

République Algérienne Démocratique et populaire
Ministère de l'Enseignement supérieur et de la recherche Scientifique



Université Mohammed Seddik BENYAHIA - Jijel
Faculté des Sciences et de la Technologie
Département d'Electronique

MEMOIRE DE FIN D'ETUDES

Présenté en vue de l'obtention du diplôme
de **MASTER** en
Systemes des télécommunications

Thème :

Rehaussement de la parole par acquisition comprimée

Encadré par :

❖ Pr. Abdellah KACHA

Réalisé par :

❖ M. Oussama Boukerkeb

❖ M. Anis Benhalla

Promotion: Juillet 2021

Remerciements

Tout d'abord nous rendons grâce à Dieu, lui qui nous a permis d'être bien portant afin d'effectuer ce travail du début jusqu'à la fin.

Nous remercions nos parents respectifs pour leurs soutiens durant notre parcours de formation.

Nos remerciements vont, à notre encadrant le professeur, Mr Abdellah KACHA qui nous a guidés avec ses orientations, ses conseils et ses critiques tout au long de ce travail de recherche en nous laissant la liberté dont on avait besoins. On ne peut que lui être reconnaissant surtout pour ses qualités intellectuelles et humaines.

Nos remerciements vont aussi au membre du jury, pour l'honneur qu'ils nous ont fait en acceptant d'évaluer ce travail et de participer à la soutenance.

Et enfin, nous sommes reconnaissants envers tous les enseignants de l'UNIVERSITE MOHAMED SADDIK BEN YAHIA de JIJEL pour leur contribution à notre formation, et également à nos camarades, amis pour leurs aides précieuses.

Dédicaces

Grâce à Dieu le tout puissant, j'ai achevé la réalisation de ce modeste travail que je tiens très chaleureusement à le dédier à :

- ✚ Ma mère chérie et mon père, qui m'ont encouragé et soutenu tout au long de mes études et pour leurs patiences, que Dieu les protège et les gardes pour moi.
- ✚ À mes adorables sœurs.
- ✚ À mes grands-mères et mes grands-pères.
- ✚ À toute la famille BOUKERKEB et DJABOUB.
- ✚ Et à mes très chers amis et camarades pour tous les moments d'échange et de débat, aux personnes qui m'ont toujours aidé et soutenue...

Oussama Boukerkeb

Dédicaces

Je dédie ce travail à mes chers parents, Karima et Zahir, pour tous les sacrifices, leur amour, leur soutien à toute épreuve et leurs prières tout au long de mes études, à mes chères sœurs MOUNA et YASMINE, pour leurs encouragements et leur soutien moral, ainsi, à toute ma famille, mes grands-mères et mon grand-père mes tantes et mes oncles, pour leur soutien tout au long de mon parcours universitaire.

Et à mes chers amis et camarades, pour leur soutien aux moments difficiles de mon travail et surtout pour leurs conseils.

Anis Benhalla

Table des matières

Remerciements	i
Dédicace.	ii
Table des matière	iv
Liste des figures	viii
Liste des tableaux	x
Liste des abréviations	xi
Introduction générale	1

Chapitre I :

Etat de l'art du rehaussement de la parole

I.1 Introduction	4
I.2 La production de la parole	4
I.2.1 L'appareil phonatoire	4
I.2.2 Modèle mécanique de production de la parole	5
I.2.3 Modèle numérique de production de la parole	6
I.2.4 Classification des sons de la parole	7
I.2.5 Propriétés du signal de la parole.....	10
I.3 Les bruits	11
I.3.1 Bruit blanc	11
I.3.2 Bruits colorés.....	11
I.3.3 Bruit musical	11
I.3.4 Bruit impulsif.....	12
I.3.5 Bruit ambiant	12
I.3.6 Bruit acoustique.....	12
I.3.7 Les modèles du bruit	12

I.4	Classification des méthodes de rehaussement de la parole	12
I.5	Etat de l'art des méthodes de rehaussement de la parole	13
I.5.1	Soustraction spectrale	14
I.5.2	Filtrage de Wiener	16
I.5.3	Débruitage par ondelettes	18
I.6	Evaluation des performances des méthodes de rehaussement.....	23
I.6.1	PESQ (évaluation perceptive de la qualité de la parole)	23
I.6.2	MOS (Note d'opinion moyen)	23
I.6.3	Rapport signal sur bruit	24
I.7	Conclusion.....	25

Chapitre II :

Acquisition comprimée

II.1	Introduction.....	26
II.2	Principe de l'acquisition comprimée	26
II.3	Parcimonie	28
II.3.1	Mesure de la parcimonie.....	28
II.3.2	Représentation parcimonieuse d'un signal.....	29
II.3.3	Transformation parcimonieuse d'un signal	30
II.4	Acquisition du signal	30
II.4.1	Incohérence.....	31
II.4.2	Propriété d'isométrie restreinte (RIP).....	32
II.4.3	Mauvaise matrice de mesure	33
II.5	Algorithme de reconstruction	34
II.5.1	Reconstruction par la minimisation L_2	34
II.5.2	Reconstruction par la minimisation L_0	35
II.5.3	Reconstruction par la minimisation L_1	35

II.5.4	Discussion.....	35
II.6	Exemple d'acquisition et de reconstruction.....	37
II.7	Applications de l'acquisition comprimée	39
II.7.1	Compression des données.....	39
II.7.2	Codage de canal.....	39
II.7.3	Acquisition des données	39
II.7.4	Imagerie médicale.....	40
II.8	Conclusion	41

Chapitre III :

Rehaussement mono-microphone par acquisition comprimée d'ensemble

III.1	Introduction.....	42
III.2	Rehaussement mono-microphone par acquisition comprimée d'ensemble.....	42
III.3	Rehaussement mono-microphone par acquisition comprimée d'ensemble et seuillage des coefficients d'ondelettes.....	48
III.3.1	Méthode	48
III.3.2	Exemple d'illustration	49
III.4	Conclusion	54

Chapitre IV :

Résultats expérimentaux

IV.1	Introduction.....	55
IV.2	Base de données.....	55
IV.3	Protocole expérimentale	55
IV.4	Evaluation des performances de la méthode de rehaussement basée sur l'acquisition comprimée d'ensemble	56

IV.5 Evaluation des performances de la méthode de rehaussement combinant l'acquisition comprimée d'ensemble et le seuillage des coefficients d'ondelettes.....	64
IV.6 Conclusion	71
Conclusion générale.....	72
Références bibliographiques.....	73
Résumé	

Liste des figures

Figure I.1: Vue générale de l'appareil phonatoire.....	5
Figure I.2 : Modèle mécanique de production de la parole.....	6
Figure I.3 : Modèle numérique de production de la parole.....	6
Figure I.4 : Spectre d'un son voisé.....	8
Figure I.5 : Spectre d'un son non voisé.....	9
Figure I.6 : Forme d'onde incluant un son voisé et un son non voisé	9
Figure I.7: Exemple d'ondelettes : (a) Chapeau mexicain (mexican hat) (b) Morlet.....	18
Figure I.8 : seuillage dur.....	22
Figure I.9 : seuillage doux.....	23
Figure II.1 : Principe de base de l'acquisition comprimée.....	27
Figure II.2 : Sphères unitaires dans R^2 pour les normes L_p	29
Figure II.3 : Processus de mesure.....	31
Figure II.4 : Exemple d'une mauvaise matrice de mesure.....	33
Figure II.5 : (a) Vecteur parcimonieux se trouvant sur un hyperplan de dimension K aligné avec les axes des coordonnées dans R^N et donc proche des axes. (b) Récupération de l'acquisition comprimée via la minimisation L_2 qui donne un vecteur non parcimonieux \hat{S} . (c) Récupération via la minimisation L_1 qui donne la bonne solution parcimonieuse S	36
Figure II.6 : Illustration de l'acquisition du signal et sa reconstruction en utilisant la minimisation L_1 . (a) Signal original et échantillons prélevés aléatoirement. (b) DCT du signal. (c) Solution L_1 notée x . (d) DCT inverse de x	38
Figure III.1 : Schéma bloc de la méthode de rehaussement par acquisition comprimée d'ensemble.....	43
Figure III.2 : Signal propre, (b) bruité et (c) rehaussé par la technique d'acquisition comprimée d'ensemble pour $M=0.4N$ et un SNR d'entrée de 0 dB.....	45
Figure III.3 : Spectrogrammes des signaux (a) propre, (b) bruité et (c) rehaussé par la technique d'acquisition comprimée d'ensemble pour $M=0.4N$ et un SNR d'entrée de 0 dB.....	46

Figure III.4 : Effet du nombre de répétitions N_r sur l'évolution du SNR de sortie du signal rehaussé par l'acquisition comprimée d'ensemble en fonction du SNR d'entrée pour une longueur de trame de 30 ms ($N=240$) et un nombre d'échantillons prélevés aléatoirement $M=0.6N$	47
Figure III.5 : Effet de la longueur de trame N sur l'évolution du SNR de sortie du signal rehaussé par acquisition comprimée d'ensemble en fonction de SNR d'entrée pour un nombre de répétitions $N_r=15$ et un nombre d'échantillons prélevés aléatoirement $M=0.6N$	47
Figure III.6 : Principe de la méthode de rehaussement par acquisition comprimée d'ensemble et seuillage des coefficients d'ondelettes.....	48
Figure III.7 : (a) Signal propre, (b) signal bruitée, (c) signal estimé par acquisition comprimée d'ensemble et (d) signal estimé acquisition comprimée d'ensemble suivi d'un débruitage des segments non voisés par seuillage des coefficients d'ondelettes pour $M=0.4N$ et un SNR d'entrée de 0 dB.....	50
Figure III.8 : Spectrogrammes des signaux propre, bruité et rehaussé pour $M=0.6N$ et un SNR d'entrée de 0 dB. (a) Signal propre. (b) signal bruité, (c) signal estimé par acquisition comprimée d'ensemble. (d) signale estimé par acquisition comprimée d'ensemble suivi d'un débruitage des segments non voisés par seuillage des coefficients d'ondelettes.....	52
Figure IV.1 : Evolution du SNR de sortie obtenu par moyennage des SNRs de sortie de toutes les phrases pour les différentes méthodes de rehaussement.....	58
Figure IV.2 : Evolution du SNRseg de sortie obtenu par moyennage des SNRseg de sortie de toutes les phrases pour les différentes méthodes de rehaussement.....	61
Figure IV.3 : Evolution du SNR de sortie obtenu par moyennage des SNRs de sortie de toutes les phrases pour la méthode de rehaussement combinant l'acquisition comprimée et le seuillage des coefficients d'ondelettes ainsi que pour la méthode de débruitage par seuillage des coefficients d'ondelettes seule.....	65
Figure IV.4 : Evolution du SNRseg de sortie obtenu par moyennage des SNRseg de sortie de toutes les phrases pour la méthode de rehaussement combinant l'acquisition comprimée et le seuillage des coefficients d'ondelettes ainsi que pour la méthode de débruitage par seuillage des coefficients d'ondelettes seule.....	68

Liste des tableaux

Tableau I.1 : Classification des méthodes de rehaussement de la parole.....13

Tableau I.2 : Valeurs de MOS et la qualité correspondante à chaque valeur.....24

Liste des abréviations

<i>CS</i>	<i>Compressed Sensing</i>
<i>CWT</i>	<i>Continuous Wavelet Transform</i>
<i>DCT</i>	<i>Discrete Cosine Transform</i>
<i>DFT</i>	<i>Discrete Fourier Transform</i>
<i>DSP</i>	<i>Power Spectral Density</i>
<i>DWT</i>	<i>Discrete Wavelet Transform</i>
<i>FIR</i>	<i>Finite Impulse Response</i>
<i>MEQM</i>	<i>Mean Squared Error</i>
<i>MOS</i>	<i>Mean Opinion Score</i>
<i>PESQ</i>	<i>Perceptual Evaluation of Speech Quality</i>
<i>RIP</i>	<i>Restricted Isometry Property</i>
<i>SNR</i>	<i>Signal-to-Noise Ratio</i>
<i>SNRseg</i>	<i>Segmental Signal-to-Noise Ratio</i>
<i>TFCT</i>	<i>Short-Time Fourier Transform</i>
<i>WT</i>	<i>Wavelet Transform</i>

Introduction
générale

Introduction générale

Les progrès technologiques ont amélioré considérablement notre capacité de communication vocale en permettant une communication directe sur une plus grande distance, un public plus large et des circonstances plus difficiles. Le signal de parole constitue un élément essentiel de la chaîne de communication. Cependant, dans beaucoup de situations, la parole est acquise ou transmise en présence de sources de perturbations appelées bruits. Ces bruits qui peuvent être additifs ou convolutifs (réverbération) dégradent la qualité et l'intelligibilité de la parole ce qui peut affecter le bon fonctionnement du système. Le rehaussement de la parole devient alors une opération nécessaire dans les systèmes de communications et de traitement de la parole. Le rehaussement de la parole trouve des applications en téléphonie mobile, dans les systèmes de reconnaissance vocale automatique, en téléconférence, etc.

Dans la plupart des applications, le but du rehaussement de la parole est de supprimer ou réduire le bruit du signal de parole bruitée en vue de produire un signal dont la qualité et l'intelligibilité sont proches à celles du signal propre. Cependant, il y a toujours un compromis entre la réduction du bruit et la distorsion du signal, une plus grande réduction du bruit s'accompagne toujours d'une plus grande distorsion du signal [1]. La qualité est hautement souhaitable car elle peut réduire la fatigue de l'auditeur [2], en particulier dans les situations dans lesquelles l'auditeur est exposé à un niveau de bruit élevé pendant une longue période (par exemple, la fabrication). En d'autres termes, les algorithmes de rehaussement de la parole réduisent ou suppriment le bruit de fond dans une certaine mesure et sont parfois appelés algorithmes de suppression du bruit. L'introduction de la distorsion de la parole due à la réduction du bruit de fond peut nuire à l'intelligibilité de la parole.

Plusieurs techniques de rehaussement de la parole ont été proposées dans la littérature. Selon le nombre de microphones utilisés, elles peuvent être classées en méthodes mono-microphones ou multi-microphones. Les méthodes mono-microphones incluent la soustraction spectrale [3], le filtrage de Wiener [4], les méthodes de sous-espace, etc. Les méthodes multi-microphones exploitent les informations spatiales des microphones. Une méthode courante pour effectuer le filtrage spatial est la formation de faisceau (beamforming).

En dépit du grand nombre de techniques de rehaussement proposées dans la littérature, le développement de méthodes de rehaussement efficaces constitue un grand intérêt. Le

principal défi est de concevoir un algorithme efficace pour supprimer le bruit sans introduire de distorsion perceptible dans le signal [5].

L'acquisition comprimée (CS : compressed sensing) est une approche, relativement nouvelle, d'acquisition et de compression simultanées qui permet une réduction potentiellement importante des coûts d'échantillonnage et de calcul au niveau du capteur pour les signaux ayant une représentation parcimonieuse dans certaines bases qui peuvent être reconstruits en utilisant un nombre d'échantillons beaucoup plus faible que celui déterminé par le théorème de Shannon-Nyquist. La CS s'appuie sur les travaux de Candes, Romberg et Tao [6] et Donoho [7].

L'idée d'appliquer la CS pour le rehaussement de la parole vient du fait que le signal de parole est parcimonieux dans certaines bases alors que les bruits ne sont pas parcimonieux. La méthode de rehaussement basée sur la CS permet donc de reconstruire seulement les signaux parcimonieux. L'avantage de cette méthode est sa simplicité de mise en œuvre et l'absence de détection d'activité vocale ou d'estimation préalable du bruit [8].

La contribution de l'acquisition comprimée au rehaussement de la parole a été abordée et étudiée dans [9]. Il a été montré qu'en moyenne, le rapport signal sur bruit (SNR : signal-to-noise ratio) dans le domaine compressé est supérieur ou égal au SNR dans le domaine non compressé. Motivé par les résultats obtenus dans [9], l'acquisition comprimée a été combinée avec la technique du beamforming pour le rehaussement multi-microphones [10]. Dans le cadre de ce travail, on se propose d'utiliser l'acquisition comprimée pour le rehaussement mono-microphone de la parole corrompue par un bruit additif. La performance de l'acquisition comprimée est limitée quand elle est appliquée au rehaussement des segments non voisés. Pour remédier à cet inconvénient, on se propose d'effectuer un débruitage des segments non voisés par seuillage des coefficients d'ondelettes afin d'améliorer le SNR des segments non voisés. La méthode de rehaussement proposée est testée sur une base de données incluant des signaux de parole corrompus par différents types de bruits et comparée avec les méthodes conventionnelles de rehaussement de la parole, à savoir la soustraction spectrale, le filtrage de Wiener et le débruitage par ondelettes.

Le mémoire est organisé comme suit :

Le premier chapitre présente l'état de l'art du rehaussement de la parole. On rappelle brièvement le mécanisme de production de la parole puis on présente les différents types de bruits ainsi que les différentes méthodes de rehaussement de la parole.

Le deuxième chapitre présente les concepts de base de la théorie de l'acquisition comprimée et ses différentes applications. La méthode est illustrée par un exemple de signal synthétique.

Le troisième chapitre présente la méthode de rehaussement mono-microphone basée sur l'acquisition comprimée et sa version améliorée en effectuant un débruitage des segments non voisés par seuillage des coefficients d'ondelettes.

Le quatrième chapitre présente les résultats expérimentaux obtenus en utilisant une base de données incluant des signaux de parole perturbés par différents types de bruits. Les performances de la méthode proposée seront comparées à celles de la soustraction spectrale, le filtrage de Wiener et le débruitage par ondelettes.

Enfin, on termine par une conclusion.

Chapitre I

Etat de l'art du

rehaussement de la parole

I.1 Introduction

Lors de sa transmission, la parole est généralement corrompue par des différents bruits qui peuvent être additifs ou convolutifs (réverbération). Cette distorsion affecte la qualité et l'intelligibilité de la parole ce qui rend désagréable la communication pour l'inter locuteur qui reçoit ces bruits. La question est de savoir comment séparer au mieux le bruit (information inutile) du signal de parole (information utile). Dans le but d'améliorer la qualité et l'intelligibilité du signal de parole, plusieurs méthodes de rehaussement ont été proposées [11].

Dans ce chapitre, nous présentons les notions fondamentales de la production de la parole et les différents bruits. Ensuite, on présente un état d'art relatif aux techniques de réduction du bruit dans le signal de parole. Nous nous intéressons particulièrement à trois familles de méthodes : la technique de soustraction spectrale, le filtrage de Wiener et le débruitage par ondelettes. A la fin du chapitre, nous présentons les différents critères d'évaluation des performances des méthodes de rehaussement de la parole.

I.2 La production de la parole

Un ensemble de processus neurologiques et musculaires produit les fluctuations de la pression d'air qui constituent le signal vocal. Celui-ci se propage dans l'air [12]. La parole est produite par quatre processus :

- La soufflerie (produite par les poumons)
- La vibration (produite par le larynx et les cordes vocal)
- La résonance (produite par les organes résonateurs c'est-à-dire les cavité buccale et nasale)
- L'articulation (lèvres supérieure et inférieure, la langue, ...)

I.2.1 L'appareil phonatoire

Techniquement parlant, la parole est une onde sonore produite par l'action volontaire et coordonnée des structures anatomiques qui forment l'appareil phonatoire humain. Ce processus est coordonné par le système nerveux central. Les sons produits sont analysés par rétroaction auditive pour assurer la qualité acoustique de la parole.

Les muscles abdominaux actionnent sur le diaphragme, en poussant l'air des poumons vers la trachée artère. Au bout supérieur de celle-ci se trouve le larynx qui module le courant d'air

sous la forme d'impulsions périodiques appliquées au conduit vocal. Celui-ci est formé d'un ensemble de cavités : la cavité pharyngienne suivie de la cavité buccale et en dérivation la cavité nasale. La luvette, qui prolonge le bord postérieur du voile du palais, contribue à la fermeture des fosses nasales. Comme résultat, pendant la production de la parole, la cavité nasale peut être couplée soit totalement, soit partiellement, ou même découplée de la cavité buccale. Autres organes anatomiques importants qui participent à la production des sons sont : la langue et les dents dans la cavité buccale, les narines dans la cavité nasale et les lèvres, comme montré dans la figure I.1 [12].

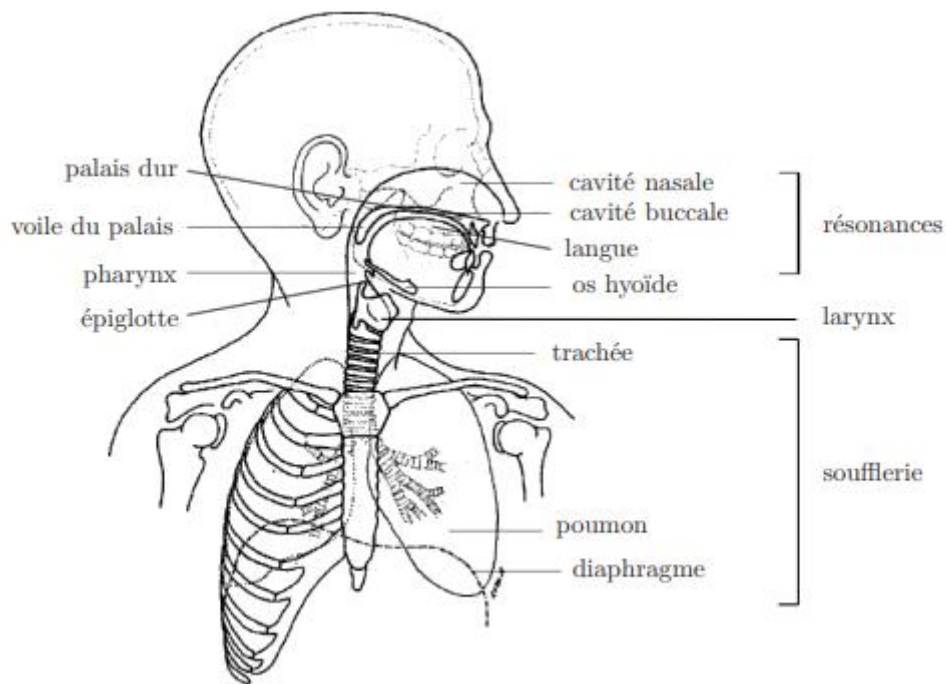


Figure 1-1 : Vue générale de l'appareil phonatoire.

I.2.2 Modèle mécanique de production de la parole

Les cordes vocales humaines sont constituées de plusieurs couches ayant chacune différentes caractéristiques mécaniques. Cette structure, dont les propriétés mécaniques sont hétérogènes et anisotropes, présente un comportement mécanique viscoélastique complexe et non linéaire. Le modèle mécanique général présenté à la figure I.2 repose sur les hypothèses selon lesquelles la loi contrainte-déformation ne dépend ni des sollicitations passées ni de la fréquence de l'excitation ou de la température [13].

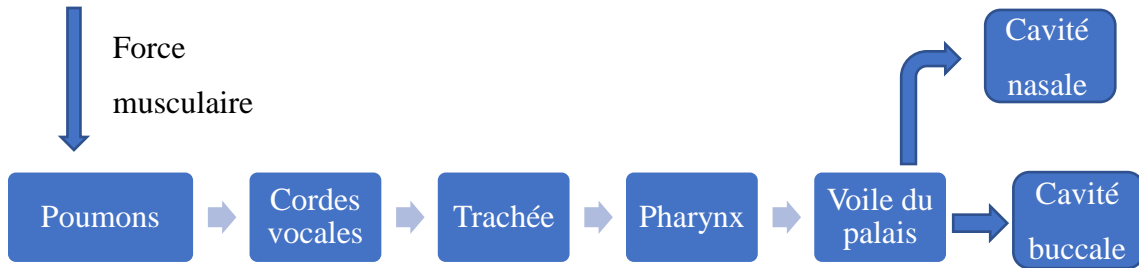


Figure I.2 : Modèle mécanique de production de la parole.

I.2.3 Modèle numérique de production de la parole

Une modélisation exhaustive de la production de la parole est très difficile et, pour des raisons pratiques, inefficace. L'idée de base dans la modélisation numérique est d'arriver à un modèle linéaire qui produit en sortie un signal équivalent au signal vocal. Le modèle est correct dans la mesure où sa sortie s'approche du signal vocal sans modéliser les phénomènes physiques intrinsèques à la production du signal vocal [14]. La figure I.3 présente un tel modèle général qui est utilisé dans le traitement numérique de la parole.

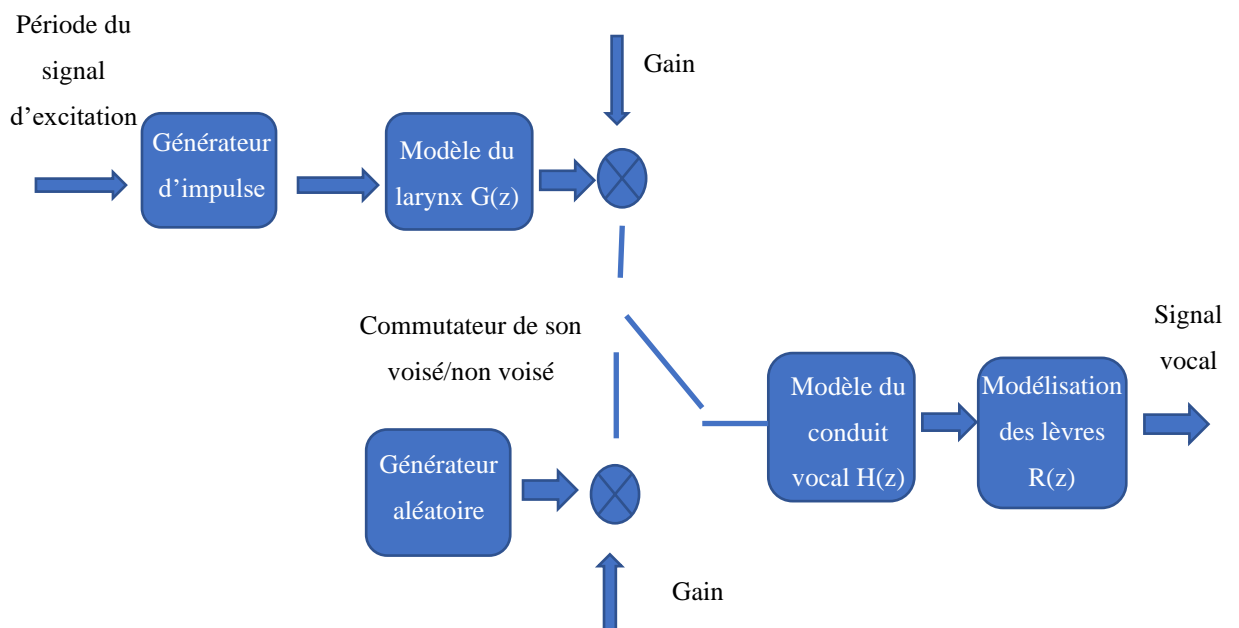


Figure I.3 : Modèle numérique de production de la parole.

