

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne démocratique et populaire

وزارة التعليم العالي و البحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique

جامعة سعد دحلب البليدة
Université SAAD DAHLAB de BLIDA

كلية التكنولوجيا
Faculté de Technologie

قسم الإلكترونيك
Département d'Électronique



Mémoire de Projet de Fin d'Études

présenté par

Ouserir Mohamed

&

Ouserir abdellah

pour l'obtention du diplôme de Master en Électronique spécialité Réseaux et
Télécommunication.

Thème

Réduction du bruit par des techniques mono-capteurs

Proposé par : **Dr DJENDI MOHAMED**

Année Universitaire 2012-2013

Remerciements

Nous tenons avant tout de remercier le bon DIEU qui nous a donnés la volonté et le courage pour la réalisation de ce travail.

Nous remercions vivement « Mr. DJENDI MOHAMED » notre promoteur pour la précieuse assistance, sa disponibilité et son soutien qu'il nous accordé tout au long de ce projet.

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

Nous tenons à exprimé notre reconnaissance à tous les enseignants de la faculté des Sciences de l'ingénieur plus particulièrement le département d'électronique qui nous ont attribués.

Sans omettre bien sur de remercier profondément tous ceux qui ont contribué de prés ou de loin à la réalisation du présent travail.

الإهداء

الحمد لله على توفيقه و فضله. لإتمام هذا العمل المتواضع, ونسأله أن يتقبله منا و أن يجعله خالصا لوجهه الكريم.

إلى من ربنتي وروتني من فيض حنانها، أعيها ندائي وسهرت الليالي من أجلي.
أمي العزيزة أطال الله في عمرها.

إلى من سعى من أجلي وكان حريصا على تعليمي وسهر الليالي من أجلي.
أبي العزيز أطال الله في عمره.

Dédicaces

Je dédie ce travail à :

- Mes parents pour leurs sacrifices et leurs encouragements.
- Mes sœurs et mes frères.
- Toute la famille Ouserir.
- Tous mes amis(es) en particulier Rida Bendoumia et Zoulikha

Meriem.

Mohamed et Abdellah

ملخص : بسبب الأهمية المتزايدة للاتصالات المتنقلة، والحاجة إلى تحسين التقاط الصوت، يصبح هذا الأمر أكثر أهمية خاصة عن طريق الحد من الإزعاج بسبب الضوضاء. وقد تم اقتراح عدة أساليب للحد من الضوضاء ذات استشعار واحد مثل تقنية الطرح الطيفية وتصفية وينر. وتستند هذه الأخيرة على الاستقلالية بين الإشارة الصوتية والضجيج، و العائق الرئيسي لهذه التقنيات هو التشويه والضوضاء الموسيقية. وقد استخدمنا تقنية القرار الموجه الذي يستند على تقدير النسبة بين إشارة الصوت و الضجيج (البداهة و البعدية). في مشروعنا، نقترح استخدام تقنية القرار الموجه مع مختلف معاملات الطرح الطيفية و معامل وينر. قمنا بعدة مقارنات بين التقنيات المقترحة وأسلوب المرجعية (بول). هذه التقنيات المقترحة تخضع لحل وسطي بين المستوى الفعال للحد من الضوضاء والتشويه المؤثر على الصوت. تظهر النتائج المحصل عليها تحسينات واضحة في مستوى الصوت مقارنة مع الطرق المرجعية أي أنها تظهر تحسينات فعالة جدا.

كلمات المفاتيح : الصوت ، الضوضاء الموسيقية ، تقنية القرار الموجه، تصفية وينر.

Résumé : Avec l'avènement des télécommunications mobiles grand public, le besoin d'améliorer la prise de son, notamment en réduisant la gêne due au bruit, s'est fait de plus en plus présent. Plusieurs techniques de réduction du bruit mono-capteurs qui sont proposées comme la soustraction spectrale et le filtrage de Wiener. Dans Ces techniques en supposant que le signal de parole est indépendant au bruit, et leur majeure inconvénient est la distorsion et le bruit musical.

Nous avons utilisé la technique de discision dirigé (DD) qui est basée sur l'estimation du RSB a priori et RSB a posteriori. Dans notre projet, nous proposons d'utiliser la technique DD avec différents gains de la soustraction spectrale (GSS) et de Wiener (GW). Nous avons fait plusieurs comparaisons entre les techniques proposées et une technique de référence (Boll). Cette technique compromis entre le niveau effectif de réduction de bruit et la distorsion qui affecte le signal de parole. Les résultats obtenus montrent des améliorations significatives par rapport aux performances des techniques de référence, ils montrent un comportement très efficace.

Mots clés : Parole ; bruit musical; technique de discision dirigé ; filtrage de Wiener.

Abstract : Because of the growing importance of mobile telecommunication, the need to enhance the sound pick-up, especially by reducing the inconvenience due to the noise, becomes more important. Several techniques have been proposed for single-sensor noise reduction as the spectral subtraction and Wiener filtering. These latter are based on the independence of speech signal and noise, and their disadvantage are the distortion and musical noise.

We have used the decision directed (DD) that is based on the estimation of a priori SNR and a posteriori SNR. In our project, we propose to use the DD technique with different gain of the spectral subtraction (SSG) and of Wiener (WG). We have done several comparisons between the proposed techniques and reference technique (Boll). These proposed techniques are subject to a compromise between the effective level of noise reduction and the distortion of the enhanced speech. The results show significant performance compared to standard techniques they show very effective comportment improvements.

Keywords : Speech; Musical noise; Decision directed; Wiener filtering.

Listes des acronymes et abréviations

Liste des symboles

$E[X]$: Espérance mathématique de X

\mathbf{R} : Matrice de corrélation.

\mathbf{r} : Vecteur d'inter-corrélation.

$\mathbf{H}(z)$: fonction de transfert

$\mathbf{R}[X]$: partie réelle de la quantité complexe X

$[\cdot]$: Partie entière

$\mathbf{x} * \mathbf{y}$: convolution de x et y

δ : distribution de Dirac

\cdot^* : Conjugué d'un vecteur, matrice, ou nombre complexe.

$\mathbf{x}(n)$: signal temporel discret

$|\mathbf{X}(p,k)|$: module de $\mathbf{X}(p,k)$

$\angle \mathbf{x}(p,k)$: phase de $X(p,k)$

$G(p,k)$: gain spectral

$\gamma_x(\mathbf{f})$: DSP de x

$\gamma_x(\mathbf{k}) = E[|X(p,k)|^2]$: DSP de X

F_e : fréquence d'échantillonnage

F_0 : fréquence fondamentale ou pitch

p : indice temporel de trame

$\mathbf{x}(p,n)$: trame p du signal $x(n)$

σ_x^2 : variance de x

μ_x : moyenne de x

f : fréquence

k : indice fréquentiel de la fréquence discrète f_k

$\mathbf{X}(p,k)$: TFCT de $x(p,n)$

Liste Abréviations

SSA: soustraction spectrale d'amplitude.

SSP : soustraction spectrale de puissance

SSG : soustraction spectrale généralisée

EQMM: erreur quadratique moyenne minimum

DD : decision dirigé

DAV : détection d'activité vocale

dB : décibel

DC : distance cepstrale

DSP : densité spectrale de puissance

FFT : fast Fourier Transform (transformée de Fourier rapide)

IFFT : inverse fast Fourier transform (transformée de Fourier inverse rapide)

OLS: technique de synthèse par overlap- and-save

OLA: technique de synthèse par overlap- and-add

RSB : rapport signal à bruit

TCD : transformée en cosinus discret

TF : transformée de Fourier

TFCT : transformée de Fourier à court terme

TFCTI : transformée de Fourier à court terme inverse

TFD : transformée de Fourier discrète(ou DFT : Discrete Fourier Transform)

TFDI : transformée de Fourier discrète inverse

Table des matières

Introduction générale	1
Chapitre 1 Généralité sur le traitement de signal	3
1.1 Introduction	3
1.2 Notion de signal.....	3
1.3 Présentation de traitement de signal.....	4
1.3.1 Théorie du signal	4
1.3.2 Traitement du signal	4
1.4 Transmission et le traitement du signal.....	5
1.4.1 Elaboration des signaux	5
1.4.2 Interprétation des signaux	5
1.5 Représentation des signaux	5
1.6 classe des signaux	5
1.6.1 signaux analogiques.....	5
1.6.2 signaux numériques	6
1.7 Notion de système.....	6
1.7.1 Système analogique	6
1.7.2 Système numérique.....	6
1.8 Les principales fonctions de traitement du signal.....	7
1.8.1 Filtrage numérique.....	7
1.8.2 la convolution	8
1.8.3 Corrélation.....	9
1.8.4 Echantillonnage	9
1.8.5 Reconstitution.....	10
1.8.6 Quantification.....	10
1.8.7 Codage	11
1.8.8 L'analyse de Fourier.....	11
1.9 Généralité sur le signal de parole et le bruit	11
1.9.1 Classification des sons de la parole	12
1.9.2 Propriétés statistiques du signal de parole	13
1.9.3 Production de signal de parole.....	13
1.9.4 Perception de la parole.....	14
1.9.5 Analyse et paramétrisation de la parole	14

1.9.6	Le bruit.....	14
1.10	conclusion.....	16
Chapitre 2	Atténuation spectrale a court terme.....	17
2.1	Introduction.....	17
2.2	Principe d'atténuation spectrale à court terme.....	18
2.3	Rapports signal à bruit.....	19
2.4	Mise en œuvre de l'atténuation spectrale a court terme.....	20
2.5	Méthodes d'atténuation spectrale à court terme.....	22
2.5.1	Soustraction spectrale.....	22
2.5.2	Filtrage du Wiener.....	25
2.6	Techniques d'estimation du bruit.....	26
2.6.1	Estimation de bruit nécessitant une DAV.....	26
2.6.2	Estimation de bruit e continu (sans DAV).....	27
2.7	L'overlap and save (OLS).....	28
2.8	L'overlap and add (OLA).....	28
2.9	conclusion.....	29
Chapitre 3	formulation des différentes techniques de réduction de bruit.....	30
3.1	introduction.....	30
3.2	Techniques améliorées.....	30
3.3	Technique décision-dirigée DD.....	30
3.4	Technique de Boll.....	33
3.5	Critères objectifs de qualité.....	34
3.6	Conclusion.....	35
Chapitre 4	Résultats des Simulations.....	36
4.1	Introduction.....	36
4.2	Simulation.....	36
4.2.1	Description des signaux de tests.....	36
4.2.2	Evaluation du RSB et la distance cepstrale.....	38
4.3	Conclusion.....	50
	Conclusion générale.....	51
	Bibliographie.....	52

Liste des figures

Figure 1. 1. Numérisation des signaux de calculateur numérique.	6
Figure 1. 2. Le principe de système.	6
Figure 1. 3. Echantillonnage d'un signal.	10
Figure 1. 4. Modèle simple de production de la parole.	13
Figure 1. 5. Fonction d'autocorrélation (à gauche) et densité spectrale de puissance (à droite) d'un bruit blanc.	15
Figure 2. 1. Schéma de principe des techniques de réduction de bruit.	21
Figure 3. 1. Schéma du principe général de l'approche DD.	31
Figure 4. 1. Signal de parole (masculin) avec sa segmentation manuelle.	37
Figure 4. 2. Signal de parole (féminin) avec la segmentation manuelle.	37
Figure 4. 3. Un signal de Bruit Blanc.	38
Figure 4. 4. Un signal de Bruit USASI.	38
Figure 4. 5. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) avec un bruit blanc à l'entrée.	40
Figure 4. 6. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) avec un bruit blanc à la l'entrée.	41
Figure 4. 7. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) avec un bruit Blanc à l'entrée.	41
Figure 4.8. Évaluation temporelle du signal de mélange (noir) et de sortie (gris) avec un bruit USASI en entrée.	43
Figure 4. 9. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) avec un bruit USASI en entrée.	44
Figure 4. 10. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) avec un bruit USASI à l'entrée.	45
Figure 4. 11. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) avec un bruit Blanc à l'entrée.	46
Figure 4. 12. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) avec un bruit Blanc à l'entrée.	47
Figure 4. 13. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) avec un bruit Blanc à l'entrée.	47
Figure 4. 14. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) avec un bruit USASI à l'entrée.	49
Figure 4. 15. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) avec un bruit USASI à l'entrée.	49
Figure 4. 16. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) av un bruit USASI à l'entrée.	50

Liste des tableaux

Table 4. 1. Evaluation du RSB avec le bruit blanc à l'entrée.....	39
Table 4. 2. Evolution de la CD avec le bruit blanc à l'entrée.....	39
Table 4. 3. Evaluation du RSB avec le bruit USASI à l'entrée.....	42
Table 4. 4. Evaluation du CD avec le bruit USASI à l'entrée.....	42
Table 4. 5. Evaluation du RSB avec le bruit Blanc à l'entrée.....	45
Table 4. 6. Evaluation du CD avec le bruit Blanc à l'entrée.....	46
Table 4. 7. Evaluation du RSB avec le bruit USASI à l'entrée.....	48
Table 4. 8. Evaluation du CD avec le bruit USASI à l'entrée.....	48

Introduction générale

La parole est un moyen de communication privilégié entre les humains, constitue une grande partie des messages transmis en télécommunications. Avec l'avènement des télécommunications mobiles grand public, le besoin d'améliorer la prise de son, notamment en réduisant la gêne due au bruit, s'est fait de plus en plus présent. Les techniques de réduction du bruit sont soumises à un compromis entre le niveau effectif de réduction et la distorsion qui affecte le signal de parole. Au vu des performances actuelles, il est souhaitable de supprimer plus de bruit tout en conservant un niveau de dégradation acceptable du signal restauré, ceci en particulier lorsque le niveau de bruit est important.

Les techniques qui ont suscité le plus d'intérêt au cours de ces 30 dernières années sont les approches par atténuation spectrale à court terme qui consistent à modifier une transformée à court terme du signal bruité en utilisant une règle de suppression. L'essor de cette famille de techniques s'explique essentiellement par le fait qu'elles permettent de respecter les contraintes de temps réel et de complexité inhérentes aux applications de communication parlée. Nous présentons ci-après les différents chapitres de ce mémoire :

Chapitre 1 est divisé en deux parties : la première partie présente les principaux outils mathématiques du traitement numérique du signal. Dans la deuxième partie, nous présentons des généralités sur la parole, le bruit et leurs différentes caractéristiques.

Chapitre 2 est consacré à la présentation d'une vision unifiée des principales techniques de réduction de bruit par atténuation spectrale à court terme (limitée aux systèmes monovoie).

Chapitre 3 présente la formulation des différentes techniques de réduction de bruit utilisées dans ce travail.

Chapitre 4 présente les simulations des techniques de réduction de bruit présentées dans le chapitre 3 à savoir la technique de soustraction spectrale de Boll, la technique de

décision dirigé et les différentes nouvelles versions de cette dernière méthode que nous proposons dans ce travail de Master. Ce présent document sera terminé par une conclusion générale.

Chapitre 1 Généralité sur le traitement de signal

1.1 Introduction

La théorie et le traitement des signaux est une discipline auxquelles elle apporte des bases théoriques fondamentales et des techniques particulières. Le traitement numérique du signal désigne l'ensemble des opérations, calculs arithmétiques et manipulations de nombres, qui sont effectués sur un signal à traiter. Les fonctions les plus variées sont réalisables de cette manière, comme le filtrage linéaire ou non linéaire, l'extraction de paramètres, détection, la modulation, l'estimation et l'analyse spectrale.

