REPUBLIQUE 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 : Systèmes des Télécommunications

Par

- **BAKHOUCHE DIABE**
- **KERAI HANANE**

Intitulé

SUPPRESSION DU BRUIT DANS LES SIGNAUX PAROLE AVEC LES METHODES DE LA SOUSTRACTION SPECTRALE

Evalué le : 15/09/2021

Par la commission d'évaluation composée de* :

Nom & Prénom	*Grade*	*Qualité*	*Etablissement*
Dr. Idris MESSAOUDENE	*MCA*	*Président*	*Univ-BBA*
Dr. Nassim ASBAI	*MCA*	*Encadreur*	*Univ-BBA*
Dr. Salim ATIA	*MCA*	*Examinateur*	*Univ-BBA*

Année Universitaire 2020/2021

Remerciements

Tout d'abord nous tenons à remercier en tout premier lieu ALLAH, le tout puissant de nous avoir donné la force et la patience et la volonté d'accomplir ce modeste travail.

Nous adressons toute notre reconnaissance à notre encadreur, Monsieur Nassim ASBAI, professeur à l'université de Bordj Bou Arréridj, pour nous avoir fait confiance et nous avoir dirigés pendant cette année. Nous tenons tout particulièrement à le remercie pour sa patience tout au long de ce travail, ainsi que pour la richesse de ses idée. Espérant d'avoir l'occasion de travail avec lui dans le futur, qu'il trouve ici l'expression de notre respectueuse gratitude.

Nous remercions les membres du jury, qui nous ont fait l'honneur de participer au jugement de ce travail.

Nos remerciements les plus sincères à toutes les personnes qui ont contribué de près ou de loin à l'élaboration de ce mémoire.

***Merci** ...*

Dédicace

Je dédie ce travail à toute

Ma famille,

Mes parents, mers frères, mes sœurs.

Diab

Dédicace

Je dédie ce modeste travail aux étoiles de ma vie

Ma mère** et **mon père

*A tous les membres de la famille **kerai***

A toutes mes amies

Et à tous ceux qui m'aiment

Hanene

Sommaire

Chapitre 2 : Méthodes de rehaussement de la parole

Chapitre 3 : Résultats et discussions

Liste des figures

Chapitre 1

Chapitre 2

Chapitre 3

Liste des abréviations

AGA	Adaptive Gain Averaging
BAK	Background Intrusiveness
DTFT	Discret Fourier Transform
FFT	Fast Fourier Transform
IDFT	Inverse Discret Fourier Transform
LLR	Log Likelihood Ratio
MOS	Mean Opinion Score
NSS	Nonlinear Spectral Subtraction
OVL	Overall Quality
PESQ	Perceptual Evaluation of Speech Quality
RAP	Reconnaissance Automatique de la Parole
RSB	Rapport Signal sur Bruit
SNR	Signal to Noise Ratio
SIG	Signal Distortion
SegSNR	Segmental Signal to Noise Ratio
SS	Subtraction Spectral
STFT	Short Time Fourier Transform
TFD	Transformée de Fourier Discrète
TF	Transformée de Fourier
VAD	Détecteur d'activité vocale
WSS	Weighted Slope Spectral

Liste des tableaux

Résumé

L'objectif de ce mémoire est d'étudier et de simuler des algorithmes de suppression du bruit dans les signaux de parole en utilisant la méthode de soustraction spectrale. Cette méthode consiste à calculer le spectre de la parole bruitée à l'aide de la transformée de Fourier rapide (FFT) et à soustraire l'amplitude moyenne du spectre de bruit du spectre de la parole bruitée. Trois méthodes ont été mises en œuvre qui sont ; Soustraction spectrale de base (SS), soustraction spectrale non linéaire (SS-NSS) et soustraction spectrale basée sur le gain moyen adaptatif (SS-AGA). Les performances de ces algorithmes sont 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, suppression de bruit, traitement de la parole.

Abstract

The objective of this memory is to study and simulate noise suppression algorithms in speech signals using the spectral subtraction method. This method consists of calculating the spectrum of the noisy speech using the fast Fourier transform (FFT) and subtracting the average amplitude of the noise spectrum from the spectrum of the noisy speech. Three methods have been implemented which are; Basic Spectral Subtraction (SS), Nonlinear Spectral Subtraction (SS-NSS) and Adaptive Average Gain Based Spectral Subtraction (SS-AGA). The performance of these algorithms is evaluated by comparing the original crisp speech and the processed speech using objective and subjective measurements and the output signal-to-noise ratio (SNR).

