

REPUBLICQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

Université de Mohamed El-Bachir El-Ibrahimi - Bordj Bou Arreridj

Faculté des Sciences et de la technologie

Département d'Electronique

Mémoire

Présenté pour obtenir

LE DIPLOME DE MASTER

FILIERE : TELECOMMUNICATION

Spécialité : SYSTEMES DES TELECOMMUNICATION

Par

- **YESSAAD NIDAL**
- **SABOUNE ROUSTOUM**

Intitulé

Une Approche Géométrique Pour La Soustraction Spectrale De La Parole

Devant le Jury composé de :

<i>Nom & Prénom</i>	<i>Grade</i>	<i>Qualité</i>	<i>Etablissement</i>
<i>Pr. AIDEL S</i>	<i>Pr</i>	<i>Président</i>	<i>Univ-BBA</i>
<i>Mr. NASSIM ASBAI</i>	<i>MCA</i>	<i>Encadreur</i>	<i>Univ-BBA</i>
<i>Mr. Dr BENDIB S</i>	<i>MCB</i>	<i>Examineur</i>	<i>Univ-BBA</i>

Année Universitaire 2021/2022

REMERCIEMENTS

Nous remercions en premier lieu le Dieu le tout puissant. C'est grâce à lui que nous avons eu la foi et la force pour accomplir ce travail.

Nous voulons remercier sincèrement Dr. N.ASBAL, Docteur à l'université de BBA, d'abord en tant qu'encadreur de ce mémoire ensuite pour ses précieux conseils, ses incessants encouragements et surtout sa grande disponibilité tout au long de la réalisation de ce travail. Nous le remercions pour toute la confiance accordée à notre égard. Ainsi que pour l'inspiration, l'aide et le temps qu'il a bien voulu nous consacrer sans quoi ce mémoire n'aurait jamais eu autant de succès.

A tous les membres de jury, Vous nous faites le grand honneur en acceptant de juger notre modeste travail, veuillez trouver ici l'expression de nos sincères gratitude et notre grand respect.

Finalement, nous tenons à remercier tous ceux qui ont contribué de près ou de loin à la finalisation de notre travail.

Dédicace

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 frères et mes sœurs.

À toute la famille YESSAAD et BOUZIANE.

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...

NIDAL

Dédicace

**Je dédie ce travail à toute Ma famille, Mes parents,
mers frères, mes sœurs.**

**Je dédie aussi ce travail à mon cher ami « Aymen
Oulha ».**

Roustoum

SOMMAIRE

Introduction générale.....	1
----------------------------	---

Chapitre 1 : concepts théoriques du signal de parole ainsi que ceux de bruit d'environnement

1.1 Introduction.....	4
1.2 Traitement du signal vocal.....	4
1.2.1 Echantillonnage et quantification.....	4
1.2.2 Analyse fréquentielle.....	5
1.2.3 Analyse de fourier a court terme.....	7
1.2.4 Analyse temporelle a court terme.....	7
1.2.5 Blocage de trame.....	8
1.2.6 Fonctions de fenêtrage.....	8
1.3 Bruit dans le signal de parole.....	9
1.3.1 Différents types de bruit.....	9
1.4 Conclusion.....	12

Chapitre 2 : une approche géométrique pour le rehaussement de la parole

2.1 Introduction.....	14
2.2.La soustraction spectacle (ss).....	14
2.3 Une approche géométrique pour la soustraction spectacle.....	16
2.4. Conclusion.....	22

Chapitre 3: résultats et discussions

3.1.Introduction.....	25
3.2.Définition des paramètres communs.....	25
3.2.1 Corpus de tests.....	25
3.2.2 Evaluation des performances.....	25
3.3 Résultats et discussions.....	27
3.4 Conclusion.....	36
Conclusion générale.....	38
Références bibliographiques.....	40

Liste Des Tableaux

Liste Des Tableaux

Titre	Page
Tableau 1.1 Types de techniques de Fourier.	06
Tableau 3.1 Échelle d'évaluation du SIG, BAK et OVL	26
Tableau 3.2 évaluation objective, subjective et SNR de sortie des méthodes SS, SS-NSS, SS-AGA et GA-SE dans un milieu corrompu par des bruits d'aéroport et Babble. Valeur moyenne en utilisant 15 phrases extraites de la base données NOIZEUS. Les meilleures performances sont indiquées en Gras.	29
Tableau 3.3 évaluation objective, subjective et SNR de sortie des méthodes SS, SS-NSS, SS-AGA et GA-SE dans un milieu corrompu par des bruits véhicule et salle d'exposition. Valeur moyenne en utilisant 15 phrases extraites de la base données NOIZEUS. Les meilleures performances sont indiquées en Gras	30
Tableau 3.4 évaluation objective, subjective et SNR de sortie des méthodes SS, SS-NSS, SS-AGA et GA-SE dans un milieu corrompu par des bruits restaurant et rue. Valeur moyenne en utilisant 15 phrases extraites de la base données NOIZEUS. Les meilleures performances sont indiquées en Gras.	31
Tableau 3. 5 évaluation objective, subjective et SNR de sortie des méthodes SS, SS-NSS, SS-AGA et GA-SE dans un milieu corrompu par des bruits train et blanc. Valeur moyenne en utilisant 15 phrases extraites de la base données NOIZEUS. Les meilleures performances sont indiquées en Gras.	32

Liste Des Figures

Liste Des Figures

Titre	page
Figure 1.1 Un signal à temps discret et son enveloppe de variation à temps continu	04
Figure 1.2 Caractéristique du quantificateur avec $n = 3$ bits [10].	05
Figure 1.3 Une fonction de fenêtre de Hamming à 250 échantillons	09
Figure.2.1 Représentation du spectre bruité $Y(\omega_k)$ dans le plan complexe comme la somme du spectre de signal propre $X(\omega_k)$ et du spectre de bruit $D(\omega_k)$	17
Figure.2.2 Triangle montrant la relation géométrique entre les phases de la parole bruitée, du bruit et des spectres de parole propres.	17
Figure 3.1. Spectrogrammes, (a) parole propre, (b) parole bruitée, (c) parole rehaussée en utilisant la méthode SS, (d) parole rehaussée en utilisant la méthode SS-NSS, (e) parole rehaussée en utilisant la méthode SS-AGA, (f) parole rehaussée en utilisant la méthode GA_SE. Cas du bruit blanc 0 dB.	34
Figure 3.2 Les allures temporelle : (a) signal parole propre, (b) signal parole bruitée, (c) signal parole rehaussée en utilisant la méthode SS, (d) signal parole rehaussée en utilisant la méthode SS-NSS, (e) signal parole rehaussée en utilisant la méthode SS-AGA, (f) signal parole rehaussée en utilisant la méthode GA_SE. Cas du bruit blanc 0 dB.	35
Figure 3.3 Evaluation SNR amélioré pour les méthodes SS, SS-NSS, SS-AGA et GA-SE en utilisant : (a) bruit d'aéroport de NOIZEUS, (b) bruit babble de NOIZEUS, (c) bruit de véhicule de NOIZEUS, (d) bruit salle d'exposition de NOIZEUS, (e) bruit restaurant de NOIZEUS, (f) bruit rue de NOIZEUS, (g) bruit train de NOIZEUS, (h) bruit blanc de NOIZEUS.	36

Liste des abréviations

Liste des abréviations

SQNR	rapport signal sur bruit de quantification
SF	Série de Fourier
TF	Transformée de Fourier
TFD	Transformée de Fourier discrète
TFTD	Transformée de Fourier en Temps Discret
RAP	Reconnaissance Automatique de Parole
RSB	Rapport Signal sur Bruit
TFCT	Transformée de Fourier à Court Terme
MMSE	Minimum Mean Square Error
VAD	détecteur d'activité vocale
GA	geometric approche
SNR	Signal to Noise Ratio
SIG	distorsion du signal
BAK	Background Intrusiveness
OVL	Overall Quality
MOS	Mean Opinion Score
segSNR	Segmental Signal to Noise Ratio
PESQ	Perceptual Evaluation of Speech Quality
WSS	Weighted Slop Spectral
LLR	Log Likelihood Ratio
SS	Soustraction Spectrale
SS-NSS	Soustraction spectrale non linéaire
SS-AGA	Soustraction Spectrale basée sur le Moyennage Adaptatif du Gain

Résumé

L'objectif de ce mémoire est d'étudier et de simuler un algorithme de suppression du bruit dans les signaux de parole en utilisant la méthode géométrique de la soustraction spectrale. Cette méthode consiste à estimer les termes croisés qui incluent les différences de phase entre les signaux bruités (et propres) et le bruit proposé. Cette méthode a été mise en œuvre sous Matlab en utilisant toutes les connaissances que nous avons acquises pendant notre cursus universitaires sur le traitement du signal. Les performances de cet algorithme ont été évaluées en comparant la parole nette d'origine et la parole rehaussée traitée à l'aide de mesures objectives et subjectives et du rapport signal sur bruit (SNR) de sortie.

Mots clés: Soustraction spectrale, approche géométrique, rehaussement de la parole.

Abstract

The purpose of this work is to study and simulate an algorithm for reducing noise in speech signals using a geometric method of spectral subtraction. The method involves estimating a cross term that contains the phase difference between the noisy (and clean) signal and the proposed noise. The method is implemented in Matlab using all the knowledge we have acquired in our university signal processing course. The performance of the algorithm is evaluated by comparing the original clean speech with the processed enhanced speech using objective and subjective measures and output signal-to-noise ratio (SNR).

Key words: Spectral subtraction, geometric approach, speech enhancement.

ملخص

الغرض من هذا البحث هو دراسة ومحاكاة خوارزمية لقمع الضوضاء في إشارات الكلام باستخدام الطريقة الهندسية للطرح الطيفي. تتضمن الطريقة تقدير المصطلحات المتقاطعة ، بما في ذلك فرق الطور بين الإشارة الصاخبة (والنظيفة) والضوضاء المقترحة. يتم تنفيذ الطريقة في Matlab باستخدام كل المعرفة التي اكتسبناها في دورة معالجة الإشارات الجامعية. يتم تقييم أداء الخوارزمية من خلال مقارنة الكلام النظيف الأصلي والكلام المحسن المعالج باستخدام قياسات موضوعية وذاتية وقياسات نسبة الإشارة إلى الضوضاء (SNR).

الكلمات المفتاحية: الطرح الطيفي ، النهج الهندسي ، تحسين الكلام.

Introduction Générale

Introduction générale

Le rehaussement de la parole est un domaine de recherche très actif qui s'est largement développé durant ces trois dernières décennies. Sa nécessité s'est manifesté en particulier dans de nombreuses situations, dans lesquelles le signal vocal provient soit d'une source bruitée, soit il est affecté directement par le bruit du canal des transmissions ou d'environnement.

En effet, le rehaussement peut être utilisé comme un prétraitement au niveau des codeurs de la parole, mis en place dans les systèmes de radiocommunications de types GSM, 3G, 4G, etc. Si ces systèmes, soumis à des bruits environnementaux typiques (véhicule, restaurant, car, rue, etc.), sont équipés de dispositifs de Reconnaissance Automatique de la Parole (RAP) (exemple : numérotation vocale), le taux de reconnaissance se voit donc décroître assez rapidement en présence de ces types de bruits [1]. Pour remédier à ces contraintes, plusieurs techniques de rehaussement de la parole ont été développées sur la base d'hypothèses de présence du bruit [2-5] et ce, en vue d'améliorer davantage le taux de reconnaissance global.

Aussi, dans le cas des communications vocales solaires, le rehaussement de la parole permet d'améliorer la qualité et l'intelligibilité de la parole du pilote qui est mélangée avec le bruit présent dans le cockpit. Pour leur part, les systèmes de téléconférence sont caractérisés par le phénomène de réverbération de la salle qui provoque la diffusion du bruit de la source sur toute la salle. Pour cela, on procède au rehaussement de la parole avant sa transmission.

Enfin, dans le cadre de la réhabilitation auditive des malentendants, le rehaussement de la parole permet de maintenir au plus haut niveau les performances des prothèses auditives qui diminuent assez rapidement dans des milieux fortement bruités.

Idéalement, les algorithmes de rehaussement doivent être en mesure d'améliorer la qualité et l'intelligibilité du signal. Cependant, en pratique, ces algorithmes agissent uniquement sur la qualité par la réduction du bruit additif et tout en introduisant des distorsions qui affectent plus ou moins l'intelligibilité du signal reçu. Pour cela, l'enjeu principal dans la conception d'un algorithme efficace de rehaussement de la parole consiste à supprimer le bruit sans pour autant introduire des distorsions perceptibles dans le signal rehaussé.