Notre travail consiste à nous intéresser plus particulièrement au domaine du traitement de la parole, La parole est la forme la plus commune de communication humaine, elle se distingue des autres sons par des caractéristiques acoustiques ayant leurs origines dans les mécanismes de production. Les sons de parole sont produits soit par les vibrations des cordes vocales (voisement), soit par l'écoulement turbulent de l'air dans le conduit vocal, soit lors du relâchement d'une occlusion de ce conduit (bruit).

1.2 Notion de signal

Le terme signal sert à désigne une grandeur physique le plus souvent de nature électrique, comme celle que l'on observe à la sortie d'un microphone, cette grandeur subit des transformations lors de son passage dans un système. Ainsi, dans une chaîne de communication, le signal est un sujet à des modifications (distorsion, affaiblissement,...) que peuvent le rendre méconnaissable. Il s'agit de comprendre cette évolution pour pouvoir récupérer dans de bonnes conditions le message initial.

Le signal est le support physique de l'information. il transporte les ordres sans les équipements de contrôle et de télécommande, il achemine sur les réseaux l'information, le parole ou l'image. Il est particulièrement fragile et doit être manipulé avec beaucoup

de soin. Son traitement réunit des théories et des méthodes relativement indépendantes du signal considéré. [1]

Soit $x(t)$ une grandeur physique mesurable dépendant d'un paramètre $t \in S$ (l'ensemble de variable t).

La nature de $x(t)$ peut représenter n'importe quel phénomène physique, soit :

- Natural (température).
- Électrique.
- Signal optique.
- Onde acoustique

La variable t peut désigner :

- Temps
- Position angulaire
- Position spatiale

1.3 Présentation de traitement de signal

1.3.1 Théorie du signal

La théorie du signal a pour objectif fondamental la description mathématique des signaux. Cette représentation commode du signal permet de mettre en évidence ses principales caractéristiques (distribution fréquentielle, énergie, etc.) et d'analyser les modifications subies lors de la transmission ou du traitement de ces signaux.

1.3.2 Traitement du signal

Le traitement du signal est la discipline technique qui, s'appuyant sur les ressources de l'électronique, de l'informatique et de la physique appliquée, a pour objet l'élaboration ou l'interprétation des signaux. Son champ d'application se situe donc dans tous les domaines concernés par la perception, la transmission ou l'exploitation des informations véhiculées par ces signaux.

1.4 Transmission et le traitement du signal

1.4.1 Elaboration des signaux

- **synthèse** : création de signaux de forme appropriée en procédant par exemple à une combinaison de signaux élémentaires.
- **Modulation, changement de fréquence** : moyen permettant d'adapter un signal aux caractéristiques fréquentielles d'une voie de transmission.
- **codage** : traduction en code binaire (quantification), etc

1.4.2 Interprétation des signaux

- **filtrage** : élimination de certaines composantes indésirables.
- **détection** : extraction du signal d'un bruit de fond (corrélation).
- **identification** : classement d'un signal dans des catégories préalablement définies.
- **analyse** : isolement des composantes essentielles ou utiles d'un signal de forme complexe (transformée de Fourier).
- **mesure** : estimation d'une grandeur caractéristique d'un signal avec un certain degré de confiance (valeur moyenne, etc). [2]

1.5 Représentation des signaux

La représentation des signaux peuvent être envisagés :

- Représentation temporelle des signaux.
- Représentation spectrale.
- Caractéristique morphologique (signal continu ou discret).

1.6 classe des signaux

Les signaux peuvent être repartis en deux classes :

1.6.1 signaux analogiques

On dit que le signal est analogique à un temps continu si la mesure est disponible de façon continue à tout instant, et définies dans un espace à deux dimensions dont leurs courbes représentatives indiquent les variations des signaux en fonction du temps. [3]

1.6.2 signaux numériques

On dit que le signal est numérique à temps discret si elle n'est observée qu'à des instants discrets particuliers en général régulièrement espacés. Afin de rendre possible le traitement d'un signal temporel $y(t)$ par un calculateur numérique (micro-ordinateur), on doit procéder à la numérisation qui est illustrée par les opérations de la Figure (I.1).

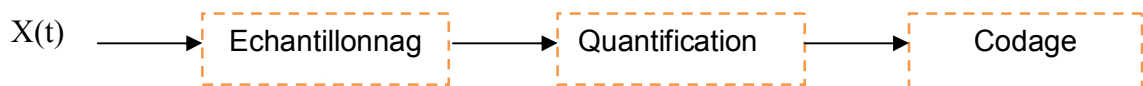


Figure 1. 1. Numérisation des signaux de calculateur numérique.

1.7 Notion de système

Un système (S) est un dispositif ou un phénomène physique qui transforme un signal d'entrée $E(kT)$ (cause, excitation) en un signal de sortie $S(kT)$ (effet, réponse). La nature d'un système peut être : électronique, biologique (parole), optique, économique, etc.



Figure 1. 2. Le principe de système.

1.7.1 Système analogique

Un système est dit analogique si les signaux traités par ce système sont analogiques (exemple : circuit RLC).

1.7.2 Système numérique

Un système est dit numérique si les signaux traités par ce système sont numériques (Exemples : porte logique, unité arithmétique et logique).

1.8 Les principales fonctions de traitement du signal

- Filtrage
- Convolution
- Corrélation
- Modulation, démodulation.
- Echantillonnage, reconstitution.
- Quantification, codage, décodage.
- Représentation spectrale (Transformée de Fourier)
- Détection, estimation, décision
- Reconstruction d'un signal

1.8.1 Filtrage numérique

Lorsqu'un système est utilisé pour modifier la distribution fréquentielle des composantes d'un signal selon des spécifications données, il s'appelle filtrage numérique.

Le problème général du filtrage numérique consiste à déterminer une fonction de transfert $H(z)$ qui représente d'une part la réponse fréquentielle voulue et d'autre part, se prête à une réalisation efficace. [4]

Les filtres numériques se divisent en deux classes importantes selon la durée de la réponse impulsionnelle.

a Filtre non récursif

Un filtre à réponse impulsionnelle de durée finie est toujours stable pour autant que les valeurs $h(n)$ de la réponse impulsionnelle soient toutes finies.

Son fonctionnement est régi par une équation de convolution portant sur un nombre fini de termes :

$$y(n) = \sum_{i=0}^{N-1} h(i) x(n - i) \quad (1.1)$$

Et sa transformée en z est donnée par :

$$H(z) = \frac{Y(z)}{X(z)} = \sum_{n=0}^{N-1} h(n) z^{-n} \quad (1.2)$$

De cette relation on a $h(n) = 0$ pour $n > N - 1$ avec N est le nombre de coefficients du filtre. Ce type de filtre ne tient pas compte des réponses précédentes, d'où l'absence de bouclage de la sortie avec l'entrée, pour cela on l'appelle non récursif.

b Filtre récursif

Ce sont des systèmes linéaires invariants dans le temps, leur fonctionnement est régi par une équation de convolution portant sur une infinité de termes :

$$y(n) = \sum_{i=0}^M b_i x(n-i) - \sum_{i=1}^L a_i y(n-i) \quad (1.3)$$

En appliquant la transformée en z à $y(n)$ on obtient :

$$Y(z) = \sum_{i=0}^M b_i X(z) z^{-i} - \sum_{i=1}^L a_i Y(z) z^{-i} \quad (1.4)$$

$$\Rightarrow Y(z) \left[1 + \sum_{i=1}^L a_i z^{-i} \right] = \sum_{i=0}^M b_i X(z) z^{-i}$$

D'où la réponse impulsionnelle de $H(z)$ devient :

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{i=0}^M b_i z^{-i}}{1 + \sum_{i=1}^L a_i z^{-i}} \quad ; \quad \text{tel que } M \leq L. \quad (1.5)$$

Ce filtre garde une trace des réponses précédentes durant une durée infinie, ce sont des filtres à mémoire, pour cela on les appelle filtre récursif.

1.8.2 la convolution

Soient deux signaux discrets $x(n)$ et $y(n)$, on définit la convolution linéaire par la relation suivante:

$$z(n) = x(n) * y(n) = \sum_{m=-\infty}^{+\infty} x(m) \cdot y(n-m) \quad (1.6)$$

1.8.3 Corrélation

On peut définir une certaine fonction de corrélation entre deux signaux $x(n)$ et $y(n)$. Cette fonction exprime l'influence d'un signal sur un autre. Son rôle consiste à déterminer à partir d'une méthode de comparaison s'il existe une relation entre les deux signaux, le résultat de cette comparaison est un réel appartenant à l'intervalle $[-1,1]$. On dit qu'il y a une forte corrélation si le résultat se rapproche de 1. [5]

On distingue deux opérations :

a Auto-corrélation

Elle consiste à comparer une fonction $x(n)$ avec elle-même durant un intervalle de temps, dont l'une est décalée d'une certaine valeur τ , elle est définie par [6] :

$$R_{xx}(\tau) = \sum_{i=-\infty}^{+\infty} x(i) \cdot x^*(i - \tau) \quad (1.7)$$

b Inter-corrélation

Elle consiste à comparer deux fonctions différentes $x(n)$ et $y(n)$ dont l'une est décalée d'une certaine valeur τ , telle que. [6]

$$R_{xy}(\tau) = \sum_{i=-\infty}^{+\infty} x(i) \cdot y^*(i - \tau) \quad (1.8)$$

1.8.4 Echantillonnage

L'échantillonnage d'un signal déterministe ou aléatoire consiste à remplacer un signal analogique $x(t)$ par la suite des valeurs $x_k = x(t_k)$ périodiques, de période T appelée pas d'échantillonnage et son inverse est appelée fréquence (cadence) d'échantillonnage et sera noté $f_e = 1/T$. c'est-à-dire qu'à chaque instant t espacé de T_0 , on définit une valeur $x(t)$. [7]

Shannon a développé la notion de l'échantillonnage. Il s'agit de remplacer un signal continu dans le temps par un autre défini à certains instants seulement (à temps discrets) équidistants et suffisamment rapprochés pour qu'il contient la même information.

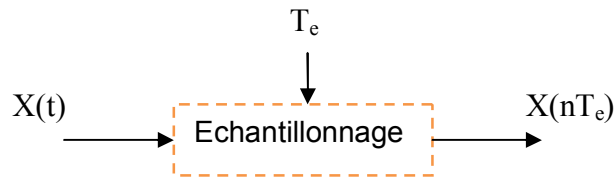


Figure 1. 3. Echantillonnage d'un signal.

Notons par $x(t)$ le signal analogique de départ. La constitution du signal échantillonné $x_e(t)$ est obtenu par multiplication de ce signal analogique par un train d'impulsions de Dirac d'amplitude unité et équi-espacées de T . Ce train est noté $p(t)$ et est défini par

$$p(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT)$$

Le signal échantillonné correspond donc à

$$x_e(t) = x(t) p(t) = x(t) \sum_{n=-\infty}^{+\infty} \delta(t - nT_0)$$

En vertu de la propriété qui veut que $x(t)\delta(t) = x(0)\delta(t)$ on a encore

$$x_e(t) = \sum_{n=-\infty}^{+\infty} x(nT) \delta(t - nT) \quad (1.9)$$

1.8.5 Reconstitution

La reconstitution d'un signal a pour but de redonner un signal continu temporellement à partir des valeurs discrètes connues à des instants discrets nT_0 . Pour reconstituer la continuité du signal, il faut effectuer une interpolation entre deux instants de discrétisation.

Afin de reconstituer le signal analogique sans erreurs il faut l'échantillonner à une fréquence $F_e \geq 2f_m$ où f_m est la plus grande fréquence du spectre et F_e représente la fréquence d'échantillonnage.

1.8.6 Quantification

En traitement du signal, la quantification est le procédé qui permet d'approximer un signal continu (ou à valeurs dans un ensemble discret de grande taille) par des valeurs d'un ensemble discret d'assez petite taille. [8]

1.8.7 Codage

Le codage consiste à associer à un ensemble de valeurs discrètes un code composé d'éléments binaires. Les codes les plus connus: code binaire naturel, code binaire décalé, code complément à 2, code DCB (Binary Coded Decimal), code Gray.

1.8.8 L'analyse de Fourier

L'analyse de Fourier est l'instrument majeur de la théorie du signal. Elle permet la décomposition d'un signal en une somme de signaux élémentaires, qui ont la propriété d'être facile à mettre en œuvre et à observer. La représentation spectrale des signaux déterministes est obtenue grâce à la transformation de Fourier généralisée par l'emploi des distributions, cette représentation exprime la composition fréquentielle de l'amplitude, de la phase et de l'énergie ou de la puissance des signaux considérés. [8]

On appelle Transformée de Fourier Discrète (TFD ou DFT : Discrete Fourier Transform) d'un signal défini par N échantillons $x(n)$, la suite de N termes $X(k)$ définie par :

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi nk/N} \quad (1.10)$$

1.9 Généralité sur le signal de parole et le bruit

La parole est un signal réel, continu, d'énergie finie et non stationnaire. Sa structure est complexe et variable avec le temps.

Deux limitations fondamentales méritent d'être prises en compte : les limitations du système auditif et celles du système vocal chez l'être humain.

Le système auditif humain est surtout sensible dans une gamme de fréquence située entre 800 Hz à 8.000 Hz; les limites extrêmes sont respectivement 20 et 20.000 Hz.

Par contre, le système vocal est encore plus limité, en résumé, pour des sons vocaliques à des fréquences au-dessus de 4 kHz, les hautes fréquences sont plus de 40 dB en dessous du sommet du spectre.

1.9.1 Classification des sons de la parole

On peut distinguer deux grandes classes de sons émis par l'appareil phonatoire :

- **Les sons voisés (ou sons sonores)**

Les sons voisés sont des signaux quasi-périodiques très riches en harmoniques d'une fréquence fondamentale, appelée pitch. Ce qui leur donne un caractère assez facilement prévisible. Ils sont 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 (ou sons sourds)**

Les sons non-voisés sont des signaux qui ne présentent pas de structure périodique. Ils ont les caractéristiques spectrales d'un bruit légèrement corrélé. 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.

- **Le silence**

Les silences sont tout simplement des intervalles où le signal utile est absent. En pratique il s'agit de bruits, d'origines diverses, d'énergie négligeable devant celle du signal utile.

Il est important de remarquer que le signal vocal n'est pas stationnaire, son évolution suit les déformations du conduit vocal. Cependant, ces déformations sont suffisamment lentes pour que le signal vocal puisse être considéré comme stationnaire sur des périodes allant de 20 à 40ms. On considère donc ce signal comme quasi-stationnaire sur de tels intervalles ce qui motive le choix du traitement par trames (de 20 à 40ms) du signal vocal pour de nombreuses applications, notamment pour la réduction de bruit et le codage de la parole.

1.9.2 Propriétés statistiques du signal de parole

À court terme, le signal vocal peut être considéré comme la réalisation particulière d'un processus aléatoire non-stationnaire. Ses statistiques moyennes doivent être estimées à long terme (au moins plusieurs secondes) et moyennées pour plusieurs locuteurs afin d'être fiables. Le signal de parole étant quasi-stationnaire on peut aussi définir des statistiques à court terme sur la durée d'une trame. Ces deux types de statistiques sont couramment utilisées dans les techniques de réduction de bruit, La parole est 70% du temps (pseudo-) périodique (bruit ou silence le reste du temps). On peut avoir un signal de parole à large bande (il remplit toute la largeur de bande) ou à bande limitée (0-8000 Hz essentiellement).