Key words: Spectral subtraction, noise suppression, speech processing.

ملخص

الهدف من هذه المذكرة هو دراسة ومحاكاة خوارزميات قمع الضوضاء في إشارات الكلام باستخدام طريقة الطرح الطيفي. تتكون هذه الطريقة من حساب طيف الكلام الصاخب باستخدام تحويل فورييه السريع (FFT) وطرح متوسط اتساع طيف الضوضاء من طيف الكلام الصاخب. تم تنفيذ ثلاث طرق: الطرح الطيفي الأساسي (SS) والطرح الطيفي غير الخطي (SS-NSS) والطرح الطيفي القائم على الكسب التكيفي (SS-AGA). يتم تقييم أداء هذه الخوارزميات من خلال مقارنة الكلام الأصلي الواضح والكلام المعالج باستخدام قياسات موضوعية وذاتية وقياسات نسبة الإشارة إلى الضوضاء (SNR).

الكلمات المفتاحية: الطرح الطيفي ، قمع الضوضاء، معالجة الكلام.

Introduction générale

Le signal de parole, sous sa forme physique, vibration de la pression de l'air, doit être transformé en premier lieu en un signal électrique et puis, plus souvent, numérisé pour être stocké ou transmis afin de permettre tout traitement éventuel (débruitage, codage, reconnaissance, compression,...) [1].Le traitement de la parole doit faire face à de nombreux problèmes, entre autres le problème de bruit. La présence d'un bruit superposé au signal utile dégrade la qualité et l'intelligibilité de la parole.

Pour remédier au problème de la dégradation de la qualité de la parole bruitée, la réduction du bruit acoustique ou le rehaussement de la parole est nécessaire, afin d'améliorer la qualité et l'intelligibilité du signal vocal et par conséquent améliorer les performances des applications vocales. Les techniques de rehaussement sont très nombreuses parmi elles : les techniques de soustraction spectrale [2], les techniques paramétriques [3], les techniques à base de modèles statistiques [4] et les techniques de sous-espace [5]. Dans ce mémoire, nous sommes intéressés à l'étude de trois méthodes de rehaussement de parole basées sur la soustraction spectrale à savoir : méthode de soustraction spectrale de base (SS), méthode de soustraction spectrale non linéaire (SS-NSS) et méthode de soustraction spectrale basée sur le moyennage adaptative du gain (SS-AGA), dans le but de réduire le bruit dans le signal de parole.

Les trois techniques de soustraction spectrale ou les techniques d'estimation du spectre à court terme étudiées dans ce travail, cherchent à estimer le spectre à court-terme du signal propre à partir de signal bruitée. Ces techniques sont très connues pour leur simplicité, leur faible complexité et la "bonne" qualité du signal d'ébruité [6].

Pour évaluer la qualité de rehaussement de parole ou comparer les performances des différentes techniques étudiées, on mesure la qualité des signaux débruités, en utilisant un ou plusieurs critères d'évaluation, les critères d'évaluation subjectifs sont obtenus en utilisant des tests d'écoute dans lesquels des participants humains évaluent la qualité de la parole sous trois aspects : la distorsion du signal, le bruit et la qualité globale du signal. Ces critères sont plus fiable en terme de précision de l'évaluation mais ils sont très couteux en temps et en équipements expérimentaux. Une solution alternative aux tests d'écoute consiste à développer

des critères objectifs, qui sont des mesures mathématiques entre le signal original et le signal débruité. Dans ce sens plusieurs critères ont été établis : Rapport signal sur bruit segmental (segSNR) [7], Evaluation Perceptive de la Qualité de la Parole (PESQ) [8], Pente Spectrale Pondérée (WSS) [9], Logarithme du Rapport de Vraisemblance (LLR) [10].

Ce mémoire est organisé en trois chapitres, deux chapitres pour la partie théorique (chapitre 1 et2) 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 de trois méthodes de rehaussement de la parole : méthode de soustraction spectrale (SS), méthode de soustraction spectrale non linéaire (SS-NSS) et la méthode de soustraction spectrale basée sur le Moyennage Adaptatif du Gain (SS-AGA).

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

Chapitre 1

Généralités sur les signaux parole et de bruit

1.1. Introduction

La parole joue un rôle majeur dans le système de communication humain. Le développement des outils mathématiques et les moyens techniques, électroniques et informatiques, a permis de classer le traitement de la parole comme une composante fondamentale des sciences de l'ingénieur et une discipline à part pour les traiteurs de signaux.