Dans ce modèle général, on utilise deux sources d'excitation. Pour les sons non voisés, la source d'excitation est un bruit blanc. Pour la production des sons voisés, la source d'excitation est un train périodique d'impulsions qui traverse un filtre passe bas d'ordre 2 de fonction de transfert $G(z)$ donnée par [12] :

$$G(z) = \frac{A}{(1+b_1z^{-1})(1+b_2z^{-1})} \quad (\text{I.1})$$

Le conduit vocal peut être modélisé par une succession de tubes acoustiques élémentaires. Chaque tube ou résonateur mécanique est assimilé à un filtre numérique d'ordre 2. La transmittance globale du modèle est de la forme [12] :

$$H(z) = \frac{B}{\prod_{k=1}^K (1+b_{1k}z^{-1}+b_{2k}z^{-2})} \quad (\text{I.2})$$

Au bout du conduit vocal le son passe à travers l'ouverture des lèvres. Celles-ci sont vues comme une composante qui transforme le débit volumique dans une onde de pression à une certaine distance. Dans le domaine spectral ; le rayonnement des lèvres a l'effet d'un filtrage passe haut. Le plus simple filtre numérique qui a cette propriété est [12] :

$$R(z) = C(1-z_0z^{-1}) \quad (\text{I.3})$$

En conclusion la fonction de transfert globale est de la forme [13] :

$$T(z) = \begin{cases} H(z)R(z) & \text{pour les sons non voisés} \\ G(z)H(z)R(z) & \text{pour les sons voisés} \end{cases} \quad (\text{I.4})$$

I.2.4 Classification des sons de la parole

La distinction la plus fondamentale entre les types de sons dans la parole est la distinction entre les sons voisés et les sons non voisés. Les sons voisés, y compris les voyelles, sont

caractérisés par une structure temporelle et fréquentielle plus ou moins régulière que les sons non voisés [15].

I.2.4.1 Les sons voisés

Les sons voisés sont généralement produits par le passage de l'air des poumons à travers la trachée, qui met en vibration les cordes vocales, ces vibrations sont provoquées par les rapprochements des deux cordes vocales, c'est-à-dire l'air sous pression essaye de sortir et à chaque ouverture, il passe et il fait bouger les cordes vocales qui se ferment puis s'ouvrent de nouveau et ainsi de suite.

Le son voisé possède un caractère périodique et des propriétés particulières dues à la forme de l'onde de débit glottique. Les sons voisés sont généralement quasi-périodiques [16]. Ce type de sons est caractérisé par une énergie plus élevée que les sons non voisés.

Au niveau du spectrogramme, les voyelles, ont une structure temporelle et fréquentielle plus régulière que les sons non voisés (Figure I.4) [15].

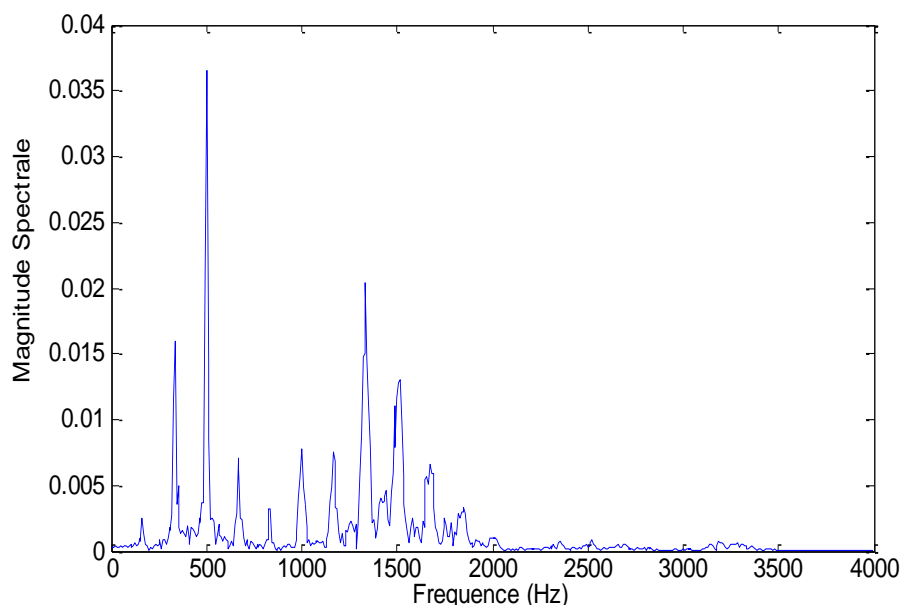


Figure I.4 : Spectre d'un son voisé.

I.2.4.2 Les sons non voisés

Les sons non voisés sont de nature apériodique (Figure I.5), ils sont produits par un air turbulent qui traverse les cordes vocales alors que celles-ci sont écartées. L'énergie de ce

type de sons est concentrée dans les hautes fréquences et correspond à du bruit [17]. La figure I.6 montre une forme d'onde incluant un son voisé et un son non voisé.

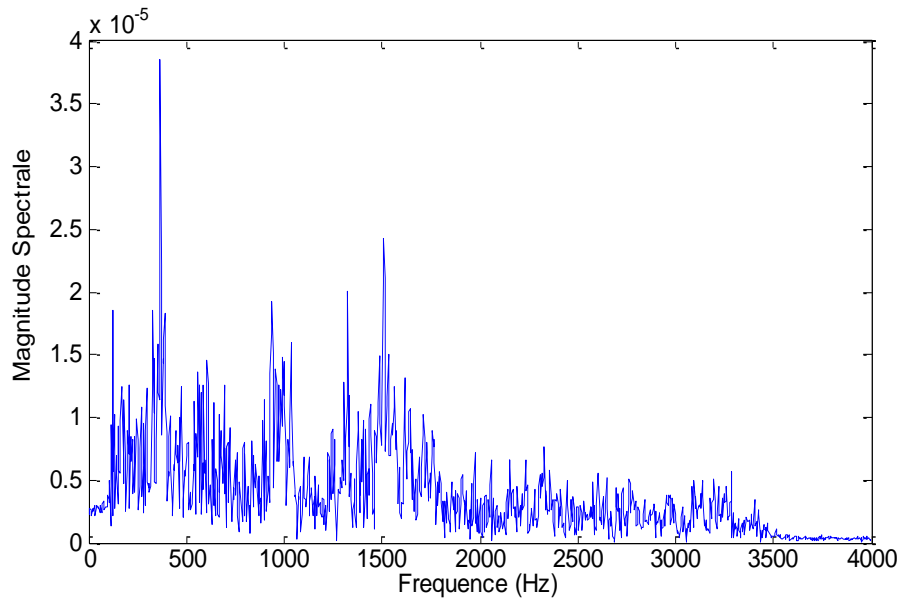


Figure I.5 : Spectre d'un son non voisé.

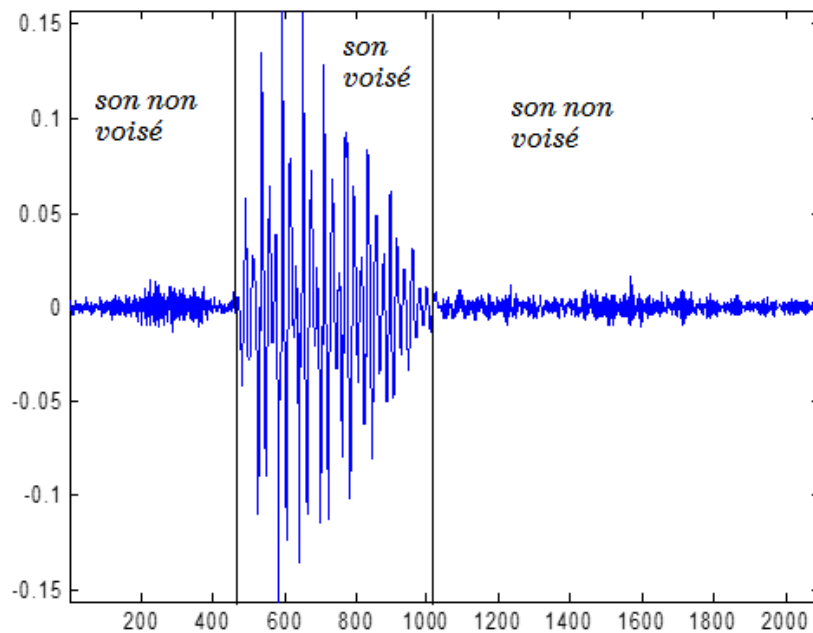


Figure I.6 : Forme d'onde incluant un son voisé et un son non voisé.

I.2.5 Propriétés du signal de la parole

I.2.5.1 Fréquence fondamentale

Pour le signal de parole, sa fréquence fondamentale F_0 n'est rien d'autre que la fréquence du cycle d'ouverture/fermeture des cordes vocales, déterminée par la tension des muscles qui contrôlent celles-ci. La plage de variation moyenne de cette fréquence varie d'un locuteur à un autre en fonction de son âge et de son sexe. Elle s'étend approximativement de 80 Hz à 200 Hz chez les hommes, de 150 Hz à 450 Hz chez les femmes, et de 200 Hz à 600 Hz chez les enfants. La détection de F_0 joue un rôle essentiel dans le domaine de traitement de la parole et doit être, si possible, pour les applications modernes, calculée en temps réel [18].

I.2.5.2 Energie

Elle est représentée par l'intensité du son qui est liée à la pression de l'air en amont du larynx. L'amplitude du signal de la parole varie au cours du temps selon le type de son, et son énergie dans une trame est donnée par :

$$E = \sum_{n=0}^{N-1} s^2(n) \quad (\text{I.5})$$

où N est la taille ou la longueur de la trame [19].

I.2.5.3 Enveloppe spectrale

L'enveloppe spectrale ou spectre représente l'intensité de la voix en fonction de la fréquence, elle est généralement obtenue par une analyse de Fourier à court terme.

La quasi stationnarité du signal de parole permet de mettre en œuvre des méthodes efficaces d'analyse et de modélisation utilisées pour le traitement à court terme du signal vocal sur des fenêtres de durée généralement comprise entre 20 ms et 30 ms appelées trames, avec un recouvrement entre ces fenêtres qui assure la continuité temporelle des caractéristiques de l'analyse.

La transformée de Fourier à court terme (TFCT) d'un signal échantillonné est par définition la transformée du signal pondéré.

$$\hat{S}(k) = \hat{S}\left(f = \frac{k}{N}\right) = \sum_{n=0}^{N-1} s(n) \cdot w(n) \cdot \exp(-j2\pi nk/N) \quad 0 \leq k \leq N-1 \quad (\text{I.6})$$

où N est le nombre de points prélevés, $s(n)$ est le segment analysé, $w(n)$ est la fenêtre de pondération et $S(k)$ est le spectre complexe.

Le spectre de puissance (appelé aussi densité spectrale de puissance) est donné par [20] :

$$|\hat{S}(k)|^2, 0 \leq k \leq N/2. \quad (\text{I.7})$$

I.3 Les bruits

Les bruits sont des perturbations qui peuvent altérer la qualité de la communication, dénaturer le message communiqué et rendre difficilement perceptible l'information. Ils constituent donc une gêne dans la compréhension du signal utile, qui est dans notre cas, la parole [20].

I.3.1 Bruit blanc

Le bruit blanc est un bruit composé de toutes les fréquences au même niveau statistique. Il présente la même énergie pour toutes les fréquences [21].

I.3.2 Bruits colorés

Dans le cas où la densité spectrale de puissance (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 par leurs spectres. Il existe plusieurs types de bruits colorés, comme le bruit rose et le bruit brun [22].

I.3.3 Bruit musical

Le bruit musical est un bruit résiduel perpétuellement 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 particulièrement tonal, d'où le caractère musical [20].

I.3.4 Bruit impulsif

Comme son nom l'indique, ce type de bruit est à caractère impulsif, il se présente sous forme de tensions perturbatrices de valeur élevée mais de durée brève. Ces bruits sont très gênants pour la transmission des données, car le signal perturbateur modifie la forme du signal reçu à des instants quelconques (aléatoires) [23].

I.3.5 Bruit ambiant

Le bruit ambiant est la somme du bruit résiduel et du bruit particulier émis par la source. Il est composé de l'ensemble des bruits émis par toutes les sources proches et lointaines [23].

I.3.6 Bruit acoustique

Il est généré par les mouvements des sources telles que les voitures, les ventilateurs, la circulation, le vent, la pluie, ... etc [21].

I.3.7 Les modèles du bruit

Pour simuler ou analyser des signaux réels, d'autres modèles, basés sur la densité de probabilité du bruit, peuvent être considérés. Ainsi, en particulier, deux modèles du bruit sont assez répandus :

- Le bruit dit gaussien dont la densité de probabilité a une répartition de type gaussien caractérisée par une valeur moyenne et un écart type.
- Le bruit dit périodique formé d'une somme de signaux sinusoïdaux sans référence de phase.

I.4 Classification des méthodes de rehaussement de la parole

Les méthodes de débruitage de la parole peuvent être classées de plusieurs façons (Tableau I.1). Elles peuvent être mono-microphones ou multi-microphones selon le nombre de microphones utilisés. Elles peuvent également être temporelles ou fréquentielles en fonction du domaine de traitement. La troisième méthode de classification peut être basée sur le type d'algorithme utilisé (adaptatif ou non adaptatif) [24].

Tableau I.1 : Classification des méthodes de rehaussement de la parole

Algorithme	Adaptif ou non adaptif
Nombre de microphones	Un seul ou plusieurs
Domaine de traitement	Temporel ou fréquentiel

I.5 Etat de l'art des méthodes de rehaussement de la parole

Une des premières techniques de rehaussement de la parole dans le domaine spectral est la soustraction spectrale introduite par Boll en 1979 [3]. La soustraction spectrale consiste à atténuer plus ou moins les composantes spectrales du signal bruité en fonction de l'estimation du niveau de bruit. Un des gros désavantages de la soustraction spectrale telle que présentée est que l'estimation du bruit induit un nouveau bruit qui contient une certaine musicalité, nommé "bruit musical" [25]. Virag [26] a présenté une généralisation de la soustraction spectrale, qui est faite par le mixage des algorithmes de Boll [3] et de Bertouli [27] avec la généralisation de Lim [28]. Cette généralisation a pour intérêt de trouver un compromis entre la réduction du bruit et la distorsion du signal amélioré [29].

Le filtre de Wiener est parmi les méthodes de débruitage classiques les plus utilisées dans la littérature. Il a été introduit pour essayer d'améliorer la qualité de la trace recueillie dans les potentiels évoqués. Le problème de ce type de filtrage est qu'il n'est pas applicable sur une moyenne d'acquisition mais pour chaque trace. Doyle [30], propose une modification pour pouvoir l'adapter à la moyenne (le calcul dans ce cas est beaucoup plus rapide), il n'y a pas besoin de filtrer chaque trace. Cependant, il faut considérer que le bruit est stationnaire dans ce cas. Paliwal [31] a proposé une méthode de filtrage de Wiener non stationnaire pour le rehaussement de la parole, où le filtre est conçu pour chaque segment de parole de courte durée en utilisant une procédure des moindres carrés.

Le débruitage par ondelettes a connu un grand succès. Une ondelette est une fonction qui oscille comme une onde mais qui est rapidement atténuée d'où son nom ondelette qui veut dire petite onde [32]. Le succès de l'analyse en ondelettes est dû à sa relation avec l'analyse multi-résolution développée par Stéphane Mallat [33] et Yves Meyer [34]. Cela a permis aux chercheurs et aux mathématiciens de construire leurs propres familles d'ondelettes [35-37].

I.5.1 Soustraction spectrale

Nous allons introduire dans cette section, l'algorithme de Boll [3] considéré comme une méthode de soustraction spectrale de référence. Soit un signal bruité $x(t)$ composé d'un signal propre $s(t)$ corrompu par un bruit additif $n(t)$ non corrélé à $s(t)$. Le signal bruité peut être exprimé par :

$$x(t) = s(t) + n(t) \quad (\text{I.8})$$

Nous désignons par $\hat{s}(t)$ le signal rehaussé qui est une estimation de $s(t)$. Nous supposons que le signal de parole est quasi-stationnaire sur des fenêtres d'analyse de 20 ms à 30 ms. La transformée de Fourier à court terme (TFCT) de $x(t)$ est donnée par :

$$X(\omega) = S(\omega) + N(\omega) \quad (\text{I.9})$$

où ω est la fréquence angulaire et $X(\omega)$, $S(\omega)$ et $N(\omega)$ désignent respectivement le spectre de $x(t)$, $s(t)$ et $n(t)$.

$X(\omega)$ peut être exprimé sous forme polaire comme suit :

$$X(\omega) = |X(\omega)| e^{i\Phi_x(\omega)} \quad (\text{I.10})$$

$|X(\omega)|$ et $\Phi_x(\omega)$ représentent, respectivement, l'amplitude et la phase de $X(\omega)$.

On peut également exprimer le bruit par $N(\omega) = |N(\omega)| e^{i\Phi_n(\omega)}$. L'amplitude du bruit $|N(\omega)|$ est inconnue, ce qui nous conduit donc à la remplacer par sa valeur moyenne, estimée durant la période d'absence d'activité vocale.

Dans les méthodes de soustraction spectrale, il est supposé que le bruit n'affecte pas la phase du signal. Pour cela, on considère uniquement l'amplitude spectrale à court terme du bruit. La valeur estimée de $\Phi_n(\omega)$ sera donc remplacée par $\Phi_x(\omega)$. En remplaçant $X(\omega)$ et $N(\omega)$ dans (I.10), la valeur estimée du spectre du signal propre est donnée par :

$$\hat{S}(\omega) = [|X(\omega)| - |\tilde{N}(\omega)|] e^{i\Phi_x(\omega)} \quad (\text{I.11})$$

où $|\tilde{N}(\omega)|$ est la valeur estimée du spectre d'amplitude du bruit calculée durant la période d'absence d'activité vocale. Le signal rehaussé peut être obtenu en calculant la transformée de Fourier inverse de $\hat{S}(\omega)$. On peut remarquer à partir de (I.11) que $|\hat{S}(\omega)| = |X(\omega)| - |\tilde{N}(\omega)|$ peut prendre des valeurs négatives en raison des erreurs d'estimation du bruit. Plusieurs solutions sont envisageables pour remédier à ce problème [3],[27]. Parmi ces techniques, on peut citer la rectification demi-onde qui permet de mettre à zéro ces valeurs négatives, comme suit :

$$\begin{cases} |\hat{S}(\omega)| = |X(\omega)| - |\tilde{N}(\omega)| & \text{si } X(\omega) > \tilde{N}(\omega) \\ 0 & \text{si } X(\omega) \leq \tilde{N}(\omega) \end{cases} \quad (\text{I.12})$$

L'algorithme de soustraction spectrale en amplitude (I.12) peut être étendu à la soustraction spectrale de puissance en multipliant $X(\omega)$ donné par (I.9) par sa conjuguée $X^*(\omega)$:

$$\begin{aligned} |X(\omega)|^2 &= |S(\omega)|^2 + |N(\omega)|^2 + S^*(\omega)N(\omega) + S(\omega)N^*(\omega) \\ &= |S(\omega)|^2 + |N(\omega)|^2 + 2 \operatorname{Re} \{ S^*(\omega)N(\omega) \} \end{aligned} \quad (\text{I.13})$$

Si nous supposons que $n(t)$ et $s(t)$ sont à moyennes nulles et non corrélés, les termes $E\{S(\omega).N^*(\omega)\}$ et $E\{S^*(\omega).N(\omega)\}$ deviennent nuls. La valeur estimée du spectre de puissance du signal propre devient :

$$|\hat{S}(\omega)|^2 = |X(\omega)|^2 - |N(\omega)|^2 \quad (\text{I.14})$$

L'équation (I.14) décrit l'algorithme de soustraction spectrale en puissance et peut être exprimée par :

$$\begin{aligned} |\hat{S}(\omega)|^2 &= H^2(\omega) |X(\omega)|^2 \\ H(\omega) &= \sqrt{1 - \frac{|N(\omega)|^2}{|X(\omega)|^2}} \end{aligned} \quad (\text{I.15})$$

$H(\omega)$ désigne le gain de suppression ou la fonction de suppression [38].

I.5.2 Filtrage de Wiener

On considère un modèle de formation des données de type convolution. On peut donc écrire :

$$x(t) = (h*s)(t) + n(t) \quad (\text{I.16})$$

où x et h représentent respectivement le signal observé et la réponse impulsionnelle du système, supposés connus. On considère que le bruit d'observation $n(t)$ et la quantité d'intérêt $s(t)$ sont aléatoires, indépendantes l'une de l'autre, centrées, stationnaires d'ordre 2 et caractérisées respectivement par leurs fonctions d'autocorrélation $R_n(\tau)$ et $R_s(\tau)$ ou, de manière équivalente, par leurs densités spectrales de puissance respectives $\Gamma_n(\omega)$ et $\Gamma_s(\omega)$. Tous les signaux sont supposés à valeurs réelles.

L'objectif est d'estimer $s(t)$ avec un estimateur linéaire de la forme :

$$\hat{s}(t) = (w*x)(t) \quad (\text{I.17})$$

selon un critère d'erreur quadratique moyenne minimale. On cherche donc à minimiser la quantité :

$$J \triangleq E [(s(t) - \hat{s}(t))^2] \quad (\text{I.18})$$

En raison de la linéarité de l'estimateur et du fait que, de par la structure de l'estimateur, $s(t)$ est fonction de l'ensemble des valeurs $\{x(u) ; u \in \mathbb{R}\}$, le principe d'orthogonalité s'écrit :

$$\forall u \in \mathbb{R} \quad E [(s(t) - \hat{s}(t)) x(t+u)] = 0 \quad (\text{I.19})$$

En raison de l'indépendance entre $x(t)$ et $n(t)$, on a :

$$\begin{aligned} E[s(t)x(t+u)] &\triangleq R_{sx}(u) = E[s(t)(h*s)(t+u)] \\ &= \int h(\tau) E[s(t)s(t+u-\tau)] d\tau \\ &= \int h(\tau) R_s(u-\tau) d\tau \end{aligned} \quad (\text{I.20})$$

De manière similaire $E[\hat{s}(t)x(t+u)]$ est une fonction de u seulement qui s'exprime comme :

$$\begin{aligned} E[\hat{s}(t)x(t+u)] &= E[(w*x)(t)x(t+u)] \\ &= \int w(\tau)E[x(t+u)x(t-\tau)]d\tau \\ &= \int w(\tau)R_x(u+\tau)d\tau \end{aligned} \quad (I.21)$$

En faisant le changement de variable $\tau \rightarrow -\tau$ dans la dernière intégrale, on obtient :

$$E[\hat{s}(t)x(t+u)] = \int w(-\tau)R_x(u-\tau)d\tau \quad (I.22)$$

D'après (I.19), les expressions (I.20) et (I.22) sont égales. En passant au domaine fréquentiel, on obtient donc l'identité :

$$H(\omega)\Gamma_s(\omega) = W^*(\omega)\Gamma_x(\omega) \quad (I.23)$$

Ce qui conduit à :

$$W(\omega) = \frac{H^*(\omega)\Gamma_s(\omega)}{\Gamma_x(\omega)} \quad (I.24)$$

où l'on a utilisé le fait que la densité spectrale de puissance d'un processus à valeurs réelles est réelle et paire. Il suffit maintenant d'utiliser le fait que $\Gamma_x(\omega)$ s'exprime comme :

$$\Gamma_x(\omega) = |H(\omega)|^2\Gamma_s(\omega) + \Gamma_n(\omega) \quad (I.25)$$

pour aboutir à l'expression classique du filtre de Wiener dans le domaine spectral [4] :

$$W(\omega) = \frac{H^*(\omega)}{|H(\omega)|^2 + \left(\frac{\Gamma_n(\omega)}{\Gamma_s(\omega)}\right)} \quad (I.26)$$

I.5.3 Débruitage par ondelettes

L'ondelette est une oscillation en forme d'onde avec une amplitude qui commence à zéro, elle augmente, puis diminue jusqu'à zéro, elle fournit une analyse du signal qui est localisée à la fois en temps et en fréquence [39]. Cette analyse temps-fréquence la mène à appartenir au groupe de méthodes d'analyse multi-échelles. Le principe de base consiste à convoluer le signal à analyser avec une fonction appelée ondelette (wavelet) [40]. La fonction d'ondelette est exprimée par

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right) \quad (\text{I.27})$$

Le paramètre d'échelle ou de dilatation 'a' correspond à une information de fréquence, et le paramètre 'b' se rapporte à l'emplacement de la fonction d'ondelettes comme s'il est décalé à travers le signal.

La figure I.7 montre deux exemples d'ondelettes appelées chapeau mexicain et ondelette de Morlet.

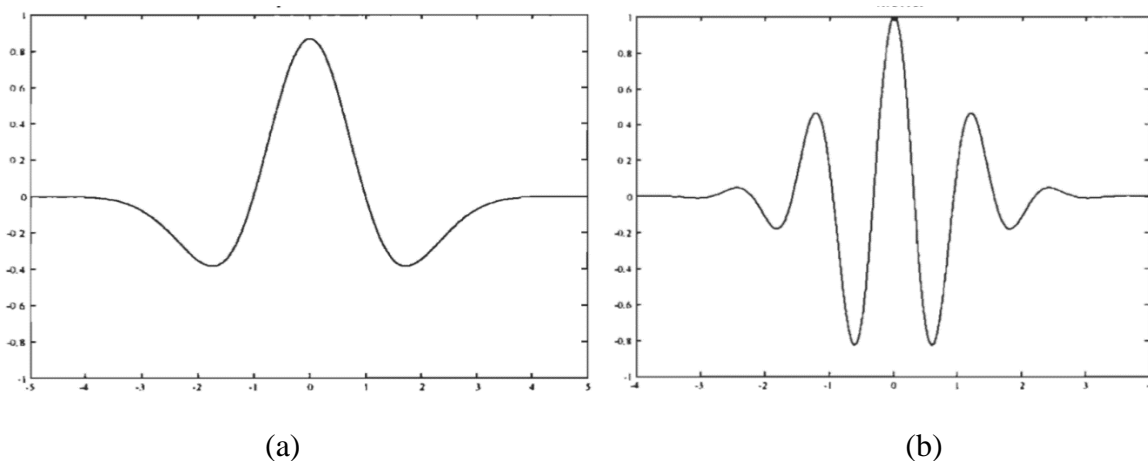


Figure I.7: Exemple d'ondelettes : (a) Chapeau mexicain (mexican hat) (b) Ondelette de Morlet.

I.5.3.1 Transformation en ondelettes continue

La transformée en ondelettes continue (CWT : continuous wavelet transform) de la fonction $g(t)$ (tel que $g(t) \in L_2$) est défini comme :

$$G(a,b) = \frac{1}{\sqrt{a}} \int \Psi^* \left(\frac{t-b}{a} \right) g(t) dt \quad (\text{I.28})$$

$\Psi^*(t)$ est la fonction conjuguée de $\Psi(t)$. L'intégral dans (I.28) est une convolution entre le signal $g(t)$ et la fonction $\Psi^*(t)$.

Un inconvénient du CWT est que la représentation du signal est souvent redondante à cause de la continuité de a et b sur L_2 [41].

I.5.3.2 Transformation en ondelettes discrète

La transformée en ondelettes discrète (DWT : discrete wavelet transform) décompose le signal en coefficients d'ondelettes à partir desquels le signal d'origine peut être reconstitué. Les coefficients d'ondelettes représentent le signal dans différentes bandes de fréquences. Ils peuvent être traités de plusieurs manières, ce qui donne au DWT des propriétés intéressantes par rapport au filtrage linéaire [42].

La transformée en ondelettes discrète du signal $x(t)$ est définie comme :

$$\text{DWT}_x(m, n) = \int x(t) \Psi_{m,n}^*(t) dt \quad (\text{I.29})$$

Tel que $a=2^{-m}$ et $b=n2^{-m}$ (m et $n \in \mathbb{Z}$, \mathbb{Z} est l'ensemble des entiers positifs).

$\Psi_{m,n}^*(t) = 2^{-m} \Psi(2^m t - n)$ est la version dilatée et translatée de l'ondelette mère $\Psi(t)$ [17].

