-1, \text{ où } R \text{ est choisie telle que } \sum_m \psi_m = 1. \text{ Le tableau (4.1)}$$

résume les différentes valeurs pour chaque itération :

Numéro d'itération	Valeur de R	Valeurs de ψ_m
Itération 0	1	$\psi_0 = 1$
Itération 1	2/3	$\psi_0 = 2/3, \psi_1 = 1/3$
Itération 2	1/2	$\psi_0 = 1/2, \psi_1 = 1/3, \psi_2 = 1/6$
Itération 3	2/5	$\psi_0 = 2/5, \psi_1 = 3/10, \psi_2 = 1/5, \psi_3 = 1/10$

Tableau 4.1 : Valeurs de la pondération triangulaire.

En ce qui concerne les contraintes spectrales au cours du temps seulement celles des équations (4.1) et (4.2) ont été programmées. Les tableaux suivants montrent les résultats de mesures objectives obtenues dans ce cas. Pour pouvoir faire la comparaison, quelques mesures objectives des variantes étudiées dans les chapitres précédents seront présentées aussi.

Mesures	0 dB				5 dB			
	LLR	WSS	SNRseg	PESQ	LLR	WSS	SNRseg	PESQ
Signal bruité	1.802	82.696	-5.081	1.539	1.545	69.972	-2.326	1.799
Wiener racine carrée	0.910	50.157	0.662	2.333	0.718	40.385	3.077	2.622
Wiener paramétrique ($\alpha = 2, \beta = 1$)	1.375	73.955	0.690	2.051	1.169	60.791	3.003	2.422
Wiener itératif_A	0.780	51.341	4.978	2.909	0.661	42.824	6.815	3.213
Wiener itératif_B	1.728	78.213	1.499	1.736	1.450	65.275	3.370	2.067
Wiener avec contraintes	1.648	74.299	2.296	1.675	1.399	62.485	3.764	2.008

Tableau 4.2 : Evaluation objective de la qualité pour un bruit blanc.

	0 dB				5 dB			
	LLR	WSS	SNRseg	PESQ	LLR	WSS	SNRseg	PESQ
Signal bruité	0.895	70.348	-4.632	1.705	0.715	56.019	-1.783	2.006
Wiener racine carrée	0.458	45.419	0.099	2.427	0.344	36.541	2.505	2.713
Wiener paramétrique ($\alpha = 2, \beta = 1$)	0.809	66.087	0.106	2.156	0.663	54.725	2.359	2.497
Wiener itératif_A	0.492	50.786	4.437	2.823	0.424	41.682	6.241	3.122
Wiener itératif_B	2.798	65.500	-0.551	1.877	2.152	51.942	0.968	2.161
Wiener avec contraintes	1.974	66.578	1.141	1.832	1.500	52.391	2.065	2.156

Tableau 4.3 : Evaluation objective de la qualité pour un bruit de parole.

D'après les mesures objectives des tableaux précédents, on remarque que :

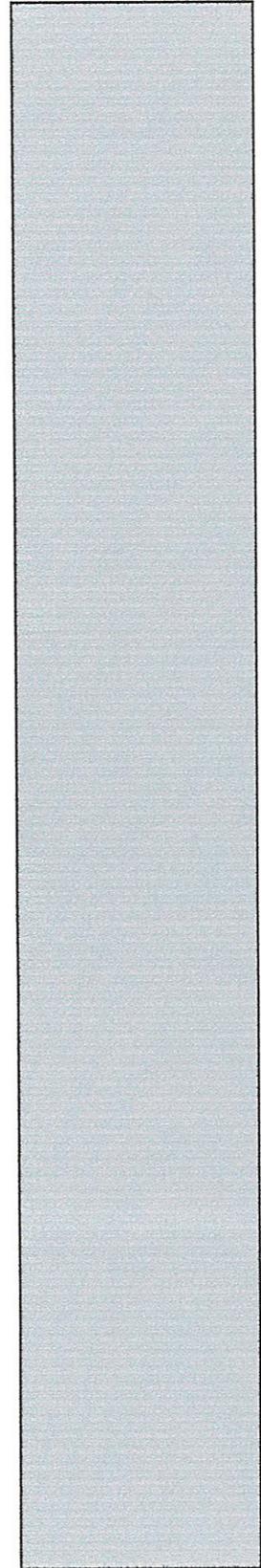
- Dans le cas du bruit blanc et pour les deux valeurs du SNR global, le filtre de Wiener itératif avec contraintes assure les meilleures performances par rapport aux versions (racine carrée, paramétrique et itérative_B).