Dans ce chapitre nous présentons des notions fondamentales sur le signal de parole, ainsi que les différents types de bruit qui peuvent affecter le signal vocal.

1.2. Techniques d'analyse

1.2.1 Analyse temporelle à court terme

La représentation à court terme dans le domaine temporel du signal consiste à extraire une représentation trame par trame du signal, L'idée de base est de travailler sur chaque trame. Dans certains cas, le signal est supposé stationnaire sur la trame donnée. La taille de la trame est un paramètre important de signal parole, et dans le cas où cette taille est fixe, elle est choisie pour contenir dans tous les cas au moins 2 à 3 période de signal, pour les signaux de parole continue, il est raisonnable de considérer que les voix d'homme descendent jusqu'a 50 Hz (période égale à 20 ms) et les voix des femmes jusqu'à 100Hz (période égale 10ms), ce qui donne des tailles de fenêtre d'environ 30 ms pour les hommes et 20 ms pour les femmes.

1.2.2. Analyse fréquentielle à court terme

D'un point de vue théorique, la théorie des série de Fourier indique qu'un signal parfaitement périodique peut être décompose en une somme d'exponentielles dont les fréquences sont toutes multiple d'une fréquence particulière, cette fréquence est appelée fréquence fondamentale, et correspondant à l'inverse de la plus petite période du signal : les différentes exponentielles sont appelées harmoniques du signal, et chacune d'entre elle est associée à une amplitude. En fait, les signaux étudiés n'étant pas périodique, la décomposition en série de Fourier doit être utilisée sur des signaux non périodiques et qui fournit quand même les harmoniques dans le cas d'un signal périodique.

D'un point de vue pratique, la seule analyse réalisable est une analyse à court terme (et a support temporel borne) dans le domaine fréquentiel, cette analyse nécessite l'utilisation

d'une fenêtre sur le signal, ce qui modifie la décomposition (de séparation) en Harmoniques. Néanmoins, si la taille de la fenêtre est suffisante, alors l'analyse fréquentielle sera en générale capable de séparer les harmoniques de signal et donc d'en estimer les caractéristiques.

1.2.3. Fonctions de fenêtrage

Le signal de parole est fortement non-stationnaire. Afin de pouvoir utiliser des opérateurs, tel qu'une transformée de Fourier (qui suppose la stationnarité du signal), il est nécessaire de séparer le signal en trames de telle sorte que chaque trame puisse être considérée comme étant un signal stationnaire. Les fonctions de fenêtrage sont des fonctions de poids qui s'appliquent directement sur chaque trame. Il existe plusieurs types de fenêtres ayant chacune des propriétés différentes.

Dans ce cas, la fenêtre Hanning a été choisie. Elle a la forme d'un cycle d'une onde cosinusoïdale plus un décalage, donc elle est toujours positive.

La fonction de Hanning fait partie de la famille des fenêtres de type cosinus. Les avantages de cette famille est la facilité d'obtenir les propriétés spectrales de façon analytique et qu'elle soit nulle à ces extrémités. Son équation pour un signal discret est donnée par :

$$W_{Hann}[n] = 0.5 + 0.5\cos(2n\pi/N) \tag{1.1}$$

Où

N : Longueur de fenêtre (nombre d'échantillons).

n : L'index temporel discret d'échantillon.

Cette fenêtre débute et termine à zéro, ce qui permet de couper toute discontinuité. D'un point de vue spectral, elle n'a pas une atténuation aussi rapide que la fenêtre de Hamming proche du premier lobe, mais elle a une meilleure atténuation plus on s'éloigne du premier lobe [11].