I.5.3.3 Propriétés des ondelettes

Les propriétés les plus importantes des ondelettes sont [34] :

- Support compact
- Symétrie
- Nombre de moments nuls
- Régularité

a) Support compact

La plupart des ondelettes sont à support compact dans le domaine temporel, ce qui veut dire qu'elles sont à durée finie (filtres FIR) et se distinguent par leur atténuation rapide. Un support compact permet une complexité de calcul réduite, une meilleure résolution dans le domaine temporel donne une résolution faible en fréquence. Comme exemple, on peut citer les ondelettes de Daubechies, Symlets, Coiflets [34], [36], [43], etc. Par dualité, les ondelettes à bande étroite sont des ondelettes à support compact dans le domaine fréquentiel mais pas dans le domaine temporel. Les ondelettes de Meyer en sont un exemple.

b) Symétrie

Les ondelettes symétriques [34] donnent naissance à des filtres à phase linéaire. Daubechies [36], [43] a montré que, pour qu'une ondelette soit symétrique, le filtre h doit être à phase complexe linéaire, et que le seul filtre miroir conjugué symétrique à support fini est le filtre de Haar [44] qui correspond à une ondelette discontinue à un seul moment nul. Mis à part l'ondelette de Haar, il n'y a donc pas d'ondelette réelle orthogonale symétrique à support compact.

C'est une propriété qui permet d'éviter des déphasages difficiles à interpréter ultérieurement.

c) Moments nuls

Par définition, une fonction $\Psi(t)$ possède p moments nuls si et seulement si elle vérifie [34] :

$$\int_{-\infty}^{\infty} t^k \Psi(t) dt = 0 \quad \text{Pour } 0 \leq k \leq p \quad (\text{I.30})$$

L'intérêt de cette propriété est que si une ondelette a un nombre de moments nuls suffisant, on obtiendra alors plus de coefficients à petites valeurs. Elle permet de caractériser aussi l'ordre des singularités d'un signal. Si nous désignons par p le nombre de moments nuls, alors la taille du support est d'au moins de $2p-1$.

d) La régularité

La régularité d'une ondelette [34] est la propriété permettant de localiser les singularités dans un signal. Cette propriété se traduit sur les coefficients d'ondelettes par une amplitude importante caractérisant une singularité dans le signal par la décroissance des coefficients avec l'échelle de résolution. La régularité est une propriété importante pour obtenir des coefficients d'ondelettes les plus petits possibles afin de les annuler.

I.5.3.4 Algorithme de seuillage

Le seuillage des coefficients d'ondelettes permet de conserver la régularité locale du signal. En général, le seuillage des ondelettes comprend trois étapes [45] :

(i) Décomposition

L'algorithme de décomposition du signal en coefficients d'ondelette est basé sur un banc de filtre. Le banc de filtre décompose le signal discret dans une base orthogonale discrète. La fonction d'ondelette $\Psi_{j,k}(n)$ et la fonction d'échelle $\Phi_{j,k}(n)$, appartiennent à la base orthogonale B

$$B = [\{\Psi_{j,k}[n]\}_{L < j \leq J, 0 \leq k < 2^{-j}}, \{\Phi_{j,k}[n]\}_{0 \leq k < 2^{-j}}] \quad (\text{I.31})$$

Le paramètre 2^j varie de $2^L = N^{-1}$ à $2^J < 1$, où N est la fréquence d'échantillonnage du signal x.

(ii) Seuillage

Après la décomposition du signal, un seuil est choisi. Un estimateur de seuillage dans la base B s'écrit comme suit :

$$F = \sum_{j=L+1}^J \sum_{k=0}^{2^{-j}} p_t (< X, \Psi_{j,k} >) \Psi_{j,k} + \sum_{k=0}^{2^{-j}} p_t (< X, \Phi_{j,k} >) \Phi_{j,k} \quad (\text{I.32})$$

où p_t est soit un seuil dur, soit un seuil doux. Normalement, le seuil sélectionné est appliqué à tous les coefficients, à part les coefficients contenant l'énergie de fréquence la plus basse. Ceci a pour but de conserver la régularité du signal reconstruit.

a) Seuillage dur (hard)

Le seuillage dur propose l'annulation de toutes les valeurs en dessous d'un seuil S , les valeurs supérieures restent inchangées [46]. Le seuillage dur s'exprime comme suit (Figure I.8) :

$$y_{\text{dur}}(x,T) = \begin{cases} 0 & \text{si } |x| < T \\ x & \text{si } |x| \geq T \end{cases} \quad (\text{I.33})$$

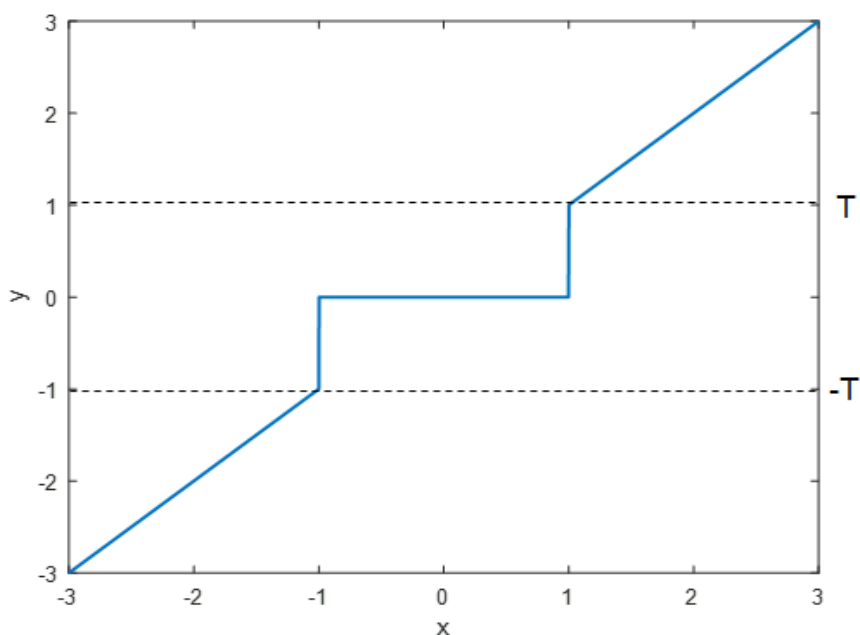


Figure I.8 : Seuillage dur.

b) Seuillage doux

Le seuillage doux permet en plus de l'annulation, une soustraction du seuil T aux valeurs restantes, supérieures au seuil, afin de réduire le nombre de discontinuités dans le signal débruité [46]. Le seuillage doux s'exprime comme suit (Figure I.9) :

$$y_{\text{doux}}(x,T) = \begin{cases} 0 & \text{si } |x| \leq T \\ x - T & \text{si } |x| > T \\ x + T & \text{si } |x| > -T \end{cases} \quad (\text{I.34})$$

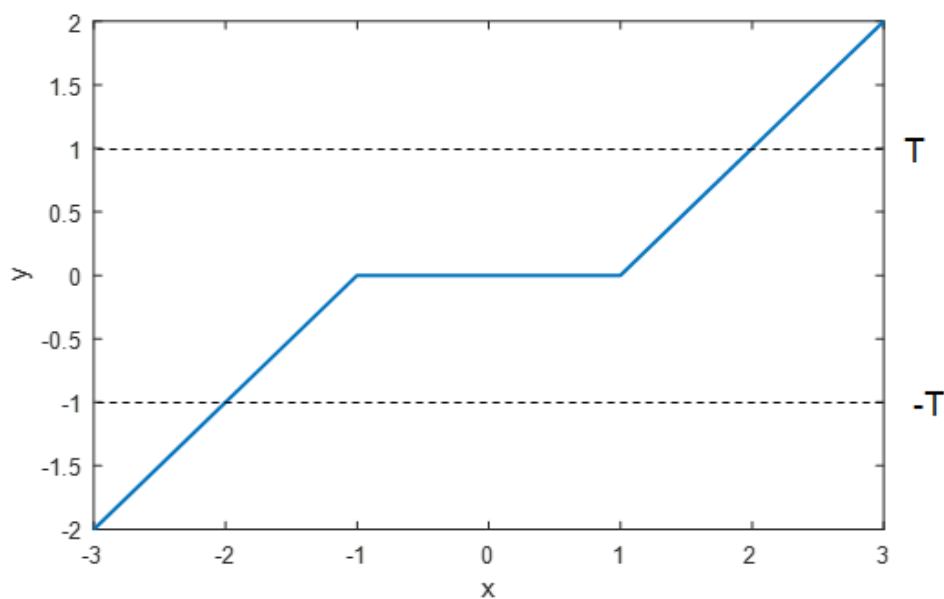


Figure I.9 : Seuillage doux.

(iii) Reconstruction

Après le seuillage, tous les coefficients sont reconstruits pour former le signal débruité [45].

I.6 Evaluation des performances des méthodes de rehaussement

Plusieurs indices sont utilisés pour évaluer et comparer les performances des méthodes de rehaussement de la parole.

I.6.1 PESQ (évaluation perceptive de la qualité de la parole)

C'est l'un des plus fiables méthodes pour évaluer les performances des méthodes de rehaussement de la parole. Il permet de détecter la dégradation du signal. La plage de PESQ se situe entre 0.4 et 4.5, avec les valeurs les plus basses interprétées comme une mauvaise qualité de la parole [47].

I.6.2 MOS (Note d'opinion moyen)

MOS est une mesure subjective, définie pour exprimer la qualité des médias. Il résulte des avis des utilisateurs lors des tests d'évaluation de la qualité. En outre, il peut être utilisé pour la comparaison de différents services ou pour la comparaison des performances des codecs dans différentes conditions du réseau.

Les valeurs du MOS vont de 1 à 5, 1 étant le pire et 5 le meilleur (Tableau I.2). Non seulement les nombres entiers sont représentatifs, mais les nombres décimaux sont aussi utilisés. Pour exemple, les valeurs inférieures à 3 sont considérées comme inacceptables par de nombreux utilisateurs [48].

Tableau I.2 : Valeurs de MOS et la qualité correspondante à chaque valeur

MOS	Qualité
5	Excellent
4	Bien
3	Agaçant
2	Très agaçant
1	Impossible de communiquer

I.6.3 Rapport signal sur bruit

La mesure du rapport signal sur bruit (SNR : signal-to-noise ratio) que l'on notera par la suite SNR est donnée par :

$$\text{SNR} = 10 \log_{10} \frac{\sum_{i=0}^{N-1} s^2}{\sum_{i=0}^{N-1} (s(i) - \hat{s}(i))^2} \quad (\text{I.35})$$

La qualité du signal estimé est d'autant meilleure que le SNR est grand.

Lors du calcul du SNR, les segments de faible amplitude contribue peu dans la valeur finale. Pour remédier à cet inconvénient, le SNR segmental est utilisé. Le SNR segmental est défini comme étant la moyenne des SNRs de segments de courte durée (10 à 20 ms) :

$$\text{SNR}_{\text{seg}} = \frac{10}{M} \sum_{m=0}^{M-1} 10 \log_{10} \frac{\sum_{i=mN}^{mN+N-1} s^2}{\sum_{i=mN}^{mN+N-1} (s(i) - \hat{s}(i))^2} \quad (\text{I.36})$$

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

Le SNR segmental souffre de deux limitations : d'abord si le signal de parole contient des segments de silence, ce qui est très probable, le $s(i)$ sera nul et n'importe quelle quantité de bruit entraînera un SNR en dB négatif pour ce segment ; du coup le SNR total sera biaisé par cette quantité. Ce problème peut être résolu partiellement en choisissant un seuil d'énergie au-delà duquel le SNR segmental sera calculé. Ensuite, il faut nécessairement que les deux signaux comparés soient alignés temporellement car ce critère est très sensible aux déphasages [49].

I.7 Conclusion

Dans ce chapitre, nous avons présenté les notions de base liées au mécanisme de production de la parole et les différents types de bruits qui l'affecte. Pour réduire le bruit qui vient corrompre le signal, on a présenté les différentes méthodes de rehaussement de la parole (le filtre de Wiener, soustraction spectrale et le débruitage par ondelettes). Nous avons aussi présenté les différents indices pour évaluer les performances des méthodes de rehaussement et la qualité du signal estimé.

Chapitre II

L'acquisition comprimée

II.1 Introduction

L'acquisition comprimée (CS : compressed sensing) est une nouvelle théorie visant à optimiser le processus d'acquisition du signal. L'idée principale derrière la technique CS réside dans la possibilité de reconstruire le signal en utilisant un nombre d'échantillons beaucoup plus faible que celui déterminé par le théorème de Shannon-Nyquist [50]-[51]. Le théorème de Shannon-Nyquist exige que les échantillons soient prélevés d'un signal à bande limitée à des intervalles de temps équidistants au plus de $1/2f_{\max}$, où f_{\max} est la fréquence maximale contenue dans le signal [50]. Sous certaines conditions, la technique de CS réduit le taux d'échantillonnage et évite une compression ultérieure du signal. Un plus petit nombre d'échantillons signifie un traitement et une transmission plus rapides du signal.

Outre la sélection aléatoire des échantillons du signal, la CS nécessite que le signal soit parcimonieux dans un certain domaine de transformation. Les signaux parcimonieux sont caractérisés par des informations importantes contenues dans un petit nombre de coefficients non nuls dans le domaine de parcimonie [50], [52]. Selon le type de signal, divers domaines où les signaux ont une représentation parcimonieuse peuvent être utilisés. La transformée de Fourier discrète (DFT : discrete Fourier transform), la transformée en cosinus discrète (DCT : discrete cosine transform) et la transformée en ondelettes (WT : wavelet transform) sont des exemples de transformations utilisées par la CS. Les observations (échantillons) sont prélevées aléatoirement dans le domaine où le signal a une représentation dense. Le prélèvement aléatoire des échantillons assure la propriété d'incohérence à satisfaire. L'incohérence et la parcimonie sont deux exigences qui doivent être satisfaites pour pouvoir appliquer l'approche de CS. L'acquisition comprimée est l'un des domaines émergents les plus attrayants de ces dernières années. Elle a montré des résultats prometteurs dans de nombreuses applications, telles que l'estimation de canal ou de spectre [14], [53], les radiocommunications cognitives [54], la détection de motifs de couleurs [55, 56] et la détection centrée sur les personnes [57].

Le but de ce chapitre est de fournir les concepts de base de la théorie de l'acquisition comprimée et ces différentes étapes.

II.2 Principe de l'acquisition comprimée

La théorie de CS affirme que l'on peut récupérer certains signaux et images à partir d'un nombre d'échantillons ou de mesures plus petit que celui utilisé dans les méthodes traditionnelles. Le principe de base de l'acquisition comprimée est illustré à la Figure II.1.

Il se compose de deux parties principales : l'émetteur et le récepteur. Le signal d'entrée côté émetteur x est composé de N échantillons. Tout d'abord, x doit être transformé dans un domaine tel que la DFT ou la DCT dans lequel il a une représentation parcimonieuse. Le signal transformé devient alors K -parcimonieux, où K est le nombre des plus grands coefficients significatifs obtenus par seuillage. Ces K plus grands coefficients contiennent la plupart des informations sur le signal. Ensuite, il est multiplié par la matrice de détection Φ . Du côté du récepteur, différentes techniques d'optimisation sont utilisées pour la reconstruction du signal d'origine. La première étape est la multiplication du signal par la matrice de détection ce qui résulte en N échantillons à partir de M mesures. Ensuite, des techniques d'optimisation convexe sont utilisées pour récupérer le signal K -parcimonieux. Une fois de plus, la parcimonie inverse est appliquée pour obtenir le signal original [58].

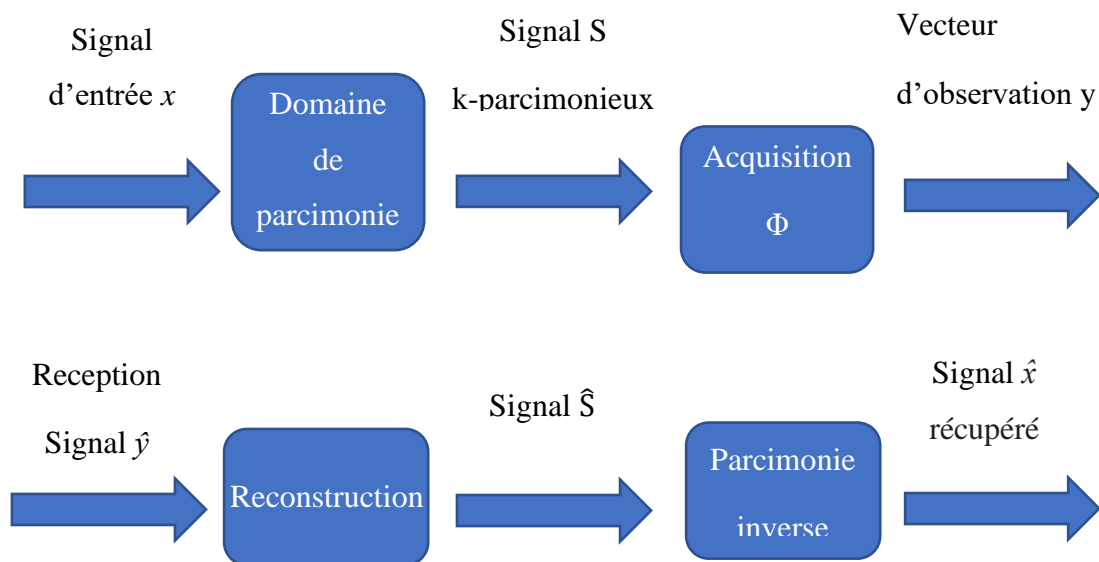


Figure II.1 : Principe de base de l'acquisition comprimée.

Le processus de CS se compose de trois parties principales :

- Représentation parcimonieuse.
- Acquisition utilisant une matrice de détection pour obtenir les mesures CS.
- Reconstruction à l'aide d'un algorithme de récupération pour récupérer le vecteur parcimonieux.

II.3 Parcimonie

Quel que soit le domaine d'application et la technique utilisée, l'optimisation des résultats nécessite de tirer profit de toute l'information a priori. Ces dernières années, la parcimonie a émergé comme un a priori fondamental. La parcimonie est reliée à la notion d'épargne. Dans le domaine informatique, ce terme est appliqué aux vecteurs et aux matrices dont quasiment tous les éléments sont égaux et dont quelques éléments ont une valeur non-standard. Une matrice est alors parcimonieuse s'il est plus économique de la décrire en donnant la valeur et la position de chaque élément non-standard, plutôt qu'en donnant la valeur de tous les éléments. En traitement du signal, une représentation d'un signal est dite parcimonieuse si elle est économique. En d'autres termes, il existe une façon plus économique de décrire le signal que de donner la valeur de tous ses éléments, cela signifie que la plupart des coefficients sont nuls, et seuls quelques coefficients ont des valeurs non-nuls.

Il existe de nombreuses façons de représenter un signal. Nous pouvons citer comme exemple la représentation du signal dans le domaine fréquentiel. On change de domaine car la nouvelle représentation permet d'observer le signal sous un autre angle en révélant des propriétés inaccessibles dans le domaine original. Le but de certaines représentations est de faire ressortir l'information utile par rapport à l'environnement du signal. Pour atteindre cet objectif, il faut pouvoir représenter le signal dans un domaine où il n'y a que peu de coefficients non nuls, de manière à concentrer son énergie dans ces quelques coefficients. La mesure du nombre de coefficients non nuls est appelée parcimonie du signal dans le domaine considéré. L'intégration de la parcimonie a donné naissance à des techniques plus performantes dans différents domaines tels que le débruitage [59], la séparation de source [60], l'inpainting [61] ou même l'acquisition [62].

II.3.1 Mesure de la parcimonie

Nous traiterons les signaux comme des fonctions à valeurs réelles soit continués soit discrètes, et soit de longueur finie ou infinie. Ces hypothèses seront précisées si nécessaire dans ce chapitre. On s'intéressera typiquement aux espaces vectoriels normés, c'est-à-dire aux espaces vectoriels dotés d'une norme. Dans le cas d'un domaine fini discret, nous pouvons voir les signaux comme des vecteurs dans un espace euclidien à N dimensions, noté \mathbb{R}^N . En traitant des vecteurs dans \mathbb{R}^N , nous utiliserons fréquemment les normes L_p représentées dans la figure II.2 et définies pour $p \in [1, \infty]$ comme [63] :

$$\|S\|_p = \begin{cases} (\sum_{i=1}^N |S_i|^p)^{\frac{1}{p}} & p \in [1, \infty) \\ \max_{i=1,2,\dots,N} |S_i| & p = \infty \end{cases} \quad (\text{II.1})$$

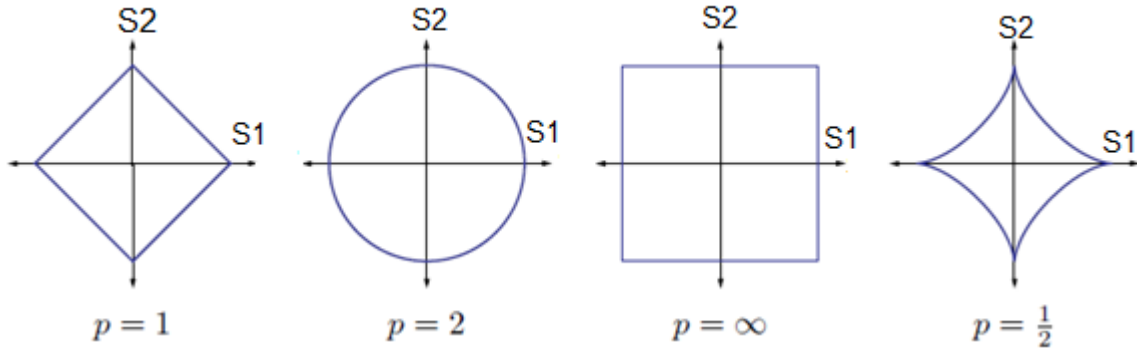


Figure II.2 : Sphères unitaires dans \mathbb{R}^2 pour les normes L_p .

II.3.2 Représentation parcimonieuse d'un signal

Un ensemble $\{\psi_i\}_{i=1}^N$ est appelé base de \mathbb{R}^N si les vecteurs de l'ensemble couvrant \mathbb{R}^N sont linéairement indépendants. Cela implique que chaque vecteur dans l'espace a une représentation unique en tant que combinaison linéaire de ces vecteurs de base. Plus précisément, pour tout $x \in \mathbb{R}^N$ il existe des coefficients (uniques) $\{S_i\}_{i=1}^N$ tels que

$$x = \sum_{i=1}^N S_i \psi_i \quad (\text{II.2})$$

Si on dénote par Ψ la matrice $N \times N$ à colonnes données par ψ_i et par \mathbf{S} le vecteur de longueur N ayant pour éléments S_i , alors nous pouvons représenter cette relation de manière plus compacte comme :

$$x = \Psi \mathbf{S} \quad (\text{II.3})$$

Un cas important d'une base est la base orthonormée définie comme étant un ensemble de vecteurs satisfaisant :

$$\langle \psi_i \psi_j \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad (\text{II.4})$$

Une base orthonormée a l'avantage que les coefficients S_i peuvent être facilement calculés comme :

$$S_i = \langle x, \psi_i \rangle \quad (\text{II.5})$$

ou

$$S_i = \Phi^T x \quad (\text{II.6})$$

où Φ est une matrice de mesure aléatoire de dimension $M \times N$ [52].

II.3.3 Transformation parcimonieuse d'un signal

La parcimonie est la première contrainte fondamentale pour la récupération en CS. La plupart des signaux n'ont pas de représentation parcimonieuse dans le domaine temporel, mais ils sont parcimonieux lorsqu'ils sont exprimés dans une base parcimonieuse Ψ ou un dictionnaire approprié D .

Les signaux vocaux varient fortement dans le temps, ils ne sont donc naturellement pas parcimonieux dans le domaine temporel. Plusieurs transformées peuvent être utilisées pour rendre parcimonieux les signaux de parole, comme la transformée en cosinus discrète, la transformée de Fourier discrète, les distributions temps-fréquence quadratiques, les ondelettes, etc. Pour les signaux compressibles, les petits coefficients peuvent être éliminés (en utilisant le seuillage) pour augmenter la parcimonie [64].

II.4 Acquisition du signal

L'acquisition comprimée exploite le fait que de nombreux signaux réels ont une représentation parcimonieuse une fois qu'ils sont exprimés dans une base appropriée Ψ . Ces signaux peuvent être reconstruits à partir d'un nombre limité d'échantillons ou mesures sur une deuxième base Φ incohérente avec la première.

L'acquisition comprimée d'un signal x correspond à la collecte de quelques mesures linéaires de S . Cela peut être représenté comme :

$$y_i = \langle x, \varphi_i \rangle \quad i = 1, 2, \dots, M \quad (\text{II.7})$$

où φ_i est la $i^{\text{ème}}$ ligne de la matrice Φ de dimensions $M \times N$ et $\langle \rangle$ désigne le produit scalaire.

Ou sous forme matricielle :

$$y = \Phi x = \Phi \Psi S = AS \quad (\text{II.8})$$

y est le vecteur de mesure du signal, S est le vecteur des coefficients du signal K -parcimonieux et $M \ll N$ le nombre de mesure requis pour une reconstruction parfaite [65].

Le processus de mesure représenté par La figure II.3 est réalisé avec une matrice de mesure Φ aléatoire gaussienne, le vecteur S est K -parcimonieux avec $K=3$.

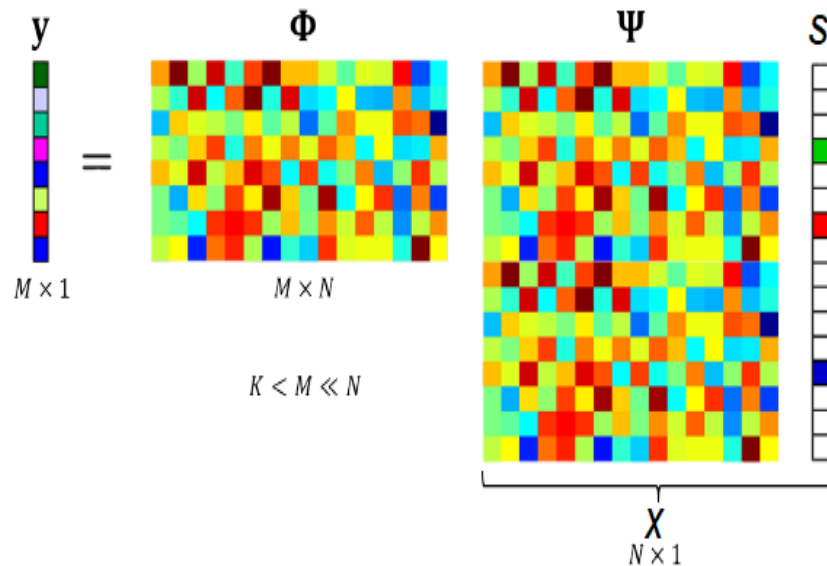


Figure II.3 : Processus de mesure.

Une reconstruction parfaite nécessite que Φ et Ψ obéissent à deux conditions appelées (i) condition d'incohérence et (ii) propriété d'isométrie restreinte (RIP : restricted isometry property).