- Dans le cas du bruit babble et le SNR = 0 dB, la même remarque précédente mais pour un SNR = 5 dB le filtre de Wiener itératif avec contraintes présente des performances meilleures par rapport au filtre de Wiener itératif_B mais très proches de celles des variantes racine carrée et paramétrique.
- La version du filtre de Wiener itératif_A reste toujours la méthode la plus performante par rapport aux autres versions.

4.5 Conclusion

Les contraintes spectrales au cours du temps et au cours des itérations imposées sur le spectre du signal augmentent d'avantage les performances du filtre de Wiener itératif. A noter que des contraintes sont difficiles à mettre en œuvre et nécessitent une attention particulière lors de la simulation.

Conclusion générale



Conclusion générale

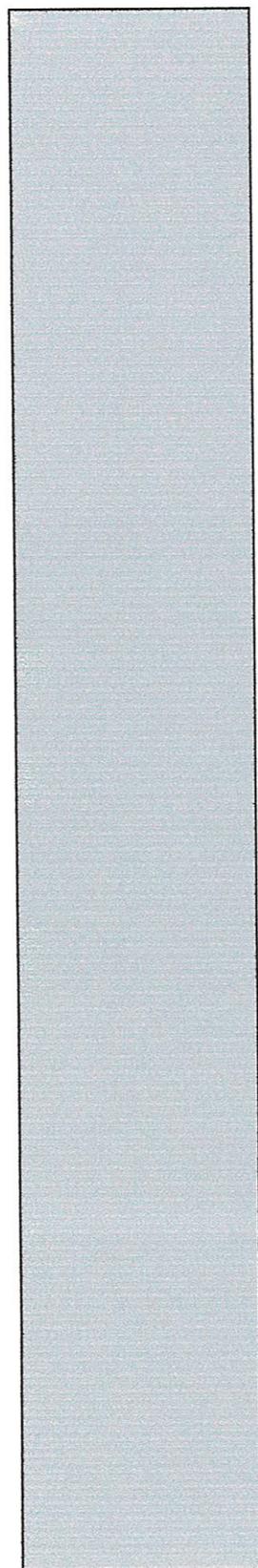
Ce manuscrit a traité un problème qui concerne les méthodes de rehaussement de la parole en situation monovoie d'un signal acoustique perturbé par un bruit additif. On s'est intéressé principalement par l'étude globale de la théorie du filtre de Wiener classique, ses versions paramétrique et racine carrée. Suivi par ses méthodes itératives sans et avec contraintes spectrales.

L'étude et l'implémentation de la méthode de réduction de bruit du filtre de Wiener classique avec sa version paramétrique, confirment l'efficacité de ces algorithmes à la réduction du bruit résiduel et montrent clairement que la version paramétrique est meilleure par rapport à la variante de base de Wiener et que la version racine carrée est la plus performante.

Dans la catégorie du filtre de Wiener itératif, le calcul itératif appliqué directement sur le spectre du signal donne les meilleures performances par rapport aux différentes méthodes simulées dans ce mémoire. Son inconvénient est qu'il nécessite temps de calcul élevé. Cependant l'utilisation de l'analyse LPC pour l'estimation du spectre du signal rehaussé à partir des coefficients LPC permet de réduire le temps de calcul, mais des contraintes doivent être imposées sur les paramètres pour avoir des mesures équivalentes ou meilleures.

Enfin, les contraintes spectrales au cours du temps et au cours des itérations ajoutées au filtre de Wiener itératif augmentent d'avantage ses performances.

Bibliographie
et
Webographie



Bibliographie et Webographie

- [1] R. Boite, et M. Kunt, *Traitement de la parole*, Presses Polytechniques Romandes, Complément au Traité d'Electricité, Première édition, 1987.
- [2] Procédés de synthèse de la parole,
<http://www.iict.ch/Tcom/Laboratoires/digivox2000/chap/chap5/synthese.htm>
- [3] René Boite, Hervé Boulard, Thierry Dutoit, Joil Hancq et Henri Leich, *Traitement de la parole*, Presses Polytechniques et Universitaire Romandes, Novembre 1999.
- [4] F. J. Harris, "On the use of windows for harmonic analysis with discrete Fourier transform", Proc. of IEEE, vol. 66, pp. 51-83, Jan. 1978.
- [5] Anne Bonneau and Yves Laprie. "Selective acoustic cues for French voiceless stop consonants", the Journal of the Acoustical Society of America, vol. 123, pp. 4482-4497, 2008.
- [6] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system", IEEE Trans. Speech and Audio Processing, vol. 7, pp. 126-137, 1999.
- [7] Marie. Guéguin, "Évaluation objective de la qualité vocale en contexte de conversation," Thèse de Doctorat, Université de Rennes 1, Décembre 2006.
- [8] Scott Keagy, *Integrating voice and data networks : Practical solutions for the new world of packetized voice over data networks*, Cisco Press, 816 pages, 2000.
- [9] S. Wang, A. Sekey and A. Gersho, "An objective measure for predicting subjective