Par ailleurs, la mise en œuvre d'une méthode de rehaussement de la parole nécessite a priori des connaissances sur le comportement du bruit [1], [6] (niveau de pression sonore, Rapport Signal à Bruit (RSB ou SNR), masquage du bruit, etc.). Aussi, l'estimation du bruit

peut avoir un impact majeur sur la qualité du signal rehaussé, au sens où si cette estimation est trop faible, le bruit résiduel devient audible et si elle est trop élevée, la parole sera distordue engendrant ainsi une perte d'intelligibilité.

Aussi la plupart des techniques de rehaussement de la parole basées sur la soustraction spectrale, supposent que les termes croisés dans l'équation de la soustraction spectrale de puissance sont nuls. Alors qu'en réalité ils sont toujours là présents, du coup, ceci mène à la présence du bruit résiduel (bruit musical) dans le signal rehaussé. Donc, pour palier au problème du bruit musical dans les signaux rehaussés, nous nous intéressons dans ce travail de fin d'étude, à l'étude d'une méthode de soustraction spectrale basée sur l'aspect géométrique du vecteur (signal) bruité dans le plan complexe, afin de prendre en compte les termes croisés présents dans toutes les soustractions spectrales de puissance. Donc, éliminer au maximum l'effet du bruit musical dans les signaux améliorés.

Ce mémoire est organisé en trois chapitres, deux chapitres pour la partie théorique (chapitre 1 et 2) et un chapitre pour la partie pratique (chapitre 3).

Le premier chapitre comprend des notions fondamentales sur le signal de parole, la TF des signaux et la TF à court terme, ainsi que le fenêtrage et les différents types de bruit qui peuvent affecter le signal de parole.

Le deuxième chapitre comporte la présentation en détail la méthode de rehaussement de la parole : méthode géométrique (GA).[7]

Le troisième chapitre est réservé à la partie simulation et présentation des mesures et résultats des méthodes de rehaussement de parole étudiées pour la discussion et l'évaluation.

CHAPITRE 1 :

*Concepts théoriques du signal de parole
ainsi que ceux de bruit d'environnement*

1.1 Introduction

Ce chapitre présente les bases théoriques fondamentales du traitement numérique du signal de parole, il englobe les principales techniques d'analyse qui sont utilisées dans la phase de traitement (ou prétraitement) du signal de parole. Ce chapitre met aussi l'accent sur l'étude de différentes caractéristiques temporelles et fréquentielles des bruits d'environnements qui affectent la parole.

1.2 Traitement du signal vocal

En traitement du signal, un signal n'est rien d'autre qu'une mesure expérimentale de tout événement [8].

Il existe deux types de signaux : analogiques et numériques. Un signal avec un domaine et une plage continue est appelé un signal analogique. Les signaux analogiques sont difficiles à stocker sur un ordinateur pour le traitement. Par conséquent, un signal analogique est d'abord converti en signal numérique, puis analysé sur un système informatique, où les plages et les domaines sont discrets [9].

1.2.1 Échantillonnage et quantification

L'échantillonnage et la quantification sont les deux premières étapes de tous les systèmes de traitement numérique du signal et de communication numérique dotés d'entrées analogiques. Un signal doit être échantillonné à un taux de plus de deux fois le contenu de fréquence le plus élevé du signal, sinon le processus d'échantillonnage entraînera une perte d'informations et une distorsion.

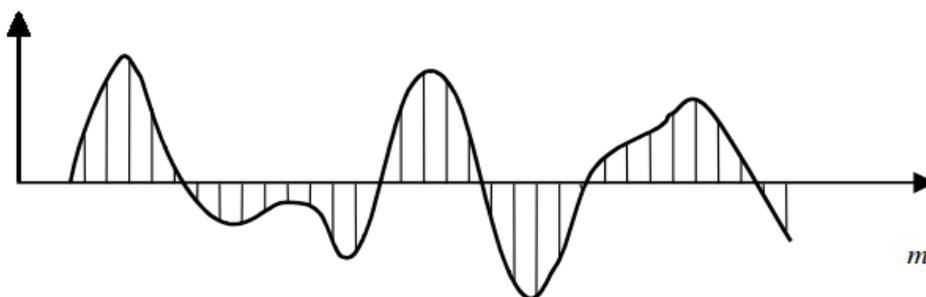


Figure 1.1 Un signal à temps discret et son enveloppe de variation à temps continu. [10]

Par contre la quantification est le processus de conversion de chaque échantillon de valeur continue pour indiquer une valeur discrète qui peut être attribuée à un mot de code numérique

unique. Pour le traitement du signal numérique, des échantillons d'amplitude continue pendant le temps discret de l'échantillon sont quantifiés, conservés et mappés en chiffrements binaires à n -bits avant d'être stockés et traités [10].

Pour la quantification en mots chiffrés de n -bits, la plage d'amplitude du signal est divisée en 2^n niveaux de quantification. Chaque échantillon est un élargissement continu il est quantifié au niveau de quantification le plus proche, puis mappé au code binaire n -bit attribué niveau.

Le mappage entre un échantillon analogique $x_a(m)$ et son quantifié la valeur $x(m)$ peut être exprimée comme :

$$x(m) = Q[x_a(m)] \tag{1.1}$$

Où Q est la fonction de quantification.

Les performances d'un quantificateur sont mesurées par le rapport signal sur bruit de quantification (SQNR).

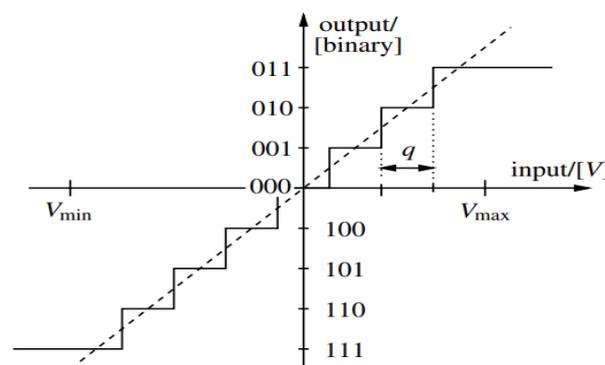


Figure 1.2 Caractéristique du quantificateur avec $n = 3$ bits [11].

1.2.2 Analyse fréquentielle

L'analyse de Fourier est une technique mathématique qui est appliquée pour transformer un signal du domaine temporel vers un domaine fréquentiel. Selon la nature et la périodicité du signal, on peut choisir la technique de Fourier appropriée. [9]

Tableau 1.1 Types de techniques de Fourier.

Propriétés du domaine temporel	Périodique	Apériodique
Continue	Série de Fourier (SF)	Transformée de Fourier (TF)
Discret	Transformée de Fourier discrète (TFD)	Transformée de Fourier en Temps Discret (TFTD)

La série de Fourier (SF) d'un signal périodique $x(t)$ de période T est définie comme :

$$x(m) = Q[x_a(m)] \tag{1.2}$$

$$x(t) = \sum_{k=-\infty}^{\infty} c_k e^{j2\pi kt/T} \tag{1.3}$$

La transformée de Fourier (TF) d'un signal apériodique $x(t)$ est définie comme :

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \tag{1.4}$$

$$x(t) = \int_{-\infty}^{\infty} X(\omega) e^{j\omega t} d\omega \tag{1.5}$$

La Transformée de Fourier Discrète (TFD) d'un signal acoustique périodique discret $x[n]$ avec une période N est définie comme :

$$c_k = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-j2\pi n/N}; \tag{1.1}$$

$$x[n] = \sum_{k=0}^{N-1} c_k e^{j2\pi n/N}; \tag{1.2}$$

La Transformée de Fourier en Temps Discret (TFTD) d'un signal apériodique numérisé $x[n]$ est définie comme :

$$X(\omega) = \sum_{n=-\infty}^{\infty} x[n] e^{-j\omega n}; \tag{1.8}$$

$$x[n] = \frac{1}{2\pi} \int_{-2\pi}^{2\pi} X(\omega) e^{j\omega n} d\omega; \tag{1.3}$$

1.2.3 Analyse de Fourier à court terme

En utilisant l'analyse de Fourier d'un signal $x[n]$, à partir de l'Eq (1.8), TFTD de la trame $x_m[n]$ est expliqué comme :

$$X(m, \omega) = X(\omega) = \sum_{n=-\infty}^{\infty} x_m[n] e^{-j\omega n} = \sum_{n=-\infty}^{\infty} x[n] w[m-n] e^{-j\omega n}; \quad (1.10)$$

Cette Eq (1.10) définit la TFTD à court terme pour le signal de parole $x[n]$.

1.2.4 Analyse temporelle à court terme

Les signaux de parole sont de nature non stationnaire, ce qui signifie que leurs paramètres statistiques tels que l'intensité et la variance varient dans le temps [8]. Les signaux vocaux peuvent être stationnaires pendant une période plus courte, mais lorsqu'ils sont considérés sur une durée plus longue, ils sont apériodiques. Par conséquent, la transformée de Fourier n'est pas une technique appropriée pour l'analyse de la parole car elle nécessite un signal périodique pendant un temps infini. Une technique appelée analyse à court terme est utilisée. Dans cette technique, le signal est divisé en trames ou segments courts, en supposant que le signal est stationnaire dans cette trame courte et en analysant chaque trame ou segment séparément. La durée de chaque trame est d'environ 10-20 ms, suffisamment courte pour satisfaire l'hypothèse. Le spectrogramme est un exemple d'analyse de temps court qui est discuté dans la section suivante. La transformée de Fourier en temps discret est appliquée à chaque image, ce qui donne des spectres dans le temps [9].

Pour un signal $x[n]$ donné, le signal temps court $x_m[n]$ de la trame m est représenté par :

$$x_m[n] = x[n]w_m[n]; \quad (1.4)$$

Où $w_m[n]$ est une fonction de fenêtre, au-delà d'une région spécifique, sa valeur est zéro. Pour cette technique, $w_m[n]$ doit être égal pour toutes les images. Donc les calculs sont comme :

$$w_m[n] = w[m-n] \quad (1.12)$$

$$w[n] = \begin{cases} \hat{w}[n] & |n| \leq \frac{N}{2} \\ 0 & |n| > \frac{N}{2} \end{cases}; \quad (1.13)$$

Où N est la longueur de la fenêtre.

1.2.5 Blocage de trame

Un signal de parole est généralement défini comme un signal non stationnaire avec un changement rapide dans le temps et l'espace. Par conséquent, pour surveiller son comportement statistique, il est nécessaire de diviser l'ensemble du signal traité $x(n)$ en petites parties. Chaque fragment est appelé une trame, dans laquelle l'échantillon de parole est supposé être approximativement constant avec un comportement statistique constant. La parole semble être fixée en choisissant une taille de trame comprise entre 8 ms et 30 ms. Le processus de segmentation d'un échantillon de parole en petites trames est appelé blocage (découpage en) de trame [11].

L'échantillon de parole est bloqué de sorte qu'il y ait une certaine interférence avec la trame suivante dans les échantillons bloqués [12].

1.2.6 Fonctions de fenêtrage

La fonction de fenêtre réduit le début et la fin de la trame à zéro et la distorsion spectrale des deux côtés. Ainsi, lorsque le signal de parole $s(n)$ est multiplié par la fenêtre $w(n)$, il produit un ensemble d'échantillons de parole $x(n)$ qui sont pondérés par l'intervalle de temps et la forme de la fenêtre (fonction rectangulaire simple).

La fenêtre de Hamming est une autre fonction couramment utilisée dont la partie non nulle est une impulsion cosinus surélevée décrite comme:

$$w[n] = \begin{cases} 0.54 - 0.46 \cos \frac{2\pi n}{N-1} & 0 \leq n \leq N-1 \\ 0 & \text{ailleurs} \end{cases}; \quad (1.5)$$

Où n est le nombre d'échantillons dans une trame donnée et N représente le nombre total d'échantillons dans la trame. La fonction de fenêtre de Hamming est représentée sur la Figure 1.3, où l'axe des x représente le nombre d'échantillons et l'axe des y indique l'amplitude des échantillons.