1.9.3 Production de signal de parole

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

Ce mouvement vibratoire correspond à une succession de cycles d'ouverture et de fermeture de la glotte. Le nombre de ces cycles par seconde correspond à la fréquence fondamentale F_0 . Quant aux signaux non-voisés, l'air passe librement à travers la glotte (du moins pas dans tout le conduit vocal) sans provoquer de vibration des cordes vocales. [9]

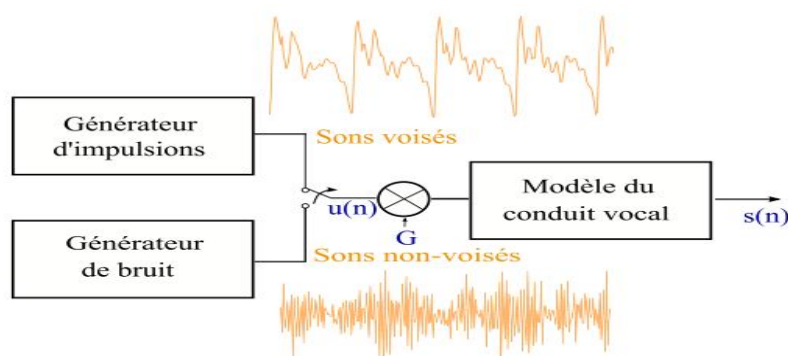


Figure 1. 4. Modèle simple de production de la parole.

1.9.4 Perception de la parole

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

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

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

Le dernier trait acoustique est l'énergie correspondant à l'intensité sonore. Elle est habituellement plus forte pour les segments voisés de la parole que pour les segments non voisés. [9]

1.9.5 Analyse et paramétrisation de la parole

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

1.9.6 Le bruit

On dit bruit 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.

a le bruit blanc

On appelle bruit blanc un processus aléatoire dont la densité spectrale de puissance est constante, quel que soit f :

$$S_{ww}(f) = \sigma_b^2$$

Le bruit à DSP constante est dit blanc par analogie avec la lumière blanche qui contient toutes les longueurs d'onde de la lumière visible. Le bruit blanc correspond à un modèle purement théorique.

En effet, il est physiquement irréalisable car il contient des fréquences infinies et sa puissance moyenne est infinie. C'est cependant un modèle très pratique et utile pour représenter des signaux dont le spectre est constant, au moins sur une large bande de fréquence.

La fonction d'autocorrélation d'un bruit blanc est une impulsion de Dirac représentée dans la figure (1.5). En effet, par transformée de Fourier inverse, on a :

$$S_{ww}(f) = \sigma_b^2 \quad \leftrightarrow \quad R_{ww}(\tau) = \sigma_b^2 \delta(\tau).$$

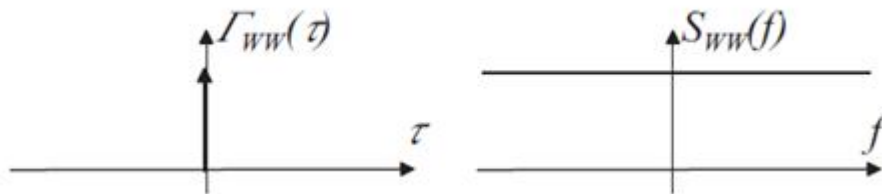


Figure 1. 5. Fonction d'autocorrélation (à gauche) et densité spectrale de puissance (à droite) d'un bruit blanc.

b Les bruits colorés

Dans le cas où la DSP n'est pas constante en fonction de la fréquence, le signal aléatoire est alors appelé bruit coloré. Pour cette représentation spectrale, les principaux types de bruits colorés se distinguent: Le bruit rose, le bruit brun...

c Le bruit musical

Le bruit musical est un bruit résiduel perceptuellement gênant qui apparaît suite au Débruitage de la parole par des algorithmes d'atténuation spectrale à court terme tels que la soustraction spectrale ou le filtrage de Wiener. Le spectre du bruit musical est

D'où le caractère musical. Son énergie moyenne est plus faible que celle du bruit initial,

Parmi les raisons d'apparition de ce type on cite : [10]

- Le traitement non linéaire des composantes négatives du signal débruité,
- L'estimée non précise de la densité spectrale de bruit,
- La variabilité de la fonction de gain appliquée au signal bruité,
- La variance des estimateurs locaux de la densité spectrale des signaux...

1.10 conclusion

Dans ce chapitre, nous avons présenté les notions de base du traitement du signal, et quelques définitions sur la parole, leurs propriétés et ces différents segments de la parole. Nous avons également étudié les différents types de bruit qui interviennent dans un traitement numérique du signal.

Dans le prochain chapitre, nous allons présenter quelques types de technique de réduction de bruit.

Chapitre 2 Atténuation spectrale a court terme

2.1 Introduction

L'oreille humaine a des capacités impressionnantes pour reconnaître et distinguer la parole du bruit. Mais, pour le bien être de l'auditeur et dans le souci de limiter sa fatigue, on cherche à améliorer la qualité de l'écoute à travers le débruitage de la parole (pour des applications telles que la téléphonie mobile et la téléphonie mains-libres). Plusieurs méthodes de l'atténuation spectrale à court terme (débruitage) de la parole ont été proposées dans la littérature pour répondre aux besoins des utilisateurs et des applications dédiées à la parole. On fera une étude de certaines d'entre elles présentant à chaque fois le principe, les limitations et les améliorations apportées.

Ce chapitre est consacré à la présentation d'une vision unifiée des principales techniques de réduction de bruit par atténuation spectrale à court terme (limitée aux systèmes monovoie) dont l'esprit sera conservé tout au long de ce mémoire. Tout d'abord, le principe général de ces approches est exposé dans la (partie 2.2), à cette occasion les notations et les hypothèses classiques seront introduites. Ensuite, la (partie 2.3) permet de définir les quantités utilisées dans le calcul des fonctions de réduction de bruit. Ces quantités sont fondamentales dans la mesure où elles conditionnent la qualité du traitement. Les principales étapes de la mise en œuvre de ces techniques, communes à la majorité des approches, sont ensuite exposées dans la (partie 2.4). Au vu des très nombreuses approches qui existent, l'inventaire proposé dans la (partie 2.5) ne se veut pas exhaustif, seules sont présentées et analysées les approches majeures et il faut garder à l'esprit qu'il existe de nombreuses variantes pour chacune d'entre elles. Finalement, une sélection de techniques d'estimation du bruit est présentée dans la (partie 2.6). Cette étape est en effet fondamentale car nécessaire au calcul des paramètres présentés dans la (Partie 2.2).

2.2 Principe d'atténuation spectrale à court terme

Dans le cas monovoie considéré, l'objectif consiste à estimer le signal de parole utile $s(n)$, celui-ci étant perturbé par un bruit additif $b(n)$ supposé indépendant du signal de parole, à partir du seul signal observé $x(n)$:

$$x(n) = s(n) + b(n) \quad (2.1)$$

Les approches basées sur l'atténuation spectrale à court terme réalisent la réduction de bruit dans le domaine fréquentiel (ou spectral). Si les signaux sont stationnaires alors à partir de la relation temporelle (2.1) on peut écrire :

$$\gamma_x(f) = \gamma_s(f) + \gamma_b(f) \quad (2.2)$$

Où les $\gamma_x(f)$ et $\gamma_s(f), \gamma_b(f)$ représentent les densités spectrales de puissance (DSP) respectives des signaux $x(n)$ et $s(n), b(n)$. Cette représentation sous forme de DSP n'est malheureusement pas exploitable à cause de la non-stationnarité du signal de parole. En effet, s'il est acceptable de considérer le bruit stationnaire, la parole ne peut être comme telle sur de courtes durées. Il devient alors possible d'exploiter la quasi-stationnarité de la parole sur des trames d'une durée de (20 à 40) ms. C'est entre autre pour cette raison qu'une majorité des techniques de réduction de bruit par atténuation spectrale sont basées sur une analyse à court terme de signaux traités. De nombreuses transformations à court terme sont disponibles mais si l'on prend en compte la nécessité qu'une transformée inverse existe ou bien encore la possibilité de modifier le signal bruité dans le domaine qu'une transformée alors le panel des transformées envisageables se réduit considérablement. La transformée de Fourier (TF) à court terme (TFCT) remplit les conditions requises et possède des mises en œuvre rapides et numériquement peu coûteuses ce qui explique qu'elle est très souvent utilisée, ce qui sera également le cas dans la suite de ce document. Chaque trame du signal temporel $x(t)$ peut donc être représentée dans le domaine fréquentiel par son module $|X(p, k)|$ et sa phase associée $\phi_x(p, k)$, où p représente l'indice temporel de la trame d'analyse courante et k le canal fréquentiel d'indice k ou autrement dit la fréquence discrète f_k . Dans le domaine fréquentiel, l'équation (2.1) considérée à la trame (p) peut donc s'exprimer ainsi :

$$|X(p, k)|e^{i\phi_x(p, k)} = |S(p, k)|e^{i\phi_s(p, k)} + |B(p, k)|e^{i\phi_b(p, k)} \quad (2.3)$$

L'objectif de l'atténuation spectrale consiste à estimer le spectre à court terme du signal de parole $\hat{S}(p, k)$. On suppose qu'il est toujours possible d'estimer la DSP du bruit, soit sur une partie du signal d'observation ne contenant que du bruit soit de façon continue (partie 2.6). Sans cette connaissance a priori de la DSP du bruit, généralement supposé stationnaire, il ne serait pas possible de réaliser une technique de réduction de bruit pour un système monovoie.

Sans faire aucune hypothèse supplémentaire sur les caractéristiques des signaux, il est possible d'obtenir des estimateurs du signal utile à partir d'une optimisation de certains critères. Il est aussi possible de poser des hypothèses supplémentaires sur les statistiques du signal de parole ou du bruit pour obtenir d'autres types d'estimateurs.

2.3 Rapports signal à bruit

En réalité : Il n'existe pas de solution simple à l'estimation spectrale de $S(p, k)$, donc généralement un gain spectral $G(p, k)$ qui dépend de (RSB) est obtenu puis est appliqué au spectre bruité $X(p, k)$:

$$\hat{S}(p, k) = G(p, k) X(p, k) \quad (2.4)$$

Le gain $G(p, k)$ spectral a toujours le comportement asymptotique suivant :

- Une valeur importante du RSB indique qu'une composante de parole est présente par rapport au niveau de bruit, le gain $G(p, k)$ doit être proche de 1 pour préserver cette composante.

- une valeur faible du RSB indique qu'une parole est absente ou très faible par rapport au niveau de bruit, le gain $G(p, k)$ doit donc apporter une atténuation importante

($G(p, k) \ll 1$) pour réduire l'effet de bruit.

Tout le problème donc consiste à estimer ce RSB. Selon les hypothèses choisies pour exprimer la gain spectral, deux types de RSB utilisés [11]

$$RSB_{\text{post}}(p, k) = \frac{|X(p, k)|^2}{\gamma_b(k)} = \frac{|X(p, k)|^2}{E[|B(p, k)|^2]} \quad (2.5)$$

$$RSB_{\text{prio}}(p, k) = \frac{\gamma_s(k)}{\gamma_b(k)} = \frac{E|S(p, k)|^2}{E[|B(p, k)|^2]} \quad (2.6)$$

La quantité $RSB_{post}(p, k)$ représente le RSB de la trame courante en tenant compte du module carré du signal bruité et dépend donc du temps. La quantité $RSB_{prior}(p, k)$ quant à elle ne dépend pas du temps car elle exprime le RSB à long terme en supposant les statistiques du signal de parole utile connues a priori. En pratique, la DSP du signal de parole est bien évidemment amenée à évoluer au cours du temps, on fera donc par la suite référence à la grandeur $RSB_{prior}(p, k)$ tout en sachant que ses caractéristiques évoluent lentement par rapport à celles du $RSB_{post}(p, k)$.

A partir du RSB a posteriori, on peut également définir le RSB instantané qui correspond à une estimée locale (ou à court terme) du RSB a priori par soustraction directe de la DSP du bruit au module carré du signal bruité [12]

$$RSB_{inst}(p, k) = \frac{|X(p, k)|^2 - \gamma_b(k)}{\gamma_b(k)} = RSB_{post}(p, k) - 1 \quad (2.7)$$

Ces trois expressions de RSB restent théoriques dans la mesure où seule la quantité $|X(p, k)|^2$ est connue. D'une part, la quantité $E[|B(p, k)|^2]$ doit être estimée à partir du signal bruité. Ceci est décrit dans la (partie 2.6) qui est consacrée aux techniques d'estimation de la DSP du bruit. Outre part, il faut aussi estimer la quantité $E[|S(p, k)|^2]$, en utilisant les techniques d'atténuation spectrale à court terme.

2.4 Mise en œuvre de l'atténuation spectrale a court terme

D'un point de vue très général, les méthodes de réduction de bruit par atténuation spectrale à court terme sont mises en œuvre de la façon suivante :

- Le signal temporel bruité $x(t)$, est découpé en trames pouvant se chevaucher puis chaque trame est fenêtrée. Nous verrons qu'un recouvrement est en fait nécessaire, mais pas suffisant, pour éviter certains artefacts liés au traitement par blocs. Chaque trame d'analyse est alors transformée dans un domaine où l'étape de réduction de bruit est réalisable. Généralement une TFD est utilisée ce qui implique que le signal est exploité dans le domaine fréquentiel. Cette première étape constitue la TFCT.

- Une estimation de la densité spectrale de puissance du bruit à long terme est réalisée soit en période de bruit seul ce qui nécessite alors une détection d'activité vocale (DAV), soit de façon continue même pendant l'activité vocale.
- Une atténuation spectrale à court terme ou autrement dit un gain spectral évalué à partir d'une règle de suppression des composantes spectrales de bruit est appliquée au module du signal bruité.

Le calcul du gain spectral requiert l'estimation du RSB (a posteriori et/ou a priori). Ces quantités sont estimées à partir du module carré du signal bruité $|X(p,k)|^2$ et de la DSP du bruit $E|B(p,k)|^2$

- Le module du signal de parole ainsi estimé $|\hat{s}(p,k)|$ et la phase du signal bruité sont alors utilisés pour revenir dans le domaine temporel en utilisant une TFD inverse (TFDI). Le signal de sortie est finalement synthétisé à partir d'une technique de type OLS (pour over lap and save) ou OLA (pour over lap and add) qui seront détaillées dans la partie (2.7) et (2.8) Cette dernière étape constitue la TFCT inverse (TFCTI). Le schéma de principe de la figure (2.1) résume le fonctionnement général de la réduction de bruit par atténuation spectrale à court terme. Comme le montre ce schéma, seul le module de la TFD subit un traitement, la phase étant réutilisée sans modification. Ceci est principalement dû au fait que, d'un point de vue pratique, la phase du signal original est très difficile à estimer et que l'oreille humaine est peu sensible aux distorsions de phase [13].