- quality of speech coders,” IEEE J. on Select. Areas in Commun, vol. SAC-10, pp.819-829, Sept 1992.
- [10] Y. Hu and P. Loizou, “Incorporating a psychoacoustic model in frequency domain speech enhancement”, IEEE Signal Processing Letters, vol. 11(2), pp. 270-273, Feb. 2004.
- [11] N. Ma, M. Bouchard and R. A. Goubran, “Perceptual Kalman filtering for speech enhancement in colored noise”, In Proc. ICASSP’04, Montreal, Canada, volume 4, pp. 1045-1048, 2004.
- [12] Y. Hu and P. Loizou, “A comparative intelligibility study of speech enhancement algorithms”, IEEE Signal Processing Letters, vol. 4, pp. 561-564, April 2007.
- [13] ITU P.861, Objective quality measurement of telephone-band (300-3400 Hz) speech codecs, ITU-T Recommendation P.861, 1998.
- [14] ITU P862, Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs, 2000.
- [15] S. Wang, A. Sekey and A. Gersho, “Modified bark spectral distortion measure which uses noise masking threshold”, IEEE Speech Coding Workshop, vol. SAC-10, 1997.
- [16] A. Rix, J. Beerends, M. Hollier and A. Hekstra, “Perceptual evaluation of speech quality (PESQ) – A new method for speech quality assessment of telephone networks and codecs”, Proc. IEEE ICASP, vol. 2, pp. 749-752, 2001.
- [17] D.H. Klatt, “Prediction of perceived phonetic distance from critical-band spectra: a first step”, in Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP), 1982.
- [18] Kotta Manohar, “Single channel enhancement of noisy speech”, M.Tech. Credit Seminar Report, Electronic Systems Group EE Dept, IIT Bombay, November 2002.

- [19] N. Wiener, *Extrapolation, Interpolation and smoothing of stationary time series with engineering applications*, Cambridge, MA: MIT Press, 1949.
- [20] N. Wiener and E. Hopf, *On a class of singular integral equations*, Proc. Russ. Acad. Math. Phys., 696 pages, 1931.
- [21] J. Chen, J. Benesty, Y. Huang and S. Doclo, “New insights into the noise reduction Wiener filter”, IEEE Trans. Speech Audio Process., vol. 14(4), pp. 1218-1234, 2006.
- [22] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,” IEEE Transactions on Acoustics, Speech and Signal Processing, vol. ASSP-32, no.6, pp. 1109-1121, December 1984.
- [23] J. Lim and A.V. Oppenheim, “Enhancement and bandwidth compression of noisy speech”, Proc. IEEE, vol. 67(12), pp. 1586-1604, 1979.
- [24] A noisy speech corpus (NOIZEUS), <http://www.utdallas.edu/~loizou/speech/noizeus/>
- [25] J. Lim and A.V. Oppenheim, “All-pole modeling of degraded speech”, IEEE Trans. Acoust. Speech Signal Processing, vol. 26(3), pp. 197-210, 1978.
- [26] S. Kay, *Fundamentals of statistical signal processing: Estimation theory*, Upper Saddle River, NJ: Prentice Hall 1993.
- [27] J. Hansen, “Analysis and compensation of stressed and noisy speech with application to robust automatic recognition”, Ph.D Thesis, Georgia Institute of Technology, 1988.
- [28] J. Hansen and M. Clements, “Constrained iterative speech enhancement with application to speech recognition”, IEEE Trans. Signal Process., vol. 39(4), pp. 795-805, 1991.
- [29] J.H.L. Hansen and M. Clements, “Constrained iterative speech enhancement with application to speech recognition”, Proc. IEEE Int. Conf. Acoust. Speech Signal Process, vol. 1, pp. 561-564, 1988.

- [30] T. Sreenivas and P. Kirnapure, "Codebook constrained Wiener filtering for speech enhancement", *IEEE Trans. Acoust. Speech Signal Processing*, vol. 4(5), pp. 383-389, 1996.

- [31] S. Furui, *Digital speech processing, synthesis and recognition*, Second edition, New York: Marcel Dekker, 2001.

- [32] B. Pellom and J. Hansen, "An improved (Auto:I, LSP:T) constrained iterative speech enhancement for colored noise environments", *IEEE Trans. Speech Audio Processing*, vol. 6(6), pp. 573-579, 1998.