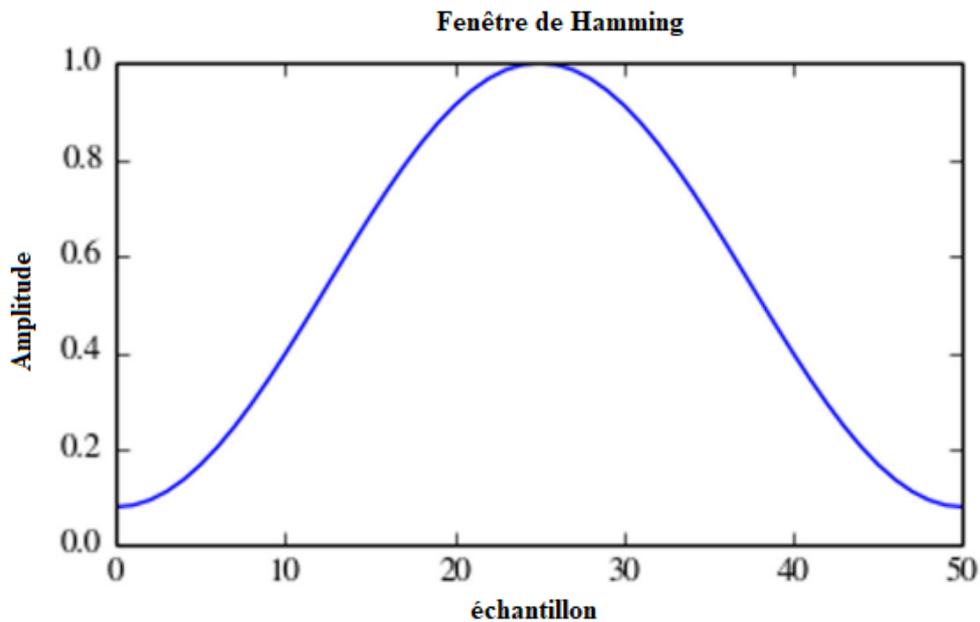


Figure 1.3 Une fonction de fenêtre de Hamming à 250 échantillons [9].

1.3 Bruit dans le signal de parole

Dans le traitement du signal, le bruit est un terme général désignant les modifications indésirables (et, en général, inconnues) qu'un signal peut subir lors de la capture, du stockage, de la transmission, du traitement ou de la conversion. Le bruit dans le signal audio est un son indésirable considéré comme désagréable, fort ou perturbateur pour l'ouïe. D'un point de vue physique, il n'y a pas de distinction entre le bruit et le son désiré, car les deux sont des vibrations à travers un milieu, tel que l'air ou l'eau. La différence survient lorsque le cerveau reçoit et perçoit un son.

1.3.1 Différents types de bruit

- **Bruit continu** : c'est le bruit qui est produit en continu, par exemple, par des machines qui continuent de fonctionner sans interruption. Cela peut provenir de l'équipement d'usine, du bruit du moteur ou des systèmes de chauffage et de ventilation.

- **Bruit intermittent** : Le bruit intermittent est un niveau de bruit qui augmente et diminue rapidement. Cela peut être dû au passage d'un train, à un équipement d'usine qui fonctionne par cycles ou à un avion en vol.

- **Bruit impulsif :** Le bruit impulsif est le plus souvent associé à l'industrie de la construction et de la démolition. Ces éclats de bruit soudains peuvent vous surprendre par leur nature rapide et surprenante. Les bruits impulsifs sont généralement créés par des explosions ou des équipements de construction.

- **Bruit à basse fréquence :** Le bruit à basse fréquence est généralement défini sur les bandes communes à 500 Hz ou moins. Cela signifie que les gens sont plus susceptibles de ressentir les vibrations du son plutôt que de l'entendre. Le bruit à basse fréquence a également des longueurs d'onde plus longues, peut parcourir de longues distances et a une endurance élevée. Les bruits produits par les moyens de transports se caractérisent par une très forte stationnarité qui correspond à la vitesse de fonctionnement des organes moteurs

- **Bruit à haute fréquence :** Les sons à haute fréquence contiennent de courtes ondes sonores d'une fréquence de 5 000 HZ ou plus. Les ondes sonores à haute fréquence sont réfléchies lorsqu'elles rencontrent des objets minces et ne se plient pas autant autour des barrières.

- **Bruits rythmiques :** Ce sont des bruits correspondants à la répétition d'une tâche de nature productive. Ils sont très souvent des bruits périodiques, produits par les systèmes industriels. Ce bruit est assez intense, mais l'énergie des raies n'est pas beaucoup plus importante. Il est, en outre possible d'observer la présence de bruits aléatoires supplémentaires en basses fréquences, entre 0 et 3000 Hz. Ce bruit est cependant très étendu [11].

- **Bruits aléatoires :** Ce sont des bruits associés aux ateliers de fabrication et aux usines. Ces bruits peuvent générer de nombreuses erreurs dans les systèmes de Reconnaissance Automatique de Parole (RAP). L'univers de la production est généralement très bruité. Un exemple des bruits aléatoire est le bruit de soudures qui correspond à des raies verticales à des instants aléatoires dans le temps.

- **Les bruits additifs :** Un bruit additif est causé par plusieurs systèmes de communication dans le même environnement. De nombreux émetteurs et de nombreux récepteurs peuvent être confinés dans un même espace, ainsi les messages de tous les émetteurs peuvent être en concurrence sur le même canal sans que les récepteurs disposent d'un mécanisme infaillible pour isoler le message qui leur est attribué. L'émetteur et le récepteur peuvent également être en présence d'un ou plusieurs dispositifs générant un bruit de fond de puissance variable. Le bruit supplémentaire peut être divisé en trois groupes selon l'endroit où il peut être rencontré [13] :

- **Bruits des systèmes industriels**
- **Bruits des moyens de transport**
- **Bruits des milieux administratifs et urbains**

• **Les bruits convolutionnels :** Les bruits convolutionnels (ou multiplicatifs) sont dus à la distorsion induite par la voie de communication. Ils résultent de la mauvaise qualité d'un ou de plusieurs éléments de support du message ou de son étroitesse en bande passante. Les moyens de communication à longue distance (la téléphonie, la radiophonie et la radiotéléphonie) sont élaborés à partir d'un compromis coût/efficacité. La parole lorsqu'elle est transmise est forcément dégradée en qualité et en intelligibilité. Un des champs possibles d'application de la RAP sont les serveurs vocaux accessibles par les lignes téléphoniques. Mais la parole transmise par téléphone souffre de déformations variables induites par la qualité de la connexion. Une transmission peut ainsi souffrir de l'étroitesse de la bande passante, de la mauvaise qualité des microphones de certains terminaux téléphoniques. La parole enregistrée dans tous les corpus utilisés pour la recherche est, en effet, toujours bruitée puisque le microphone utilisé effectue toujours un filtrage linéaire [13].

• **Les bruits physiologiques :** D'autres bruits peuvent également être considérés dans le domaine du traitement de la parole, mais ils n'ont pas la généralité des bruits de type additifs ou convolutionnels car ils sont spécifiques à l'être humain lors de sa phase de production de parole. La plupart des systèmes de RAP (Reconnaissance Automatique de Parole) fonctionnent mal en milieu bruité car les contraintes posées par de tels environnements n'ont pas été prises en compte dès le départ. L'homme essaie, lui, de s'adapter aux conditions sonores rencontrées en modifiant sa méthode de production de parole. Un des phénomènes les plus remarquables de modification de production de la parole par l'homme est l'effet Lombard. Lorsqu'un locuteur est placé dans un environnement bruité, il modifie sa voix, et son effort vocal, en "haussant le ton" de manière à ce que la parole produite conserve un bon RSB (Rapport Signal sur Bruit) par rapport à l'environnement. Cette accentuation de la voix pose cependant un problème majeur aux systèmes de RAP car les spectres de tous les phonèmes peuvent être modifiés ce qui a pour effet de nettement amoindrir les taux de mots isolés ou de la parole continue masqués par du bruit. Il faut enfin noter qu'il existe des situations où la parole est modifiée sans que l'homme ne modifie sa façon de parler de manière volontaire. Ceci peut arriver lorsque la parole est produite par une personne se trouvant en contact avec un appareil en phase vibratoire [13].

1.4 Conclusion

Dans ce chapitre, nous avons généralement parlé des signaux audio, en particulier du signal parole et de la manière dont ils sont analysés à l'aide de différents outils mathématiques du traitement de signal. Nous avons également discuté en détail de la plupart des types de bruits pouvant affecter le signal parole. Dans le prochain chapitre, nous allons présenter une méthode de réduction de bruit dans le signal de parole, en se basant sur la géométrie du signal bruité.

Chapitre 2 :

Une approche géométrique
pour le rehaussement de la
parole

2.1 Introduction

Dans un large domaine d'applications vocales, à savoir, les systèmes de communication mobile et de reconnaissance vocale, le rehaussement de la parole est nécessaire. De nombreuses méthodes d'amélioration de la parole ont été établies depuis des décennies.

Ces techniques de rehaussement de la parole peuvent être classées en méthodes de domaine temporel et en méthodes de domaine spectral. Sinon, les techniques de rehaussement de la parole les plus utilisées surtout dans un téléphone portable, sont les méthodes du domaine spectral. Dans ce chapitre, nous présentons l'extension de la méthode spectrale basée sur la projection géométrique (GA) du signal (parole) bruité sur les axes complexes.

2.2 La Soustraction Spectrale (SS)

Le processus de soustraction spectrale consiste à atténuer les composantes spectrales du signal parole dans une mesure plus ou moins grande, selon la puissance du bruit. Ci en bas, nous aborderons l'algorithme de Berouti [14], qui est une technique considérée comme une référence pour la soustraction spectrale.

Soit un signal bruité x constitué d'un signal propre s qui a été altéré par un bruit additif n qui n'est pas corrélé. Une façon de représenter le signal bruité est la suivante :

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

Nous faisons référence au $\hat{s}(t)$ comme signal rehaussé (amélioré), qui est une estimation du $s(t)$.

Nous partons du principe que le signal vocal est quasi stationnaire sur des fenêtres d'analyse de 20 à 30 ms. La Transformée de Fourier à Court Terme (TFCT) de $x(t)$ est donnée par:

$$X(\omega) = S(\omega) + N(\omega) \quad (2.2)$$

Où $X(\omega)$, $S(\omega)$ et $N(\omega)$ font référence au spectre des $x(t)$, $s(t)$ et $n(t)$ respectivement.

$X(\omega)$ s'exprime sous forme polaire comme suit:

$$X(\omega) = |X(\omega)|e^{j\phi_x(\omega)} \quad (2.3)$$

$|X(\omega)|$ et $\phi_x(\omega)$ constituent respectivement l'amplitude et la phase de $X(\omega)$

$N(\omega) = |N(\omega)|e^{j\phi_n(\omega)}$ peut également être utilisé pour exprimer le bruit. L'amplitude du bruit $|N(\omega)|$ est inconnue, donc on la remplace par sa valeur moyenne estimée en l'absence d'activité vocale.

Dans les méthodes de soustraction spectrale, on suppose que le bruit n'a aucun effet sur la phase du signal. Seule l'amplitude spectrale à court terme du bruit est considérée dans ce cas. En conséquence, la valeur estimée de $\phi_n(\omega)$ sera remplacée par $\phi_x(\omega)$. En remplaçant

$X(\omega)$ et $N(\omega)$ dans (2.2), la largeur spectrale estimée du signal est donnée par :

$$\hat{S}(\omega) = [|X(\omega)| - |\hat{N}(\omega)|]e^{j\phi_x(\omega)} \quad (2.4)$$

Alternativement, $|\hat{N}(\omega)|$ est la valeur prévue du spectre d'amplitude du bruit lorsqu'il n'y a pas d'activité vocale. Le calcul de la transformée de Fourier inverse de $\hat{S}(\omega)$ entraînera l'acquisition du signal amélioré (2.4). Il est possible que des erreurs d'estimation du bruit entraînent $\hat{S}(\omega) = |X(\omega)| - |\hat{N}(\omega)|$ à produire des valeurs négatives, comme le montre (2.4). Plusieurs autres approches pourraient être utilisées pour régler ce problème [1], [14-21].