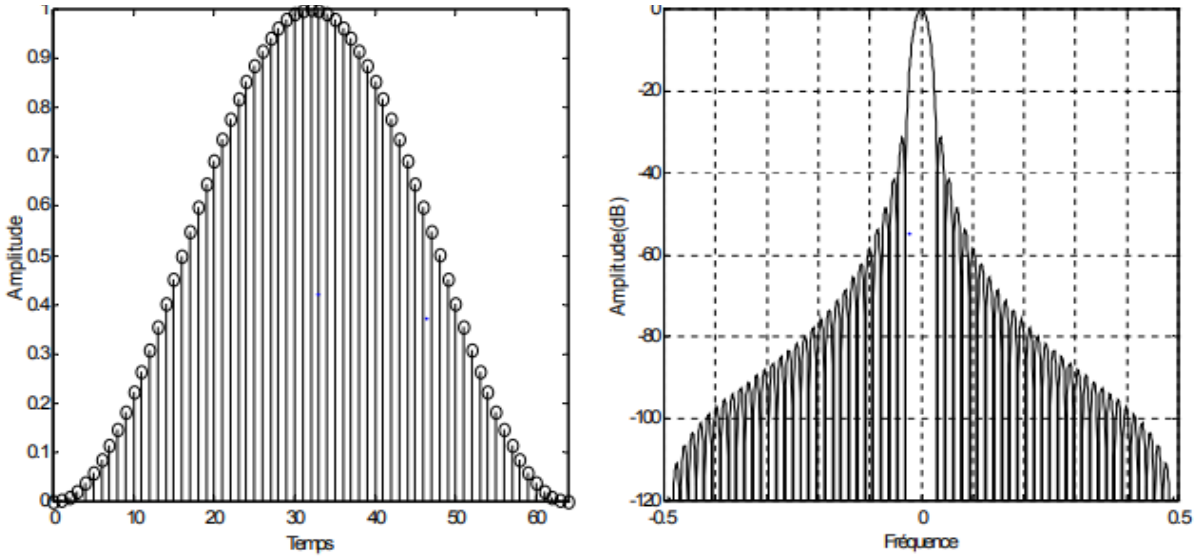


Figure 1.1 Fenêtre de Hanning

1.2.4. Analyse de fréquence

La conversion d'un signal dans le domaine fréquentiel est le résultat de la décomposition du signal en composantes sinusoïdales. Pour cela, deux outils mathématiques sont utilisés, selon la nature continue ou discrète du signal.

Pour les signaux continus, la transformation de Fourier rapide (FFT) est utilisée, tandis que pour les signaux discrets ou les séquences, l'outil doit être la transformation de Fourier discrète (DFT). Ce projet se concentre sur la DFT, car les signaux audio numériques sont des signaux discrets ou des séquences.

- ✓ **Transformée de Fourier discrète (DFT)**

La transformée de Fourier discrète, ou transformée de Fourier d'une suite $x[n]$, est une fonction $X(e^{jw})$, continue et périodique, de période 2π. Il peut être calculé avec l'expression suivante :

$$X\left(e^{jw}\right) = \sum_{n=-\infty}^{\infty} x[n]\, e^{-jwn} \tag{1.2}$$

La transformée de Fourier inverse en séquences (IDFT) renverra la séquence d'origine, étant son expression appelée synthèse :

$$x[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} X\left(e^{jw}\right)e^{jwn}dw \quad (1.3)$$

Pour synthétiser à nouveau $x[n]$, il faut utiliser la transformée de Fourier inverse.
Comme $X\left(e^{jw}\right)$ c'est une fonction complexe avec des composants réels et imaginaires, elle peut être représentée comme :

$$X\left(e^{jw}\right) = X_R\left(e^{jw}\right) + jX_I\left(e^{jw}\right) \quad (1.4)$$

Ou, sous forme polaire, le module et la décomposition de phase :

$$X\left(e^{jw}\right) = \left|X\left(e^{jw}\right)\right|.e^{j\angle X\left(e^{jw}\right)} \quad (1.5)$$

Où sont$\left|X\left(e^{jw}\right)\right|$ le module et $\angle X\left(e^{jw}\right)$la phase.

- ✓ **Transformée de Fourier à court terme**

La transformée de Fourier à court terme ou la STFT (Short Time Fourier Transforme) est un outil puissant pour le traitement du signal vocal [12], cette transformation consiste à faire des analyses locales d'un signal. Nous balayons ce dernier par des fenêtres étroites, tel que les parties du signal vu à travers ces fenêtres sont en effet stationnaires. Dans le but de réaliser le meilleur compromis entre les résolutions temporelles et fréquentielles**,** la définition mathématique de la STFT est montrée dans la relation suivante [2] :

$$X_m(w) = \sum_{n=-\infty}^{\infty} x(n)w(n - mR)e^{-jwn} \quad (1.6)$$

Où

$x(n)$: Signal d'entrée à l'instant n, n: l'indice de temps discret,

$w(n)$: Fonction de fenêtre de longueur M (avec M c'est le nombre d'échantillons),

Ex : fenêtre de hanning,

$X_m(w)$: DTFT de signal fenêtrée centré sur le temps mR,

R : Taille de saut en échantillons, entre les DTFT successives.