II.4.1 Incohérence

L'incohérence signifie que la cohérence mutuelle entre la matrice d'acquisition Φ et la base parcimonieuse Ψ est faible. La cohérence mutuelle mesure la plus grande corrélation entre deux éléments quelconques de Φ et Ψ . En d'autres termes, il s'agit de la valeur absolue

maximale des corrélations croisées entre les colonnes de Φ et Ψ . Elle peut être définie comme suit [66]

$$\mu(\Phi, \Psi) = \alpha \max_{1 \leq k, i \leq N} |\langle \phi_k, \psi_j \rangle| \quad (\text{II.9})$$

où ϕ_k, ψ_i sont les $k^{\text{ième}}$ et $j^{\text{ième}}$ colonnes de Φ et Ψ , respectivement. Selon divers auteurs, α peut prendre les valeurs N, \sqrt{N} ou une valeur normalisée à un. La valeur de α dépend des propriétés algébriques des matrices Φ et Ψ choisies, ce qui donne une plage de variation de la cohérence entre 1 et \sqrt{N} [62],[67]. Pour les matrices incohérentes, μ est proche de la borne inférieure. La cohérence sert de caractérisation grossière du degré de similitude (ou « corrélation ») entre les systèmes de parcimonie et de mesure. La valeur $\mu = 1$ est le minimum (valeur la plus basse) de la cohérence, ce qui signifie que les lignes de la matrice de mesure Φ doivent être réparties dans le domaine Ψ . En d'autres termes, plus la cohérence est faible, plus l'incohérence est grande. Cette idée apparaît dans l'acquisition comprimée lors du choix de la matrice de mesure qui doit être incohérente avec la base dans laquelle la parcimonie est exprimée [62],[66].

II.4.2 Propriété d'isométrie restreinte (RIP)

Lorsque les mesures sont incohérentes, la matrice Φ satisfait la propriété d'isométrie restreinte (RIP) pour les vecteurs K -parcimonieux

$$(1 - \delta_K) \|S\|_2^2 \leq \|\Phi \Psi S\|_2^2 \leq (1 + \delta_K) \|S\|_2^2 \quad (\text{II.10})$$

où δ_K est la constante d'isométrie restreinte définie comme le plus petit nombre qui satisfait l'inégalité (II.10) pour tous les vecteurs S K -parcimonieux.

Lorsque δ_K est petit, alors $\Phi \Psi$ agit comme une isométrie proche sur les vecteurs S K -parcimonieux. En pratique, il est difficile de calculer δ_K directement ; de plus, la matrice de mesure Φ peut être choisie aléatoire, de sorte qu'il est plus souhaitable d'obtenir des propriétés statistiques sur les limites de δ_K pour une famille de matrices de mesure Φ , plutôt que de calculer δ_K pour une matrice Φ spécifique. Généralement, le nombre de mesures

diminuera la constante δ_K , améliorant la propriété de Φ à agir de manière isométrique sur des vecteurs parcimonieux. Lorsqu'il y a suffisamment de mesures incohérentes, il est possible de déterminer avec précision les K éléments non nuls du vecteur S de longueur N . Dans ce cas, il existe des limites sur la constante δ_K qui garantissent une reconstruction exacte du signal pour des données sans bruit [68].

Le RIP est la condition suffisante pour les matrices de CS qui garantit les performances de la reconstruction du signal en termes d'efficacité et de robustesse [62]. Comme les matrices aléatoires avec des entrées indépendantes et identiquement distribuées (iid) remplissent la condition RIP avec une forte probabilité, ce sont les matrices de mesure les plus utilisées dans la CS [69].

II.4.3 Mauvaise matrice de mesure

Jusqu'à présent, nous avons décrit comment prendre de bonnes mesures compressées. La figure II.4 montre un choix particulièrement mauvais de la matrice de mesure Φ correspondante aux M dernières colonnes de la base parcimonieuse Ψ . Dans ce cas, le produit $A = \Phi\Psi$ est une matrice identité $M \times M$ complétée par des zéros à gauche. Dans ce cas, tout signal S qui n'est pas actif dans les M dernières colonnes de Ψ est dans l'espace nul de A , et est totalement invisible aux mesures y . Dans ce cas, ces mesures entraînent des pertes d'informations importantes pour de nombreux vecteurs parcimonieux [68].

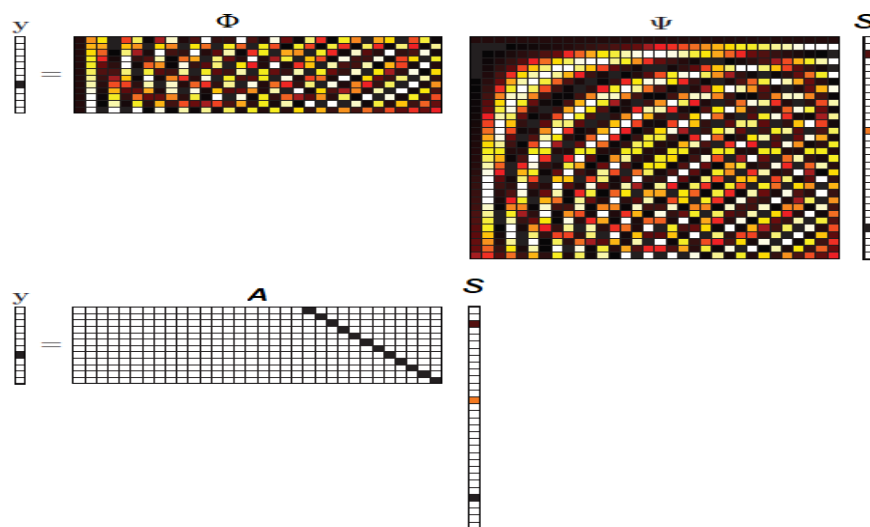


Figure II.4 : Exemple d'une mauvaise matrice de mesure.

II.5 Algorithme de reconstruction

L'algorithme de reconstruction du signal doit prendre les M mesures dans le vecteur y , la matrice de mesure aléatoire Φ (ou la graine aléatoire qui l'a générée) et la base Ψ et reconstruire le signal x de longueur N , ou d'une manière équivalente, son vecteur de coefficients parcimonieux S . Pour les signaux K -parcimonieux, puisque $M < N$ dans (II.8) il y a une infinité de vecteurs S qui satisfont

$$AS = y \quad (\text{II.11})$$

En effet, si $AS = y$ alors $A(S + r) = y$ pour tout vecteur r dans l'espace nul $N(A)$ de A . Par conséquent, l'algorithme de reconstruction du signal vise à trouver le vecteur de coefficients parcimonieux du signal dans l'espace nul ($N - M$) traduit en dimension $H = N(A) + S$.

Par conséquent, notre objectif est de trouver le vecteur de coefficients parcimonieux du signal S dans l'espace nul traduit.

II.5.1 Reconstruction par la minimisation L_2

L'approche classique des problèmes inverses de ce type consiste à trouver le vecteur dans l'espace nul traduit avec la plus petite norme L_2 (énergie) par résolution :

$$\hat{S} = \operatorname{argmin} \|S\|_2; \text{ tell que } AS = y \quad (\text{II.12})$$

La solution de ce problème d'optimisation est de la forme fermée :

$$\hat{S} = A^T(AA^T)^{-1}y \quad (\text{II.13})$$

Malheureusement, la minimisation L_2 ne trouvera presque jamais de solution K -parcimonieux, renvoyant à la place un vecteur \hat{S} non parcimonieux avec de nombreux éléments non nuls [70].

II.5.2 Reconstruction par la minimisation L_0

Puisque la norme L_2 mesure l'énergie du signal et non sa parcimonie, considérons la norme L_0 qui compte le nombre d'entrées non nulles dans S . (Par conséquent, un vecteur K -parcimonieux à une norme 0 égale à K .) L'optimisation modifiée donne :

$$\hat{S} = \operatorname{argmin} \|S\|_0 ; \text{ tell que } AS = y \quad (\text{II.14})$$

On peut récupérer un signal K -parcimonieux exactement avec une probabilité élevée en utilisant uniquement des mesures gaussiennes $M = K+1$ [71]. Malheureusement, la résolution de (II.14) est à la fois numériquement instable et NP-complète, ce qui nécessite une énumération exhaustive de tous les emplacements possibles $\binom{N}{K}$ des entrées non nulles de S [70].

II.5.3 Reconstruction par la minimisation L_1

De manière surprenante, en utilisant l'optimisation basée sur la norme L_1

$$\hat{S} = \operatorname{argmin} \|S\|_1 ; \text{ tell que } AS = y \quad (\text{II.15})$$

on peut récupérer exactement des signaux K -parcimonieux et des signaux compressibles très approximatifs avec une probabilité élevée en utilisant uniquement des mesures gaussiennes iid, le nombre d'échelles de mesures définie comme $M \geq CK \log(N/K)$. Le multiplicateur constant C définit exactement le nombre de mesures nécessaires. Il s'agit d'un problème d'optimisation convexe qui se réduit commodément à un programme linéaire connu sous le nom de poursuite de base [72, 73] dont la complexité de calcul est d'environ $O(N^3)$. Pour résumer, un système d'acquisition de données compressive consiste en des mesures aléatoires suivies d'une reconstruction par programmation linéaire pour obtenir x .

II.5.4 Discussion

La géométrie du problème d'acquisition par compression dans \mathbb{R}^N nous aide à visualiser pourquoi la reconstruction L_1 réussit là où la reconstruction L_2 échoue. Tout d'abord, il faut

noter que l'ensemble de tous les vecteurs K -parcimonieux S dans \mathbb{R}^N est un espace fortement non linéaire constitué de tous les hyperplans à K dimensions qui sont alignés avec l'ensemble des vecteurs qui sont alignés avec les axes des coordonnées (figure II.5). Ainsi, les vecteurs parcimonieux vivent à proximité des axes des coordonnées dans \mathbb{R}^N .

Pour visualiser pourquoi la reconstruction L_2 échoue, notons que l'espace nul traduit $H = N(A) + S$ est de dimension $(N - M)$ et orienté selon un angle aléatoire en raison du caractère aléatoire de la matrice A (Figure II.5.b), mais attention, en pratique $N, M, K \gg 3$ nous devons donc extrapoler notre intuition à des dimensions élevées. Le minimiseur \hat{S} de (II.12) au sens L_2 est le point sur H le plus proche de l'origine. On peut trouver ce point en faisant exploser une hypersphère (la boule L_2) jusqu'à ce qu'elle touche H . En raison de l'orientation aléatoire de H , le point S le plus proche vivra avec une forte probabilité loin des axes des coordonnées et ne sera donc ni clairsemé ni proche de la bonne solution S .

Contrairement à la boule L_2 , la boule L_1 de la figure II.5 (c) est "pointue", les points étant alignés le long des axes des coordonnées (et elle devient plus pointue à mesure que la dimension ambiante N augmente). Par conséquent, lorsque nous faisons exploser la boule L_1 , elle touchera d'abord l'espace nul translaté H en un point proche des axes des coordonnées, ce qui est précisément l'endroit où se trouve le vecteur parcimonieux S [70].

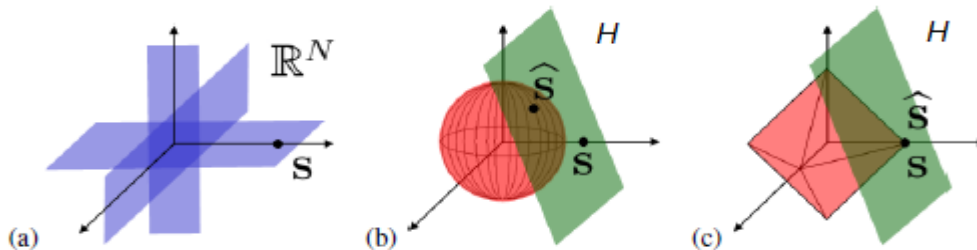


Figure II.5 : (a) Vecteur parcimonieux se trouvant sur un hyperplan de dimension K aligné avec les axes des coordonnées dans \mathbb{R}^N et donc proche des axes. (b) Récupération de l'acquisition comprimée via la minimisation L_2 qui donne un vecteur non parcimonieux \hat{S} . (c) Récupération via la minimisation L_1 qui donne la bonne solution parcimonieuse S .

II.6 Exemple d'acquisition et de reconstruction

Notre exemple utilise la transformée en cosinus discrète (DCT) comme base. Le signal généré est la somme de deux sinusoïdes

$$f(t) = \sin(1394 \pi t) + \sin(3266 \pi t) \quad (\text{II.15})$$

Le signal est échantillonné pendant 1/8 de seconde à 40000 Hz, le résultat est un vecteur f de longueur $N = 5000$. La figure II.6 (a) montre une partie de ce signal, ainsi que certains des $M = 500$ échantillons aléatoires que nous avons pris. Les figures II.6 (b) et II.6 (c) montrent, respectivement, les coefficients c de la DCT de f et la solution x obtenue dans le domaine de la DCT par minimisation L_1 . La figure II.6 (d) montre le signal reconstruit en prenant la transformée en cosinus discrète inverse de x .

Pour la minimisation L_1 , le programme Matlab *l1-magic*, écrit par Justin Romberg et Emmanuel Candès a été utilisé [74].

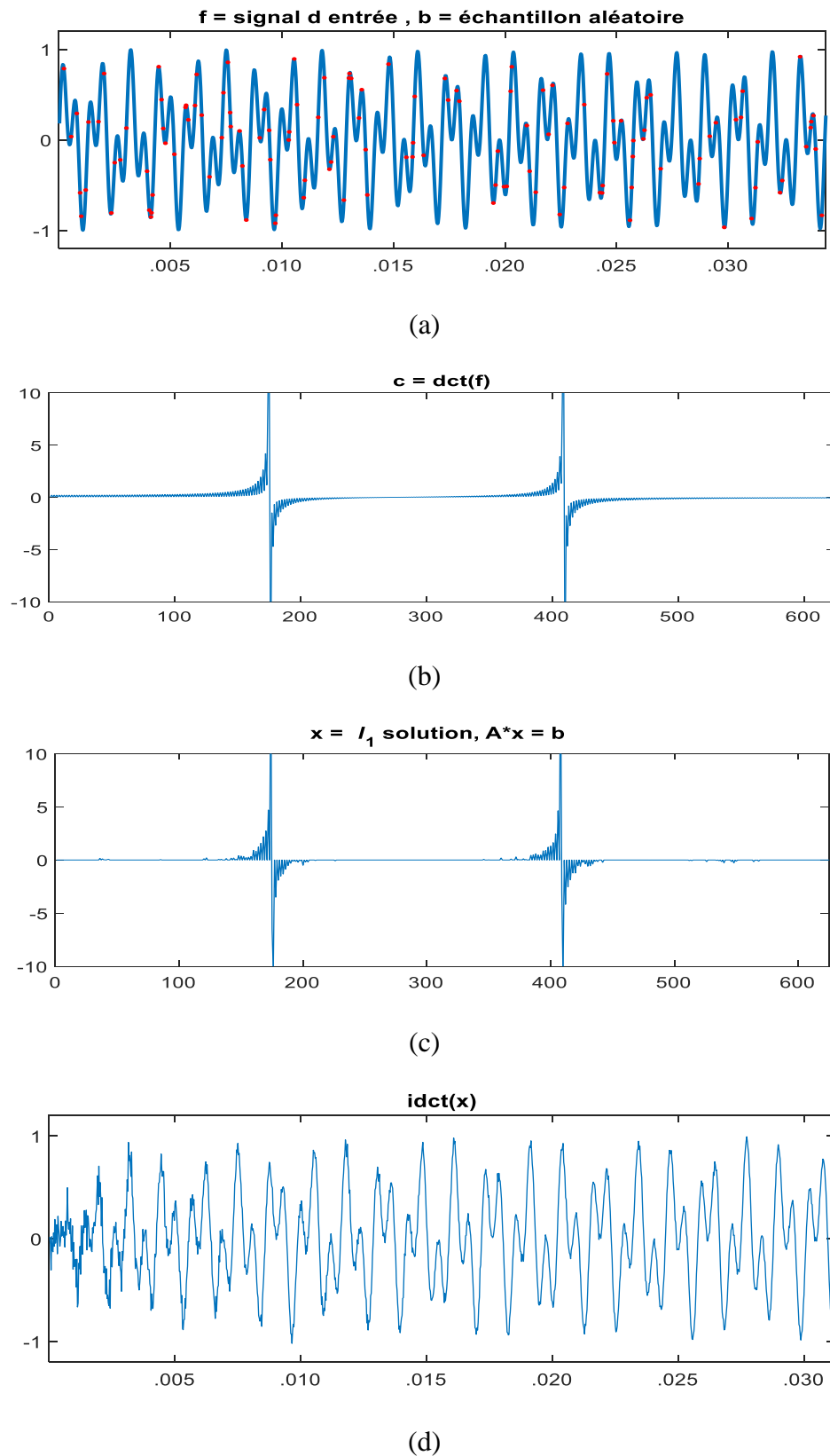


Figure II.6 : Illustration de l'acquisition du signal et sa reconstruction en utilisant la minimisation L1. (a) Signal original et échantillons prélevés aléatoirement. (b) DCT du signal. (c) Solution L1 notée x . (d) DCT inverse de x .

II.7 Applications de l'acquisition comprimée

Depuis son apparition en 2005 [7],[75], l'acquisition comprimée a fait l'objet de recherches intensives. Le potentiel des applications est énorme, cela concerne le traitement du signal et des images. Il s'applique principalement dans le cas où l'acquisition complète du signal est une procédure coûteuse, longue, difficile, dangereuse ou impossible.

De plus, il devrait y avoir des raisons de supposer que le signal en question est parcimonieux dans une base convenable connue et incohérente avec la base d'acquisition. Pratiquement, ceci s'applique à plusieurs types de signaux, et par suite ceci mène à des implications importantes qui concernent un certain nombre d'applications possibles. Nous citons entre autres :

II.7.1 Compression des données

Dans certaines situations, la base de reconstruction Ψ , dans laquelle le signal possède une représentation parcimonieuse, peut être inconnue à l'encodeur (près du dispositif de mesure), ou il est peu pratique de la mettre en œuvre. Cependant, une matrice d'acquisition Φ conçue de manière aléatoire peut être considérée comme une stratégie de codage universel [66], parce qu'elle ne doit pas être conçue en relation avec la structure de Ψ . Ainsi, le choix, la connaissance, et l'aptitude à mettre en œuvre Ψ ne sont requises que pour le décodage ou la reconstruction du signal. Cette universalité peut être particulièrement utile pour le codage de sources distribuées tels que les réseaux de capteurs [71], ainsi que dans les applications astronomiques [76].

II.7.2 Codage de canal

Les principes du CS (parcimonie, probabilité, optimisation convexe) peuvent être manipulés et appliqués à la conception des codes correcteurs d'erreurs rapides, et par suite corriger les erreurs de la transmission [77].

II.7.3 Acquisition des données

Dans certains cas, l'acquisition complète des échantillons discrets d'un signal analogique peut être difficile à obtenir, ou peut-être difficile à compresser sur place [27]. L'acquisition peut être donc effectuée à une faible fréquence d'échantillonnage des mesures incohérentes du signal analogique incident. Dans d'autres cas tels que les applications de la résonance

magnétique nucléaire [74], [76], [78, 79] (e.g. spectroscopie, imagerie), le processus d'acquisition des données est intrinsèquement lent. L'application du CS offre potentiellement d'importantes réductions des du temps d'acquisitions, permettant ainsi l'amélioration du confort du patient et la réduction des artefacts reliés aux mouvements respiratoires. Notons ici que l'imagerie par résonance magnétique obéit naturellement à deux conditions nécessaires à la réussite de l'application du CS :

- Elle est naturellement compressible dans une base appropriée, *e.g.* les transformations d'ondelettes, les différences finies.
- Son acquisition correspond à des échantillons codés linéairement (codage espace fréquence), plutôt qu'à une acquisition directe des intensités des pixels de l'image.

II.7.4 Imagerie médicale

L'acquisition comprimée est activement recherchée pour l'imagerie médicale, notamment en imagerie par résonance magnétique (IRM). Les images par résonance magnétique, comme les angiogrammes, ont des propriétés de parcimonie dans des domaines tels que la base de Fourier ou d'ondelettes. Généralement, l'IRM est un processus coûteux et long en raison de son processus de collecte de données qui dépend des contraintes physiques et physiologiques. Cependant, l'introduction de techniques basées sur la CS a amélioré la qualité de l'image en réduisant le nombre de mesures collectées et en tirant parti de leur parcimonie implicite. L'IRM est un domaine de recherche actif pour la communauté de CS et dans un passé récent, un certain nombre d'algorithmes de CS ont été spécifiquement conçus pour cela [79-83].

Nous n'avons donné que quelques exemples des applications possibles du CS, et qui font qu'elle détient, en tant que procédé d'acquisition et de traitement, une grande promesse pour les applications actuelles et futures.

II.8 Conclusion

Dans ce chapitre, nous nous sommes intéressés à la théorie de l'acquisition comprimée, ses origines, ses conditions d'application ainsi que ces différentes étapes de l'acquisition du signal à sa reconstruction. L'acquisition comprimée, s'appuyant sur la parcimonie de la grande majorité des signaux pratiques, permet de s'affranchir des limitations inhérentes au théorème de Shannon. On a présenté aussi un aperçu de ses applications et un exemple d'illustration de l'acquisition comprimée et de la reconstruction en utilisant l'algorithme d'optimisation l1-magic.

Chapitre

III

*Rehaussement mono-
microphone par acquisition
comprimée d'ensemble*

III.1 Introduction

Le rehaussement de la parole par l'acquisition comprimée a été proposé dans [9] comme alternative aux techniques de rehaussement conventionnelles. Il a été montré, qu'en moyenne, le SNR dans le domaine comprimé est supérieur ou égal au SNR calculé dans le domaine non comprimé. Dans [84], l'acquisition comprimée utilisant une matrice de mesure aléatoire et combinée avec le filtre de Wiener a été proposée pour réduire le bruit et produire ainsi un rapport signal sur bruit élevé. Motivé par les résultats obtenus dans [9], l'acquisition comprimée dite d'ensemble a été combinée avec la technique du beamforming pour le rehaussement multi-microphones [10]. Dans cette approche de rehaussement, le signal de parole est estimé comme la moyenne d'ensemble de plusieurs réalisations. Les résultats obtenus dans [10] montrent que la technique de rehaussement de la parole combinant le beamforming avec l'acquisition comprimée présente des performances supérieures à celles du beamforming seul.

Dans ce chapitre, on se propose d'utiliser l'acquisition comprimée d'ensemble pour le rehaussement mono-microphone de la parole. Dans son principe, le rehaussement de la parole en utilisant l'acquisition comprimée est basé sur le fait que les signaux vocaux sont de nature parcimonieux dans le domaine temps-fréquence alors que les signaux de bruit ne sont pas parcimonieux [9]. Cependant, les signaux de parole contiennent des segments voisés et des segments non voisés, la condition de parcimonie n'est satisfaite que par les segments de parole voisés, les segments non voisés ne peuvent être considérés comme parcimonieux parce que leur spectre présente une structure semblable à celle du bruit. Comme conséquence, la performance de l'acquisition comprimée est limitée quand elle est appliquée au rehaussement des segments non voisés. Pour remédier à cet inconvénient, on se propose d'utiliser un post-traitement des segments non voisés en procédant à un débruitage par seuillage des coefficients d'ondelettes afin d'améliorer le SNR des segments non voisés.

III.2 Rehaussement mono-microphone par acquisition comprimée d'ensemble

III.2.1 Méthode

La méthode de rehaussement par acquisition comprimée d'ensemble [9] consiste à appliquer l'algorithme de l'acquisition comprimée au signal d'entrée plusieurs fois. Pour cela, nous utilisons une matrice de mesure aléatoire lors de l'acquisition du signal. En raison du caractère aléatoire de la matrice de mesure, chaque application de l'algorithme d'acquisition comprimée résulte en une réalisation du signal. Le signal de parole est ensuite estimé comme

étant la moyenne d'ensemble de toutes les réalisations. Le principe de la méthode de rehaussement par acquisition comprimée d'ensemble est illustré par la figure III.1. Les différentes étapes de l'algorithme peuvent être résumées comme suit :

1. Décomposer le signal d'entrée en trames de N échantillons.
2. Appliquer l'algorithme de l'acquisition comprimée aux différentes trames en utilisant une matrice de mesure aléatoire.
3. Reconstruire et concaténer les différentes trames pour obtenir une réalisation du signal.
4. Répéter N_r fois les étapes 2 et 3.
5. Estimer le signal de la parole comme étant la moyenne d'ensemble des N_r réalisations.

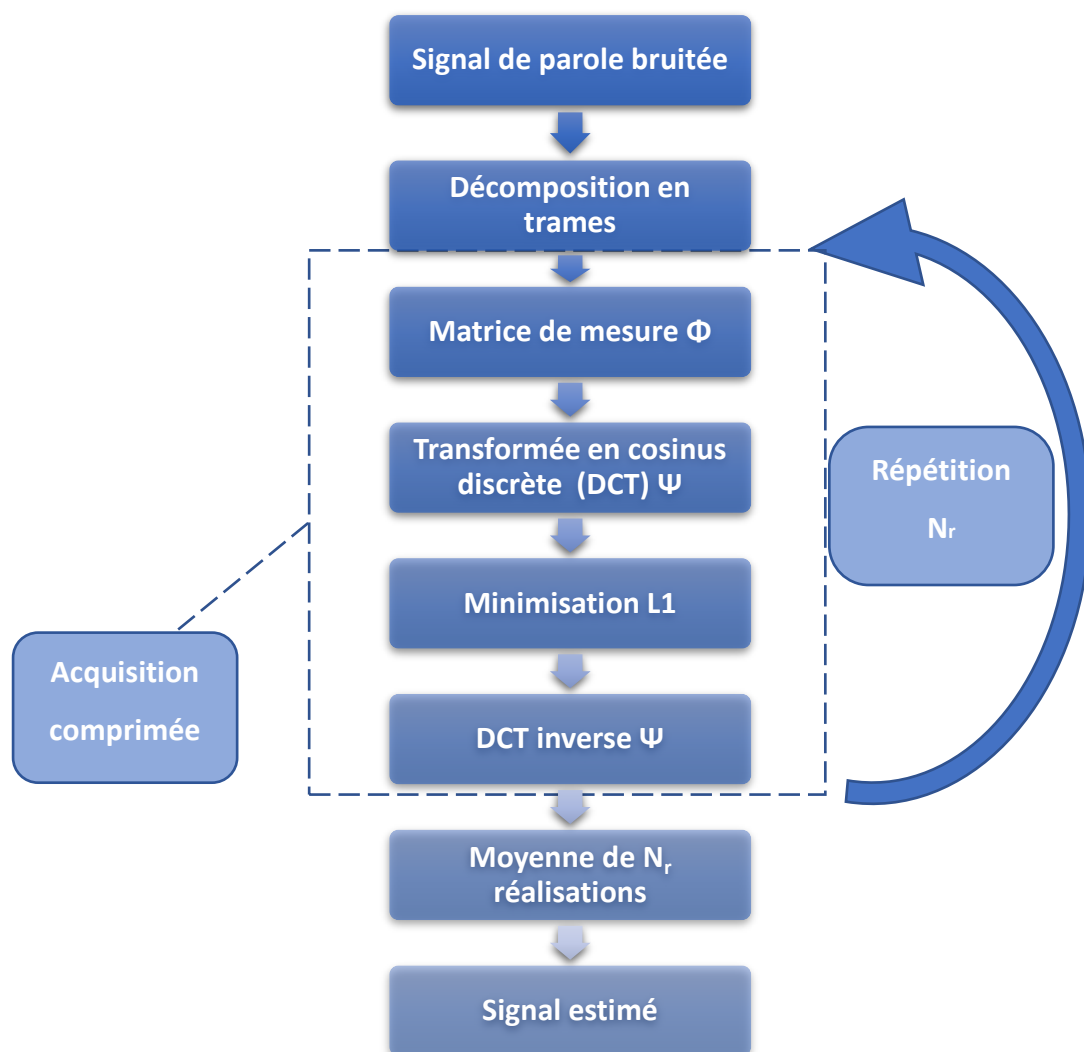


Figure III.1 : Schéma bloc de la méthode de rehaussement par acquisition comprimée d'ensemble.