Une de ces méthodes est appelée rectification demi-onde, et c'est cette méthode qui permet à ces valeurs négatives d'être remises à zéro de la manière suivante :

$$\{\hat{S}(\omega) = |X(\omega)| - |\hat{N}(\omega)| \quad \text{si } |X(\omega)| > |\hat{N}(\omega)|\} \quad (2.5)$$

0

ailleurs

En multipliant $X(\omega)$ (2.2) par son conjugué $X^*(\omega)$, la méthode de soustraction spectrale d'amplitude (2.4) peut être étendue à la soustraction de puissance spectrale. Cela peut se faire de la façon suivante :

$$|X(\omega)|^2 = |S(\omega)|^2 + |N(\omega)|^2 + S(\omega) \cdot N^*(\omega) + S^*(\omega) \cdot N(\omega) \quad (2.6)$$

$$= |S(\omega)|^2 + |N(\omega)|^2 + 2 \cdot \text{Re} \{S(\omega) \cdot N^*(\omega)\}$$

Les termes $|N(\omega)|^2$, $S(\omega)$, $N^*(\omega)$ et $S^*(\omega)$. $N(\omega)$ ne peut être obtenu directement, mais approximativement par $E\{|N(\omega)|^2\}$, $E\{S(\omega) \cdot N^*(\omega)\}$ et $E\{S^*(\omega) \cdot N(\omega)\}$, où $E\{\cdot\}$ désigne l'opérateur de prévision mathématique. Généralement $E\{|N(\omega)|^2\}$ est estimé pendant la période d'absence d'activité de la voix est dénommé $|\hat{N}(\omega)|^2$.

Si nous supposons que $n(t)$ et $s(t)$ n'ont pas de moyen et ne sont pas corrélés avec $s(t)$, alors les termes $E\{S(\omega) \cdot N^*(\omega)\}$ et $E\{S^*(\omega) \cdot N(\omega)\}$ deviennent tous deux nuls. Voici les valeurs prévues pour le spectre de puissance du signal propre :

$$|\hat{S}(\omega)|^2 = |X(\omega)|^2 + |\hat{N}(\omega)|^2 \quad (2.7)$$

L'équation (2.7) décrit l'algorithme de soustraction spectrale de puissance et peut être exprimée sous la forme suivante :

$$|\hat{S}(\omega)|^2 = H^2(\omega)|X(\omega)|^2 \quad (2.8)$$

Ou

$$H(\omega) = \sqrt{1 - \frac{|\hat{N}(\omega)|^2}{|X(\omega)|^2}} \quad (2.9)$$

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

2.3 Une approche géométrique pour la Soustraction spectacle

Le spectre bruité $Y(\omega_k)$ à la fréquence (ω_k) est créé en ajoutant deux spectres à valeur complexe à la fréquence (ω_k) , comme le montre l'équation (2.10) :

$$Y(\omega_k) = X(\omega_k) + D(\omega_k) \quad (2.10)$$

En conséquence, sur le plan complexe, $Y(\omega_k)$ peut être représenté géométriquement comme la somme de deux nombres complexes, $X(\omega_k)$ et $D(\omega_k)$. Ceci est démontré à la Figure. 1, qui décrit $Y(\omega_k)$ comme une addition vectorielle de $X(\omega_k)$ et $D(\omega_k)$ dans le plan complexe.

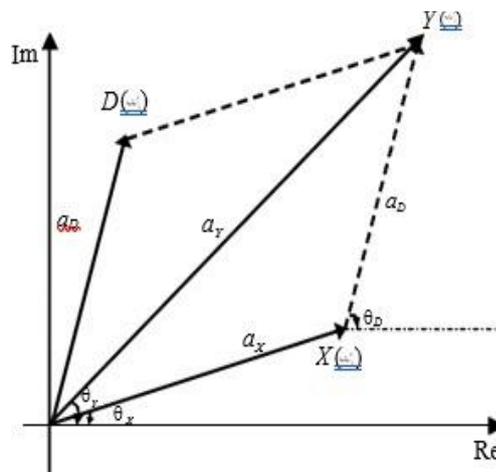


Figure.2.1 Représentation du spectre bruité $Y(\omega_k)$ dans le plan complexe comme la somme du spectre de signal propre $X(\omega_k)$ et du spectre de bruit $D(\omega_k)$. [1]

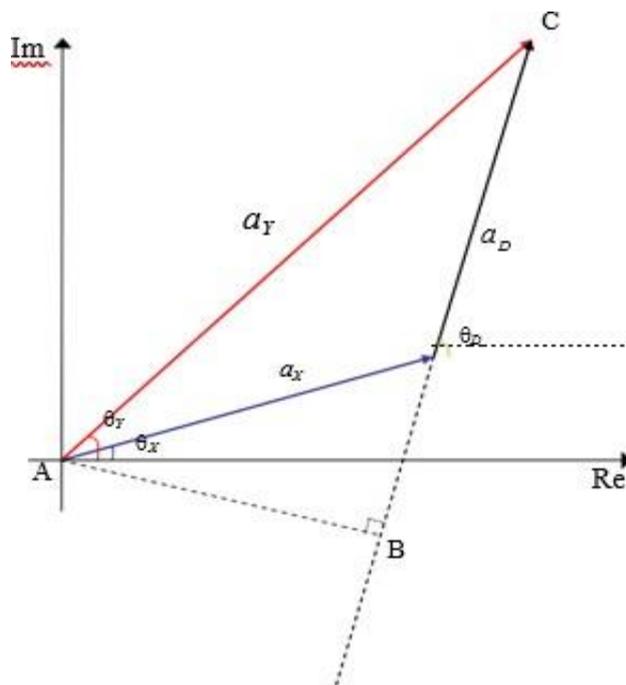


Figure.2.2 Triangle montrant la relation géométrique entre les phases de la parole bruitée, du bruit et des spectres de parole propres. [1]

L'équation (2.11) donne la fonction de gain généralement utilisée des méthodes de soustraction du spectre de puissance, qui est dérivée en supposant que les termes croisés sont zéro:

$$H(\omega k) = \sqrt{1 - \frac{|\widehat{D}(\omega k)|^2}{|Y(\omega k)|^2}} = \sqrt{\frac{\gamma(k)-1}{\gamma(k)}} \quad (2.11)$$

ou la différence de phase $[\theta_x(k) - \theta_D(k)]$ est égale à $\pm\pi/2$. La fonction de gain universel pour la soustraction spectrale est alors dérivée, ce qui ne fait aucune hypothèse concernant l'ampleur de la différence de phase entre les signaux bruités et propres. Nous commençons par réécrire l'éq (2.10) sous forme polaire comme suit :

$$a_y e^{i\theta_y} = a + a_D e^{i\theta_D} \quad (2.12)$$

Où $\{a_y, a_x, a_D\}$ sont les amplitudes et $\{\theta_Y, \theta_X, \theta_D\}$ sont les phases des spectres bruités, propres et de bruit. Par souci de simplicité, nous n'utiliserons plus l'indice de fréquence k .

Considérons le triangle montré dans Figure. 2. En utilisant la loi de Sinus, ou le triangle droit ABC avec $\underline{AB} \perp \underline{BC}$, nous avons

$$\begin{aligned} \underline{AB} &= a_y \sin(\theta_D - \theta_Y) = a_x \sin(\theta_D - \theta_X), \\ &= a_y^2 \sin^2(\theta_D - \theta_Y) = a_x^2 \sin^2(\theta_D - \theta_X), \\ &= a_y^2 [1 - \cos^2(\theta_D - \theta_Y)] = a_x^2 [1 - \cos^2(\theta_D - \theta_X)], \\ &= a_y^2 (1 - C_{YD}^2) = a_x^2 (1 - C_{XD}^2) \end{aligned} \quad (2.13)$$

Nous pouvons obtenir la nouvelle fonction de gain à partir de l'équation donnée.

$$H_{GA} = \frac{a_x}{a_y} = \sqrt{\frac{1 - C_{YD}^2}{1 - C_{XD}^2}} \quad (2.14)$$

Parce que les termes C_{YD} et C_{XD} sont limités par un, la fonction de gain ci-dessus est toujours réelle et positive ($H_{GA} \geq 0$). Contrairement à la fonction de gain de soustraction spectrale de puissance Eq. (2.14), qui est toujours positif et inférieur (ou égal à) un, la fonction de gain décrite ci-dessus peut être supérieure à un si $|C_{YD}| < |C_{XD}|$ [19]. L'une des équations qui peut être obtenue en utilisant les principes trigonométriques est Eq. (2.11).

Notamment, la fonction de suppression ci-dessus se réduit à la fonction de suppression de l'approche de soustraction spectrale de puissance Eq.(2.11) si $C_{XD} = 0$, c'est-à-dire si le signal et les vecteurs de bruit sont orthogonaux. Statistiquement, le signal et le bruit ne sont pas corrélés s'ils sont orthogonaux ($E[(\omega_K) \cdot D(\omega_K)] = 0$) et ont une moyenne nulle [20].

Démontrer que la fonction de suppression susmentionnée est réduite à l'éq. (2.11) lorsque ($C_{XD} = 0$), il est évident d'après la Figure. 1 que lorsque les vecteurs de signal bruités et propres sont orthogonaux les uns aux autres ($C_{XD} = 0$), la fonction de suppression se réduit à Eq.(2.11).

$$C_{XD} = \frac{a_D}{a_Y} \quad (2.15)$$

En substituant l'équation précédente à l'équation (2.14), nous obtenons l'équation (2.11). À la lumière de l'analyse ci-dessus, nous pouvons conclure que la règle de suppression prévue dans l'Eq. (2.14) est la règle de suppression réelle et précise pour les algorithmes de soustraction spectrale si aucune hypothèse statistique n'est faite sur la relation entre le signal et le bruit. En revanche, la méthode de suppression fournie dans l'Eq. (2.11) n'est qu'une approximation puisqu'elle présuppose que $C_{XD} = 0$, c'est-à-dire que le signal propre et les vecteurs de bruit sont orthogonaux les uns par rapport aux autres pendant de courtes périodes (20 à 30 ms). Même si nous avons accès au spectre de magnitude de bruit correct $|D(\omega_K)|$, multiplier le signal bruité par la fonction de suppression décrite dans l'équation (2.14) ne produirait pas le spectre de magnitude de signal propre. En revanche, la multiplication du spectre de magnitude bruité (a_Y) par la fonction de suppression décrite dans l'équation (2.14) donnerait le spectre de magnitude précis du signal propre (a_X).

La fonction de suppression décrite ci-dessus est basée sur l'estimation des différences de phase entre les signaux bruités (ou propres) et les signaux de bruit. Cependant, déterminer précisément les valeurs de ces phases est un problème difficile en soi, et aucune technique n'existe pour le moment. Une option consiste à utiliser des concepts trigonométriques pour développer et appliquer des corrélations explicites entre les phases de signaux bruités et propres.

$$C_{YD} = \frac{a^2_Y + a^2_D + a^2_X}{2a_Y a_D} \quad (2.16)$$

$$C_{XD} = \frac{a^2_Y + a^2_X + a^2_D}{2a_X a_D} \quad (2.17)$$

La difficulté fondamentale à utiliser les équations susmentionnées pour estimer les différences de phase entre le signal et les signaux parole est leur dépendance à l'amplitude du signal propre, que nous ne connaissons pas. Nous pouvons obtenir une autre équation pour C_{YD} et C_{XD} en divisant le numérateur et le dénominateur des éq. (2.16) et (2.17) par a^2_D .

$$C_{YD} = \frac{\gamma + 1 + \xi}{2\sqrt{\gamma}} \quad (2.18)$$

$$C_{XD} = \frac{\gamma - 1 + \xi}{2\sqrt{\xi}} \quad (2.19)$$

Par conséquent, les variables γ et ξ sont définies comme suit :

$$\gamma = \frac{a^2_Y}{a^2_D} \quad (2.20)$$

$$\xi = \frac{a^2_X}{a^2_D} \quad (2.21)$$

Il convient de noter que les termes c et n représentent les versions instantanées des SNRs posteriori et a priori, qui sont utilisées dans les algorithmes MMSE, respectivement [16],[18]. En remplaçant les éq (2.18) et (2.19) par l'éq (2.14), on obtient la fonction de suppression en

γ et ξ comme suit:

$$H_{GA}(\xi, \gamma) = \sqrt{\frac{1 - \frac{(\gamma + 1 - \xi)^2}{4\gamma}}{1 - \frac{(\gamma - 1 - \xi)^2}{4\xi}}} \quad (2.22)$$

En théorie, la fonction de suppression décrite ci-dessus peut être supérieure à un. La fonction de gain ci-dessus, comme la fonction de gain des algorithmes d'amélioration basés sur le MMSE [16], dépend de deux paramètres, c et n . Pour mieux comprendre la relation entre ces deux paramètres et la suppression, nous présentons $H_{GA}(\xi, \gamma)$, comme une fonction de $(c - 1)$ pour des valeurs fixes de n . Pour comparaison, les courbes de suppression de l'algorithme MMSE (Ephraim et Malah, 1984) sont superposées. Les fonctions de gain de l'algorithme GA suivent clairement le modèle des fonctions de gain MMSE pour la plupart. Pour les valeurs $(\gamma - 1)$ inférieures à 5 dB, les fonctions de gain de l'approche GA reflètent étroitement les fonctions de

gain MMSE, puis partent. Lorsque $(\gamma - 1)$ est supérieur à 5 dB, les fonctions de gain de l'algorithme GA deviennent plus suppressives que les fonctions de gain MMSE.