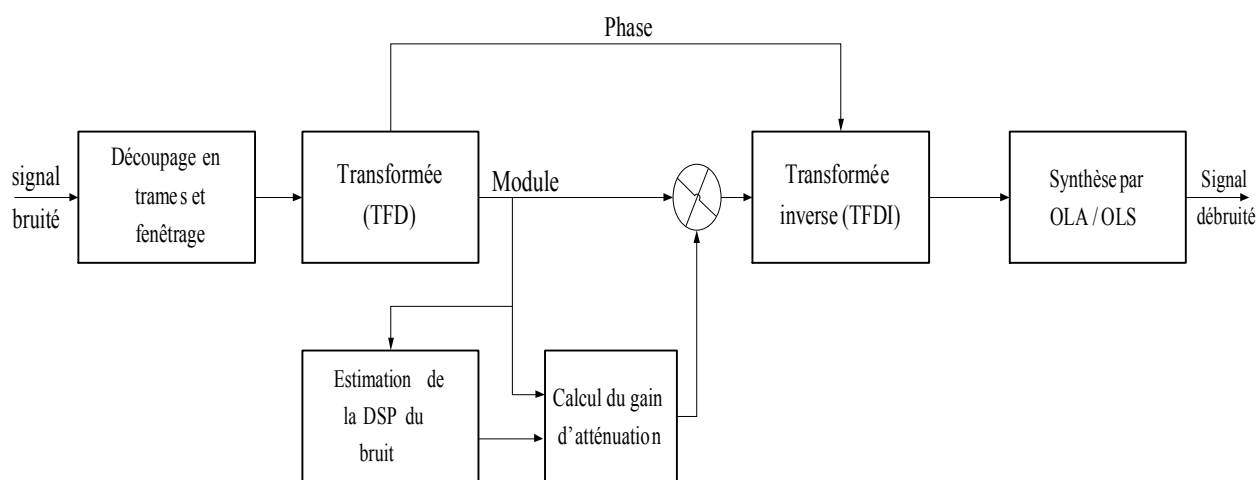


Figure 2. 1. Schéma de principe des techniques de réduction de bruit.

2.5 Méthodes d'atténuation spectrale à court terme

Les techniques de réduction de bruit par atténuation spectrale à court terme peuvent être classées de différentes manières. Dans cet état de l'art, nous allons différencier les approches ne nécessitant pas de modèle statistique pour les signaux traités de celles qui en requièrent.

2.5.1 Soustraction spectrale

la soustraction spectrale est une approche qui ne nécessite pas de modèle statistique et la méthode de débrouillage la plus ancienne. Elle est introduite dans [14]. Elle opère dans le domaine fréquentiel et a pour principe de soustraire une estimation du bruit à partir du signal observé. Le bruit est supposé additif, stationnaire ou légèrement variant ce qui nous permet de l'estimer pendant les périodes de silence.

a Soustraction spectrale d'amplitude (SSA)

Le module de spectre à court terme est estimé comme ceci :

$$|\hat{S}(p, k)| = |X(p, k)| - \sqrt{E[|B(p, k)|^2]} \quad (2.8)$$

Avec :

$|X(p, k)|$: Le module de spectre bruité (module de spectre Instantané).

$\sqrt{E[|B(p, k)|^2]}$: La valeur moyenne de bruit

$|\hat{S}(p, k)|$: Le module du spectre du signal estimé.

Puisque une valeur négative n'ayant pas de signification physique on a besoin d'un simple seuillage :

$$|\hat{S}(p, k)| = \begin{cases} |X(p, k)| - \sqrt{E[|B(p, k)|^2]} & \text{si } |X(p, k)| \geq \sqrt{E[|B(p, k)|^2]} \\ 0 & \text{sinon} \end{cases} \quad (2.9)$$

Le passage dans le domaine temporelle se réalise par la transformée de Fourier inverse en gardant la phase de signal bruité pour donner l'estimation de spectre de signal utile

$$|\hat{s}(t)| = \text{IFFT} [|\hat{S}(p, k)| e^{i\phi(X(p, k))}] \quad (2.10)$$

Avec : $\emptyset X(p, k)$ représent lea phase de signal bruité.

Autre forme de SSA :

Le spectre estimé de signal utile peut être obtenu par l'équation suivant :

$$\hat{S}(p, k) = G_{SSA}(p, k)X(p, k) \quad (2.11)$$

Avec :

$$G_{SSA}(p, k) = \begin{cases} \mathbf{1} - \sqrt{\frac{1}{RSB_{post}(p, k)}} & \text{si } RSB_{post}(p, k) \geq \mathbf{1}, \\ \mathbf{0} & \text{sinon} \end{cases} \quad (2.12)$$

b Soustraction spectrale de puissance (SSP)

La soustraction spectrale en puissance(SSP) a été proposée par [15] :

$$|\hat{S}(p, k)|^2 = |X(p, k)|^2 - E[|B(p, k)|^2] \quad (2.13)$$

De la même façon que pour la SSA, les valeurs négatives du module carré $|\hat{S}(p, k)|^2$ sont à proscrire ce qui amène à contraindre le résultat :

$$|\hat{S}(p, k)|^2 = \begin{cases} |X(p, k)|^2 - E[|B(p, k)|^2] & \text{si } |X(p, k)|^2 \geq E[|B(p, k)|^2], \\ \mathbf{0} & \text{sinon} \end{cases} \quad (2.14)$$

La SSP peut aussi s'exprimer en fonction du RSB posteriori :

$$\hat{S}(p, k) = G_{SSP}(p, k)X(p, k) \quad (2.15)$$

Avec :

$$G(p, k) = \begin{cases} \sqrt{\mathbf{1} - \frac{1}{RSB_{post}(p, k)}} & \text{si } RSB_{post}(p, k) \geq \mathbf{1} \\ \mathbf{0} & \text{sinon} \end{cases} \quad (2.16)$$

- L'avantage de la soustraction spectrale en générale et la simplicité de sa mise on œuvre. et l'inconvénient Engendre un bruit résiduel : Ce bruit musical est dû au fait que le spectre moyen d'énergie du bruit $E[|B(p, k)|^2]$ est soustrait à un spectre d'énergie instantané $|X(p, k)|^2$ possédant une forte variance par rapport à cette moyenne (en période de bruit seul). Le spectre d'énergie résiduel possède donc des pics fréquentiels localisés de façon aléatoire en fréquence et qui

ne dépassent pas statistiquement la durée d'une trame. Bien que la puissance du bruit résiduel soit inférieure à celle du bruit original,

Donc, des solutions ont été proposées pour limiter le bruit musical résiduel, elles sont englobées sous l'appellation de soustraction spectrale généralisée

c soustraction spectrale généralisé (SSG)

La soustraction spectrale généralisée (SSG) [16] pour but de supprimer le bruit musical (le bruit musicale résiduel) généré par l'approche de SSP. Elle propose des paramètres par deux facteurs, le premier pour soustraire une surestimation du bruit et le second pour éviter que le résultat de la soustraction dépasse un seuil de tolérance fonction de l'estimée de la puissance de bruit. On obtient l'expression suivante :

$$|\hat{S}(p, k)|^2 = \begin{cases} D(p, k)^{\frac{1}{\gamma}} & \text{si } D(p, k)^{\frac{1}{\gamma}} \geq \beta E[|B(p, k)|^2], \\ \beta E[|B(p, k)|^2] & \text{sinon} \end{cases} \quad (2.17)$$

Avec :

$$D(p, k) = (|X(p, k)|^2)^{\gamma} - \alpha (E[|B(p, k)|^2])^{\gamma} \quad (2.18)$$

Ou $\alpha \geq 1$ et $0 \leq \beta \leq 1$ et $\gamma \geq 0$

α : Le coefficient de surestimation :

Il permet de surestimer la DSP du bruit et ainsi de réduire le niveau de bruit musical résiduel. Cependant, cette surestimation introduit une distorsion du signal de parole restauré. Pour limiter cet effet, le coefficient de surestimation peut être rendu adaptatif en fonction de la fréquence et suivant le type de bruit [17]

β : Le plancher spectrale : ce plancher permet de conserver une certaine quantité du bruit original ce qui va permettre de masquer le bruit musical. Cependant, pour supprimer totalement le bruit musical il faut utiliser une valeur de suffisamment grande ce qui a pour effet de réintroduire une grande partie du bruit.

γ : Détermine le type de soustraction spectrale (agressivité)

De la même façon que pour la SSA et la SSP, la SSG peut s'exprimer sous forme de gain :

$$\hat{S}(p, k) = G_{SSG}(p, k)X(p, k) \quad (2.19)$$

Avec

$$G_{SSG}(p, k) = \begin{cases} \left(1 - \alpha \frac{1}{RSB_{post}^\gamma(p, k)}\right)^{\frac{1}{2\gamma}} & \text{si } (RSB_{post}^\gamma(p, k) - \alpha)^{\frac{1}{\gamma}} \geq \beta \\ \sqrt{\frac{\beta}{RSB_{post}(p, k)}} & \text{sinon} \end{cases} \quad (2.20)$$

Remarque :

Si $\alpha = 1$ et $\gamma = 0.5$, il s'agit d'une soustraction spectrale d'amplitude.

Si $\alpha = 1$ et $\gamma = 1$, il s'agit d'une soustraction spectrale de puissance.

2.5.2 Filtrage du Wiener

Le filtrage de Wiener [18] est le Filtre linéaire optimale au sens de (EQMM), c'est-à-dire minimise la fonction d'erreur suivante :

$$E[(S(p, k) - G_W(p, k)X(p, k))^2] \quad (2.20)$$

Avec les hypothèses classiques posées dans la partie II.1, le filtre de Wiener peut s'exprimer ainsi :

$$G_W(p, k) = \frac{E[|S(p, k)|^2]}{E[|X(p, k)|^2]} \quad (2.21)$$

Les quantités intervenant dans l'expression du filtre de Wiener étant calculées à long terme, on peut l'exprimer en fonction du RSB a priori :

$$G_W(p, k) = \frac{RSB_{prio}(p, k)}{1 + RSB_{prio}(p, k)} \quad (2.22)$$

En pratique, il est possible de remplacer la DSP du signal utile $E[|S(p, k)|^2]$ par une estimée obtenue directement à partir du spectre bruité instantané selon le principe de la SSP : $|X(p, k)|^2 - E[|B(p, k)|^2]$ bien que cela n'ait pas de fondement théorique. On s'éloigne alors des hypothèses de stationnarité nécessaires au filtre de Wiener. C'est pourquoi on parle de filtre pseudo-Wiener selon l'appellation proposée dans [19] Il est alors possible d'exprimer le filtre pseudo-Wiener en fonction du RSB a posteriori:

$$G_W(p, k) = 1 - \frac{1}{RSb_{post}(p, k)} \quad (2.23)$$

On peut d'ailleurs remarquer que dans ce cas $G_W(p, k) = G_{SSP}^2(p, k)$ attestent du fait que ce filtre fait partie de la famille des méthodes de soustraction spectrale.

2.6 Techniques d'estimation du bruit

Bien que la DSP du bruit soit estimée à long terme, elle est calculée à partir de trames à court terme et l'indice de trame p sera conservé pour faciliter la présentation des différents estimateurs. Dans cette partie, l'estimée de cette quantité sera donc notée

$$\hat{\gamma}_b(p, k).$$

2.6.1 Estimation de bruit nécessitant une DAV

Dans ce cas de figure, la DSP du bruit est estimée à long terme et uniquement pendant l'inactivité vocale, c'est-à-dire qu'une DAV [20] renseigne l'estimateur sur la présence ou non de parole pour la trame courante. Les périodes d'inactivité vocale doivent être suffisamment longues pour obtenir une estimée de la DSP du bruit possédant une faible variance ce qui ne pose pas de problème en pratique. Si la trame courante contient uniquement du bruit alors l'estimée de la DSP du bruit est obtenue par un lissage exponentiel sinon, i.e. la trame courante contient de la parole, l'estimation de la DSP du bruit est figée [21] :

$$\begin{cases} \hat{\gamma}_b(p, k) = \lambda_B \hat{\gamma}_b(p-1, k) + (1 - \lambda_B) |X(p, k)|^2 & \text{si la trame } p \text{ contient uniquement du bruit} \\ \hat{\gamma}_b(p, k) = \hat{\gamma}_b(p-1, k) & \text{sinon} \end{cases} \quad (2.24)$$

Le paramètre λ_B contrôle le degré de lissage désiré et peut éventuellement varier au cours du temps. En général $0,50 \leq B \leq 0,99$ et peut se calculer comme ceci :

$$B = \exp(-R/(F_e \tau_b)) \quad (2.25)$$

où R est le nombre de points de recouvrement entre deux trames successives, F_e la fréquence d'échantillonnage et τ_b une constante de temps fixée selon le degré de stationnarité du bruit.

2.6.2 Estimation de bruit e continu (sans DAV)

Des approches permettant de poursuivre les non-stationnarités du bruit en continu, i.e. indifféremment pendant l'inactivité ou l'activité vocale, ont été développées. L'aperçu qui va suivre n'est pas exhaustif mais permet de donner un éventail des différentes approches possibles.

➤ Estimateur bornée

La méthode la plus simple a été proposée dans [22]. L'estimation se fait par lissage exponentiel à chaque trame selon la première expression de l'équation (2.24) mais également pendant l'activité vocale. Cependant, l'évolution de la DSP du bruit est contrainte de la manière suivante:

$$1.006 \sqrt{\hat{\gamma}_b(p-1, k)} \leq \sqrt{\hat{\gamma}_b(p, k)} \leq 0.978 \sqrt{\hat{\gamma}_b(p, k)} \quad (2.26)$$

Ainsi, l'estimée de la DSP du bruit ne peut pas augmenter de plus de 3dB par seconde et diminuer de plus de 12dB par seconde avec les paramètres de traitement donnés dans [22]. L'estimée de la DSP du bruit augmente lentement pendant l'activité vocale mais a la capacité de retourner rapidement à une valeur correcte pendant l'inactivité vocale. Cette approche a l'avantage d'être très simple, cependant, l'estimée de la DSP du bruit est biaisée pendant l'activité vocale ce qui engendre des distorsions du signal restauré. Elles permettent de suivre les non-stationnarités du bruit pendant les périodes d'activité.

On conclure que Il ressort de cette partie des techniques d'estimation de la DSP du bruit que les méthodes d'estimation en continu(sans DAV) présentent deux avantages par rapport aux techniques à base de DAV. D'une part, il n'est pas nécessaire de créer une DAV robuste, ce qui est très difficile à faible RSB. D'autre part, vocale. Ceci peut s'avérer indispensable dans certains cas où la DAV est inutilisable. On peut noter par exemple les applications sans instant de silence (messagerie vocale, musique) ou encore plus généralement les applications où le RSB est très faible. Néanmoins, aucune technique d'estimation de la DSP du bruit ne permet de réaliser un suivi fin des évolutions du bruit trame par trame. En effet, il y a toujours un délai de réponse plus ou moins long à une variation du niveau de bruit.

2.7 L'overlap and save (OLS)

La méthode OLS [23] consiste à ne conserver que les échantillons valides à l'issue de l'opération de filtrage. Les autres échantillons sont considérés non valides car soit ils sont obtenus à partir d'un filtre tronqué (filtrage temporel) soit ils subissent le phénomène de repliement temporel (filtrage fréquentiel).

Historiquement, l'OLS s'implémente dans le domaine fréquentiel (application de convolution rapide par FFT). L'utilisation de l'OLS pour la réduction de bruit nécessite cependant d'élargir le cadre classique. En effet, dans l'application qui nous intéresse, chaque trame est fenêtrée et les trames se recouvrent (avec un taux minimum de 50%). Donc étant donné que l'OLS ne conserve que les échantillons valides à l'issue du filtrage il est impossible de satisfaire à la contrainte de reconstruction parfaite par sommation des trames filtrées, cette sommation impliquant des échantillons repliés et donc non valides. Par contre il est possible de contourner ce problème en filtrant directement les échantillons non fenêtrés.