III.2.2 Exemple d'illustration

Dans cet exemple, la méthode de rehaussement de la parole par acquisition comprimée d'ensemble (Figure III.1) est appliquée à un signal de parole bruitée échantillonné à une fréquence de 8 kHz. Le signal correspond à la phrase /We find joy in the simplest things/ prononcée par un locuteur masculin corrompu par un bruit de hall d'exposition. La longueur de trame est fixée à 30 ms, ce qui correspond à un nombre d'échantillons $N=240$, le nombre d'échantillons pris aléatoirement est $M=0.4N$ et le nombre de répétitions N_r de l'algorithme est fixé à 10. La figure III.2 montre le signal propre, le signal bruité à un SNR de 0 dB ainsi que le signal rehaussé en utilisant la méthode basée sur l'acquisition comprimée d'ensemble et la figure III.3 montre leurs spectrogrammes respectifs. Le SNR du signal rehaussé est de 3.4 dB.

La figure III.4 montre l'effet du nombre de répétitions N_r sur l'évolution du SNR de sortie du signal rehaussé par acquisition comprimée d'ensemble en fonction du SNR d'entrée pour une longueur de trame de 30 ms, soit $N=240$, et un nombre d'échantillons prélevés aléatoirement $M=0.6N$. L'application de l'acquisition comprimée une seule fois ($N_r=1$) n'améliore pas le SNR. Cependant, en augmentant le nombre de répétitions N_r , le SNR de sortie augmente. A partir d'un nombre de répétitions $N_r=10$, le SNR n'augmente que légèrement.

La figure III.5 montre l'effet de la longueur de trame N sur l'évolution du SNR de sortie du signal rehaussé par acquisition comprimée d'ensemble en fonction du SNR d'entrée pour un nombre de répétitions $N_r=15$ et un nombre d'échantillons prélevés aléatoirement $M=0.6N$. On observe qu'à un SNR d'entrée de 15 dB, le SNR de sortie varie de moins de 2 dB lorsqu'on varie la longueur de trame de 5 ms à 30 ms. Cette variation diminue avec le SNR d'entrée. A un SNR d'entrée de 0 dB, la variation de la longueur de trame n'affecte presque pas le SNR de sortie.

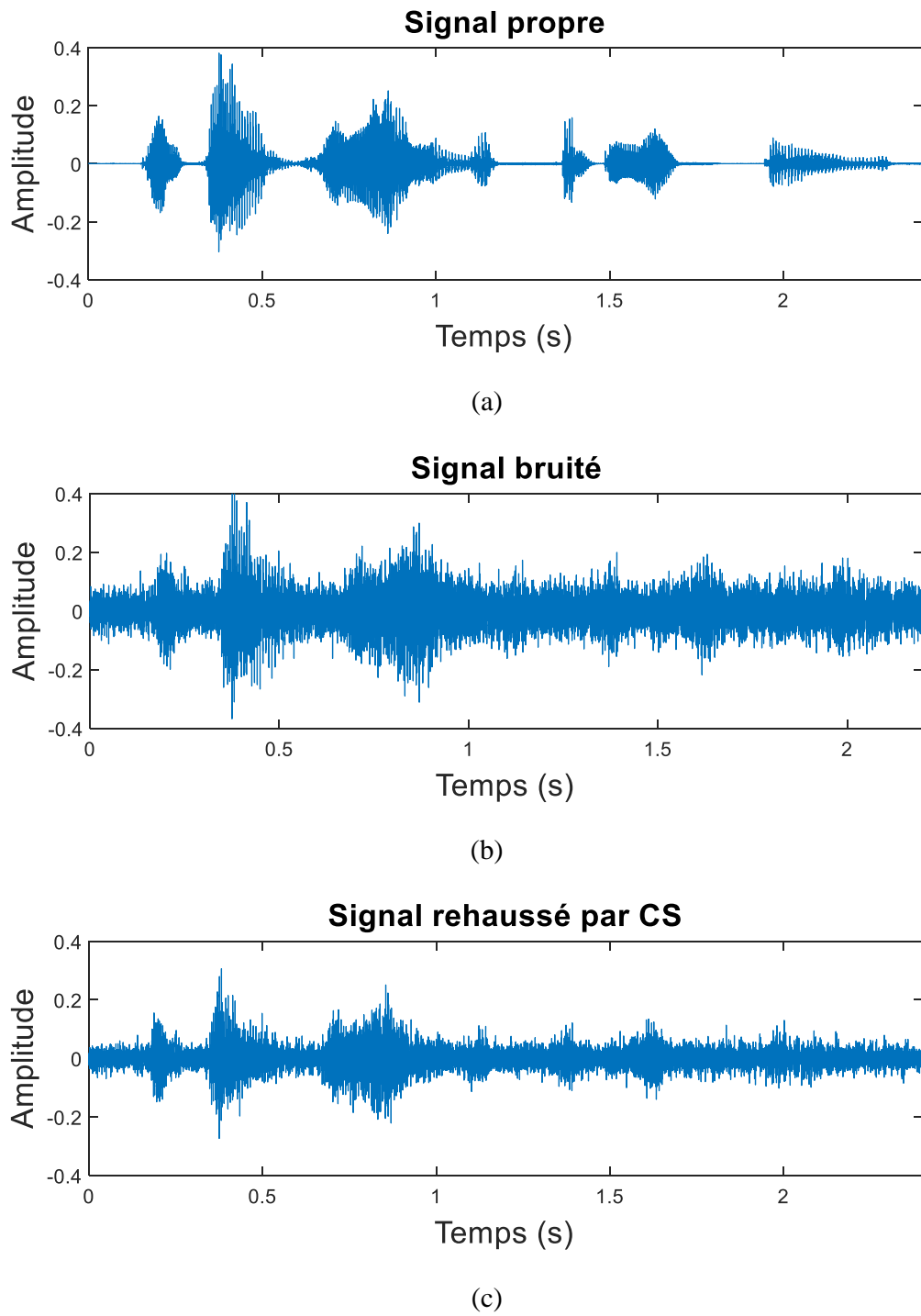


Figure III.2 : Signal (a) propre, (b) bruité et (c) rehaussé par la technique d'acquisition comprimée d'ensemble pour $M=0.4N$ et un SNR d'entrée de 0 dB.

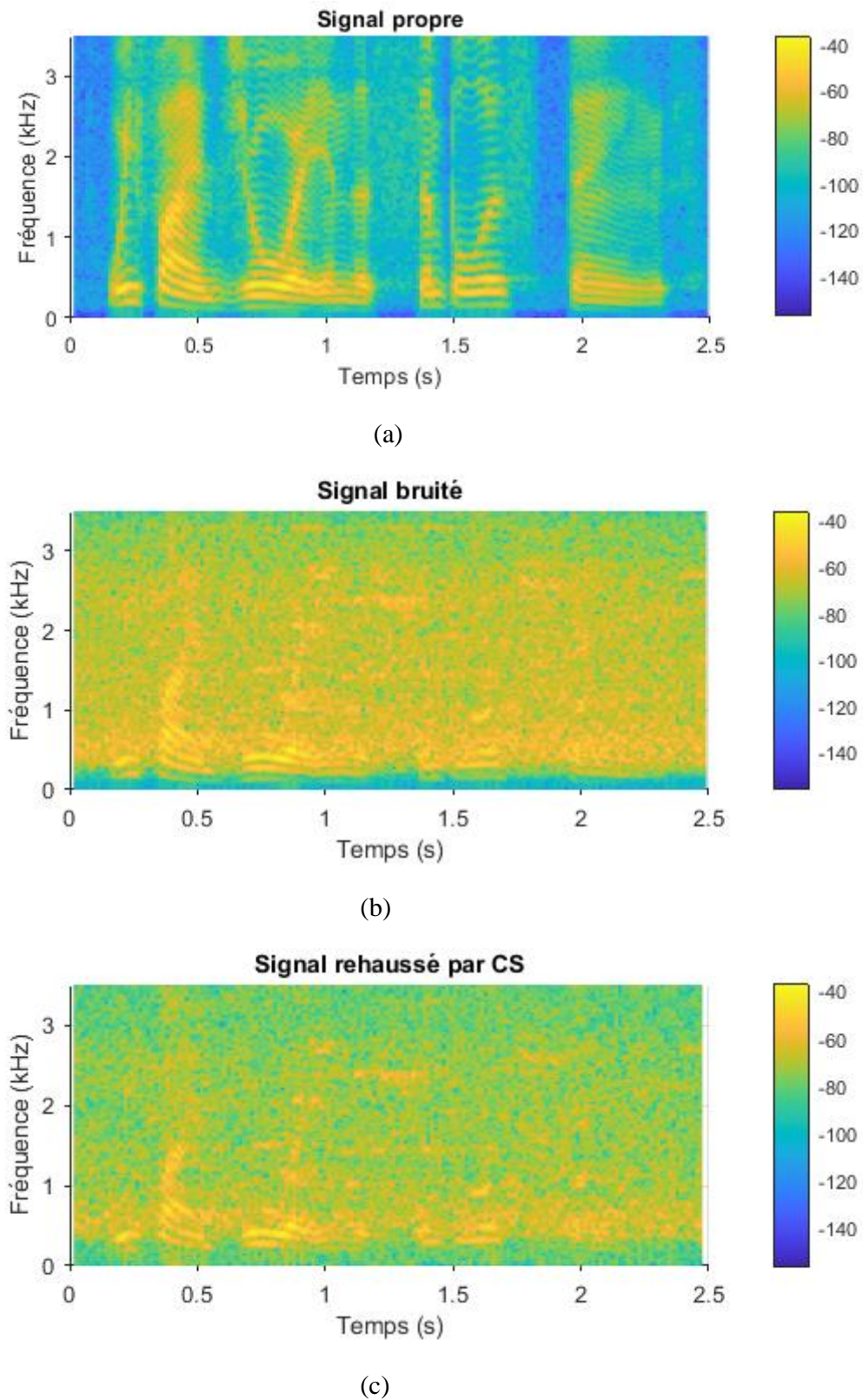


Figure III.3 : Spectrogrammes des signaux (a) propre, (b) bruité et (c) rehaussé par la technique d'acquisition comprimée d'ensemble pour $M=0.4N$ et un SNR d'entrée de 0 dB.

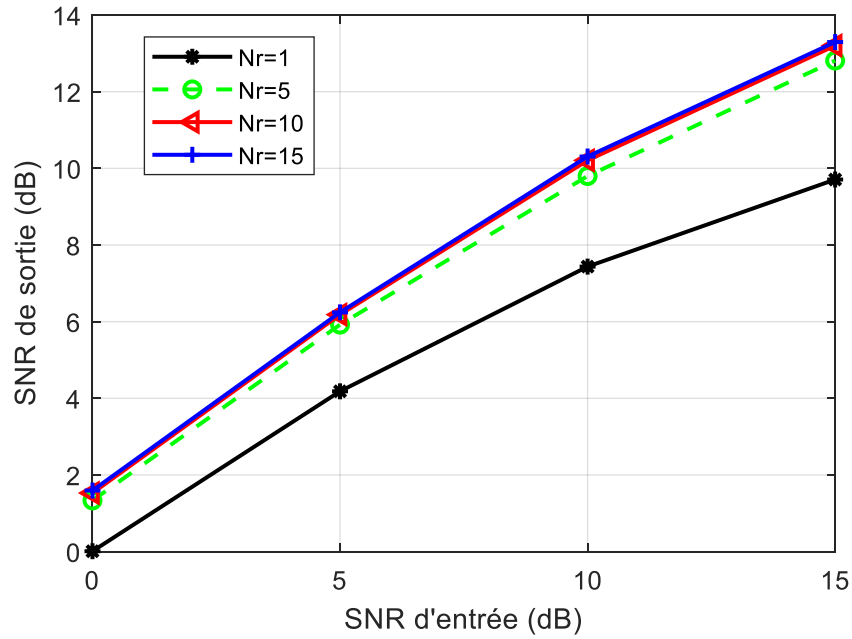


Figure III.4 : Effet du nombre de répétitions N_r sur l'évolution du SNR de sortie du signal rehaussé par l'acquisition comprimée d'ensemble en fonction du SNR d'entrée pour une longueur de trame de 30 ms ($N=240$) et un nombre d'échantillons prélevés aléatoirement $M=0.6N$.

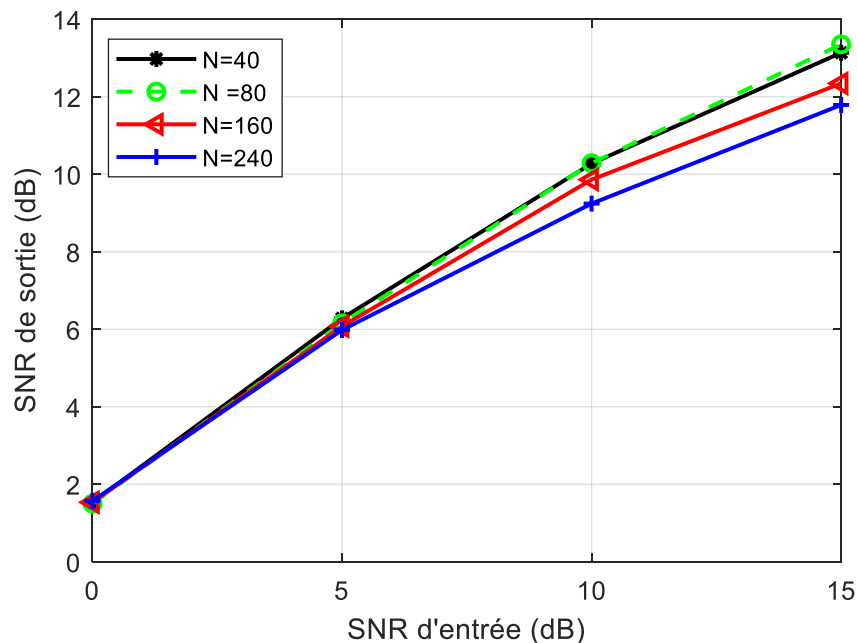


Figure III.5 : Effet de la longueur de trame N sur l'évolution du SNR de sortie du signal rehaussé par acquisition comprimée d'ensemble en fonction du SNR d'entrée pour un nombre de répétitions $N_r=15$ et un nombre d'échantillons prélevés aléatoirement $M=0.6N$.

III.3 Rehaussement mono-microphone par acquisition comprimée d'ensemble et seuillage des coefficients d'ondelettes

III.3.1 Méthode

A partir de la figure III.2 (c), on observe que l'acquisition comprimée n'élimine pas tout le bruit dans les segments non voisés. Pour cela, on se propose un post-traitement qui consiste à effectuer un débruitage par seuillage des coefficients d'ondelettes des segments non voisés du signal rehaussé par acquisition comprimée d'ensemble afin d'améliorer le SNR. La détection de voisement est effectuée en comparant l'énergie des segments à un seuil prédéfini. Le segment est classé non voisé si son énergie est inférieure au seuil prédéfini.

Le principe de la méthode de rehaussement combinant l'acquisition comprimée d'ensemble et le débruitage par seuillage des coefficients d'ondelettes des segments non voisés est illustré par la figure III.6.

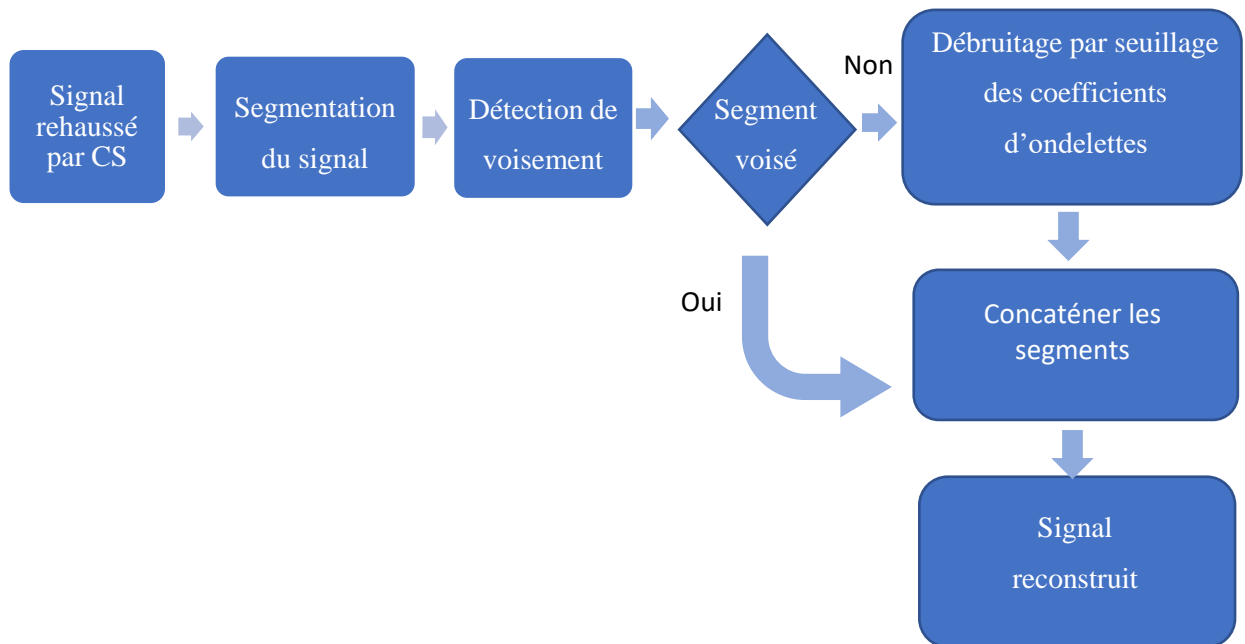
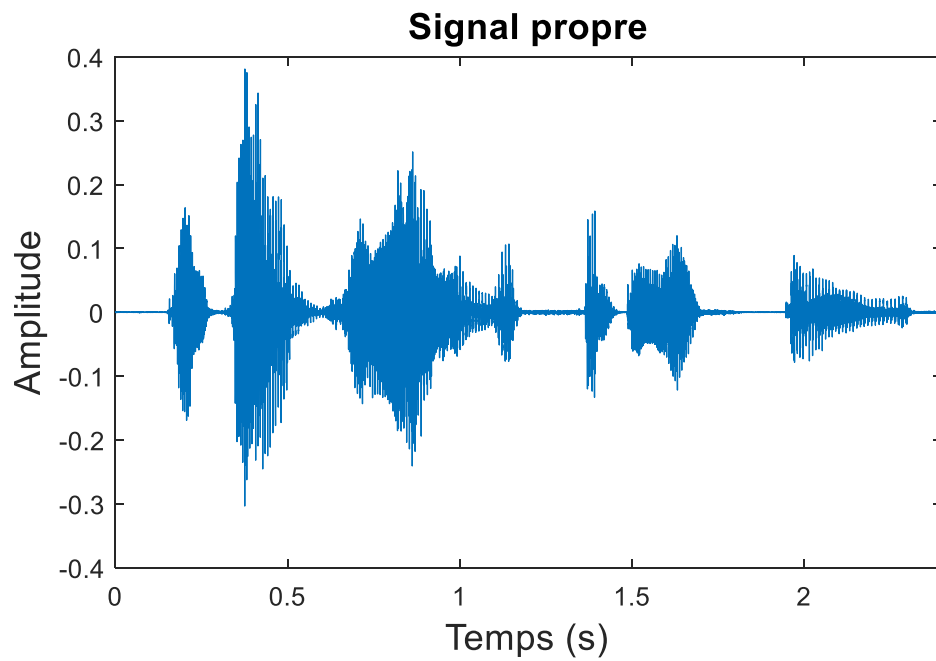


Figure III.6 : Principe de la méthode de rehaussement par acquisition comprimée d'ensemble et seuillage des coefficients d'ondelettes.

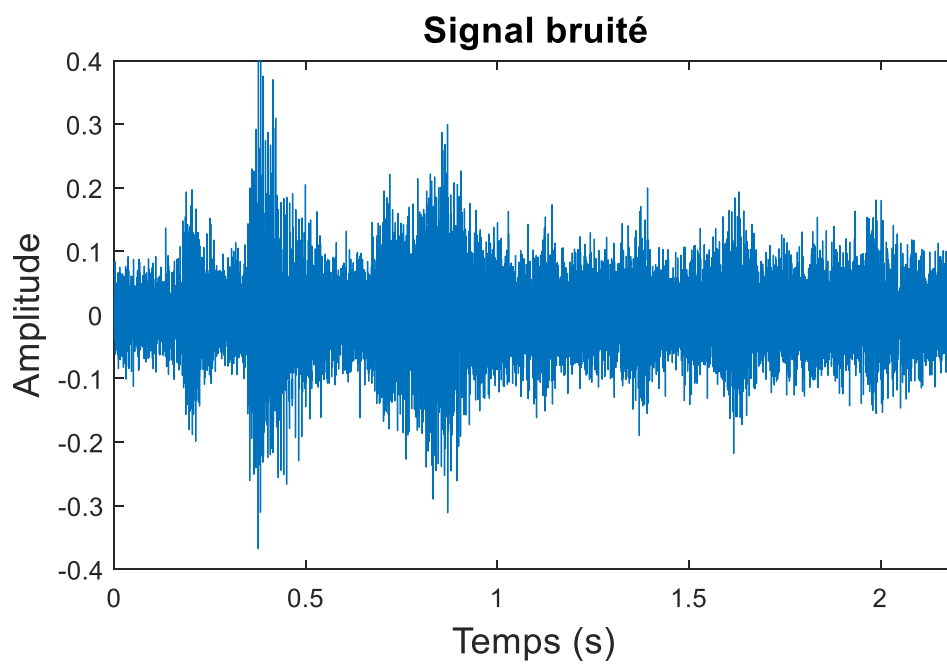
III.3.2 Exemple d'illustration

Dans cet exemple, la méthode de rehaussement de la parole par acquisition comprimée d'ensemble suivi d'un débruitage par seuillage des coefficients d'ondelettes des segments non voisés (Figure III.6) est appliquée au signal de parole bruité de l'exemple précédent. Le signal est segmenté en trames de 10 ms afin de détecter les segments non voisés qui peuvent être de courte durée. Pour la détection des segments non voisés, le seuil est fixé à 5% de l'énergie maximale de tous les segments du signal. La figure III.7 montre le signal propre, le signal bruité, le signal estimé par acquisition comprimée d'ensemble ainsi que le signal estimé par acquisition comprimée d'ensemble suivi d'un débruitage des segments non voisés par seuillage des coefficients d'ondelettes pour $M=0.4N$ et un SNR d'entrée de 0 dB. La figure III.8 montre les spectrogrammes respectifs de ces signaux.

Comme on peut l'observer, comparativement au rehaussement par l'acquisition comprimée d'ensemble seule, le bruit a été considérablement réduit dans les segments non voisés lorsque le rehaussement par acquisition comprimée d'ensemble est suivi d'un débruitage des segments non voisés par seuillage des coefficients d'ondelettes. Le SNR de sortie du signal rehaussé est de 4.9674 dB. Ces observations suggèrent la possibilité de produire un meilleur rehaussement du signal de la parole par la combinaison de l'acquisition comprimée d'ensemble et le débruitage des segments non voisés par seuillage des coefficients d'ondelettes.

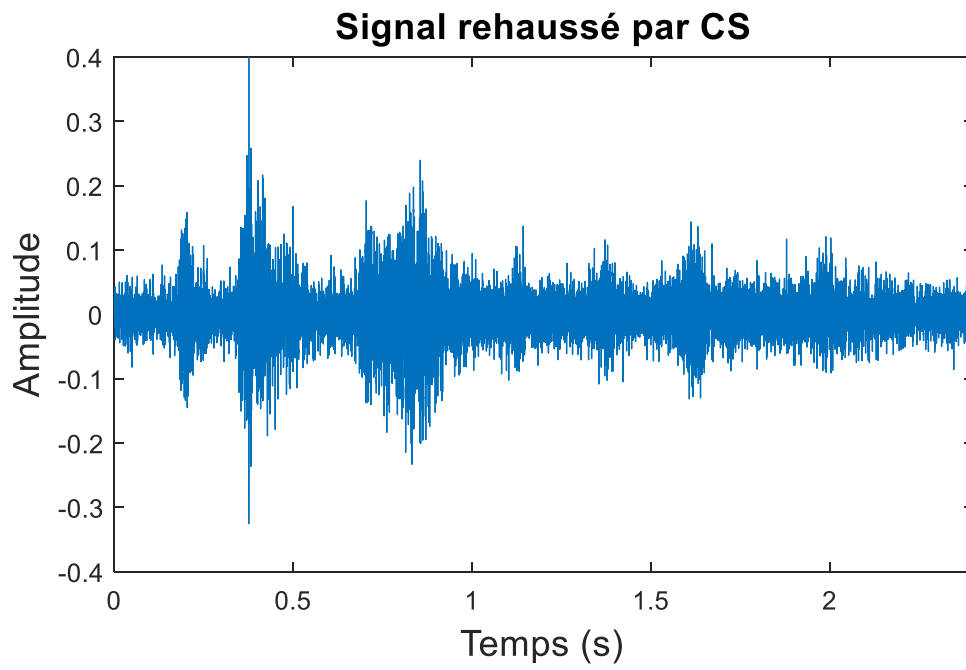


(a)

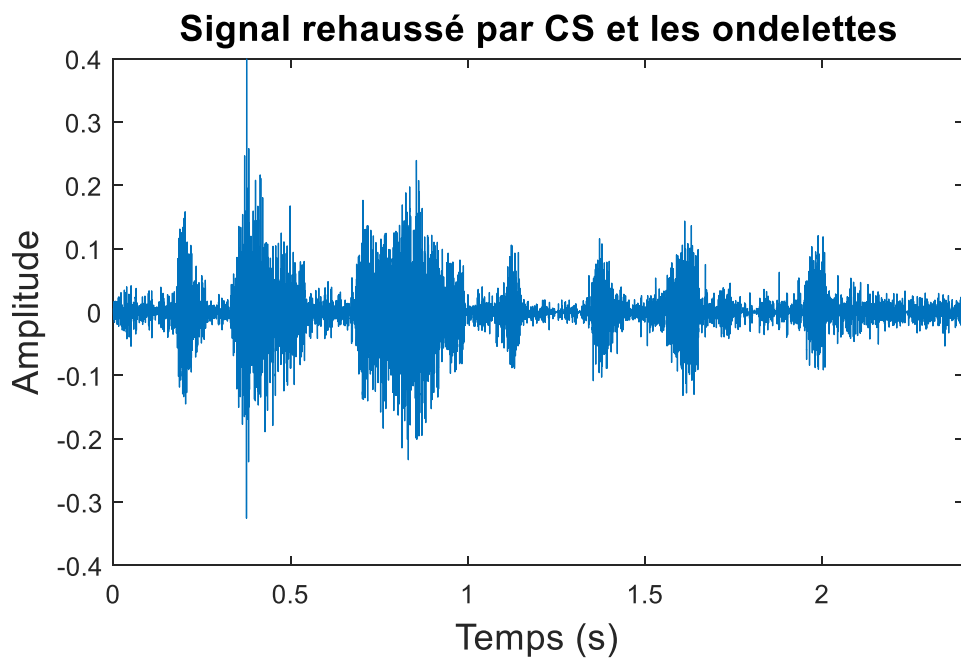


(b)

Figure III.7 : (a) Signal propre, (b) signal bruitée, (c) signal estimé par acquisition comprimée d'ensemble et (d) signal estimé par acquisition comprimée d'ensemble suivi d'un débruitage des segments non voisés par seuillage des coefficients d'ondelettes pour $M=0.4N$ et un SNR d'entrée de 0 dB.