La fonction de gain Wiener ($H_W = \xi / (\xi + 1)$) et la fonction de gain MMSE sont également affichées. Globalement, avec la même valeur $(\gamma - 1)$, la fonction de gain GA correspond étroitement à la fonction de gain MMSE pour les valeurs petites et négatives de $(\gamma - 1)$. En revanche, la fonction de gain GA devient plus suppressive que la fonction de gain MMSE pour $(\gamma - 1)$ des valeurs supérieures à 5 dB.

Le fait que la fonction de gain de l'algorithme GA possède des caractéristiques comparables à celles des algorithmes MMSE montre qu'il hérite des attributs et du comportement de l'algorithme MMSE [15]. Le paramètre a posteriori γ , comme la méthode MMSE, fonctionne comme un paramètre de correction qui n'affecte l'atténuation que lorsque ξ est faible. La correction, d'autre part, est faite dans la direction opposée, une atténuation importante est utilisée lorsque γ est grand, plutôt que lorsque c est petit, comme on pourrait s'y attendre. Ce comportement illogique n'est pas un artefact de l'algorithme, mais il est avantageux lorsqu'il s'agit de parties de la parole à faible énergie. Dans les segments avec bruit de fond, les valeurs γ sont déraisonnablement élevées dans certaines images, et ces images sont affectées une plus grande atténuation. Ceci est dû à la règle de suppression mettant une plus grande "confiance" dans les valeurs, qui sont minuscules dans ces cadres par rapport aux valeurs γ . Comme l'atténuation de l'algorithme MMSE est principalement contrôlée par la valeur lissée du SNR a priori, l'atténuation ne fluctuera pas beaucoup d'une image à l'autre. En conséquence, le bruit musical sera minimisé ou entièrement supprimé. L'approche typique de soustraction spectrale de puissance, par contre, est basée sur l'estimation du SNR a posteriori, qui peut varier considérablement d'une image à l'autre. Le bruit musical est créé en conséquence. En résumé, le comportement de lissage de la méthode "décisionnelle" combiné à la règle de suppression du MMSE est responsable de la réduction de l'impact du bruit musical dans l'algorithme du MMSE. Nous pouvons prévoir peu ou pas de bruit musical avec l'algorithme GA puisqu'il hérite du comportement de la méthode MMSE.

Il y a deux différences majeures entre l'algorithme GA suggéré et la méthode MMSE qui devraient être soulignées ici. D'abord et avant tout, la méthode GA est déterministe et ne repose sur

aucun modèle statistique. Le spectre de magnitude propre est considéré comme inconnu mais prévisible. Par conséquent, aucune hypothèse concernant les distributions statistiques de la parole et du bruit n'est formulée. Les coefficients de transformée de Fourier sont utilisés dans la méthode MMSE. Deuxièmement, les paramètres γ et ξ représentent des valeurs instantanées plutôt que des valeurs moyennes statistiques à long terme. Par conséquent, d'autres méthodes doivent être utilisées pour estimer ces paramètres, qui sont présentés ci-après. Pour être complet, nous évaluons et comparons la performance de l'algorithme GA proposé en utilisant les données γ et ξ moyennes immédiates et à long terme.

2.4 Conclusion

Dans ce chapitre nous avons présenté une approche géométrique (Algorithme GA) pour la soustraction spectrale de la parole basée sur des règles géométriques. Contrairement à la soustraction spectrale de puissance conventionnelle. Algorithme qui suppose que les termes qui se croisent incluent la différence de phase entre le signal et le bruit est nulle, l'algorithme étudié ne fait pas de telles hypothèses. C'est elle était étayée par une analyse d'erreur qui indiquant que il est prudent d'ignorer les termes croisés lorsque le SNR spectrale approche de 0 dB.

CHAPITRE 3

Résultats et Discussions

3.1. Introduction

Ce chapitre est formé de plusieurs modules. Le premier est consacré à l'acquisition des données. Cela implique l'enregistrement de la parole de différents types sous différents bruits d'environnement [1], [6] Le deuxième module examine le détecteur d'activité vocale (VAD). Étudier et créer un algorithme approprié qui agirait à la base de l'algorithme de soustraction spectrale. Un taux d'erreur de classification élevé dans le VAD conduira à de mauvaises estimations des spectres de bruit. Cela fait ressortir une grave détérioration des performances de suppression du bruit. Le troisième module se concentre sur la soustraction spectrale à la base de l'algorithme GA choisi. Sa mise en œuvre a été faite et toutes les améliorations appropriées ont été insérées.

Le dernier module est le test. Obtenir une qualité subjective et une qualité objective des mesures permettant d'évaluer la performance de l'algorithme analysé (GA) est sa principale cible.

3.2. Définition des paramètres communs

3.2.1 Corpus de tests

Dans notre travail ou mémoire nous avons utilisé des corpus vocaux bruités NOIZEUS [22] pour faciliter le test et la comparaison des algorithmes d'amélioration (rehaussement) de la parole que nous avons étudiés dans le deuxième chapitre, la base de données bruitées comprend 30 phrases IEEE produites par trois (03) hommes et trois (03) femmes, ces phrases sont corrompues par huit différents bruits du monde réel à différents SNR 0, 5 et 10 dB. Le bruit a été tiré de la base des données AURORA [23] et comprend le bruit d'aéroport, le bavardage à plusieurs locuteurs (babble), bruit véhicule, rue, restaurant, salle d'exposition, train et bruit blanc gaussien.

3.2.2 Evaluation des performances

➤ Définition de l'intelligibilité la parole

L'intelligibilité de la parole est une mesure apparue à la fin des années 1920, avec le besoin pour les ingénieurs d'évaluer différents systèmes de transmission téléphonique [24]. Et est une mesure du taux de transmission de la parole, utilisée pour évaluer les performances de systèmes de télécommunication, de sonorisation, de salles, ou encore de personnes. La qualité et l'intelligibilité de la parole peuvent être quantifiées à l'aide de mesures subjectives et objectives.

- Mesures subjectives

Les mesures subjectives de la qualité de la parole sont obtenues en utilisant des tests d'écoute dans lesquels des participants humains évaluent la qualité de la parole. Chaque écoute porte sur un ensemble de trois évaluations, pour la première évaluation, les sujets doivent s'occuper uniquement du signal vocal et donner une note de qualité traduisant le niveau de distorsion. Pour la deuxième évaluation, les sujets s'occupent uniquement du bruit de fond en donnant une note traduisant l'importance de ce dernier. Pour la troisième évaluation, les sujets doivent évaluer globalement à la fois la qualité du signal vocal et la quantité du bruit de fond en donnant une note de qualité globale [25].

En moyennant les différentes notes attribuées aux évaluations de tests, on calcule la note finale évaluant la distorsion du signal appelée SIG, la note finale évaluant le bruit résiduel appelée BAK et la note finale évaluant la qualité globale appelée OVL ou MOS. Comme il est indiqué sur le tableau :

Tableau 3.1 Échelle d'évaluation du SIG, BAK et OVL

Note	SIG : le signal vocal est	BAK : le bruit de fond est	OVL : la séquence vocale est
5	dépourvu de distorsion	imperceptible	Excellente
4	légèrement distordu	légèrement imperceptible	Bonne
3	quelque peu distordu	perceptible mais non gênant	Passable
2	assez distordu	quelque peu gênant	Médiocre
1	très distordu	très gênant	mauvaise

- Mesures objectives

Les mesures de qualité objective prédisent la qualité de la parole perçue sur la base d'un calcul de la distance numérique ou de la distorsion entre le signal vocal propre et le signal vocal traité ou rehaussée.

On a plusieurs mesures objectives de la qualité de la parole, dans notre travail nous avons choisi les mesures suivantes :

- Rapport signal sur bruit segmental (segSNR : Segmental Signal to Noise Ratio) [26],
- Evaluation Perceptive de la Qualité de la Parole (PESQ : Perceptual Evaluation of Speech Quality) [27],

- Pente Spectrale Pondérée (WSS : Weighted Slop Spectral)[28],
- Logarithme du Rapport de Vraisemblance (LLR : Log Likelihood Ratio)[29].

3.3 Résultats et discussions

Nous avons extrait à partir de la base de données NOIZEUS, 15 phrases mélangées avec des bruits de aéroport, babble, véhicule, salle d'exposition, restaurant, rue, train et de bruit blanc gaussien à des niveaux de RSB de 0, 5 et 10 dB. Le signal bruité est ensuite échantillonné à 8KHz puis segmenté, à travers la fenêtre de Hamming, en plusieurs trames de 30 ms avec un recouvrement de 40%.

L'évaluation des méthodes ; Soustraction Spectrale (SS), Soustraction spectrale non linéaire (SS-NSS) , Soustraction Spectrale basée sur le Moyennage Adaptatif du Gain (SS-AGA) et l'approche géométrique (GA-SE) étudiée dans notre travail ont été effectuées par les mesures objectives et subjectives et SNR de signal rehaussé illustrés dans les tableaux 2, 3, 4 et 5, et par les spectrogrammes, les allures temporelles de signaux et la mesure SNR amélioré illustrées dans les Figures 1, 2 et 3 respectivement.

À partir des tableaux 2, 3, 4 et 5 :

Nous remarquons, que la méthode de rehaussement de la parole GA-SE donne de meilleurs résultats en terme SIG, BAK et OVL par rapport aux trois autres méthodes SS, SS-NSS et SS-AGA à l'exception de cas suivant :

- Pour le bruit salle d'exposition à 10 dB, la valeur de OVL de la méthode SS-AGA est la plus grande.

Aussi nous remarquons, que la méthode de rehaussement de la parole SS donne de meilleurs résultats en terme WSS par rapport aux trois autres méthodes SS-NSS, SS-AGA, et GA-SE.

La méthode de rehaussement de la parole SS-AGA donne de meilleurs résultats en terme LIR par rapport aux trois autres méthodes, pour les bruits d'aéroport, babble, restaurant et les cas à 10 dB pour les bruits véhicule, salle d'exposition et rue.

Les autres cas restants, en terme LIR , la méthode de rehaussement de la parole SS donne de meilleurs résultats par rapport aux autres méthodes (bruit de train, blanc et les cas à 0 et 5 dB pour le bruit de véhicule, salle d'exposition et rue).

La méthode de rehaussement de la parole SS donne de meilleurs résultats en terme SNR sortie par rapport aux autres méthodes à l'exception de cas suivant :

- Pour le bruit babble, restaurant et train à 0 dB, la valeur de la méthode GA-SE est la plus grande.

Pour le bruit Babble, le bruit le plus utilisé dans notre vie en générale et à partir du tableau 2 : En termes de PESQ et segSNR, nous remarquons que la méthode GA-SE présente les meilleures valeurs par rapport aux autres méthodes SS, SS-NSS et SS-AGA, sauf les cas suivants :

- La méthode SS-AGA donne meilleure valeur en termes de PESQ à 10 dB.
- La méthode SS donne meilleur valeur en terme segSNR à 10 dB.

Dans cette section nous nous concentrons davantage sur le bruit le plus utilisée dans le domaine de la communication (le bruit blanc) et à partir du tableau 5 :

- En terme SIG, BAK, OVL et PESQ, nous remarquons que la méthode GA-SE donne des meilleures valeurs par rapport aux trois autres méthodes et à travers cette remarque nous pouvons dire que cette méthode (GA-SE) est efficace dans une certaine mesure, grâce aux bons résultats qu'elle a donnés (GA-SE) en deux termes de qualité (PESQ) et (OVL).
- En terme segSNR, WSS, LIR et SNR sortie, nous remarquons que la méthode SS donne de meilleurs résultats par rapport aux autres méthodes SS, SS-AGA et GA-SE, et pour cette remarque en peut dire que cette méthode (SS) est basée sur le rapport signal sur bruit et le SNR sortie mais elle a perdu les standards de qualité.

Après analyse des tableaux précédents, la méthode de rehaussement de la parole GA-SE est la meilleure dans l'ensemble et dans la majorité des valeurs et particulièrement en termes d'évaluation perceptive de la qualité du parole et en termes de la qualité général et ça c'est un objectif visée et un avantage pour cette méthode, par contre et comme toute autre méthode, elle a ses inconvénients et ici on parle par exemple de termes de distorsion du signal qui il est déjà développé par cette méthode.

Et concernant le terme de SNR sortie, on trouve que la méthode de rehaussement de la parole SS c'est la meilleure dans le cas du bruit blanc.