1.3. Signal parole

1.3.1. Définition

Le flux d'air en provenance des poumons est modulé par les cordes vocales, créant des ondes de pression qui se propagent à travers le conduit vocal. Ce dernier, constitué de cavités buccales et nasales, se comporte comme un filtre caractérisé par des fréquences de résonance appelées formants [13]. La parole est un signal continu, d'énergie finie, non stationnaire. Sa structure est complexe et variable dans le temps.

1.3.2. Classification des sons de la parole

Une décomposition simplifiée du signal de la parole doit ressortir deux types de sons : voisés et non voisés.

- Les sons voisés, tels que des voyelles, sont produits par le passage de l'air de poumons à travers la trachée qui met en vibration les cordes vocales. Ce mode, qui représente 80% du temps de phonation, est caractérisé en général par un quasi périodicité, une énergie élevée et une fréquence fondamentale. Typiquement, la période fondamental des différents sons voisés varie entre 2ms et 20ms.
- Les sons non voisés, comme certaines consonnes, dans ce cas les cordes vocales ne vibrent pas, l'air passe à haute vitesse entre les cordes vocales. Le signal produit est équivalent à un bruit blanc.

Dans le première type de sons, la forme de signal glottique est sensiblement triangulaire, son spectre est riche en harmoniques. La figure 1.2 représente la production d'un son voisé.

Dans le seconde type, la figure 1.3 exprime que le signal est apériodique son spectre est relativement uniforme sur une large bande de fréquence.

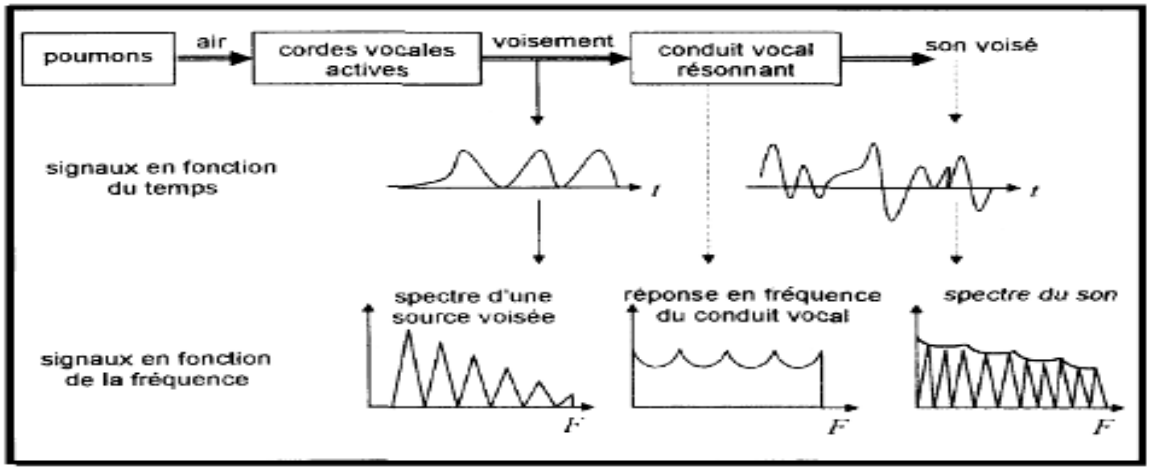


Figure 1.2 Production d'un son voisé [14].