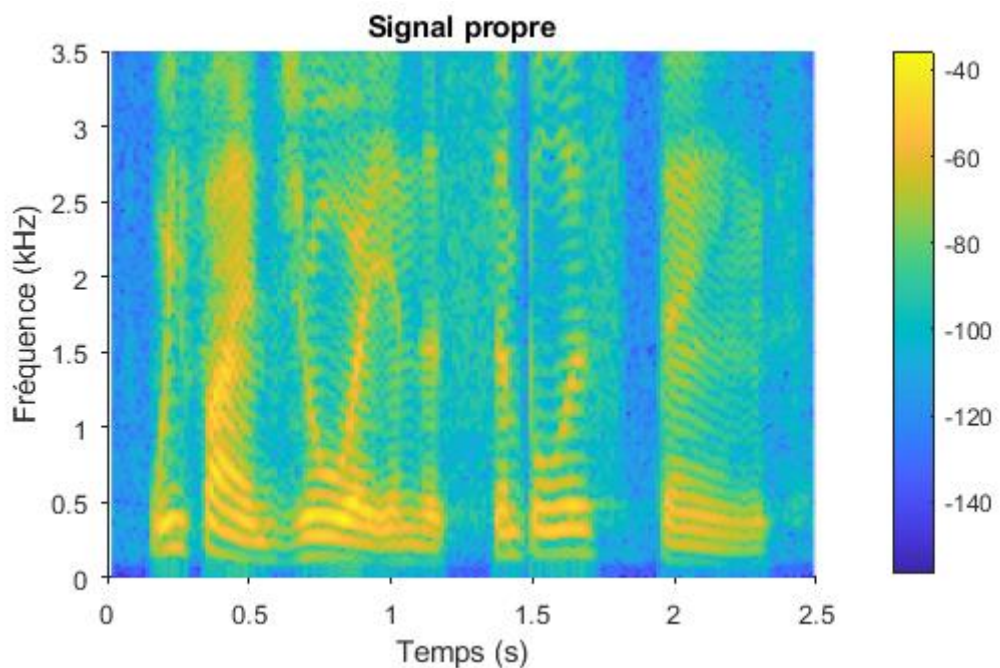


(c)

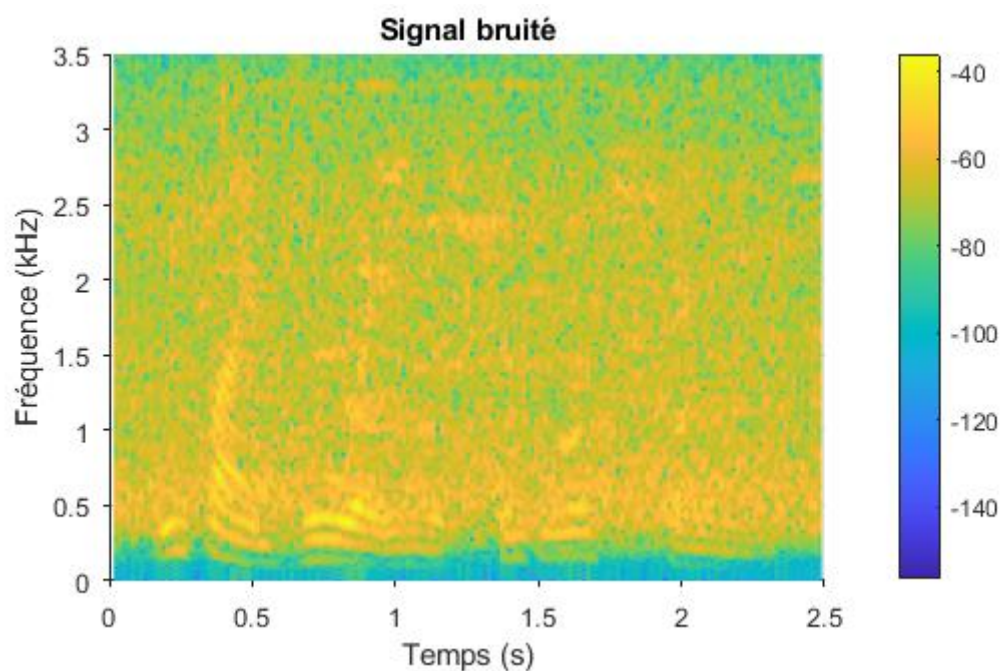


(d)

Figure III.7 (suite)

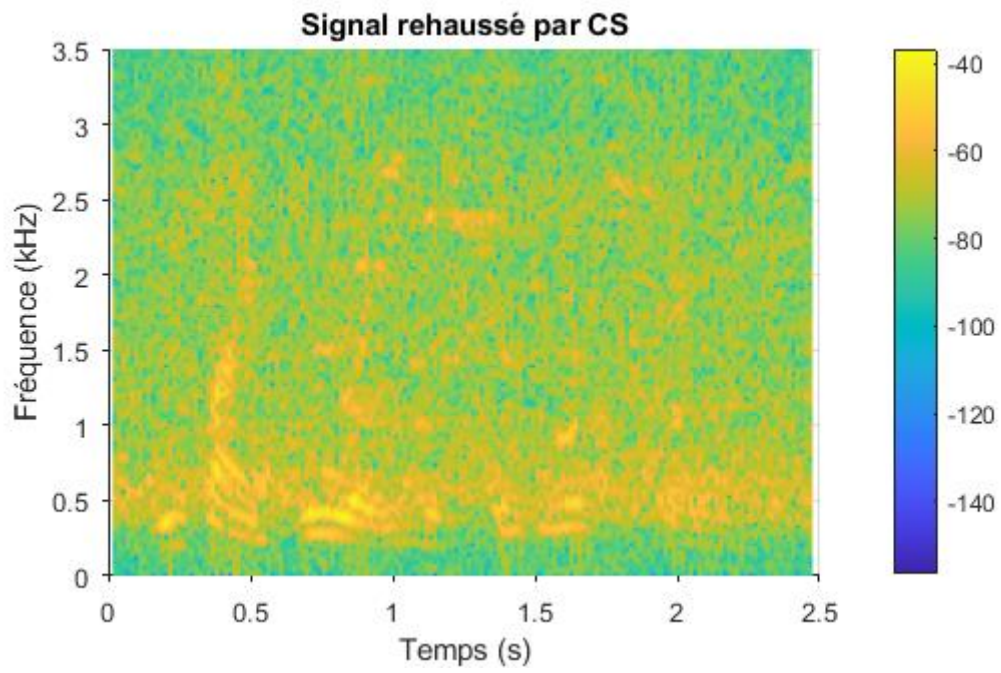


(a)

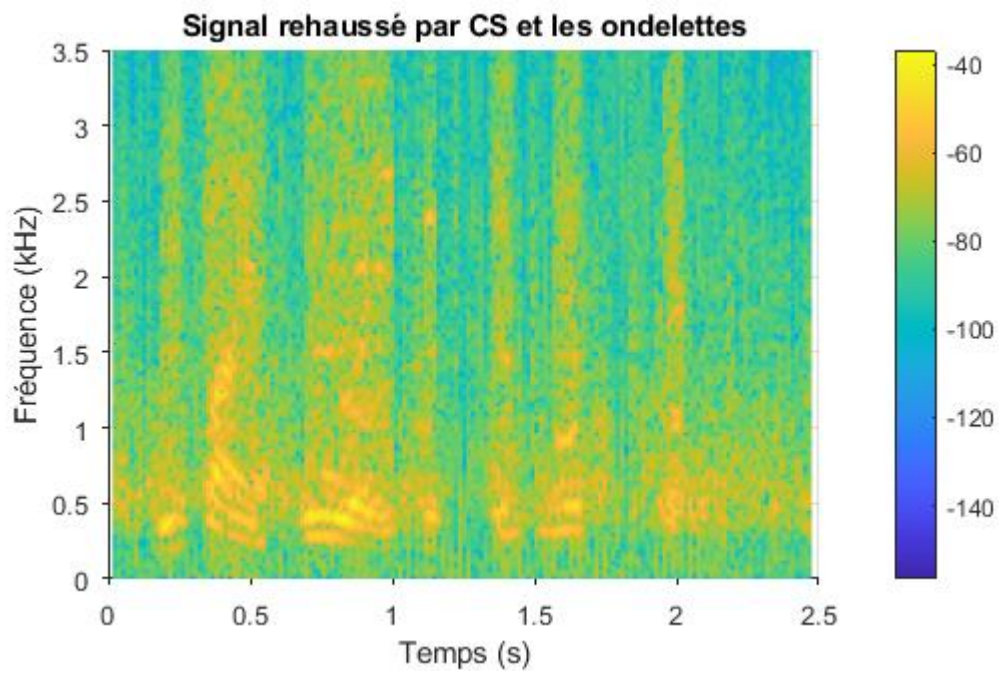


(b)

Figure III.8 : Spectrogrammes des signaux propre, bruité et rehaussé pour $M=0.6N$ et un SNR d'entrée de 0 dB. (a) Signal propre. (b) signal bruité. (c) signal estimé par acquisition comprimée d'ensemble. (d) signale estimé par acquisition comprimée d'ensemble suivi d'un débruitage des segments non voisés par seuillage des coefficients d'ondelettes.



(c)



(d)

Figure III.8 (suite)

III.4 Conclusion

Dans ce chapitre, nous avons présenté la méthode de rehaussement mono-microphone de la parole basée sur l'acquisition comprimée d'ensemble. L'effet de la longueur de trame et du nombre de répétitions sur les performances de la méthode a été aussi investigué. Afin d'améliorer la performance de la méthode de rehaussement en termes de SNR, on s'est proposé d'effectuer un post traitement du signal rehaussé par acquisition comprimé d'ensemble consistant en un débruitage des segments non voisés par seuillage des coefficients d'ondelettes. Il a été observé que ce post-traitement réduit considérablement le bruit des segments non voisé et améliore le SNR du signal rehaussé.

Chapitre IV
Résultats
expérimentaux

IV.1 Introduction

Dans ce chapitre, la méthode de rehaussement mono-microphone par acquisition comprimée ainsi que celle combinant l'acquisition comprimée et le débruitage des segments non voisés par seuillage des coefficients d'ondelettes sont testées et évaluées sur une base de données incluant des signaux de parole corrompus par différents types de bruits à différents SNR. Les résultats sont comparés à ceux obtenus par la méthode de soustraction spectrale, le filtrage de Wiener et le débruitage par seuillage des coefficients d'ondelettes.

IV.2 Base de données

La base de données NOIZEUS [85] a été utilisée pour évaluer les différentes méthodes de rehaussement de la parole. Cette base de données a été développée spécialement pour faciliter la comparaison des algorithmes de rehaussement de la parole parmi les groupes de recherche et elle est gratuitement téléchargeable à partir du site internet : (<https://ecs.utdallas.edu/loizou/speech/noizeus/>). Elle contient 30 phrases (produites par trois hommes et trois femmes) échantillonnées à 8 kHz et corrompues par huit types de bruits du monde réel à différents SNR. Le bruit provient de la base de données AURORA et comprend le bruit des trains de banlieue, de babillage, des voitures, des halls d'exposition, des restaurants, des rues, des aéroports et des gares.

IV.3 Protocole expérimentale

Les simulations ont été effectuées sous MATLAB R2020a. La boîte à outil l1-magic [77] a été utilisée pour l'optimisation l_1 . Pour chaque type de bruit et chaque SNR d'entrée, on applique la méthode de rehaussement à toutes les phrases et le SNR de sortie est calculé comme étant la moyenne des SNRs de toutes les phrases. Le signal d'entrée est divisé en trames de longueur 10 ms, ce qui correspond à un nombre d'échantillons $N=80$. Le rehaussement est réalisé trame par trame et les simulations sont effectuées pour différentes valeurs du nombre d'échantillons M prélevés aléatoirement ($0.4N$, $0.6N$, $0.8N$ et $0.9N$). Le nombre de répétitions de l'algorithme d'acquisition comprimée est fixé à $N_r=10$. Le rapports signal sur bruit (SNR) et le SNR segmental (SNRseg) sont utilisés comme indices pour la comparaison des performances des différentes méthodes de rehaussement.

IV.4 Evaluation des performances de la méthode de rehaussement basée sur l'acquisition comprimée d'ensemble

Dans cette section, on évalue les performances de la méthode de rehaussement basée sur l'acquisition comprimée en termes de SNR et SNR segmental et on les compare à celles de trois méthodes de référence qui sont la soustraction spectrale, le filtrage de Wiener et le débruitage par seuillage des coefficients d'ondelettes. Les SNRs des signaux d'entrée varient de 0 dB à 15 dB par pas de 5 dB. Les résultats sont regroupés selon le type du bruit corrompant le signal.

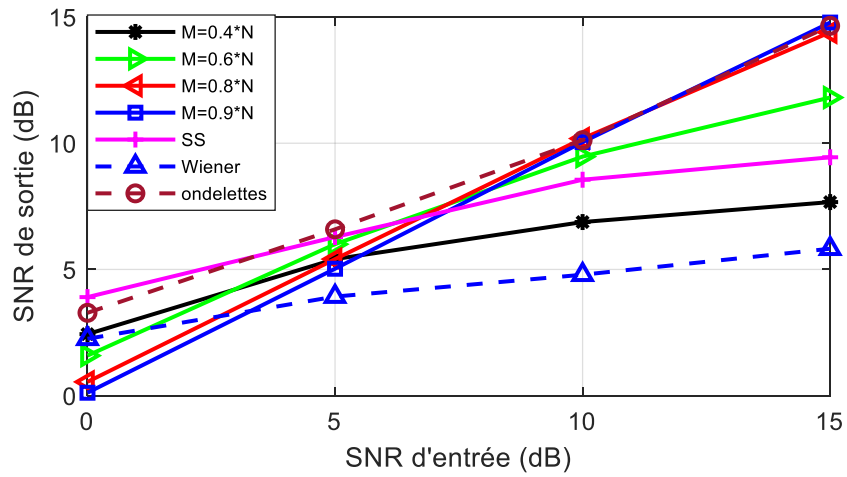
La figure IV.1 montre l'évolution du SNR global de sortie obtenu comme étant la moyenne des SNRs de toutes les phrases pour les différentes méthodes de rehaussement. Comme on peut l'observer à une faible valeur du SNR d'entrée (0 dB), lorsque le rehaussement est effectué par la méthode basée sur l'acquisition comprimée d'ensemble, le plus grand SNR de sortie est obtenu pour $M=0.4N$ et il est comparable aux SNRs obtenus par les méthodes de référence (soustraction spectrale, filtrage de Wiener et seuillage des coefficients d'ondelettes). Lorsque la valeur du SNR du signal d'entrée augmente, nous devons augmenter le nombre d'échantillons M pris aléatoirement pour améliorer le SNR de sortie fourni par la méthode de rehaussement basée sur l'acquisition comprimée d'ensemble. A partir d'un SNR d'entrée de 5 dB, on remarque que la méthode de rehaussement basée sur l'acquisition comprimée d'ensemble donne des SNRs supérieurs à ceux obtenus par le filtre de Wiener quelle que soit la valeur du nombre d'échantillons M prélevés aléatoirement. Pour un SNR d'entrée de 10 dB, la méthode de rehaussement basée sur l'acquisition comprimée présente une meilleure performance en termes de SNR que la soustraction spectrale mais très proche à celle de la méthode de débruitage par seuillage des coefficients d'ondelettes lorsque le nombre d'échantillons M prélevés aléatoirement est supérieur ou égal à $0.8N$. A grand SNR d'entrée, la méthode de rehaussement basée sur l'acquisition comprimée présente des performances meilleures à celles de la soustraction spectrale et le filtrage de Wiener et comparables à celle de la méthode de débruitage par seuillage des coefficients d'ondelettes pour les valeurs de M supérieures ou égales à $0.9N$.

La figure IV.2 montre l'évolution du SNR segmental de sortie obtenu par moyennage sur toutes les phrases pour les différentes méthodes de rehaussement.

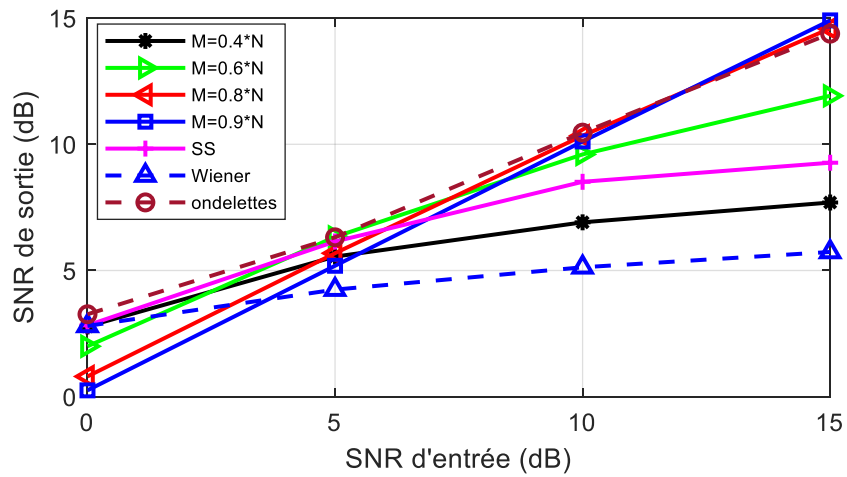
Comme on peut observer, les méthodes de référence présentent, en général, une meilleure performance que l'acquisition comprimée en termes de SNRseg. Ceci peut être expliqué comme suit :

(i) La parole continue contient des segments voisés et non voisés. Les segments voisés sont caractérisés par des grandes amplitudes et ils sont pseudo-périodiques, ils possèdent donc une représentation parcimonieuse dans le domaine de la DCT ou le domaine fréquentiel. Quant aux segments non voisés, ils sont de faibles amplitudes et leur spectre présente une structure semblable à celle du bruit et donc ils ne sont pas parcimonieux. Par conséquent, l'application de la méthode de rehaussement par acquisition comprimée permet de réduire considérablement le bruit dans les segments voisés alors que son efficacité reste très limitée pour les segments non voisés.

(ii) La limitation de l'acquisition comprimée pour le rehaussement des segments non voisés n'est pas reflétée par le SNR global parce que la valeur de ce dernier est influencée surtout par les segments de grande amplitude (segments voisés), la contribution des segments de faible amplitude (non voisés) dans la valeur du SNR global est négligeable. Cette limitation devient apparente lorsqu'on utilise le SNR segmental comme indice de performance. En effet, le SNR segmental est calculé segment par segment, sa valeur est influencée aussi bien par les segments de grandes valeurs (voisés) que par les segments de petites valeurs (non voisés), et par conséquent, la présence de segments non voisés encore bruités après rehaussement par acquisition comprimée d'ensemble affecte considérablement la valeur finale du SNR segmental.

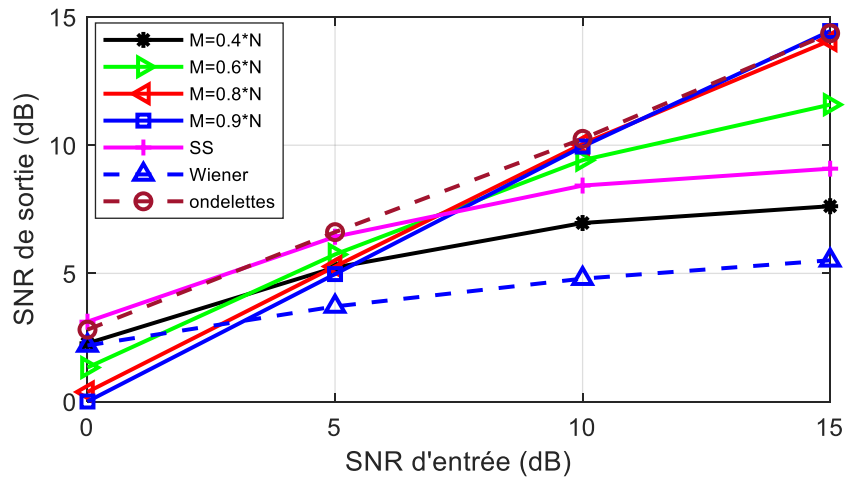


(a) Bruit de voiture

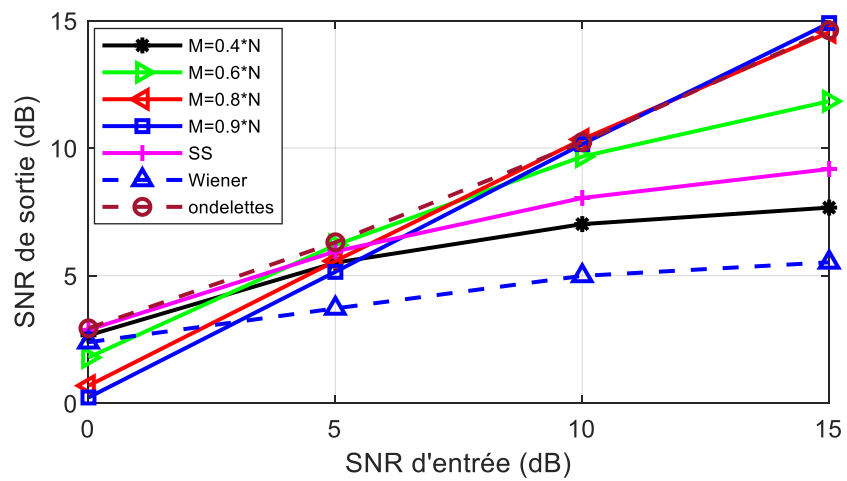


(b) Bruit de train

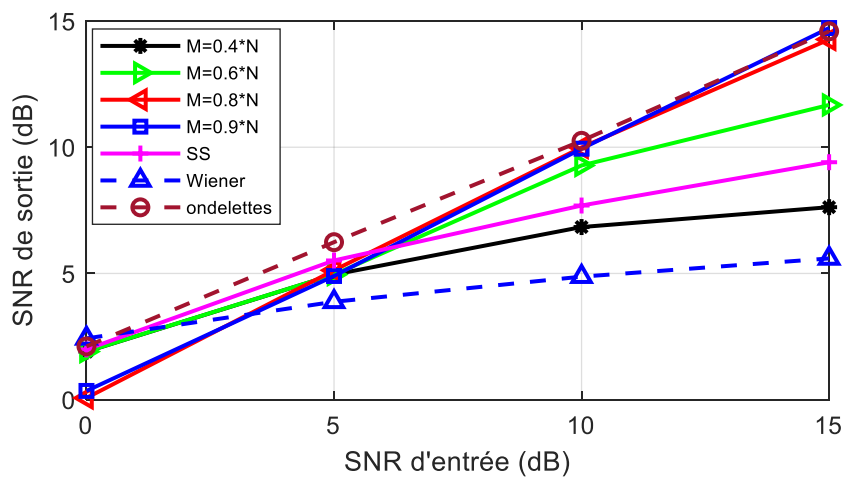
Figure IV.1 : Evolution du SNR de sortie obtenu par moyennage des SNRs de sortie de toutes les phrases pour les différentes méthodes de rehaussement.



(c) Bruit de gare

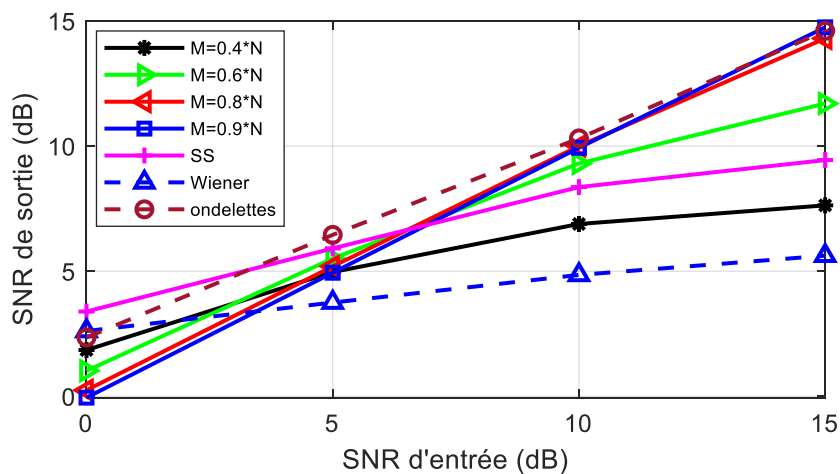


(d) Bruit de rue

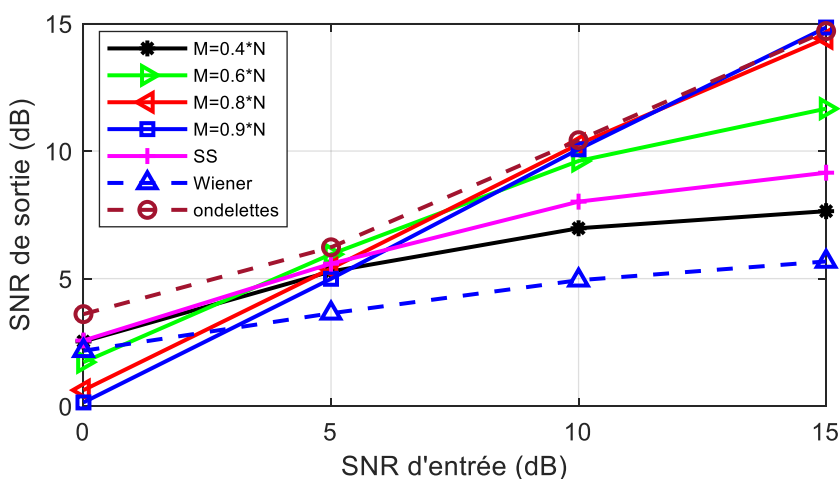


(e) Bruit de restaurant

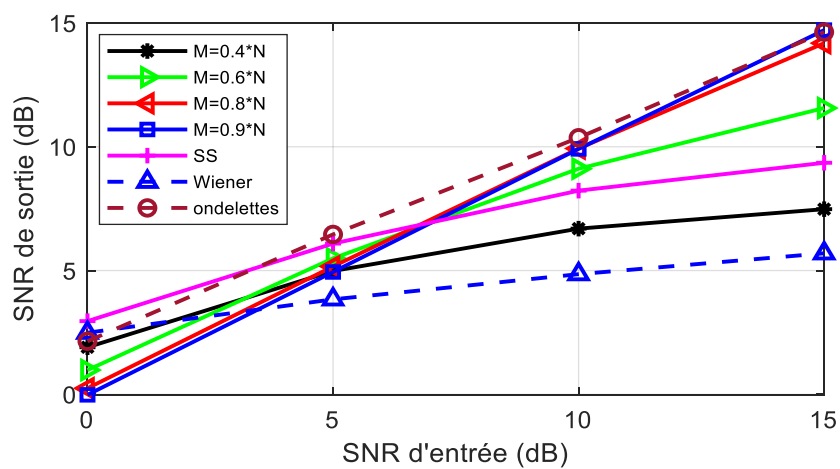
Figure IV.1 (Suite)