Pour la Figure 1 : Nous remarquons, que la méthode de rehaussement de la parole SS a récupéré presque le signal du parole propre avec un certain bruit, alors que les deux autres méthodes AGA et GA-SE réduisaient plus de bruit mais avec moins d'intelligibilité.

Pour la Figure 2 : Nous remarquons, que le signal parole rehaussé par la méthode SS c'est le meilleure par rapport autres signaux rehaussés.

La Figure 3, représente les résultats obtenus avec la mesure SNR amélioré pour huit types de bruits à des niveaux de SNR de 0, 5 et 10 dB. Ces résultats montrent que le débruitage par la méthode GA-SE donne une meilleure amélioration du SNR.

Tableau 3.2 évaluation objective, subjective et SNR de sortie des méthodes SS, SS-NSS, SS-AGA et GA-SE dans un milieu corrompu par des bruits d'aéroport et Babble. Valeur moyenne en utilisant 15 phrases extraites de la base données NOIZEUS. Les meilleures performances sont indiquées en Gras.

Mesures Objective subjective	Entrée SNR (dB)	Bruit aéroport				Bruit babble			
		SS	SS - NSS	SS - AGA	GA-SE	SS	SS - NSS	SS - AGA	GA-SE
SIG [1 à 5]	0	1.6226	1.9957	1.5302	2.2393	1.4454	1.9060	1.4903	2.2493
	5	2.3364	2.4501	2.3251	2.6635	2.1825	2.4685	2.2183	2.7835
	10	2.9770	2.9754	3.0151	3.3283	2.8910	2.9387	2.7752	3.3251
BAK [1 à 5]	0	1.4162	1.3001	0.8852	1.6721	1.3465	1.2567	0.8413	1.7231
	5	1.9060	1.5606	1.3607	2.0508	1.8337	1.5591	1.3687	2.1167
	10	2.4102	1.8456	1.8404	2.5483	2.3662	1.8336	1.7757	2.5444
OVL [1 à 5]	0	1.3644	1.7149	1.1604	1.7906	1.2383	1.6395	1.1646	1.8598
	5	1.9708	2.1478	1.9553	2.2323	1.8618	2.1575	1.8394	2.3318
	10	2.5359	2.6509	2.6461	2.8047	2.4761	2.6132	2.4083	2.8127
PESQ [0.5 à 4.5]	0	1.6023	1.7791	1.6527	1.7003	1.5347	1.7297	1.6840	1.8215
	5	1.9823	2.1073	2.1489	2.1085	1.9399	2.1105	2.1096	2.1476
	10	2.3629	2.5034	2.5351	2.4692	2.3391	2.4712	2.5249	2.4903
segSNR	0	-2.0823	-8.6012	-6.6530	-1.8450	-2.5120	-8.6567	-7.1987	-2.0868
	5	0.1923	-8.7635	-7.5229	-0.0096	-0.1471	-8.7650	-7.3210	-0.1448
	10	2.9204	-9.0827	-7.7209	2.5278	2.6206	-8.8876	-7.8517	2.3567
WSS	0	121.7839	91.7764	106.8134	94.0651	123.2636	94.0946	109.7937	92.8723
	5	98.2352	75.5172	88.9444	84.3598	102.6180	75.9355	92.7764	76.8370
	10	76.7513	58.9810	73.5976	60.7692	78.7201	60.2389	74.4015	61.2025
LLR	0	1.3027	1.3063	1.3787	1.0034	1.4224	1.3442	1.4308	1.0750
	5	1.0377	1.1991	1.2507	0.9151	1.1241	1.1795	1.2612	0.8912
	10	0.8261	1.0654	1.1110	0.6871	0.8786	1.0712	1.1477	0.6985
SNR sortie	0	3.7272	-9.1668	-6.9282	3.6003	3.3614	-8.9580	-7.3273	3.7443
	5	7.3216	-9.0962	-8.0025	5.9823	7.1899	-9.0615	-7.7269	6.2560
	10	10.6976	-9.4068	-8.4524	9.1207	10.8377	-9.1076	-8.3302	9.1253

Tableau 3.3 évaluation objective, subjective et SNR de sortie des méthodes SS, SS-NSS, SS-AGA et GA-SE dans un milieu corrompu par des bruits véhicule et salle d'exposition. Valeur moyenne en utilisant 15 phrases extraites de la base données NOIZEUS. Les meilleures performances sont indiquées en Gras

Mesures	Entrée	Bruit véhicule				Bruit salle d'exposition			
		SS	SS – NSS	SS - AGA	GA-SE	SS	SS - NSS	SS - AGA	
SIG [1 à 5]	0	1.4081	2.1718	2.0256	2.4965	1.1994	2.0454	1.6192	2.1548
	5	2.2437	2.5905	2.4690	3.0116	2.0180	2.5668	2.4463	2.7419
	10	2.8910	2.9387	2.7752	3.3251	2.6114	2.9961	3.0618	3.2152
BAK [1 à 5]	0	1.4903	1.3493	1.2836	1.6649	1.3021	1.2779	1.1004	1.7050
	5	1.9813	1.6052	1.6465	2.2493	1.8499	1.5827	1.5442	2.1288
	10	2.3662	1.8336	1.7757	2.5444	2.3108	1.8331	1.9605	2.5431
OVL [1 à 5]	0	1.2716	1.8083	1.5523	2.0363	1.0629	1.6794	1.2210	1.7295
	5	1.9467	2.2193	2.0639	2.4933	1.7633	2.1941	2.0352	2.2709
	10	2.4761	2.6132	2.4083	2.8127	2.2960	2.6246	2.7363	2.7307
PESQ [0.5 à 4.5]	0	1.6293	1.7597	1.7648	1.8481	1.4648	1.6546	1.6397	1.6190
	5	2.0290	2.0895	2.1847	2.1834	1.9223	2.0845	2.1592	2.0476
	10	2.3391	2.4712	2.5249	2.4903	2.2996	2.4435	2.5495	2.4337
segSNR	0	-1.1260	-8.3561	-4.8554	-1.4882	-1.9046	-8.1212	-5.7689	-1.6765
	5	1.1041	-8.3890	-5.9824	0.3709	0.5858	-8.2265	-7.1353	0.3527
	10	2.6206	-8.8876	-7.8517	2.3567	2.9678	-8.5470	-7.4266	2.7053
WSS	0	121.6559	85.6348	88.3062	76.9715	130.2949	90.7637	95.7393	85.3166
	5	98.8676	71.2945	75.7447	64.5368	105.6971	75.6353	87.2965	72.3084
	10	78.7201	60.2389	74.4015	61.2025	87.0512	61.4946	73.7427	60.6626
LLR	0	1.5282	1.1774	1.1845	0.9895	1.5590	1.1938	1.2530	1.1143
	5	1.1496	1.0893	1.1378	0.7942	1.2467	1.0713	1.1569	0.9087
	10	0.8786	1.0712	1.1477	0.6985	1.0542	0.9882	1.0993	0.7768
SNR sortie	0	5.4893	-7.7741	-4.8439	3.8844	3.5877	-7.7252	-6.0551	3.0376
	5	8.7768	-8.2569	-7.0553	6.2748	7.5498	-8.3000	-7.9946	5.9553
	10	10.8377	-9.1076	-8.3302	9.1253	10.8747	-8.9168	-8.5844	9.0593

Tableau 3.4 évaluation objective, subjective et SNR de sortie des méthodes SS, SS-NSS ,SS-AGA et GA-SE dans un milieu corrompu par des bruits restaurant et rue. Valeur moyenne en utilisant 15 phrases extraites de la base données NOIZEUS. Les meilleures performances sont indiquées en Gras.

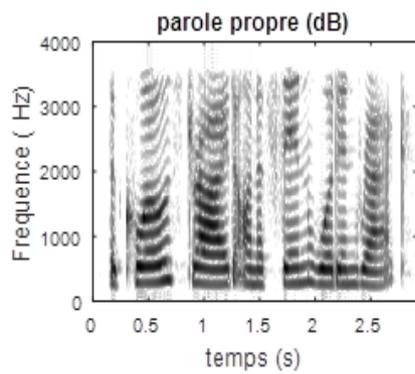
Mesures Objective Subjective	Entrée SNR (dB)	Bruit restaurant				GA-SE	Bruit rue			
		SS	SS - NSS	SS - AGA			SS	SS - NSS	SS - AGA	GA- SE
SIG [1 à 5]	0	1.6732	2.0715	2.0414	2.3234	1.3273	2.1320	2.0114	2.2928	
	5	2.3620	2.5050	2.6182	2.8150	2.0594	2.6370	2.4374	2.9114	
	10	2.9474	3.0094	2.9940	3.3313	2.7495	3.0676	2.9289	3.3452	
BAK [1 à 5]	0	1.4264	1.2954	1.1926	1.7352	1.4242	1.3499	1.2663	1.7822	
	5	1.8857	1.5334	1.5182	2.8150	1.9255	1.5898	1.5892	2.2225	
	10	2.3827	1.8457	1.8481	2.5486	2.4006	1.8719	1.8695	2.6060	
OVL [1 à 5]	0	1.3828	1.7486	1.6303	1.8980	1.2010	1.7715	1.5485	1.8672	
	5	1.9718	2.1604	2.1893	2.3241	1.8165	2.2250	2.0056	2.4196	
	10	2.5038	2.6583	2.5894	2.8101	2.4122	2.6920	2.6500	2.8286	
PESQ [0.5 à 4.5]	0	1.5555	1.7612	1.7084	1.8131	1.5678	1.7166	1.7246	1.7493	
	5	1.9533	2.0800	2.1180	2.0988	1.9421	2.0471	2.0765	2.1449	
	10	2.3227	2.4878	2.5361	2.4755	2.3622	2.4882	2.5121	2.4836	
segSNR	0	-2.3075	-8.7594	-7.4768	-2.0752	-1.7389	-	-6.1378	-	
	5	-0.0367	-8.9476	-7.9728	0.0619	0.6373	8.2322	-7.2021	1.5794	
	10	2.6608	-8.8891	-7.8432	2.4550	3.2203	-	-7.2922	0.4301	
							8.4872		2.9759	
							-			
							8.6778			
WSS	0	115.10	89.799	104.30	90.6721	121.380	83.724	93.0055	84.063	
	5	62	6	53	75.9061	5	5	78.4552	2	
	10	97.093	75.872	88.888	60.4745	96.7079	69.725	70.4546	66.257	
	0	75.590	6	4		80.7809	5		0	
	7		59.626	71.888			57.829		57.519	
		7	9	8			5		7	
LLR	0	1.2845	1.2393	1.3479	1.0173	1.5731	1.2076	1.2549	1.0675	
	5	1.0059	1.1267	1.1695	0.8361	1.2968	1.0330	1.0885	0.8539	

	10	0.8414	1.0176	1.0979	0.6901		1.0115	0.9770	1.0757	0.7073
SNR sortie	0	2.6117	-9.8208	-8.5657	3.2292		3.8138	-	-6.9270	3.6614
	5	6.6136	-9.5604	-8.4924	6.3995		7.8416	8.4406	-7.8480	6.5675
	10	10.279	-9.0707	-8.4786	9.1761		10.9168	-	-8.2857	9.2224
	2							8.6460		
								-		
							9.1236			

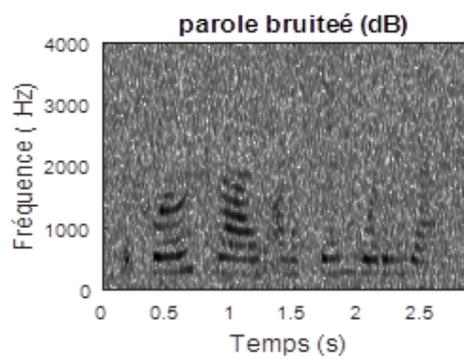
Tableau 3. 5 évaluation objective, subjective et SNR de sortie des méthodes SS, SS-NSS, SS-AGA et GA-SE dans un milieu corrompu par des bruits train et blanc. Valeur moyenne en utilisant 15 phrases extraites de la base données NOIZEUS. Les meilleures performances sont indiquées en Gras.