2.8 L'overlap and add (OLA)

La technique OLA [24] consiste à conserver les échantillons filtrés partiellement (filtre tronqué) de la trame x_p , puis à les combiner avec ceux issus des trames précédentes ou suivantes (selon que l'on considère le début ou la fin de la trame) afin d'obtenir des échantillons filtrés avec un filtre complet. Cette technique ne souffre pas des problèmes de "clics" que l'on rencontre avec la technique OLS car la transition entre les filtres des trames successives est assurée du fait de l'interpolation implicite d'une trame à l'autre.

Comme pour la technique OLS, l'implémentation peut se faire dans le domaine fréquentiel ou dans le domaine temporel. Au même titre que l'OLS, à l'origine l'OLA s'implémente dans le domaine fréquentiel (application de convolution rapide par FFT). Cependant son utilisation pour la réduction de bruit nécessite d'élargir le cadre classique.

2.9 conclusion

Au cours de ce chapitre, nous avons présenté en détails la technique d'atténuation spectrale à court terme et ses différentes approches comme la soustraction spectrale d'amplitude SSA, de puissance SSP, généralisée SSG et le filtrage du Wiener qui sont utilisées dans un système mono-capteur. Ainsi que les caractéristiques de chaque technique. Toutes les techniques sont proposées pour améliorer la qualité du signal parole, pour cet objectif, il existe les problèmes de la distorsion et le bruit musicale.

Dans le prochain chapitre nous essayons de proposer des autres techniques pour régler ces problèmes.

Chapitre 3 formulation des différentes techniques de réduction de bruit

3.1 introduction

Dans le chapitre précédent, nous avons présenté plusieurs techniques de réduction du bruit. Dans le présent chapitre nous allons présenter les différentes nouvelles techniques utilisées dans notre projet pour obtenir des solutions des problèmes posé et d'évaluer l'amélioration apportée par ces techniques à partir d'un corpus représentatif de nombreuses conditions de bruit. Les résultats objectifs seront donnés à partir de deux critères. D'une part, le RSB de sortie permettra de quantifier la réduction effective de bruit et d'autre part, la mesure de la distance cepstrale permettra de quantifier la distorsion générée par les techniques de réduction de bruit.

3.2 Techniques améliorées

Dans notre travail, nous avons basé sur la technique de réduction de bruit à décision-dirigée DD qu'est considéré au point de départ comme la meilleure solution pour réduire le bruit. Lors de ce projet, nous proposons d'utiliser d'autres gains (gain de la soustraction spectrale d'amplitude (GSSA), gain de la soustraction spectrale de puissance (GSSP), gain de la soustraction spectrale généralisé (GSSG) et le gain de Wiener (GW)) et faire une comparaison avec la technique de soustraction spectrale de Boll.

3.3 Technique décision-dirigée DD

Les performances des techniques de réduction de bruit dépendent de la fonction de gain choisie. Dans cette section, nous allons étudier toutes les étapes de l'approche de décision-dirigée.

➤ Principe de l'approche DD

Cette approche basé sur l'évaluation du rapport signal-à-bruit a posteriori $R\hat{S}B_{post}$ et a priori $R\hat{S}B_{prio}^{DD}$ a partir de l'estimation des spectres du signal de parole à la trame précédente $\hat{S}_{DD}(p-1, k)$ (introduit un retard d'une trame). Cet estimateur du RSB a priori, nommé décision-dirigée (en anglais decision-directed (DD)) ce qui signifie dirigé par la décision et son comportement est contrôlé par le paramètre β (théoriquement toujours proche de 1). La Figure 3.1 représente le principe général de l'approche DD.

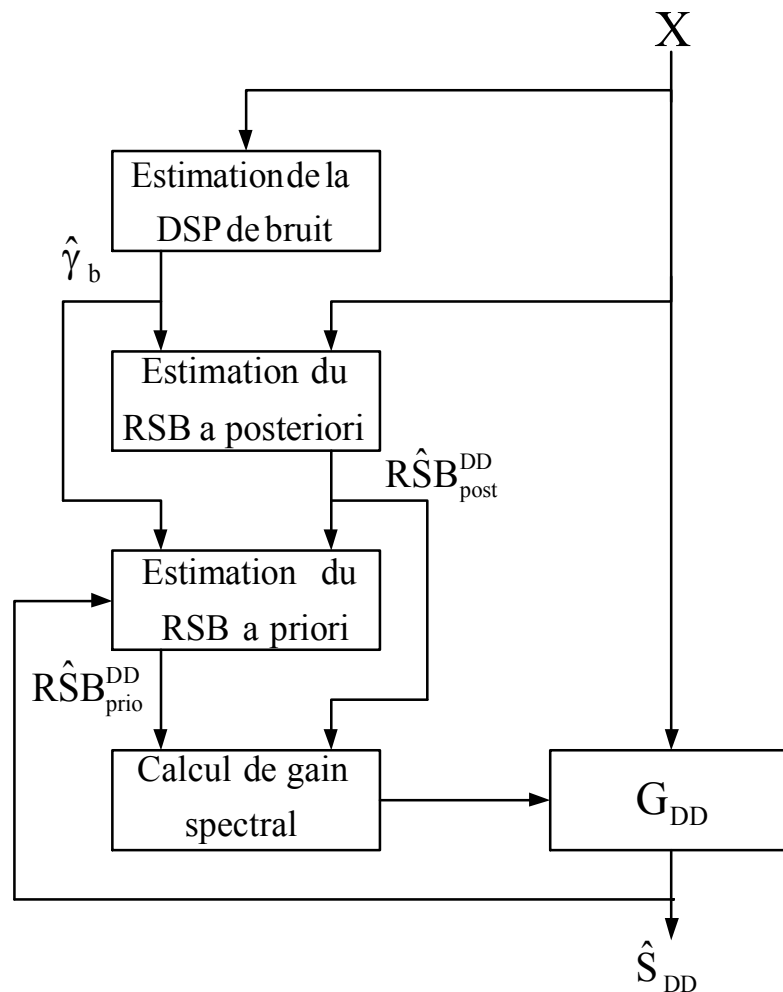


Figure 3 1. Schéma du principe général de l'approche DD.

- **Estimation de la DSP de bruit $\hat{\gamma}_b(k)$**

Pour estimer la densité spectrale de puissance DSP du bruit, en utilisant un mécanisme de détection d'activité vocale manuelle DAV, implique que le calcul de la DSP de bruit seulement dans les périodes de l'absence de signal de parole.

- **L'estimation de $RSB_{post}(p, k)$**

L'estimation de $RSB_{post}(p, k)$ est donné par la relation suivante :

$$R\hat{S}B_{post}(p, k) = \frac{|X(p, k)|^2}{\hat{\gamma}_b(k)} \quad (3.1)$$

- **L'estimation de $RSB_{prior}(p, k)$**

L'estimation de *rapport signal-à-bruit* a priori $RSB_{prior}(p, k)$ dépend de $RSB_{post}(p, k)$ qu'est exprimé par l'équation suivante :

$$R\hat{S}B_{prio}^{DD}(p, k) = \beta \frac{|\hat{S}_{DD}(p-1, k)|^2}{\hat{\gamma}_b(k)} + (1 - \beta) \max(R\hat{S}B_{post}(p, k) - 1, 0) \quad (3.2)$$

- **Le calcul du gain spectral**

Le gain spectral peut être calculé à partir des expressions suivantes :

Le gain de Wiener

$$G_{WDD}(p, k) = \frac{R\hat{S}B_{prio}^{DD}(p, k)}{1 + R\hat{S}B_{prio}^{DD}(p, k)} \quad (3.3)$$

Le gain de la soustraction spectrale de d'amplitude

$$G_{SSA}(p, k) = \begin{cases} 1 - \sqrt{\frac{1}{RSB_{post}(p, k)}} & \text{si } RSB_{post}(p, k) \geq 1, \\ 0 & \text{sinon} \end{cases} .$$

Le gain de la soustraction spectrale de puissance

$$G_{SSP}(p, k) = \begin{cases} \sqrt{1 - \frac{1}{RSB_{post}(p, k)}} & \text{si } RSB_{post}(p, k) \geq 1 \\ 0 & \text{sinon} \end{cases}$$

Gain de la soustraction spectrale généralisé

$$G_{SSG}(p, k) = \begin{cases} \left(1 - \alpha \frac{1}{RSB_{post}(p, k)}\right)^{\frac{1}{2\gamma}} & \text{si } (RSB_{post}^\gamma(p, k) - \alpha)^{\frac{1}{\gamma}} \geq \beta \\ \sqrt{\frac{\beta}{RSB_{post}(p, k)}} & \text{sinon} \end{cases} .$$

Notant que dans ces trois dernière relations on remplace $RSB_{post}(p, k)$ par $RSB_{prior}(p, k)$.

- **L'estimation du signal de sortie**

généralement un gain spectral $G(p, k)$ qui dépend du RSB obtenu est appliqué au spectre bruité $X(p, k)$, donc le spectre du signal de parole estimé est donné par la relation suivante:

$$\hat{S}(p, k) = G_{DD}(p, k)X(p, k) \quad (3.4)$$

3.4 Technique de Boll

L'approche de débruitage de Boll opère dans le domaine fréquentielle et a pour but de soustraire l'estimation du bruit à partir de signal observé, qui donné par cette expression :

$$|\hat{S}(p, k)| = |X(p, k)| - \sqrt{E[|B(p, k)|^2]}$$

Puisque une valeur négative n'ayant pas de signification physique on a besoin d'une simple seuillage :

$$|\hat{S}(p, k)| = \begin{cases} |X(p, k)| - \sqrt{E[|B(p, k)|^2]} & \text{si } |X(p, k)| \geq \sqrt{E[|B(p, k)|^2]} \\ 0 & \text{sinon} \end{cases} .$$

3.5 Critères objectifs de qualité

Dans le contexte de la réduction de bruit, l'évaluation des performances représente une réelle problématique : en effet, c'est l'oreille humaine qu'on est le juge final de la qualité du système. Deux critères objectifs seront utilisés le long de ce papier : le rapport signal à bruit (RSB) et la distance cepstrale (CD).

- **Rapport signal/bruit**

La mesure du rapport signal/bruit RSB estimé à la sortie de traitement donne l'amélioration apportée par la technique simulée sur le signal de parole de sortie. Il est donné par la relation suivante :

$$RSB = 10 \log_{10} \left(\frac{\sum_{i=0}^{N-1} s^2(i+Nk)}{\sum_{i=0}^{N-1} (s(i+Nk) - \hat{s}(i+Nk))^2} \right) \quad (3.5)$$

Où N est la taille de la trame.

- Une valeur importante du RSB indique qu'une forte composante de parole est présente par rapport au niveau du bruit.
- Une valeur faible du RSB indique que la parole est absente ou très faible par rapport au niveau du bruit.

- **La mesure de la distance cepstrale (DC)**

C'est une mesure de similarité entre les cepstres réels de deux signaux qui correspond à une version tronquée de la distance spectrale logarithmique de ces spectres. Pour les deux signaux original et estimé, la distance cepstrale évoluée pour chaque trame p de N échantillons :

$$DC(p) = \sum_{i=1}^{N-1} \left[IFFT \left(\log(|S(f, p)|) - \log(|\hat{S}(f, p)|) \right) \right]^2 \quad (3.6)$$

3.6 Conclusion

Dans ce chapitre, nous avons présenté les différentes étapes pour chaque technique de Décision dirigée DD et de soustraction spectrale de Boll qui sont utilisées pour la réduction du bruit mono-capteur. Pour voir la performance des deux techniques, nous avons défini deux critères de mesure (le rapport signal à bruit RSB et la distance cepstrale DC).

Dans le chapitre suivant, nous allons présenter les résultats de simulation des toutes les techniques que nous avons présentées.

Chapitre 4 Résultats des Simulations

4.1 Introduction

Dans ce chapitre nous présentons les résultats de simulation de la réduction de bruit en utilisant les techniques suivantes :

- Technique de la soustraction spectrale (Boll) ,
- Technique de DD avec le gain de Wiener,
- Technique de DD avec le gain de la soustraction spectrale d'amplitude,
- Technique de DD avec le gain de la soustraction spectrale de puissance,
- Technique de DD avec le gain de la soustraction spectrale généralisé.

Nous présentons également l'influence des paramètres (β , le rapport signal à bruit d'entrée (RSB) et la distance cepstrale) pour les deux techniques et fera une comparaison entre eux.

4.2 Simulation

4.2.1 Description des signaux de tests

Les deux figures (1) et (2) représentent les deux signaux de parole originaux avec ses segmentations manuelles. Le premier est un signal de locuteur masculin avec une fréquence d'échantillonnage 8kHz de durée presque (4s). Et le deuxième est un locuteur féminin avec une fréquence d'échantillonnage 8kHz de durée presque (4s).

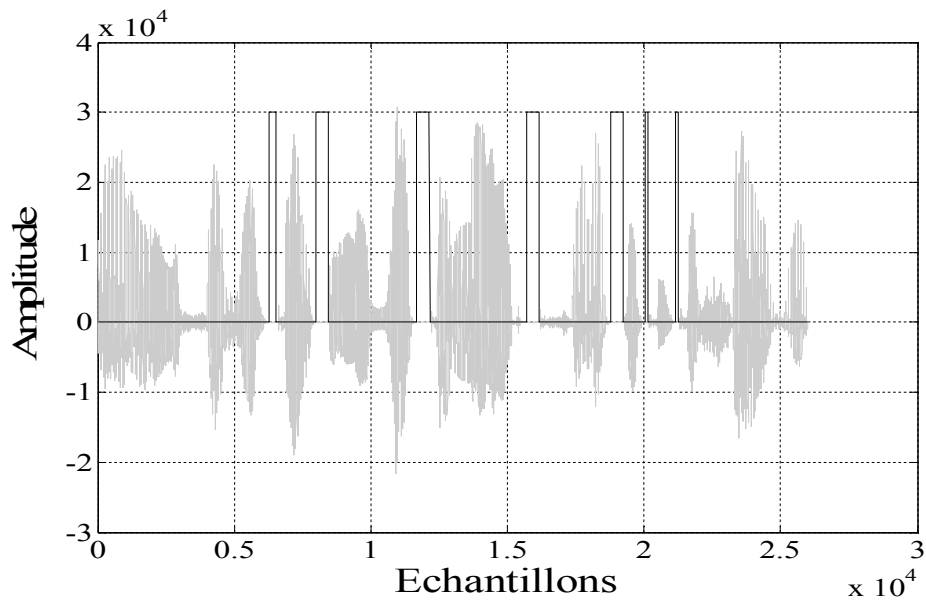


Figure 4. 1. Signal de parole (masculin) avec sa segmentation manuelle.