(f) Bruit de babillage

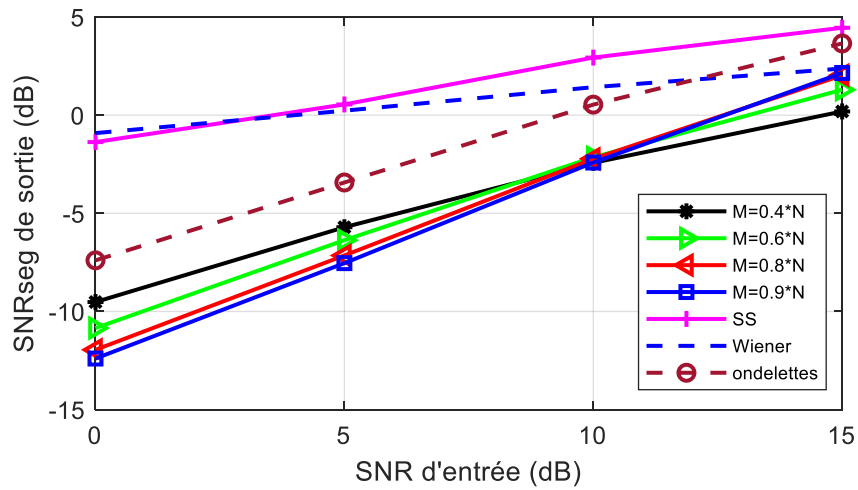


(g) Bruit de hall d'exposition

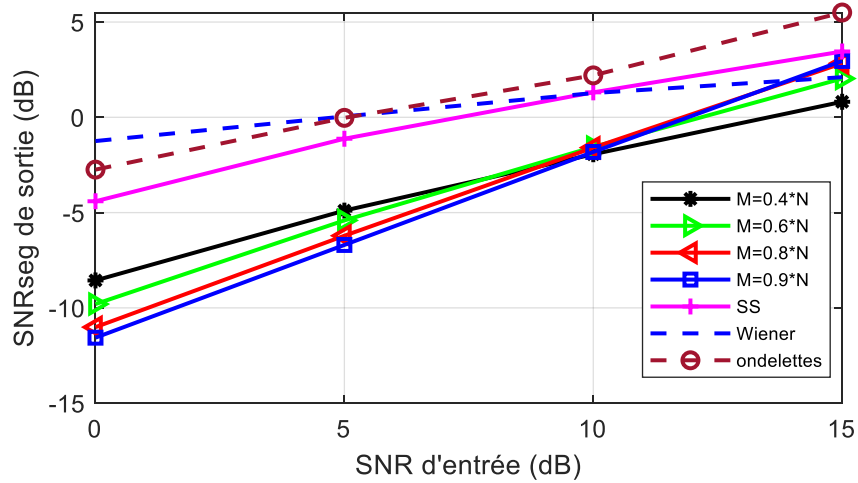


(h) Bruit d'aéroport

Figure IV.1 (Suite)

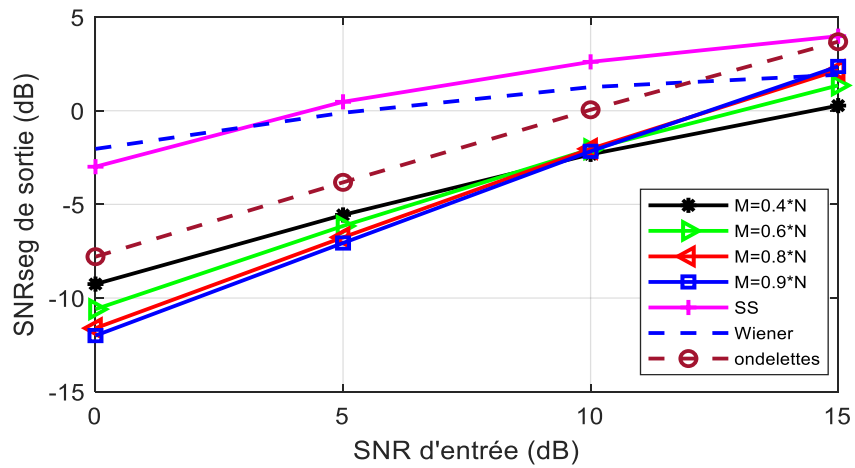


(a) Bruit de voiture

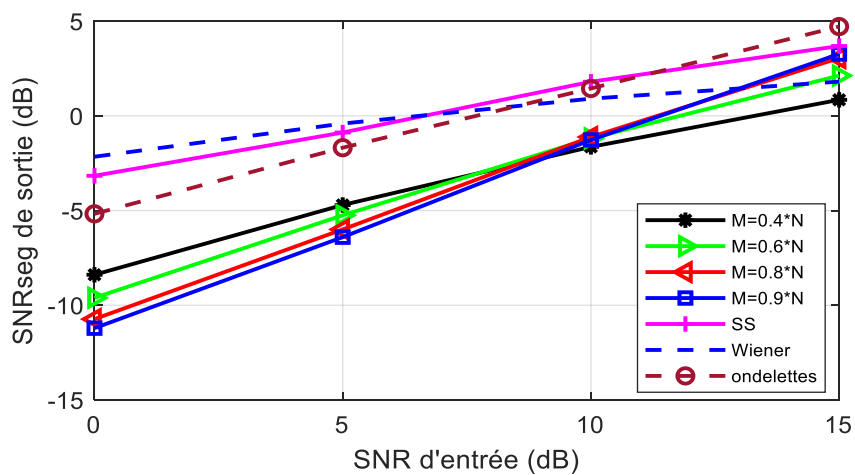


(b) Bruit de train

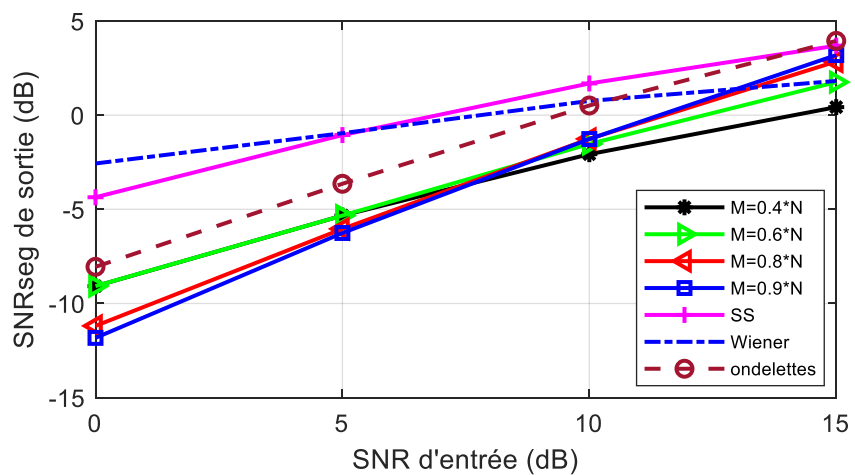
Figure IV.2 : Evolution du SNR segmental de sortie obtenu par moyennage des SNRs segmentaux de sortie de toutes les phrases pour les différentes méthodes de rehaussement.



(c) Bruit de gare

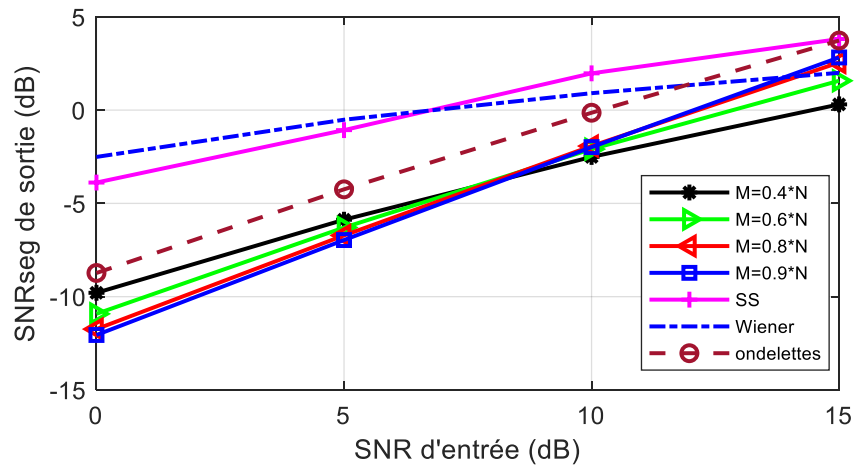


(d) Bruit de rue

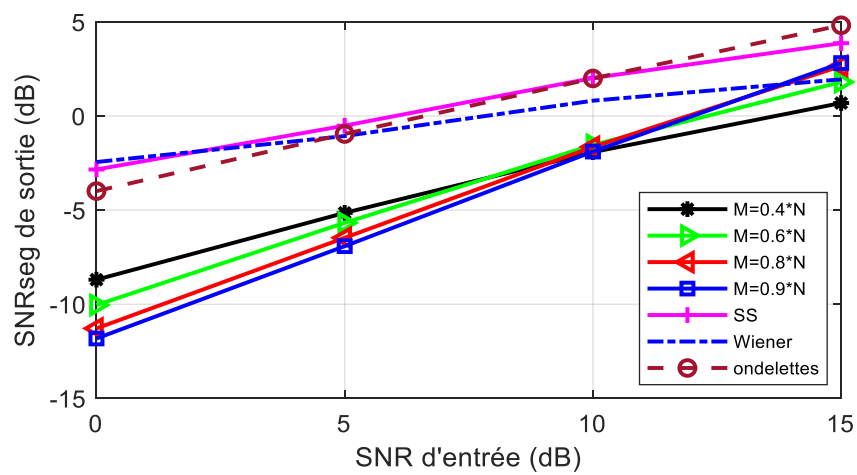


(e) Bruit de restaurant

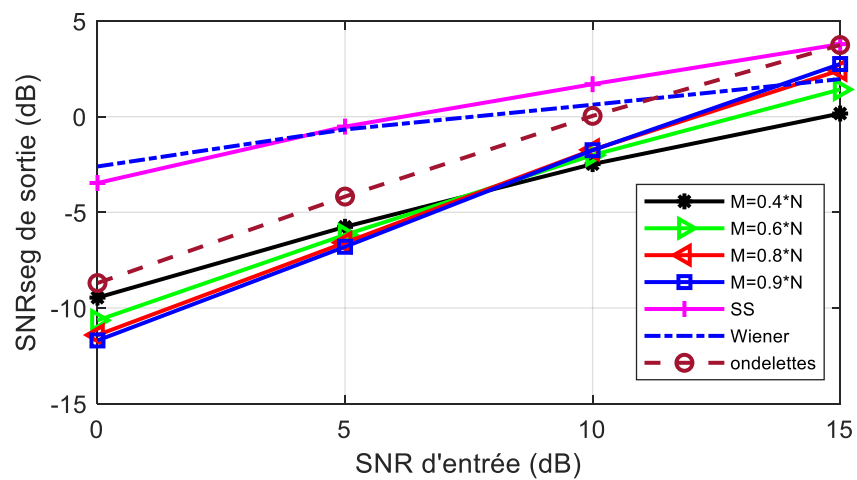
Figure IV.2 (suite)



(f) Bruit de babillage



(g) Bruit de hall d'exposition



(h) Bruit d'aéroport

Figure IV.2 (suite)

IV.5 Evaluation des performances de la méthode de rehaussement combinant l'acquisition comprimée d'ensemble et le seuillage des coefficients d'ondelettes

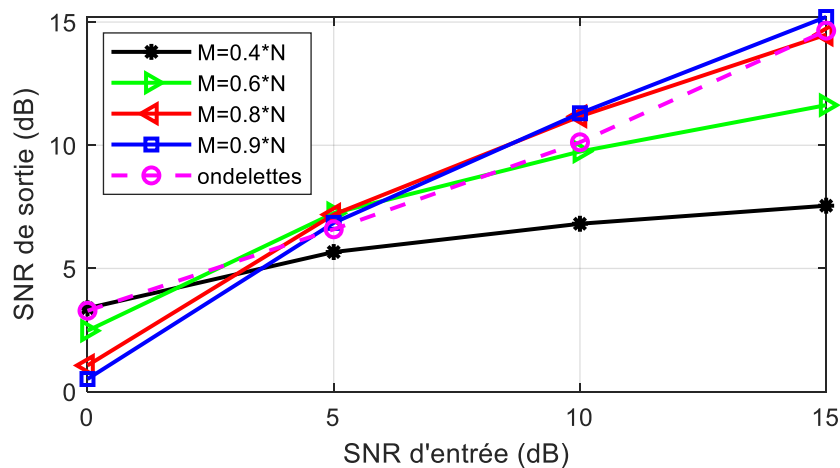
Dans cette section, on évalue les performances de la méthode de rehaussement basée sur l'acquisition comprimée d'ensemble combinée avec le seuillage des coefficients d'ondelettes des segments non voisés et on les compare à celles de la méthode de débruitage par seuillage des coefficients d'ondelettes parce que d'après les résultats précédents, c'est elle qui donne les meilleures performances en termes de SNR et SNRseg comparée à la soustraction spectrale et le filtrage de Wiener.

Le figure IV.3 montre l'évolution du SNR de sortie obtenu par moyennage des SNRs de sortie de toutes les phrases pour la méthode de rehaussement combinant l'acquisition comprimée et le seuillage des coefficients d'ondelettes des segments non voisés ainsi que pour la méthode de débruitage par seuillage des coefficients d'ondelettes seule.

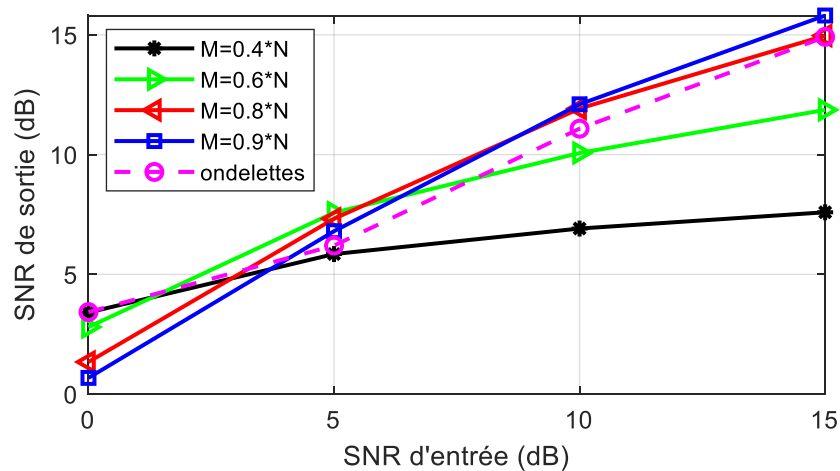
Comme on peut l'observer, quand le SNR d'entrée est égal à 0 dB et la valeur du nombre d'échantillons M prélevés aléatoirement est fixée à $0.4N$, la méthode de rehaussement par acquisition comprimée d'ensemble combinée avec le seuillage des coefficients d'ondelettes des segments non voisés donne une performance généralement meilleure ou comparable à celle de la méthode de débruitage par seuillage des coefficients d'ondelettes seule. A partir d'un de SNR d'entrée de 5 dB, la méthode de rehaussement combinant l'acquisition comprimée et le seuillage des coefficients d'ondelettes des segments non voisés présente une meilleure performance en termes de SNR que celle de la méthode de débruitage par seuillage des coefficients d'ondelettes seule lorsque le nombre d'échantillons M prélevés aléatoirement est supérieur ou égal à $0.6N$.

La figure IV.4 montre l'évolution du SNRseg de sortie obtenu par moyennage des SNRseg de sortie de toutes les phrases pour la méthode de rehaussement combinant l'acquisition comprimée et le seuillage des coefficients d'ondelettes des segments non voisés ainsi que pour la méthode de débruitage par seuillage des coefficients d'ondelettes seule.

On peut observer que lorsque le nombre d'échantillons M prélevés aléatoirement est convenablement choisi, les valeurs du SNR segmental obtenues par la méthode combinant l'acquisition comprimée et le seuillage des coefficients d'ondelettes des segments non voisés sont généralement supérieures à celles obtenues par la méthode de débruitage par seuillage des coefficients d'ondelettes seule.

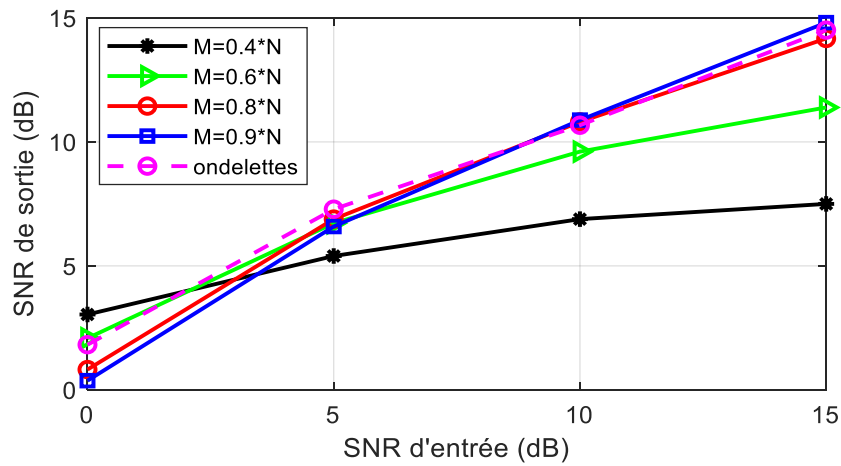


(a) Bruit de voiture

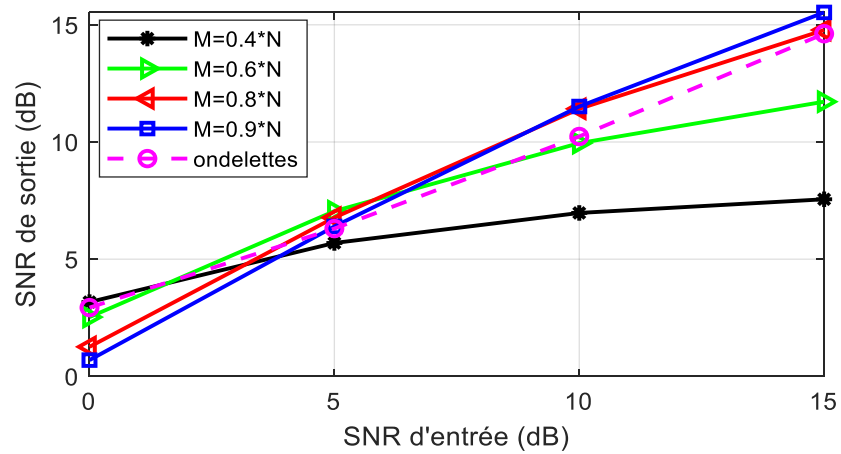


(b) Bruit train

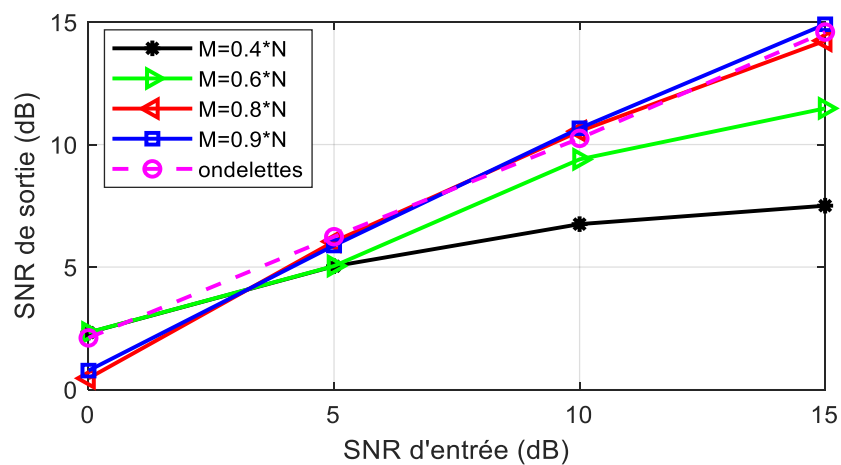
Figure IV.3 : Evolution du SNR de sortie obtenu par moyennage des SNRs de sortie de toutes les phrases pour la méthode de rehaussement combinant l'acquisition comprimée et le seuillage des coefficients d'ondelettes des segments non voisés ainsi que pour la méthode de débruitage par seuillage des coefficients d'ondelettes seule.



(c) Bruit de gare

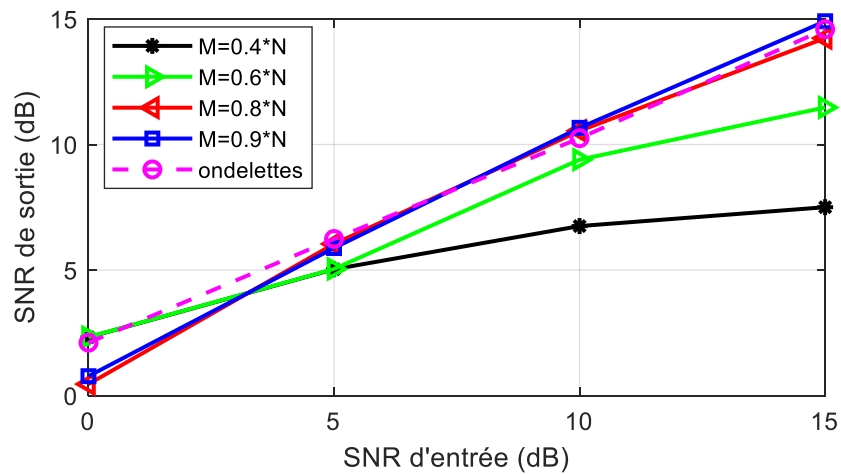


(d) Bruit rue

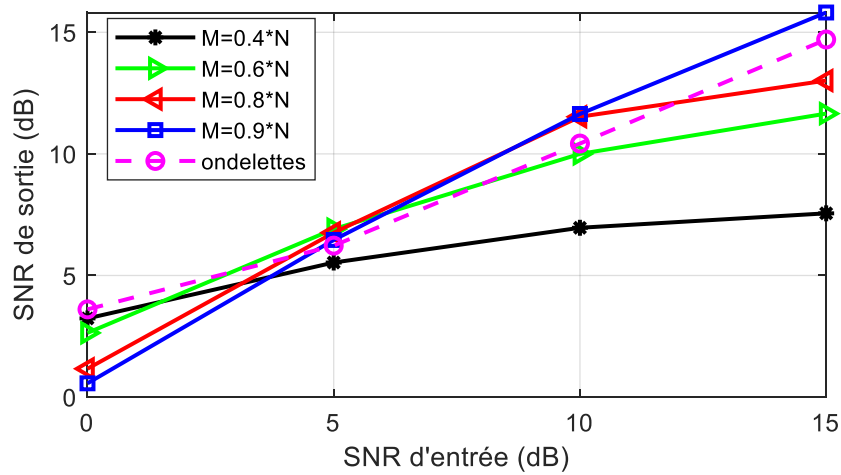


(e) Bruit de restaurant

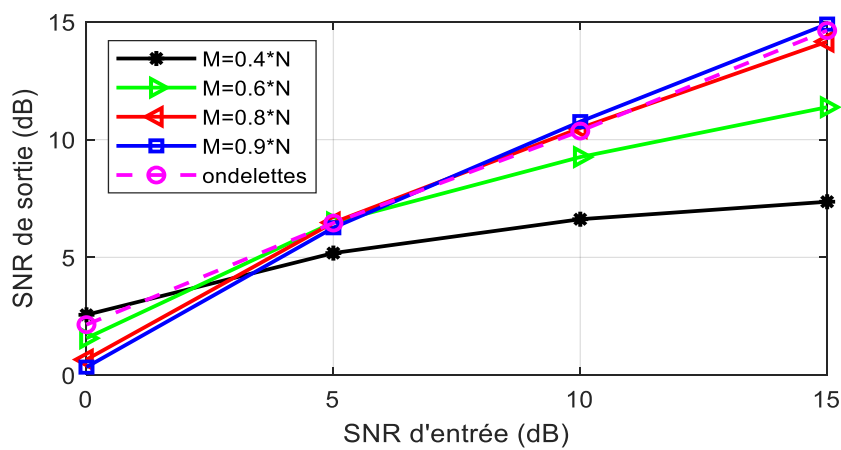
Figure IV.3 (suite)



(f) Bruit de babillage

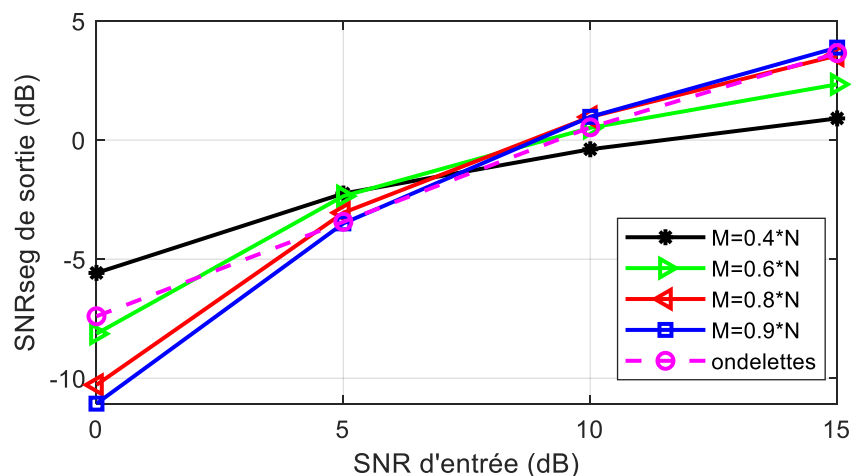


(g) Bruit de hall d'exposition

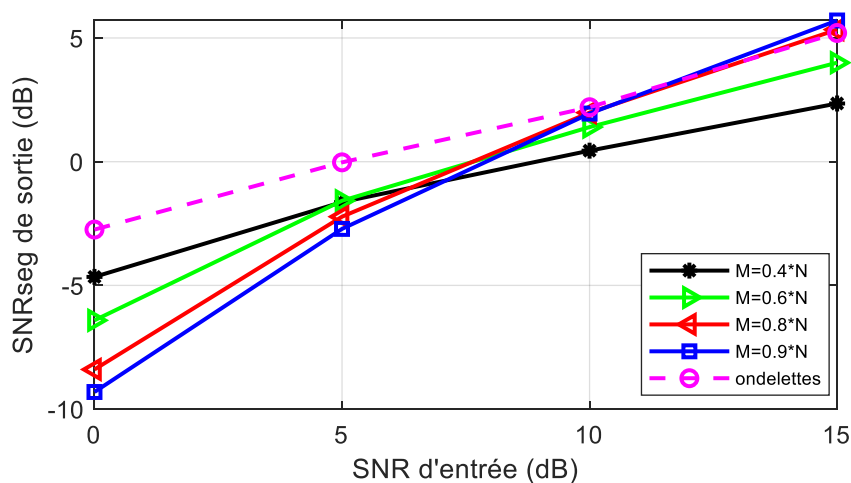


(h) Bruit d'aéroport

Figure IV.3 (suite)

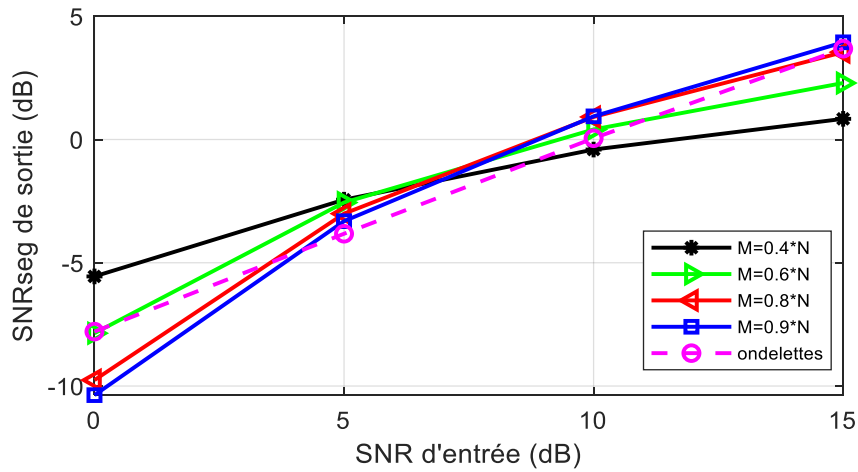


(a) Bruit de voiture

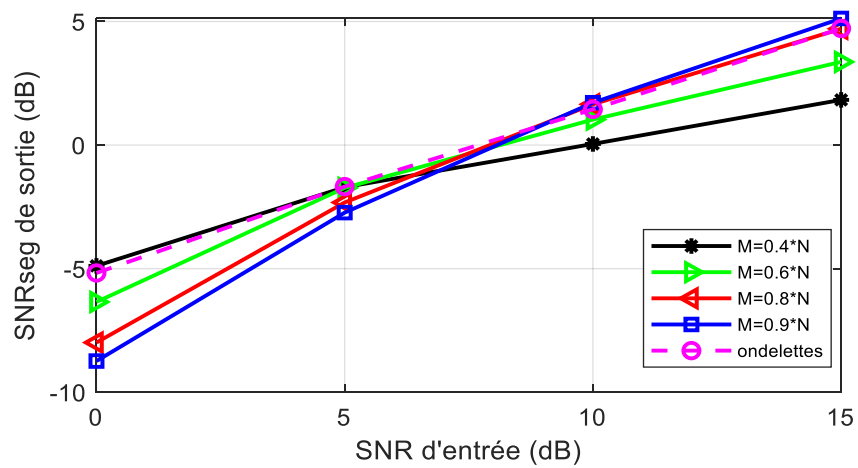


(b) Bruit de train

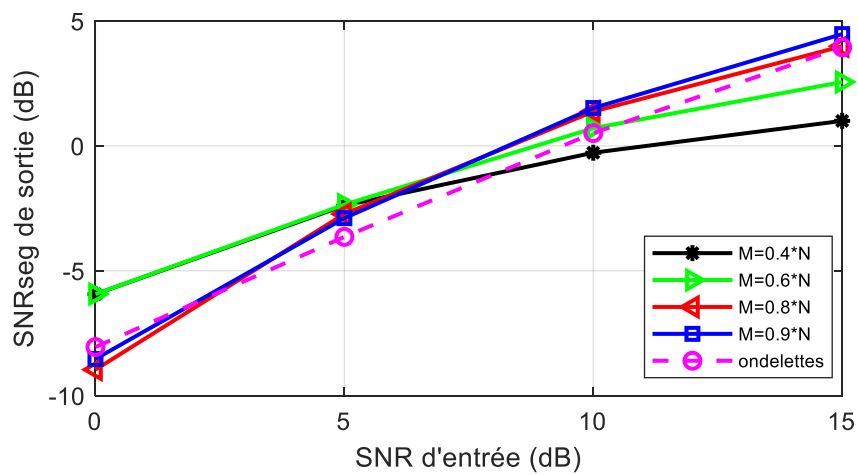
Figure IV.4 : Evolution du SNRseg de sortie obtenu par moyennage des SNRseg de sortie de toutes les phrases pour la méthode de rehaussement combinant l'acquisition comprimée et le seuillage des coefficients d'ondelettes des segments non voisés ainsi que pour la méthode de débruitage par seuillage des coefficients d'ondelettes seule.