Mesures	Entrée	Bruit train				Bruit blanc			
		SS	SS - NSS	SS - AGA	GA-SE	SS	SS - NSS	SS - AGA	GA- SE
SIG [1 à 5]	0	1.2575	2.2869	1.9857	2.4159	0.8909	1.6199	1.4120	1.9085
	5	1.9152	2.7065	2.6462	2.8830	1.6360	2.1365	1.9035	2.5027
	10	2.5747	3.1208	3.2890	3.3117	2.2772	2.6015	2.4680	2.9941
BAK [1 à 5]	0	1.4663	1.3588	1.2171	1.9139	1.4825	1.1091	1.0895	1.8390
	5	1.9330	1.6178	1.6783	2.2543	1.9571	1.4179	1.4038	2.2418
	10	2.3870	1.8642	1.9113	2.6207	2.3978	1.6852	1.7145	2.6217
OVL [1 à 5]	0	1.1340	1.8626	1.5248	1.9954	1.0496	1.4279	1.1541	1.6960
	5	1.7155	2.2698	2.2500	2.4083	1.6641	1.9070	1.5662	2.2256
	10	2.2901	2.6969	2.7983	2.8028	2.2141	2.3460	2.1718	2.6699
PESQ [0.5 à 4.5]	0	1.4379	1.7346	1.6387	1.8213	1.7856	1.6943	1.6941	1.7955
	5	1.8476	2.0613	2.1454	2.1179	2.1521	2.0574	1.9334	2.1976
	10	2.2611	2.4354	2.4906	2.4283	2.5151	2.3993	2.3714	2.5347
segSNR	0	-1.5545	-	-6.6017	-1.3819	-0.5155	-8.4696	-4.0587	-
	5	0.6620	8.4417	-7.2082	0.4167	1.6482	-8.0233	-4.4182	0.8843
	10	3.0729	-8.2873	-7.4021	2.8007	3.7903	-7.9200	-5.8962	1.0971
			-8.6205						3.2682
WSS	0	108.14	81.786	90.627	71.9423	138.928	114.454	107.260	85.358
	5	98	4	6	59.7615	2	5	8	2

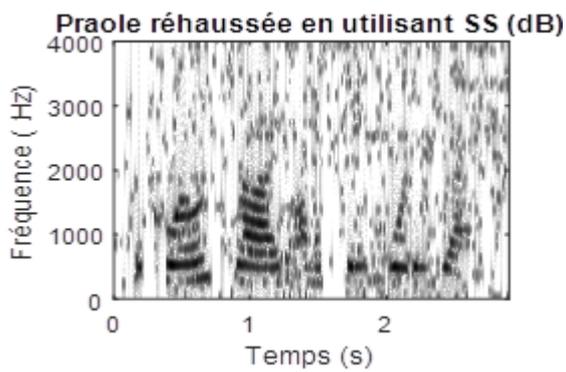
	10	89.408	68.485	75.282	50.0660		115.637	99.1495	94.7448	73.108
		3	7	1			0	85.2416	83.3017	7
		74.488	55.830	61.756			96.7401			61.402
		1	3	5						9
LLR	0	1.6805	1.0845	1.1667	1.0960		1.9713	1.4234	1.4321	1.4568
	5	1.4453	0.9845	1.0865	0.9225		1.6657	1.2680	1.3070	1.2221
	10	1.1772	0.9118	1.0072	0.7725		1.4205	1.1381	1.2082	1.0444
SNR sortie	0	4.0329	-8.8489	-7.2255	4.3569		7.2255	-8.1540	-3.2195	4.8283
	5	7.8836	-8.3137	-7.7792	6.7191		10.0530	-7.2727	-5.2123	7.2911
	10	11.162	-8.9482	-8.3193	9.4149		12.2800	-7.5339	-7.4644	9.6768
		6								



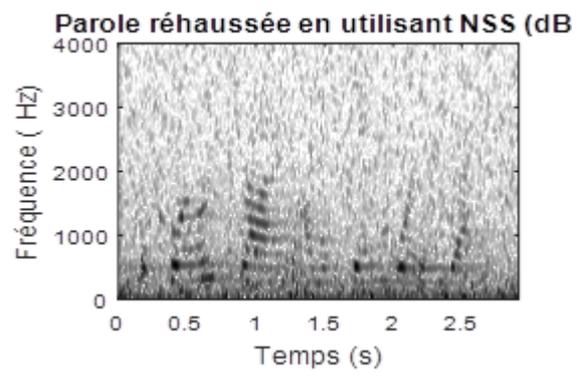
(a)



(b)



(c)



(d)

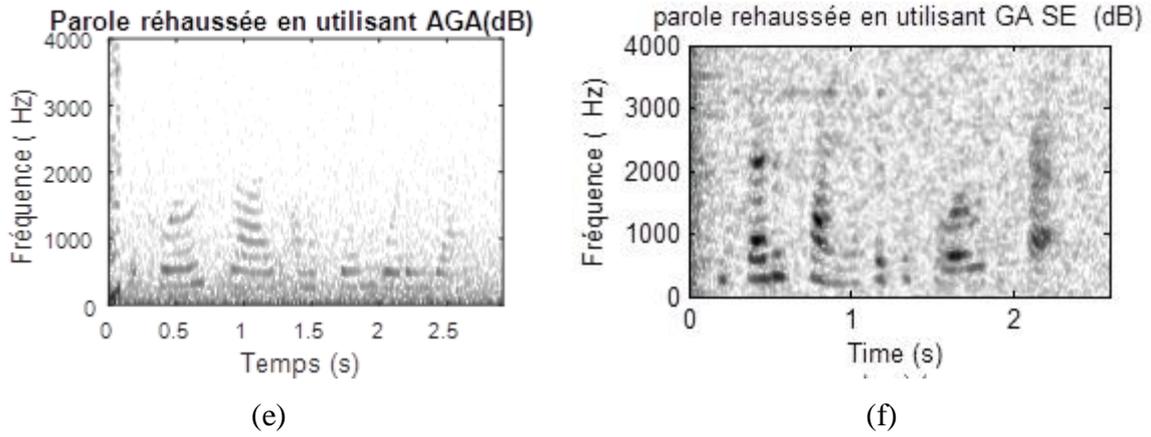
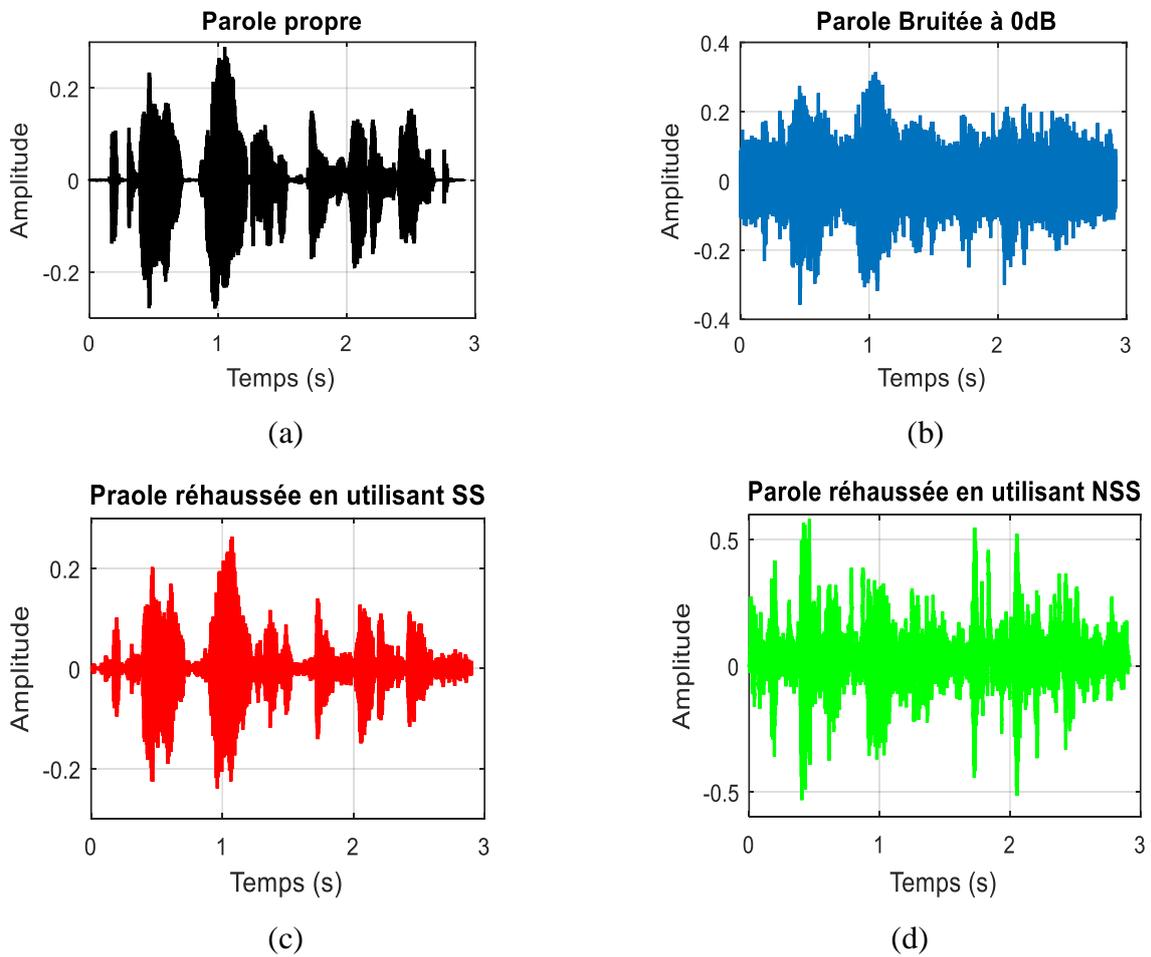


Figure 3.1. Spectrogrammes, (a) parole propre, (b) parole bruitée, (c) parole réhaussée en utilisant la méthode SS, (d) parole réhaussée en utilisant la méthode SS-NSS, (e) parole réhaussée en utilisant la méthode SS-AGA,(f) parole réhaussée en utilisant la méthode GA_SE. Cas du bruit blanc 0 dB.



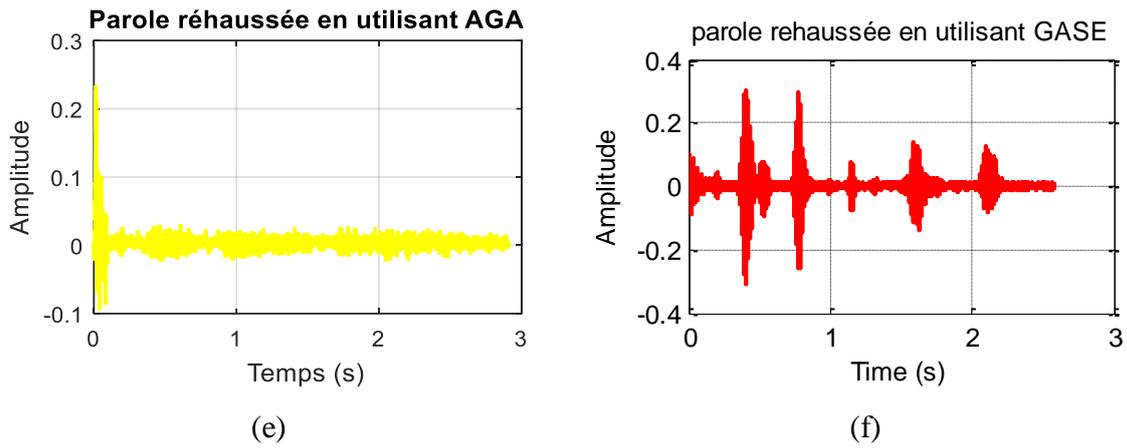
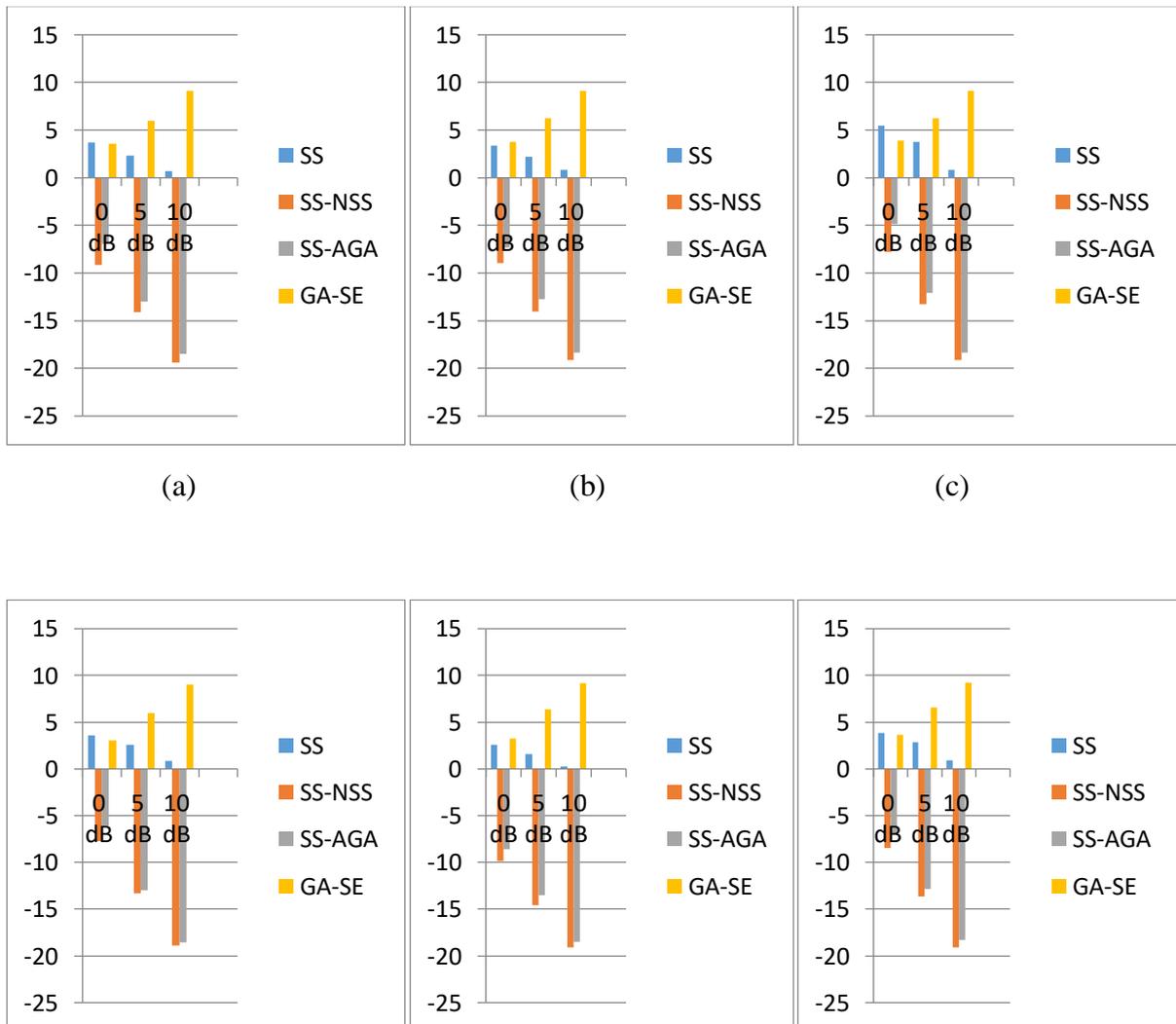