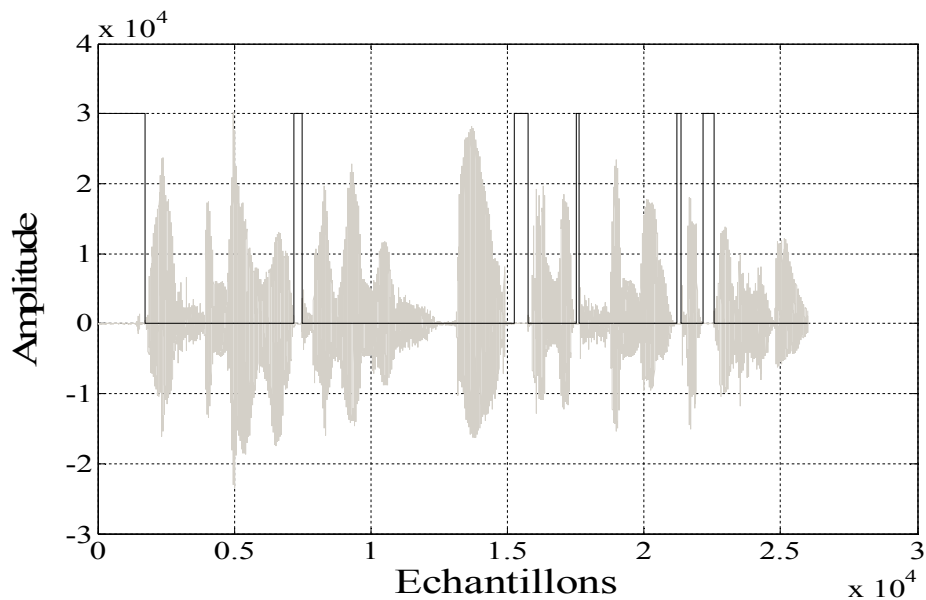


Figure 4. 2. Signal de parole (féminin) avec la segmentation manuelle.

En ce qui concerne les signaux du bruit, la figure (3) représente le bruit Blanc gaussien stationnaire qui est une réalisation d'un processus aléatoire dans lequel la densité spectrale de puissance est la même pour toutes les fréquences et sert surtout à vérifier la stabilité numérique de l'algorithme utilisé. Et la figure (4) représente le bruit USASI.

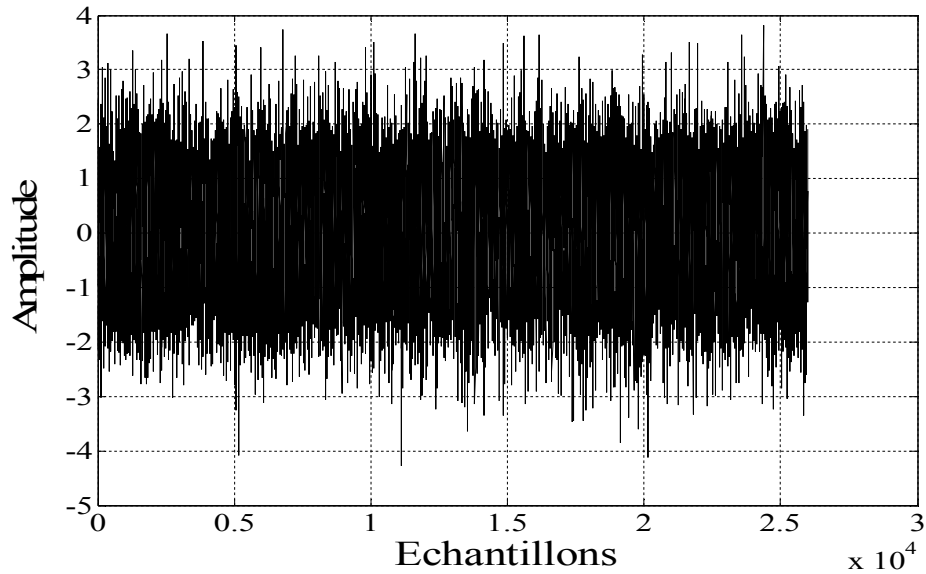


Figure 4. 3. Un signal de Bruit Blanc.

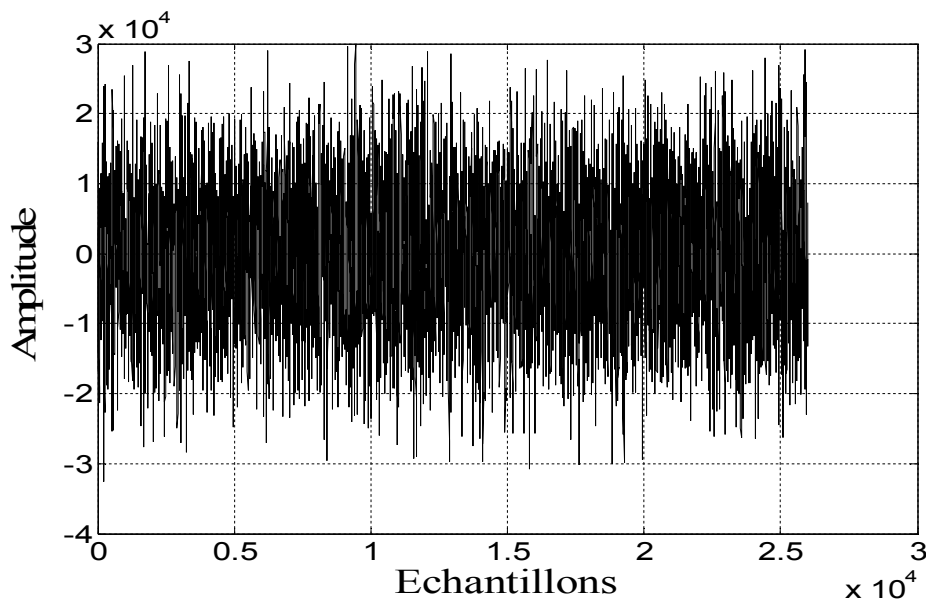


Figure 4. 4. Un signal de Bruit USASI.

4.2.2 Evaluation du RSB et la distance cepstrale

Pour la technique de décision dirigée DD avec les gains de la soustraction spectrale, les résultats obtenus avec le paramètre $\beta = 0.98$ ne sont pas de bonne qualité (le signal de sortie est distordu (cela est dû à la suppression de certaines composantes du signal de la parole) et très bruité. Nous avons mené des simulations afin de chercher la bonne valeur de ce paramètre β . D'après plusieurs essais, on a trouvé que la valeur $\beta = 0.7$ permet d'avoir de très bonne qualité du signal de parole de sortie. Cependant, et avec la

technique de Wiener (GW), la meilleure valeur du paramètre $\beta=0.98$. Avec la technique de DD combinée avec le gain GSSG, on a sélectionné les paramètres suivants :

$\alpha = 5$: Le coefficient de surestimation le mélange contenant du bruit Blanc.

$\alpha = 1$ Pour le mélange contenant du bruit USASI

$\beta = 0.0001$: Le plancher spectral.

$\gamma = 2$: Détermine le type de soustraction spectrale (agressivité).

Ces paramètres permettent d'obtenir des bons résultats avec une bonne qualité du signal de parole en sortie de traitement.

a Résultats de simulations obtenus avec un signal de parole 'Homme'

- Mélange contenant du Bruit Blanc

RSB (dB)	RSB de la décision dirigée (DD) (dB)				RSB(Boll)(dB)
	GSSA	GSSP	GSSG	GW	SS
-5	7	5	6	7	6
0	12.5	11	12.5	11.5	9
3	15	13	15	14	11
7	18	16	18	17	14
12	22	20	22	21	18

Table 4. 1. Evaluation du RSB avec le bruit blanc à l'entrée.

RSB (dB)	Distance cepstrale de la décision dirigée (DD) en (dB)				DC (Boll)
	GSSA	GSSP	GSSG	GW	SS
-5	-2.94	-2.98	-3.46	-3.27	-0.8
0	-1.4	-3.27	-3.28	-3.47	-1.73
3	-1.95	-3.58	-3.38	-3.58	-2.47
7	-4.2	-4.7	-3.56	-4.07	-3.73
12	-4.94	-5.64	-4.5	-4.88	-4.1

Table 4. 2. Evolution de la DC avec le bruit blanc à l'entrée.

- D'après les résultats obtenus par la technique de décision dirigée combinée avec le GSSA, nous remarquons qu'on obtient un gain avoisinant ($G \approx 12\text{dB}$) à la sortie pour chaque valeur de RSB d'entrée. La technique DD combinée avec le gain GSSP permet d'avoir un rehaussement de ($G \approx 9\text{dB}$). Cependant, la technique DD combinée avec le gain GSSG permet d'avoir ($G \approx 12\text{dB}$). La figure suivante représente le signal de parole estimé à la sortie de traitement et le signal de mélange (bruité) en utilisant la technique de DD combinée avec le gain de GSSA :

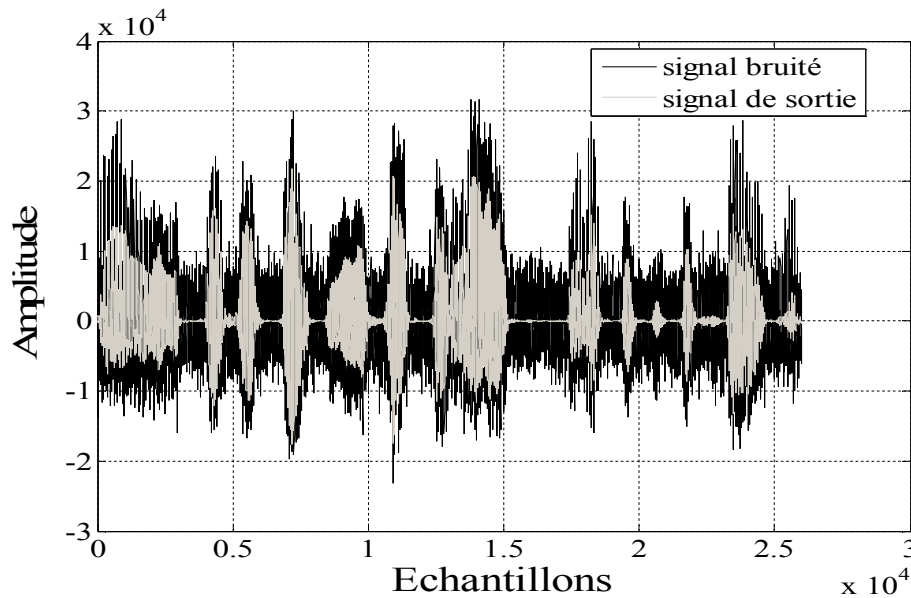


Figure 4. 5. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) avec un bruit blanc à l'entrée.

D'après cette figure et certains tests d'écoutes, nous avons remarqué que le signal de sortie contient du bruit musical qui est introduit à cause de la valeur de $\beta = 0.7$. Aussi, et d'après cette figure et les résultats de la distance cepstrale (Table 4.2), nous concluons que la distorsion du signal de sortie est faible, donc la réduction de bruit par cette technique est forte.

- **Cependant**, la simulation de la technique DD combinée avec le gain de Wiener a permis un gain en sortie de traitement de ($G \approx 11\text{dB}$) pour chaque valeur du RSB d'entrée. La figure suivante représente le signal de parole estimé à la sortie de traitement et le signal de mélange (bruité) par la technique DD combinée avec le gain de Wiener:

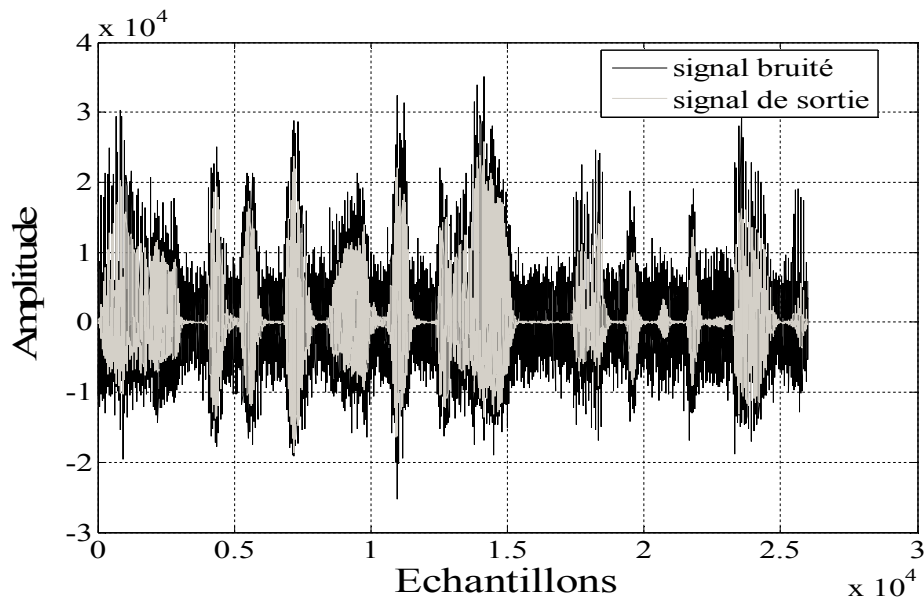


Figure 4. 6. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) avec un bruit blanc à la l'entrée.

D'après cette figure et certains tests d'écoute, le signal de sortie a montré l'absence de bruit musical, et les résultats de la distance cepstrale (Table 4.2) montrent que la distorsion de signal de sortie est faible donc la réduction du bruit par cette technique est très bonne.

- **Les résultats** de simulation de la technique de soustraction spectrale de Boll qui permet d'avoir un gain en sortie de ($G \approx 7\text{dB}$). La figure suivante représente le signal de parole estimé à la sortie de traitement et le signal de mélange (bruité) par la technique de Boll :

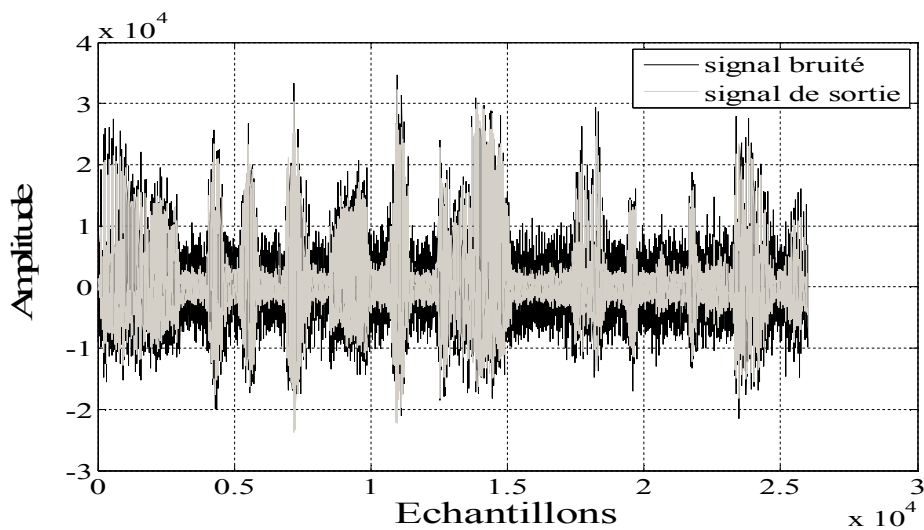


Figure 4. 7. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) avec un bruit Blanc à l'entrée.

D'après cette figure les tests d'écoute du signal de sortie estimé, on observe la présence de bruit musical, et les résultats de la distance cepstrale (Table 4.2) montrent que la distorsion de signal de sortie est très forte, donc la réduction du bruit par cette technique est faible.

- Suite aux résultats de simulation que nous avons obtenus jusqu'à maintenant, on note que la technique de DD combinée avec le gain de Wiener apporte une très bonne amélioration au niveau du signal de sortie en comparaison avec les deux techniques de DD combiné avec le gain de la SSP et la technique de Boll.