(c) Bruit de gare

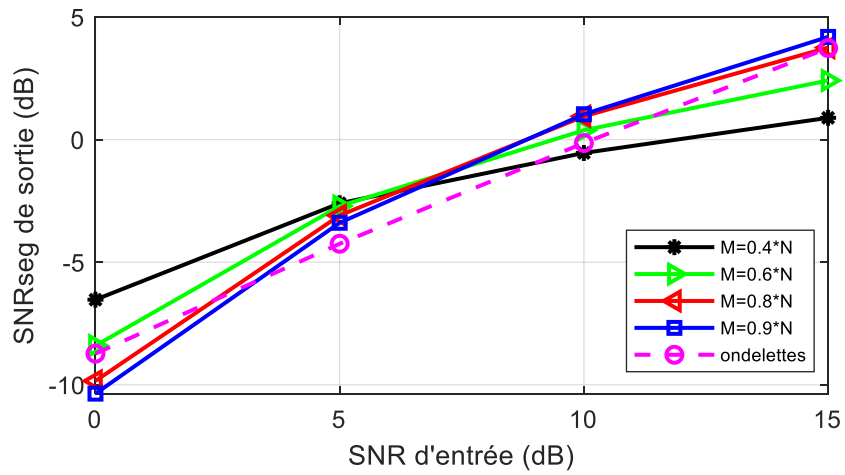


(d) Bruit de rue

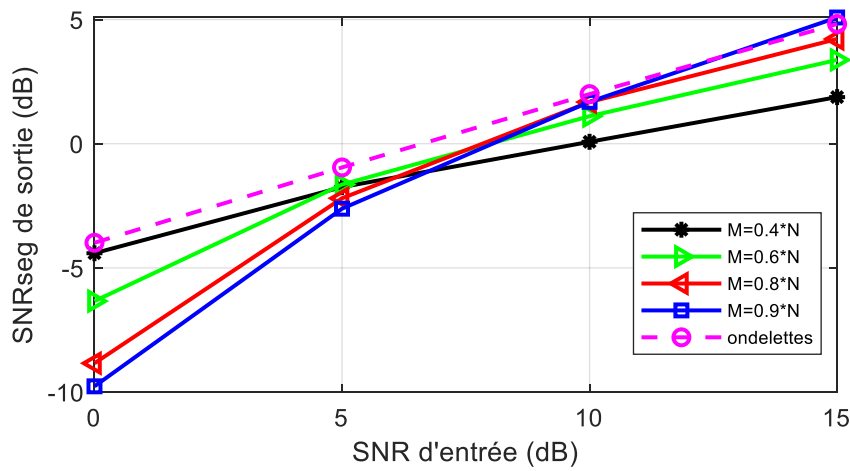


(e) Bruit de restaurant

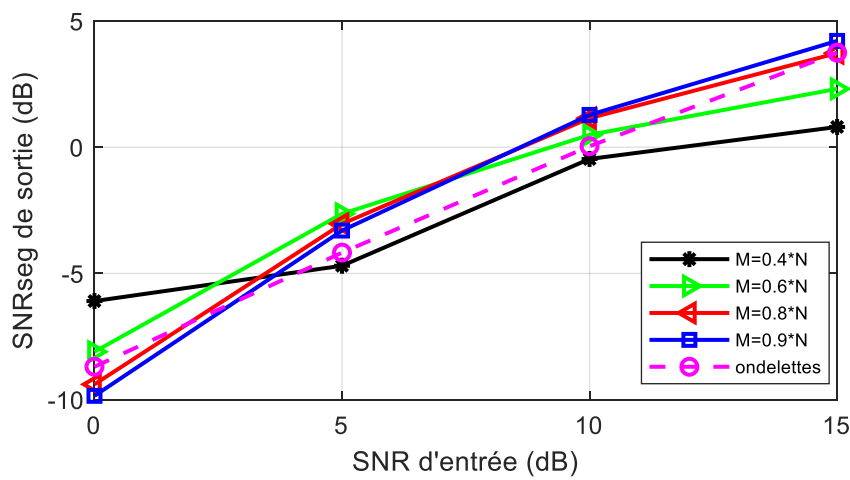
Figure IV.4 (suite)



(f) Bruit de babillage



(g) Bruit de hall d'exposition



(h) Bruit d'aéroport

Figure IV.4 (suite)

IV.6 Conclusion

Dans ce chapitre, nous avons présenté les résultats expérimentaux de la méthode de rehaussement proposée basée sur l'acquisition comprimée d'ensemble ainsi que ceux de la méthode combinant l'acquisition comprimée et le seuillage des coefficients d'ondelettes des segments non voisés. Les performances de ces méthodes de rehaussement en termes de rapport signal sur bruit global et segmental ont été analysées et comparées à celles de trois méthodes de référence qui sont la soustraction spectrale, le filtrage de Wiener et le débruitage par seuillage des coefficients d'ondelettes. Parmi les méthodes de référence, le débruitage par seuillage des coefficients d'ondelettes présente des performances comparables à celles de la méthode de rehaussement par acquisition comprimée d'ensemble. Cependant, en combinant l'acquisition comprimée et le seuillage des coefficients d'ondelettes des segments non voisés, on obtient des performances en termes de SNR et SNRseg meilleures que celles de la méthode de débruitage par seuillage des coefficients d'ondelettes seule.

*Conclusion
générale*

Conclusion générale

L'acquisition comprimée est une nouvelle théorie de compression, elle permet de reconstruire les composants parcimonieux d'un signal à partir d'un petit nombre d'échantillons ou de mesures plus faible que celui déterminé par le théorème de Shannon-Nyquist. Les échantillons sont prélevés aléatoirement pour satisfaire la propriété d'incohérence.

Une technique de rehaussement mono-microphone de la parole basée sur l'acquisition comprimée a été proposée. L'algorithme d'acquisition comprimée et reconstruction du signal est appliqué plusieurs fois sur le signal bruité résultant en plusieurs réalisations de ce dernier. Le signal propre est estimé comme étant la moyenne d'ensemble de toutes les réalisations. L'effet de la longueur du segment et du nombre de répétitions sur les performances de la méthode a été étudié. Un post-traitement des segments non voisés en effectuant un débruitage des par seuillage coefficients d'ondelette a été proposé pour améliorer la performance de la méthode de rehaussement basée sur l'acquisition comprimée.

La méthode de rehaussement basée sur l'acquisition comprimée d'ensemble a été testée sur la base de données NOIZEUS comprenant des signaux de parole corrompus par différents types de bruits à différents SNRs d'entrée. Les performances de cette méthode ont été analysées et comparées à celles de méthodes de référence. Les résultats expérimentaux montrent que la technique de rehaussement combinant l'acquisition comprimée et le débruitage des segments non voisés par seuillage des coefficients d'ondelettes présente des performances en termes de SNR et SNR segmental meilleures que celles de la soustraction spectrale, du filtrage de Wiener et du débruitage par ondelettes seul.

La méthode de rehaussement a été évaluée sur des signaux de parole en présence de bruit additif. L'application de la méthode à la déréverbération et à la réduction du bruit impulsif peut faire l'objet d'un travail futur.

*Références
bibliographiques*

Références bibliographiques :

- [1] Sudirga, R. (2009). *A Speech Enhancement System Based on Statistical and Acoustic-Phonetic Knowledge* (Thèse de doctorat, Queen's University).
- [2] Loizou, P. C. (2007). *Speech enhancement: theory and practice*. CRC press.
- [3] Boll, S. (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on acoustics, speech, and signal processing*, 27(2), 113-120
- [4] Yves Goussard,(2015). Établissement des équations du filtre de Wiener
- [5] Sulong, A., Gunawan, T. S., Khalifa, O. O., & Chebil, J. (2013, December). Speech Enhancement based on Compressive Sensing Algorithm. *In IOP Conference Series: Materials Science and Engineering*, 53(1), 012076.
- [6] Candès, E. J. (2006, August). Compressive sampling. *In Proceedings of the international congress of mathematicians*, 3, 1433-1452.
- [7] Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on information theory*, 52(4), 1289-1306.
- [8] Haneche, H., Boudraa, B., & Ouahabi, A. (2020). A new way to enhance speech signal based on compressed sensing. *Measurement*, 151, 107117.
- [9] Low, S. Y., Pham, D. S., & Venkatesh, S. (2013). Compressive speech enhancement. *Speech Communication*, 55(6), 757-768.
- [10] Chioukh, L., Boussayoud, R. (2020). *Rehaussement multi-microphones du signal de parole*, (Mémoire de Master, Université de Jijel).
- [11] Girard, A. (2010). *Reduction de bruit de signaux de parole mono-capteur basee sur la modelisation par EMD* (Thèse de doctorat, Université de Sherbrooke.).
- [12] Calliope, L., & Fant, G. (1989). *La parole et son traitement automatique*. Paris : Masson.
- [13] Deller, J.R., Hansen, J.H. L., Proakis J. G. (1999). *Discrete Time Processing of Speech Signais*, (3e éd). New Jersey: IEEE Press.
- [14] Liu, S., Zhang, Y. D., Shan, T., & Tao, R. (2018). Structure-aware Bayesian compressive sensing for frequency-hopping spectrum estimation with missing observations. *IEEE Transactions on Signal Processing*, 66(8), 2153-2166.
- [15] Rabiner, L. R. (1978). Digital processing of speech signal. *Digital Processing of Speech Signal*.

- [16] Shim, I., Soraghan, J. J., & Siew, W. H. (2001). Detection of PD utilizing digital signal processing methods. Part 3: Open-loop noise reduction. *IEEE Electrical Insulation Magazine*, 17(1), 6-13.
- [17] Dutoit, T. (2000). Introduction au traitement automatique de la parole. *Notes de cours, Faculté Polytechnique de Mons*.
- [18] Gargouri, D. (2010). *Contribution à l'estimation et à la poursuite des trajectoires de formants de parole*. Thèse, Ecole Nationale d'Ingénieurs de Sfax, Tunisie.
- [19] Bahja, F. (2013). *Détection du fondamental de la parole en temps réel: application aux voix pathologiques* (Thèse de doctorat, Université Mohammed V-Agdal UFR Informatique et Télécommunications Laboratoire LRIT Unité associée au CNRST, URAC 29, Faculté des sciences).
- [20] Aziza, Y. (2018). *Modélisation AR et ARMA de la Parole pour une Vérification Robuste du Locuteur dans un Milieu Bruité en Mode Dépendant du Texte* (Thèse de doctorat, Université de Setif).
- [21] Amehraye, A. (2009). *Débruitage perceptuel de la parole* (Thèse de doctorat, Télécom Bretagne).
- [22] F. Bouderbala, O. Chabouni. (2018). *Proposition d'un algorithme rapide à deux canaux pour la réduction du bruit dans les systèmes téléphoniques à mains-libres*, (Mémoire de Master, Université de Blida-1).
- [23] R. Bendoumia, (2014). *Annulation du bruit par les méthodes de séparation de sources aveugles. Application aux systèmes de télécommunications numériques*, (Thèse de doctorat, Université Blida-1).
- [24] Cottet, F. (2000). *Aide-mémoire de traitement du signal*, Dunod.
- [25] ANUSHA, G., & SHASHIDHAR, M. Phase Estimation for Single and Multi-Channel Speech Enhancement in Multi Source Environment, 4(4), 3307, 3400.
- [26] Virag, N. (1999). Single channel speech enhancement based on masking properties of the human auditory system. *IEEE Transactions on speech and audio processing*, 7(2), 126-137.
- [27] Berouti, M., Schwartz, R., & Makhoul, J. (1979, April). Enhancement of speech corrupted by acoustic noise. In *ICASSP'79. IEEE International Conference on Acoustics, Speech, and Signal Processing* IEEE, 4, 208-211.
- [28] Lim, J. S., & Oppenheim, A. V. (1979). Enhancement and bandwidth compression of noisy speech. *Proceedings of the IEEE*, 67(12), 1586-1604.

- [29] Chergui, L. (2018). *Débrouillage de la parole par de méthodes basées sur les transformées discrètes* (Thèse de doctorat, Université de Setif).
- [30] Doyle, D. J. (1975). Some comments on the use of Wiener filtering for the estimation of evoked potentials. *Electroencephalography and clinical neurophysiology*, 38(5), 533-534.
- [31] Paliwal, K., & Basu, A. (1987, April). A speech enhancement method based on Kalman filtering. In *ICASSP'87. IEEE International Conference on Acoustics, Speech, and Signal Processing*, IEEE, 12, 177-180.
- [32] Mitisi, M., & ITISI, Y. (2003). G. Oppenheim et J. Poggi, « les ondelettes et leur applications ».
- [33] Mallat, S. (july 1989). A theory for multiresolution signal decomposition : the wavelet representation , *IEEE, PAMI*, 11(7), 674-693.
- [34] Mallat, S. (1999). *A wavelet tour of signal processing*. Elsevier.
- [35] Graps, A. (1995). An introduction to wavelets. *IEEE computational science and engineering*, 2(2), 50-61.
- [36] Daubechies, I. (1992). *Ten lectures on wavelets*. Society for industrial and applied mathematics.
- [37] Chun-Lin, L. (2010). A tutorial of the wavelet transform. *NTUEE, Taiwan*.
- [38] Saadoun, A. (2014). *Rehaussement de la parole par les méthodes PCA-VRE* (Thèse de doctorat, Université des sciences et de la technologie Houari-Boumediène,).
- [39] Nasif, H. (2015). Wavelet Applicability in the Area of Signal Processing.
- [40] Chiodi, R. F. (2010). *Détection d'activité vocale basée sur la transformée en ondelettes* (Thèse de doctorat, Université du Québec à Trois-Rivières).
- [41] Ito, K., & Enoki, M. (2007, January). Real-time denoising of AE signals by short time Fourier transform and wavelet transform. In *Proceedings of the 6th International Conference on Acoustic Emission*, 94-99.
- [42] Merry, R. J. E. (2005). Wavelet theory and applications: a literature study. *DCT rapporten, 2005*.
- [43] Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications on pure and applied mathematics*, 41(7), 909-996.
- [44] Strang, G., & Nguyen, T. (1996). *Wavelets and filter banks*. SIAM.

- [45] Luo, G., Zhang, D., & Baleanu, D. D. (2012). Wavelet denoising. *Advances in wavelet theory and their applications in engineering, physics and technology*, 634.//Shim, I., Soraghan, J. J., & Siew, W. H. (2001). Detection of PD utilizing digital signal processing methods. Part 3: Open-loop noise reduction. *IEEE Electrical Insulation Magazine*, 17(1), 6-13.
- [46] Bendoumia, R. (2014). Annulation du bruit par les méthodes de séparation de sources aveugle. application aux système de télécommunications numériques.
- [47] The PESQ Algorithm as the Solution for Speech Quality Evaluation on 2.5G and 3G Network Technical Paper. URL: <http://www.cn.ascom.com/cn/pesq-3g.pdf>
- [48] Pol González, A. (2017). *Definition of a MENA opinion score for VP8 over real-time connections* (Thèse de doctorat, Université de Vigo).
- [49] [Derbal, Abdelhalim](#). (2012). Filtrage adaptatif de Wiener pour la réduction de bruit dans les systèmes Multi-microphones. (Mémoire de master, Université de Blida1).
- [50] Candes, E. J., Romberg, J. K., & Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(8), 1207-1223.
- [51] Haupt, J., & Nowak, R. (2006). Signal reconstruction from noisy random projections. *IEEE Transactions on Information Theory*, 52(9), 4036-4048.
- [52] Eldar, Y. C., & Kutyniok, G. (Eds.). (2012). *Compressed sensing: theory and applications*. Cambridge university press.
- [53] Akbarpour-Kasgari, A., & Ardebilipour, M. (2018). Massive mimo-ofdm channel estimation via distributed compressed sensing. *IEEE Wireless Communications Letters*, 8(2), 376-379.
- [54] Sharma, S. K., Lagunas, E., Chatzinotas, S., & Ottersten, B. (2016). Application of compressive sensing in cognitive radio communications: A survey. *IEEE communications surveys & tutorials*, 18(3), 1838-1860.
- [55] Shi, W., Jiang, F., Liu, S., & Zhao, D. (2019). Image compressed sensing using convolutional neural network. *IEEE Transactions on Image Processing*, 29, 375-388.

- [56] Rousseau, S., & Helbert, D. (2019). Compressive color pattern detection using partial orthogonal circulant sensing matrix. *IEEE Transactions on Image Processing*, 29, 670-678.
- [57] Xu, Y., Li, S., & Zhang, Y. (2019). Privacy-Aware Service Subscription in People-Centric Sensing: A Combinatorial Auction Approach. *CMC-COMPUTERS MATERIALS & CONTINUA*, 61(1), 129-139.
- [58] Siddhi, D., & Naitik, N. (2014). Evaluating performance of compressive sensing for speech signal with various basis. *International Journal of Computer Applications*, 94(11), 975-8887.
- [59] Elad, M., & Aharon, M. (2006). Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12), 3736-3745.
- [60] Bobin, J., Starck, J. L., Fadili, J., & Moudden, Y. (2007). Sparsity and morphological diversity in blind source separation. *IEEE Transactions on Image Processing*, 16(11), 2662-2674.
- [61] Fadili, M. J., Starck, J. L., & Murtagh, F. (2009). Inpainting and zooming using sparse representations. *The Computer Journal*, 52(1), 64-79.
- [62] Candès, E. J., & Wakin, M. B. (2008). An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2), 21-30.
- [63] Davenport, M. A., Duarte, M. F., Eldar, Y. C., & Kutyniok, G. (2012). Introduction to compressed sensing.
- [64] Haneche, H. (2020). *Compressed sensing in mobile systems* (Thèse de doctorat, Université des sciences et de la technologie Houari-Boumédiène).
- [65] S. Mazari K. Belloulata, M. Chikr El Mezouar (2015), Etude comparative des méthodes de reconstruction par Compressive sensing en imagerie IRM, proceeding of SCCBOV, Sidi bel abbes, 02-03.
- [66] Candès, E. J., & Tao, T. (2006). Near-optimal signal recovery from random projections: Universal encoding strategies?. *IEEE transactions on information theory*, 52(12), 5406-5425.
- [67] Donoho, D. L., & Huo, X. (2001). Uncertainty principles and ideal atomic decomposition. *IEEE transactions on information theory*, 47(7), 2845-2862.

- [68] Brunton, S. L., & Kutz, J. N. (2019). *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press.
- [69] Baraniuk, R., Davenport, M., DeVore, R., & Wakin, M. (2008). A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3), 253-263.
- [70] Baraniuk, R. (2007). Compressive sensing [lecture notes]. *IEEE signal processing magazine*, 24(4), 118-121.
- [71] Baron, D., Wakin, M. B., Duarte, M. F., Sarvotham, S., & Baraniuk, R. G. (2006). Distributed compressed sensing. *Rice University, Depart. Electrical and Computer Engineering Technical Report TREE-0612*.
- [72] Dick, C., Harris, F., & Rice, M. (2000, April). Synchronization in software radios. Carrier and timing recovery using FPGAs. *In Proceedings 2000 IEEE Symposium on Field-Programmable Custom Computing Machines* (Cat. No. PR00871), 195-204.
- [73] Valls, J., Sansaloni, T., Perez-Pascual, A., Torres, V., & Almenar, V. (2006). The use of CORDIC in software defined radios: A tutorial. *IEEE communications magazine*, 44(9), 46-50.
- [74] Candes, E., & Romberg, J. (2005). 11-magic: Recovery of sparse signals via convex programming.
URL: www.acm.caltech.edu/11magic/downloads/11magic.pdf, 4, 14.
- [75] Candès, E. J., Romberg, J., & Tao, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2), 489-509.
- [76] Poglitsch, A., Waelkens, C., Geis, N., Feuchtgruber, H., Vandenbussche, B., Rodriguez, L., ... & Wiezorrek, E. (2010). The photodetector array camera and spectrometer (PACS) on the Herschel space observatory. *Astronomy & astrophysics*, 518, L2.
- [77] Candes, E. J., & Tao, T. (2005). Decoding by linear programming. *IEEE transactions on information theory*, 51(12), 4203-4215.
- [78] Hu, S., Lustig, M., Chen, A. P., Crane, J., Kerr, A., Kelley, D. A., ... & Vigneron, D. B. (2008). Compressed sensing for resolution enhancement of

- hyperpolarized ^{13}C flyback 3D-MRSI. *Journal of magnetic resonance*, 192(2), 258-264.
- [79] Lustig, M., Donoho, D. L., Santos, J. M., & Pauly, J. M. (2008). Compressed sensing MRI. *IEEE signal processing magazine*, 25(2), 72-82.
- [80] Yin, W., Osher, S., Goldfarb, D., & Darbon, J. (2008). Bregman iterative algorithms for ℓ_1 -minimization with applications to compressed sensing. *SIAM Journal on Imaging sciences*, 1(1), 143-168.
- [81] Gamper, U., Boesiger, P., & Kozerke, S. (2008). Compressed sensing in dynamic MRI. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 59(2), 365-373.
- [82] Lustig, M., Donoho, D., & Pauly, J. M. (2007). Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(6), 1182-1195.
- [83] Jung, H., Sung, K., Nayak, K. S., Kim, E. Y., & Ye, J. C. (2009). k-t FOCUSS: a general compressed sensing framework for high resolution dynamic MRI. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 61(1), 103-116.
- [84] Sulong, A., Gunawan, T. S., Khalifa, O. O., Kartiwi, M., & Ambikairajah, E. (2016). Speech enhancement based on wiener filter and compressive sensing. *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, 2(2), 367-379.
- [85] Hu, Y., & Loizou, P. C. (2007). Evaluation of objective quality measures for speech enhancement. *IEEE Transactions on audio, speech, and language processing*, 16(1), 229-238.

Résumé

L'acquisition comprimée est une méthode d'acquisition et de compression simultanées des signaux. Elle permet la reconstruction des signaux ayant une représentation parcimonieuse dans certaines bases en utilisant un nombre d'échantillons beaucoup plus faible que celui déterminé par le théorème de Shannon-Nyquist. Dans ce mémoire, on se propose d'utiliser l'acquisition comprimée pour le rehaussement mono-microphone de la parole. La performance de l'acquisition comprimée est limitée quand elle est appliquée au rehaussement des segments non voisés. Pour remédier à cet inconvénient, on se propose d'effectuer un débruitage des segments non voisés par seuillage d'ondelettes afin d'améliorer le rapport signal sur bruit des segments non voisés. La méthode de rehaussement proposée est testée sur la base de données NOIZEUS incluant des signaux de parole corrompus par différents types de bruits et comparée avec les méthodes conventionnelles de rehaussement de la parole, à savoir la soustraction spectrale, le filtrage de Wiener et le débruitage par ondelettes.

Mots clés : acquisition comprimée, parcimonie, rehaussement de la parole, minimisation L_1 .

Abstract

Compressed sensing is a method of simultaneous acquisition and compression of signals. It allows the reconstruction of signals having a sparse representation in certain bases by using a number of samples much smaller than that determined by the Shannon-Nyquist theorem. In this thesis, we propose to use compressed sensing for mono-microphone speech enhancement. The performance of compressed sensing is limited when applied to the enhancement of unvoiced segments. To overcome this drawback, it is proposed to perform a denoising of the unvoiced segments by wavelet thresholding in order to improve the signal-to-noise ratio of the unvoiced segments. The proposed enhancement method is tested on NOIZEUS database which includes speech signals corrupted by different types of noises and compared with conventional speech enhancement methods, namely spectral subtraction, Wiener filtering and wavelet denoising.

Key words: compressed sensing, sparsity, speech enhancement, L_1 minimization.

ملخص

الاستحواذ المضغوط هو طريقة للحصول على الإشارات وضغطها في وقت واحد. يسمح بإعادة بناء الإشارات ذات التمثيل البخل في قواعد معينة باستخدام عدد من العينات أصغر بكثير من تلك التي تحددها نظرية شانون-نيكوست. في هذه الأطروحة، نقترح استخدام الاكتساب المضغوط لتحسين الميكروفون الأحادي للكلام. يكون أداء الاستحواذ المضغوط محدودًا عند تطبيقه على تحسين المقاطع غير المفوترة. لعلاج هذه السلبية، يُقترح إجراء إزالة الضوضاء من المقاطع غير المفوترة عن طريق عتبة الموجة من أجل تحسين نسبة الإشارة إلى الضوضاء للقطاعات غير المفوترة. تم اختبار طريقة التحسين المقترحة على قاعدة بيانات NOIZEUS بما في ذلك إشارات الكلام التي أفسدتها أنواع مختلفة من الضوضاء ومقارنتها بأساليب تحسين الكلام التقليدية، وهي الطرح الطيفي وتصفية Wiener وتقليل الضوضاء بواسطة الموجات.

كلمات مفتاحية: استشعار مضغوط، تناثر، تحسين الكلام، تصغير L_1 .