Figure 3.2 Les allures temporelle : (a) signal parole propre, (b) signal parole bruitée, (c) signal parole rehaussée en utilisant la méthode SS, (d) signal parole rehaussée en utilisant la méthode SS-NSS, (e) signal parole rehaussée en utilisant la méthode SS-AGA ,(f) signal parole rehaussée en utilisant la méthode GA_SE . Cas du bruit blanc 0 dB.



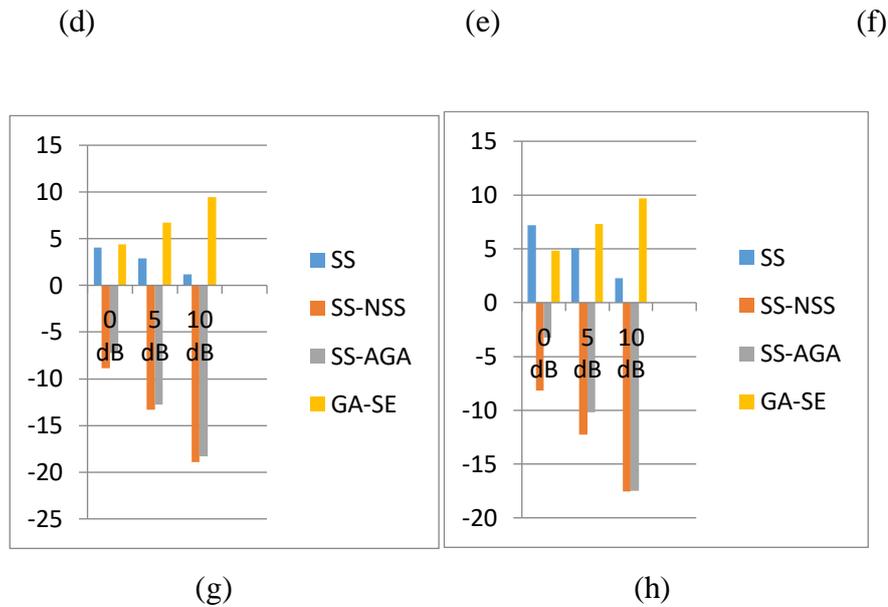


Figure 3.3 Evaluation SNR amélioré pour les méthodes SS, SS-NSS, SS-AGA et GA-SE en utilisant : (a) bruit d'aéroport de NOIZEUS, (b) bruit babble de NOIZEUS, (c) bruit de véhicule de NOIZEUS, (d) bruit salle d'exposition de NOIZEUS, (e) bruit restaurant de NOIZEUS, (f) bruit rue de NOIZEUS, (g) bruit train de NOIZEUS, (h) bruit blanc de NOIZEUS.

3.4 Conclusion

Une application capable de supprimer le bruit dans la parole en utilisant la méthode de soustraction spectrale géométrique a été évaluée avec succès. Quatre méthodes ont été mises en œuvre et leurs performances ont été examinées, en décidant laquelle est la meilleure. Comme mentionné précédemment, toujours de notre point de vue, la mesure subjective de la qualité de parole est la plus importante parmi les mesures d'évaluation. En conclusion, pour nous, les meilleurs résultats sont obtenus avec l'algorithme de la méthode géométrique. L'algorithme GA_SE a obtenu globalement une valeur autour de 2 sur l'échelle MOS (PSEQ) pour tous les bruits à SNR=0 (**milieu fortement bruité**), dont la note la plus élevée est 4.5. Cette valeur est de un demi-point au-dessus de la pire méthode (SS).

Conclusion générale

Conclusion générale

Ce travail s'articulait sur l'étude d'un algorithme de rehaussement de la parole utilisant les domaines spectraux à base de la géométrie vectorielle. Comme vu initialement, les systèmes de traitement de parole fonctionnant en environnement non contrôlé opèrent avec des signaux de parole bruités. Il est donc nécessaire pour leur bon fonctionnement de réduire au maximum la présence de bruit.

L'approche géométrique étudiée (GA) a été évaluée avec succès, elle a montré de meilleurs résultats en termes d'intelligibilité de la parole rehaussée que tous les autres algorithmes comparés dans le chapitre 3. Et ce pour tous les types de bruits étudiés. De plus, en termes de PESQ, il semble être le meilleur estimateur pour les bruits stationnaires à haut RSB (ou SNR) et est comparable aux autres algorithmes pour les autres conditions, par contre les autres méthodes de rehaussement (la soustraction spectrale de base (SS), la soustraction spectrale non linéaire (SS-NSS) et la soustraction spectrale basée sur le moyennage adaptatif du gain (SS-AGA)) semblent introduire davantage de distorsion que la méthode étudiée. Le PESQ est généralement davantage corrélé avec les distorsions alors que le RSB segmental est davantage corrélé avec la réduction de bruit. Cela semble donc indiquer que l'algorithme GA double réduit davantage le bruit que les autres algorithmes comparés, mais en contrepartie il introduit également davantage de distorsion. Ce compromis entre la réduction du bruit et l'ajout de distorsion dans la parole est bien connu dans le domaine du rehaussement de la parole.

Références bibliographiques

Références bibliographiques

- [1] P. C. Loizou, *Speech Enhancement Theory And Practice*, 1st Edition, CRC Press (2007).
- [2] A. Acero, *Acoustical And Environmental Robustness In Automatic Speech Recognition*, Kluwer academic Publisher, 1993.
- [3] Md. F.R. Chowdhury, *A Soft Computing Approach For On-Line Automatic Speech Recognition In Highly Non-Stationary Acoustic Environments*, Ph.D in Telecommunications, INRS, université du Québec, 2012.
- [4] J.C. Junqua, J. Haton, *Robustness In Automatic Speech Recognition: Fundamentals And Applications*, Kluwer Academic Publisher (1996).
- [5] M. Akbacak, J.H.L. Hansen, *Environmental Sniffing : Noise Knowledge Estimation For Robust Speech Systems*, *IEEE Trans. on ASLP*. 15(2) (2007) 465-477.
- [6] K. Pearsons, R. Bennett, S. Fidell, *Speech Levels In Various Noise Environments*, Technical Report EPA-600/1-77-025, US, Environmental Protection Agency.
- [7] Lu, Y., & Loizou, P. C. (2008). A geometric approach to spectral subtraction. *Speech communication*, 50(6), 453-466.
- [8] H. Beigi, *Fundamentals of speaker recognition*, Springer Science & Business Media, 2011.
- [9] Krishna Kant Singh, Mohamed Elhoseny, Akansha Singh, Ahmed A. Elngar - *Machine Learning and the Internet of Medical Things in Healthcare*-Academic Press (2021) p 209
- [10] Saeed V. Vaseghi - *Multimedia Signal Processing_Theory and Applications in Speech, Music and Communications*-Wiley (2007)
- [11] *The electrical engineering and applied signal processing series*, Gillian M. Davis - *Noise Reduction in Speech Applications*-CRC Press (2002)
- [12] A.M. Goberman, M.P. Robb, *Acoustic examination of preterm and full-term infant cries: The long-time average spectrum*, *J. Speech, Language, Hearing Res.* 42 (4) (1999) 850-861.
- [13] Loucif R et Ayache H. 2004. *Débruitage du signal de la parole*. Ecole militaire polytechnique. Alger.
- [14] Berouti, R. Schwartz, J. Makhoul, 1979. *Enhancement Of Speech Corrupted By Acoustic Noise*, In Proc. IEEE. ICASSP. 208-211
- [15] Cappe, O., 1994. *Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor*. *IEEE Trans. Speech Audio Process.* 2 (2), 346–349.
- [16] Ephraim, Y., Malah, D., 1984. *Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator*. *IEEE Trans. Acoust. Speech Signal Process.* ASSP-32

(6), 1109–1121

[17] H. Gustafsson, S. Nordholm, I. Claesson, 2001. Spectral subtraction using reduced delay convolution and adaptive averaging, *IEEE Trans. Speech Audio Process.*, 9(8) 799-807. AGA

[18] Loizou, P., 2005. Speech enhancement based on perceptually motivated Bayesian estimators of the speech magnitude spectrum. *IEEE Trans. Speech Audio Process.* 13 (5), 857–869.

[19] Papoulis, A., Pillai, S., 2002. *Probability Random Variables and Stochastic Processes*, fourth ed. McGraw-Hill, Inc., New York

[20] P. Lockwood, J. Boudy, 1992. Experiments With A Non-Linear Speech Subtractor (NSS), Hidden Markov Models And Projections For Robust Recognition In Cars, *Speech Comm.* 11(2-3) 215-228.

[21] S. Boll, 1979. Suppression Of Acoustic Noise In Speech Using Spectral Subtraction, *IEEE Trans. On ASSP.* 27(2) 113-120.

[22] Hu Y, Loizou P.C. ” Subjective comparaison and evaluation of speech enhancement algorithms”. available at : <https://www.ncbi.nlm.nih.gov/pmc/articles/pmc2098693/>

[23] Hirsch H, Pearce D. 2000. ” The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions”. ISCA ITRWASR. Paris, France.

[24] Fletcher H, Steinberg J. C. 1929. “Articulation Testing Methods”. *Bell Syst. Tech. J.* 8, 806-854.

[25] 2003. “Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm”. ITU-T, ITU-T Rec. P. 835. available at : https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-P.835-200311-I!!PDF-E&type=items

[26] Richards D. 1965. “Speech Transmission Performance Of PCM Systems”. *Electron Letters.* P, 40-91.

[27] (2000). “Perceptual Evaluation Of Speech Quality (PESQ) And Objective Method For End-To-End Speech Quality Assessment Of Narrowband Telephone Networks And Speech Codecs”. ITU-T. Recommendation P.862.

[28] Klatt D. 1988. “Prediction Of Perceived Phonetic Distance From Critical Band Spectra”. In *Proc. IEEE. ICASSP 7.* P, 1278-1281.

[29] Quackenbush S, Barnwell T, clements M. “Objective Measure Of Speech Quality”. Englewood Clift, NJ : Prentice Hall.