➤ **Mélange contenant du Bruit USASI**

RSB (dB)	RSB de la décision dirigé (DD_Melah) (dB)				RSB(Boll)(dB)
	GSSA	GSSP	GSSG	GW	GW
-5	4	3	3	3	1
0	10	8	8	9	5
3	12	10	10	12	8
7	15	13	13	14	12
12	18	17	17	17	16

Table 4. 3. Evaluation du RSB avec le bruit USASI à l'entrée.

RSB (dB)	Distance cepstrale de la décision dirigée (DD) en (dB)				DC Boll en(dB)
	GSSA	GSSP	GSSG	GW	SS
-5	-3.12	-3.36	-3.59	-4	-2.42
0	-3.69	-4.19	-4.26	-3.7	-3.65
3	-4.39	-5.17	-5.21	-4.19	-3.98
7	-5.54	-6.96	-7.04	-5.76	-4.12
12	-7.63	-9.71	-9.85	-8.23	-5.56

Table 4. 4. Evaluation du CD avec le bruit USASI à l'entrée.

- D'après les résultats obtenus pour la technique de décision diriger avec le GSSA permet d'avoir un gain ($G \approx 8\text{dB}$) à la sortie pour chaque valeur de RSB d'entrée. La technique DD avec le gain de GSSP permet d'avoir un gain de ($G \approx 7\text{dB}$). Cependant, la technique DD avec le gain de SSG a montré un gain de ($G \approx 6\text{dB}$). La figure suivante représente le signal de parole estimé à la sortie de traitement et le signal de mélange (bruité) en utilisant la technique de DD avec le gain de GSSA :

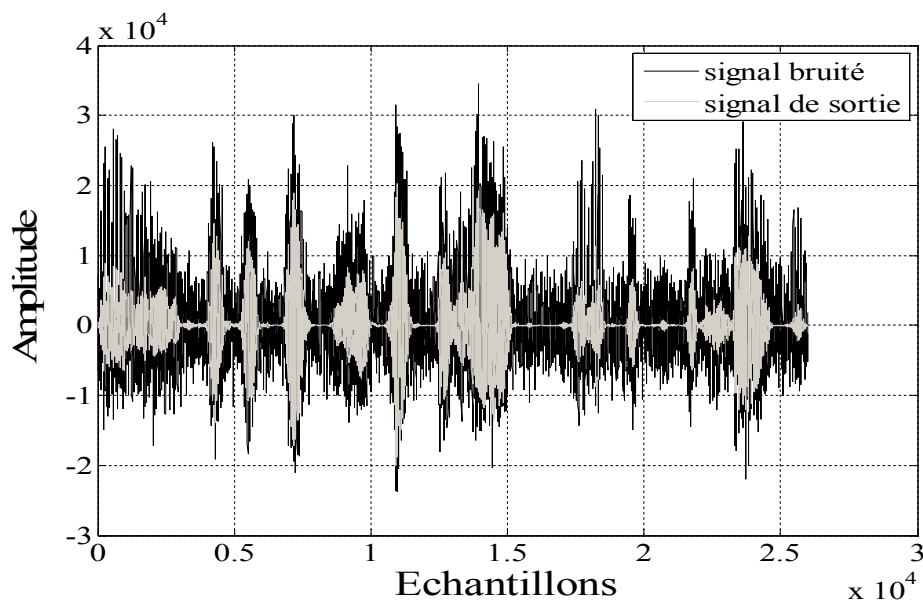


Figure 4.8. Évaluation temporelle du signal de mélange (noir) et de sortie (gris) avec un bruit USASI en entrée.

D'après les tests d'écoute que nous avons réalisé sur le signal de sortie, on note la présence de bruit musical introduit à cause de la valeur de $\beta = 0.7$. et d'après cette figure et les résultats de la distance cepstrale (Table 4.4), on montre que la distorsion de signal de sortie est faible donc la réduction du bruit pour cette technique est forte.

- **Par la suite**, la simulation de la technique DD a permis un gain en sortie de traitement de ($G \approx 8\text{dB}$) pour chaque valeur de RSB d'entrée. La figure suivante représente le signal de parole estimé à la sortie de traitement et le signal de mélange (bruité) par la technique DD combiné avec le gain de Wiener:

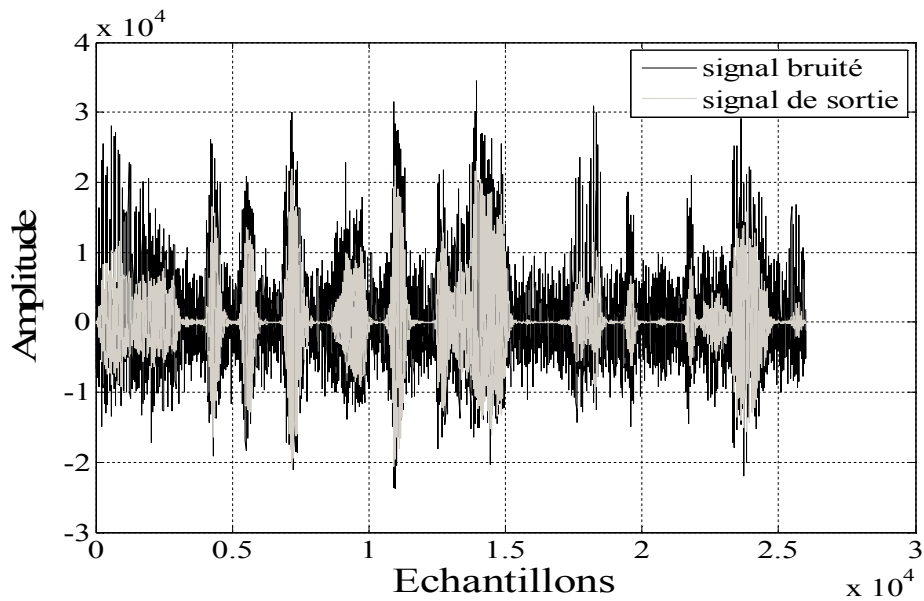


Figure 4.9. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) avec un bruit USASI en entrée.

D'après cette figure et certain test d'écoute, le signal de sortie a montré l'absence de bruit musical, et les résultats de la distance cepstrale (Table 4.4) montrent que la distorsion de signal de sortie est faible donc la réduction du bruit par cette technique est très forte.

- **Les résultats** de simulation de la technique de soustraction spectrale de Boll qui permet d'avoir un gain en sortie de ($G \approx 5\text{dB}$). La figure suivante représente le signal de parole estimé à la sortie de traitement et le signal de mélange (bruité) par la technique de Boll :

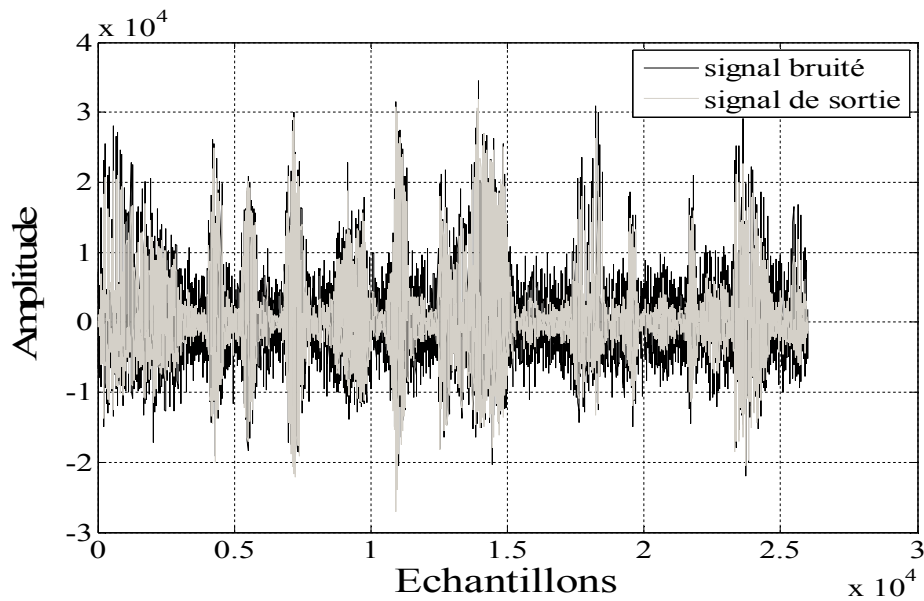


Figure 4.10. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) avec un bruit USASI à l'entrée.

D'après cette figure les tests d'écoute du signal de sortie estimé, on observe la présence de bruit musical, et les résultats de la distance cepstrale (Table 4.4) montrent que la distorsion de signal de sortie est forte, donc la réduction du bruit par cette technique est faible.

- D'après les résultats de simulation que nous avons obtenus jusqu'à maintenant, notant que la technique de DD avec le gain de Wiener reste la bonne technique de réduction de bruit.

b Résultats de simulations obtenus avec un signal de parole 'Femme'

➤ Mélange Bruit Blanc

RSB (dB)	RSB de la décision dirigé (DD_Melah) (dB)				RSB(Boll)(dB)
	GSSA	GSSP	GSSG	GW	GW
-5	10	7	10	9	6
0	14.5	12	15	14	9
3	17	14	17	16	12
7	20	17	20	19	15
12	24	21	24	23	19

Table 4. 5. Evaluation du RSB avec le bruit Blanc à l'entrée.

RSB (dB)	Distance cepstrale de la décision dirigée (DD) en (dB)				DC de Boll en(dB)
	GSSA	GSSP	GSSG	GW	SS
-5	-2.92	-2.31	-2.8	-3.03	-0.75
0	-3.68	-3.2	-3.86	-3.81	-1.7
3	-4.12	-3.82	-4.2	-4.25	-2.12
7	-4.66	-4.65	-4.52	-4.84	-3.06
12	-5.48	-5.67	-5.07	-5.55	-4.55

Table 4. 6. Evaluation du CD avec le bruit Blanc à l'entrée.

- D'après les résultats obtenus avec la technique de DD combiné avec le GSSA on obtient un gain en sortie de ($G \approx 13\text{dB}$) pour chaque valeur de RSB d'entré. Avec la technique GSSP, on obtient un gain de ($G \approx 11\text{dB}$). Cependant, la technique der DD avec le GSSG donne un gain($G \approx 13\text{dB}$). La figure suivante représente le signal de parole estimé à la sortie et le signal de mélange (bruité) par la technique de SSP :

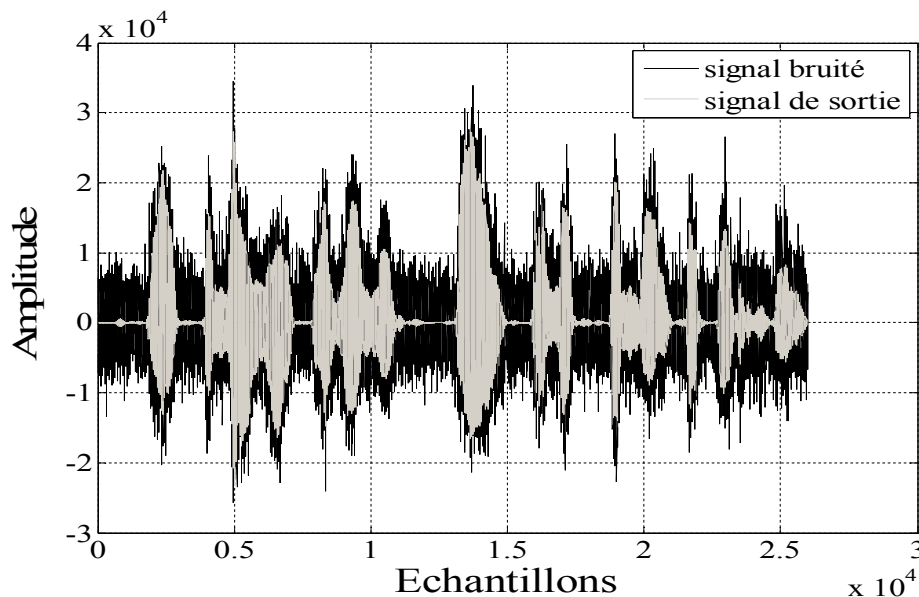


Figure 4. 8. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) avec un bruit Blanc à l'entrée.

- **Par la suite**, la simulation de la technique DD avec le GW a permet un gain en sortie de traitement de ($G \approx 12\text{dB}$) pour chaque valeur de RSB d'entré. La figure

suivante représente le signal de parole estimé à la sortie de traitement et le signal de mélange (bruité) par la technique DD combiné avec le gain de Wiener:

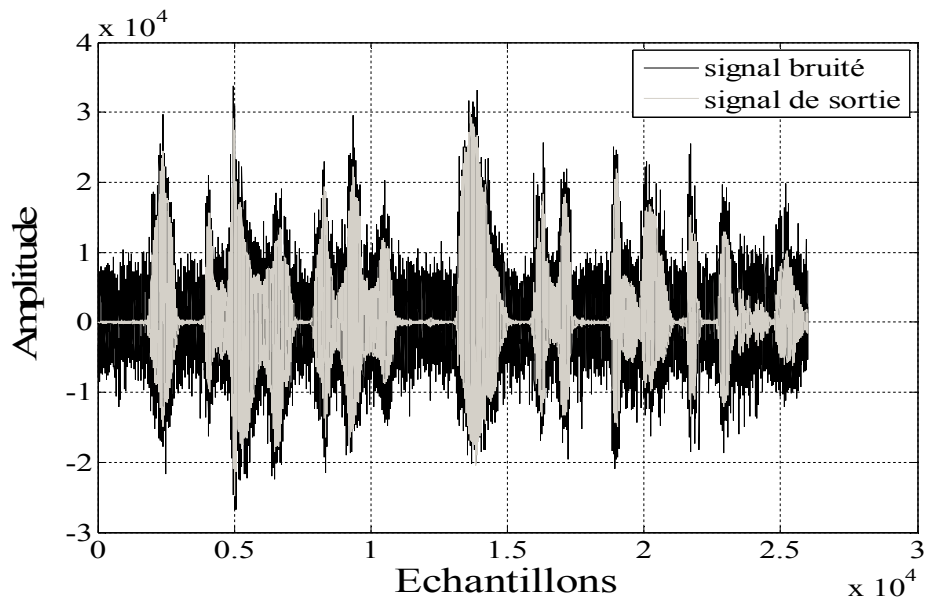


Figure 4.12. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) avec un bruit Blanc à l'entrée.

- **Les résultats** de simulation de la technique de la soustraction de Boll qui donne un gain ($G \approx 7\text{dB}$). La figure suivante représente le signal de parole estimé à la sortie de traitement et le signal de mélange (bruité) par la technique de Boll :

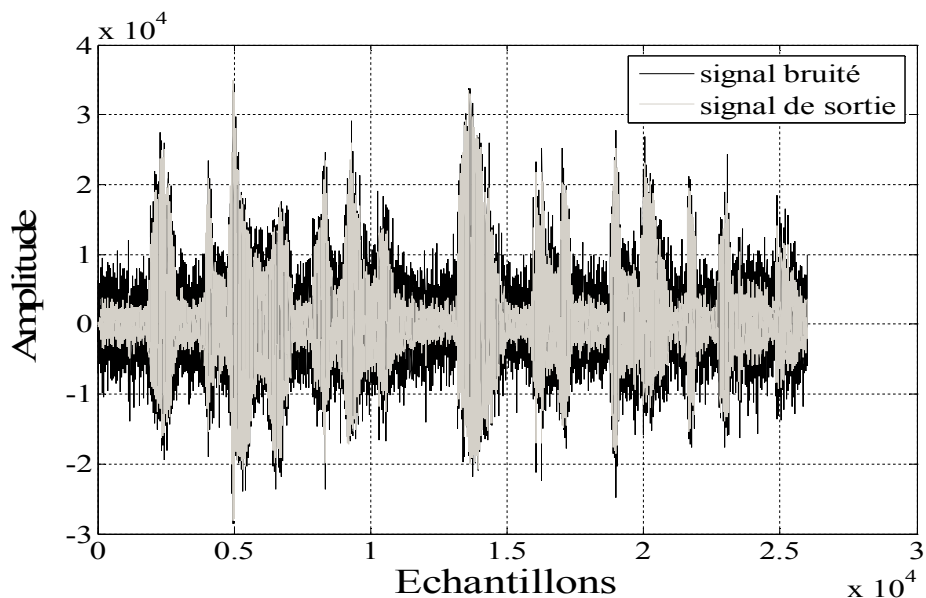


Figure 4.13. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) avec un bruit Blanc à l'entrée.

- D'après les résultats de simulation que nous avons obtenus jusqu'à maintenant, on note que la technique de DD combiné avec le gain de Wiener reste toujours la meilleure technique de réduction de bruit.

➤ **Mélange contenant du Bruit USASI**

RSB (dB)	RSB de la décision dirigé (DD_Melah) (dB)				RSB(Boll)(dB)
	GSSA	GSSP	GSSG	GW	SS
-5	4	2.5	3	3.5	2
0	10	9	9	10	8
3	13	11.5	11	13	11
7	16	15	15	16	13
12	20	18	18	19	17

Table 4. 7. Evaluation du RSB avec le bruit USASI à l'entrée.

RSB (dB)	Distance cepstrale de la décision dirigée (DD) en (dB)				(Boll)
	GSSA	GSSP	GSSG	GW	SS
-5	-2.54	-2.5	-2.63	-2.61	-2.09
0	-3.86	-4.17	-4.26	-4.07	-3.22
3	-4.85	-5.19	-5.26	-4.87	-4.12.
7	-6.03	6.37	-6.4	-5.83	-5.56
12	-7.5	-8.1	-8.09	-7.38	-6.17

Table 4. 8. Evaluation du CD avec le bruit USASI à l'entrée.

- D'après ces résultats de simulations obtenus par la technique de DD combiné avec GSSA on obtient un gain en sortie de traitement de ($G \approx 9\text{dB}$) Our chaque valeur de RSB d'entré. Le GSSP permet un gain de sortie de ($G \approx 8\text{dB}$). Cependant, la DD avec le gain du GSSG rehausse le signal de sortie avec un gain de ($G \approx 8\text{dB}$).

La figure suivante représente le signal de parole estimé à la sortie et le signal de mélange (bruité) par la technique de SSG :

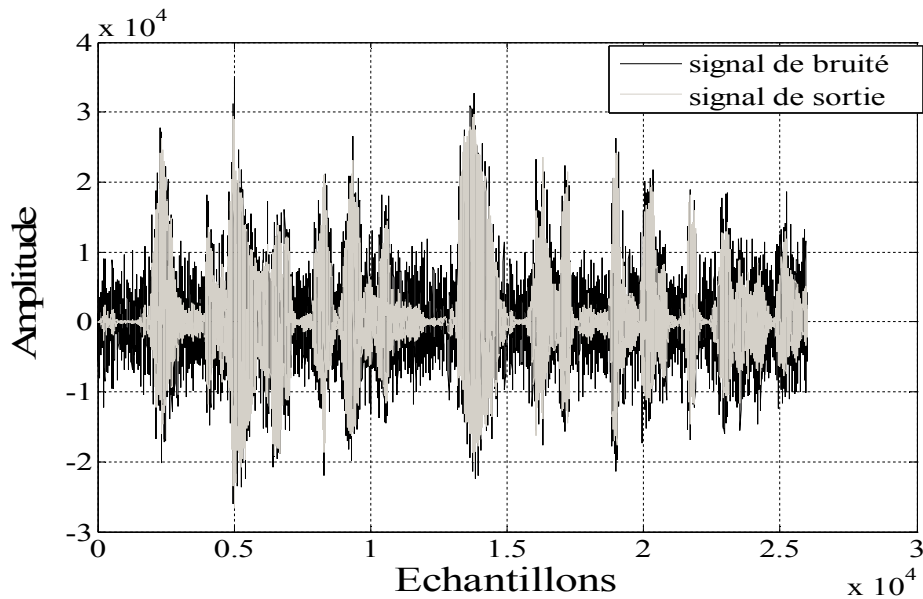


Figure 4.14. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) avec un bruit USASI à l'entrée.

- **En ce qui concerne la technique de DD combinée avec le gain de Wiener,** on obtient un gain de ($G \approx 10\text{dB}$) à la sortie pour toutes les valeurs de RSB d'entrée. La figure suivante représente le signal de parole estimé à la sortie et le signal de mélange (bruité) par la technique de DD avec le gain GW :

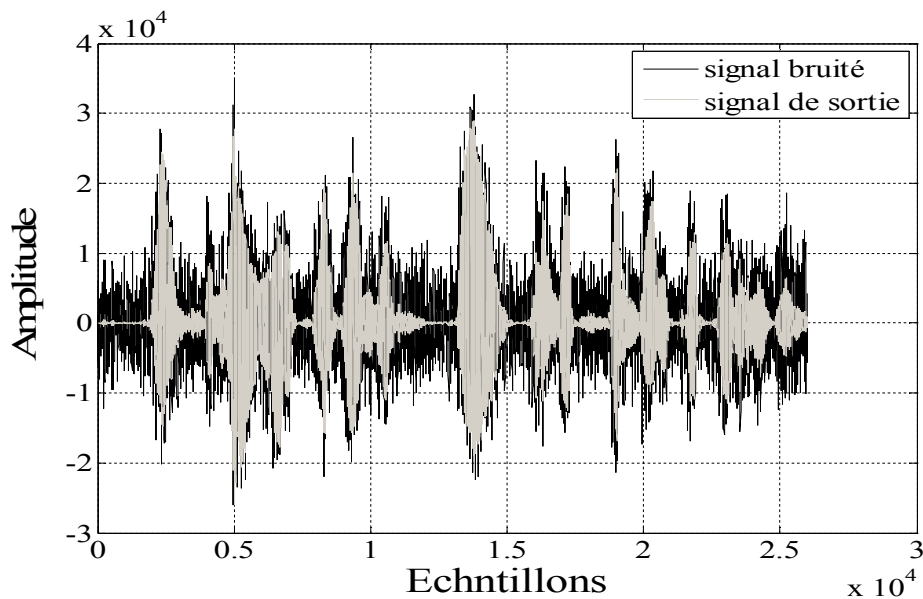


Figure 4.15. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) avec un bruit USASI à l'entrée.

- **Les résultats** de la soustraction spectrale avec la technique de Boll qui donne un $(G \approx 4\text{dB})$. La figure suivante représente le signal de parole estimé à la sortie et le signal de mélange (bruité) par la technique de Boll :

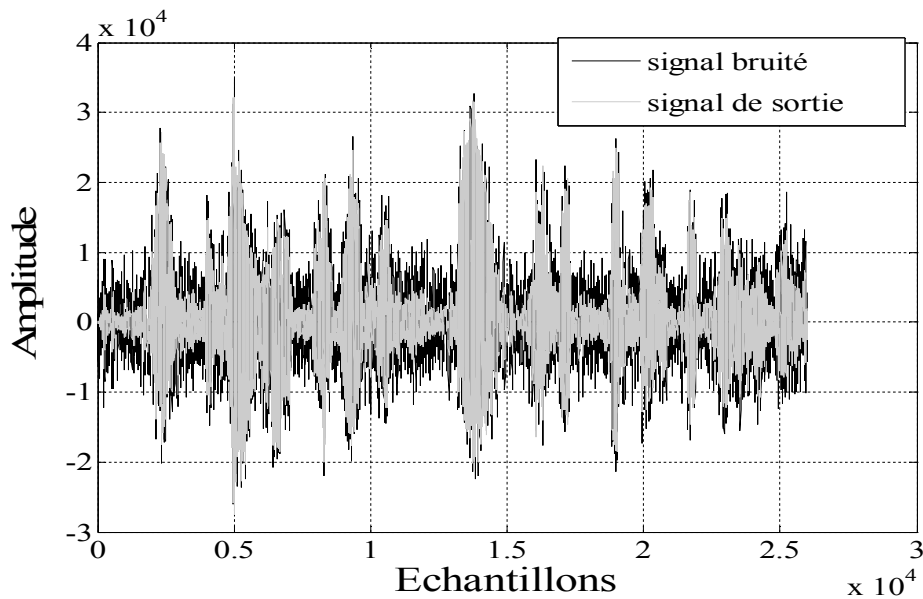


Figure 4.16. Évaluation temporelle du Signal de mélange (noir) et de sortie (gris) av un bruit USASI à l'entrée.

4.3 Conclusion

D'après toutes les simulations que nous avons faites avec toutes les techniques présentées ci-dessus, nous avons bien remarqué que la technique de DD combinée avec le gain de Wiener donne des bons résultats par apport aux autres techniques pour la réduction du bruit. Nous avons aussi remarqué que les résultats obtenus avec le bruit blanc sont beaucoup mieux que les résultats obtenus avec le bruit USASI, cela est du au fait que le bruit blanc est stationnaire est décorrélé, alors que le USASI est stationnaire mais son spectre ressemble à celui de la parole.

Conclusion générale

L'étude que nous présentons dans ce mémoire concerne la réduction du bruit et le rehaussement du signal de parole par les techniques basées sur une seule prise de son, i.e. mono-capteurs.

Nous avons, dans un premier niveau de cette thèse, réalisé un état de l'art des techniques de réduction du bruit mono-capteurs qui permettent le rehaussement du signal de parole avec le minimum de distorsions. Parmi ces méthodes qui existent dans la littérature, nous avons détaillé la méthode de soustraction spectrale de Boll (SS) et la technique de décision-dirigé d'Ephraïm and Mellah (DD). **Nous avons proposé lors de ce projet de Master trois nouvelles versions de la technique de DD. Ces nouvelles versions de la technique de DD sont obtenus en combinant la DD avec (1) le gain de la soustraction spectrale d'amplitude (GSSA), (2) le gain de la soustraction spectrale de puissance (GSSP) et (3) le gain de la soustraction spectrale généralisée (GSSG).**

Lors des simulations de ces différentes technique de réduction de bruit, nous avons constaté que le technique DD combiné avec le gain de Wiener garde une supériorité permanente sur toutes les autres technique de réduction de bruit et cela quelque soit le contexte de la simulation et le niveau et le type de du bruit qui existent dans les signaux de mélanges.

Ce projet de Master nous a permis de faire un premier pas dans le monde de la recherche scientifique de la réduction du bruit et le rehaussement de la qualité du signal de parole.

Bibliographie

- [1] André.Quinquis, “livre de traitement du signal sous matlab”.
- [2] F. Cottet, “ traitement du signal ” Livre, 2005 Dunod, Paris.
- [3] G. Dimitris Manolakis, K. Vinay Ingle, M. Stephen Kogon “Statistical and Adaptive Signal Processing” article, 2005.
- [4] Murat.Kunt , “Traitement Numérique des signaux” Presse Polytechnique Romandes 1981.
- [5] L.Herous, “Elément de base de théorie et du traitement du signa” OPU Alger 1989.
- [6] G. Dimitris Manolakis, K. Vinay Ingle, M. Stephen Kogon “Statistical and Adaptive Signal Processing” article, 2005.
- [7] F.Ykhlef, “Etude de la stabilité des algorithmes moindres carrés transversaux rapides avec des signaux parfaitement prédictibles application a la parole” Thèse de Magister, université Saad Dahlab de Blida, Janvier 2002.
- [8] M. Bellenger, “ traitement numérique du signal ”, édition 8, Dunod ,1987.
- [9] A.Amehraye, “Débruitage perceptuel de la parole” Thèse de Doctorat, l'Ecole Nationale Supérieure des Télécommunications de Bretagne, 2009.
- [10] P. Loizou, “Speech enchasement: Theory and practice”.
CRC; 1 edition, 2007
- [11] Y. Ephraïm, et D. Malah, “Speech Enhancement Using a Minimum Mean Square Er-
ror Short-Time Spectral Amplitude Estimator,” IEEE Trans. Acoust., Speech,
Signal Processing, Vol. ASSP-32, No. 6, pp. 1109–1121, Décembre 1984.
- [12] C. Plapous, C. Marro, L. Mauuary, et P. Scalart, “Two-Step Noise Reduction
Technique,” IEEE Intl. Conf. Acoust., Speech, Signal Processing, Montréal,
Canada, Vol. 1, pp. 289–292, Mai 2004

- [13] D. L. Wang, et J. S. Lim, "The Unimportance of Phase in Speech Enhancement, Trans, Acoust" Speech, Signal Processing, Vol. ASSP-30, No. 4, pp. 679–681, Août 1982.
- [14] S. F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-27, No. 2, pp. 113–120, Avril 1979.
- [15] P. Vary, "Noise Suppression by Spectral Magnitude Estimation - Mechanism and Theoretical Limits," Signal Processing, Vol. 8, pp. 387–400, 1985.
- [16] M. Berouti, R. Schwartz, et J. Makhoul, "Enhancement of Speech Corrupted by Acoustic Noise," IEEE Intl. Conf. Acoust., Speech, Signal Processing, Washington, États-Unis, pp. 208–211, Avril 1979.
- [17] R. Le Bouquin, "Traitements pour la Réduction du Bruit sur de la Parole. Application aux Communications Radio-Mobiles," Thèse de l'Université de Rennes 1, 1991
- [18] P. Scalart, et J. Vieira Filho, "Speech Enhancement Based on a Priori Signal to Noise Estimation," IEEE Intl. Conf. Acoust., Speech, Signal Processing, Atlanta, États-Unis, Vol. 2, pp.629–632, Mai 1996.
- [19] O. Cappé, "Techniques de Réduction de Bruit pour la Restauration d'Enregistrements Musicaux," Thèse de l'École Nationale Supérieure des Télécommunications, Paris, Septembre 1993.
- [20] F. Beritelli, S. Casale, G. Ruggeri, et S. Serrano, "Performance Evaluation and Comparison of G.729/AMR/Fuzzy Voice Activity Detectors," IEEE Signal Processing Lett., Vol. 9, No. 3, Mars 2002.
- [21] P. Scalart, et J. Vieira Filho, "Speech Enhancement Based on a Priori Signal to Noise Estimation," IEEE Intl. Conf. Acoust., Speech, Signal Processing, Atlanta, États-Unis, Vol. 2, pp.629–632, Mai 1996.
- [22] L. Arslan, A. McCree, et V. Viswanathan, "New Methods for Adaptive Noise Suppression," IEEE Intl. Conf. Acoust., Speech, Signal Processing, Detroit, États-Unis, Vol. 1, pp.812–815, Mai 1995.
- [23] R. E. Crochiere, et L. R. Rabiner, "Multirate Digital Signal Processing," Prentice-Hall, Première édition, 1983.
- [24] M. Kunt, "Traité d'Électricité - Traitement Numérique des Signaux," Presses Polytechniques et universitaires Romandes, Artech House, Troisième édition, 1998.