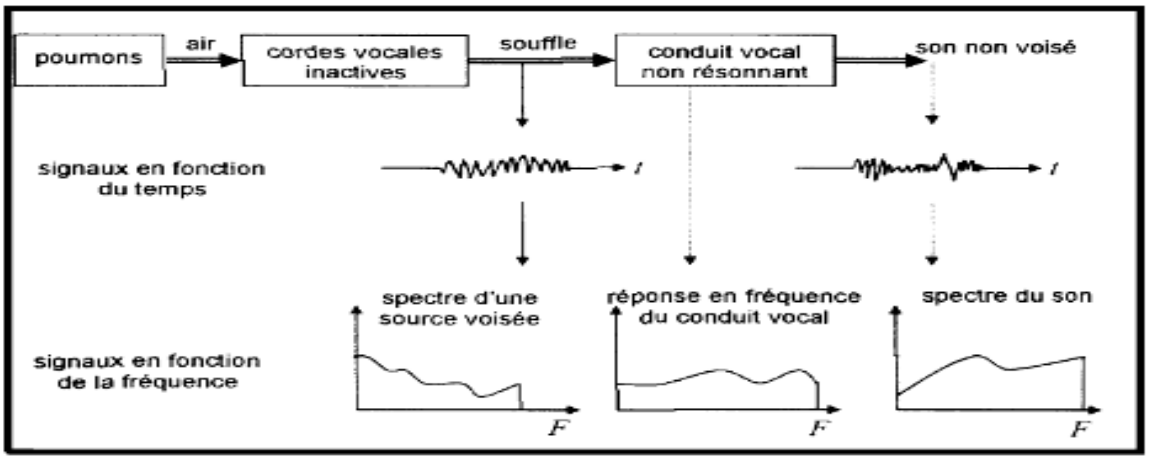


Figure 1.3 Production d'un son non voisé [14].

1.3.3. Paramètres du signal de parole

Le signal de parole est un vecteur acoustique porteur d'information d'une grande complexité, variabilité et redondance. Les caractéristique de ce signal sont appelées traits acoustiques. Chaque trait acoustique a une signification sur le plan perceptuel [15].

1.3.3.1. La fréquence fondamentale

La fréquence fondamentale est parmi les paramètres acoustiques d'un signal parole, souvent désignée par F0, fait référence à la fréquence approximative de la structure (quasi) périodique des signaux vocaux. La fréquence fondamentale est définie comme le nombre moyen d'oscillation par seconde et exprimée en hertz.

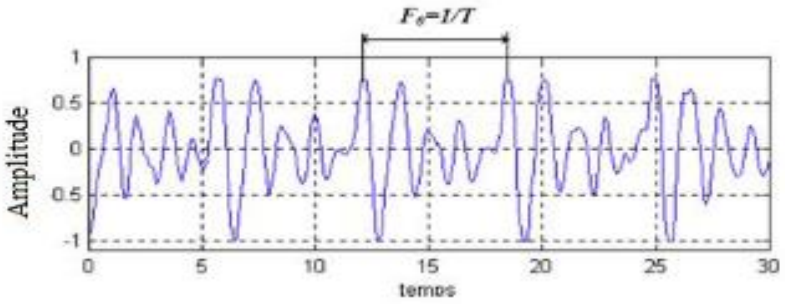


Figure 1.4 Un son voisé avec sa fréquence fondamentale 'F0' [16].

1.3.3.2. Spectre fréquentiel

Le spectre fréquentiel dont dépend principalement le timbre de la voix. Le timbre est une caractéristique permettant d'identifier une personne à la simple écoute de sa voix. Le timbre

dépend de la corrélation entre la fréquence fondamentale et les harmoniques qui sont les multiples de cette fréquence [17].

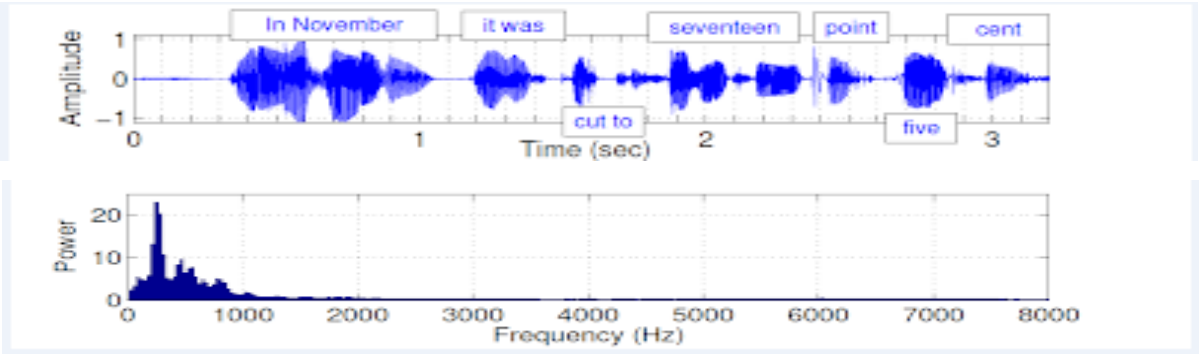


Figure 1.5 Exemple d'un signal parole et son spectre [18].

1.3.3.3. 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 :

$$E = \sum_{n=0}^{N-1} s^2(n) \tag{1.7}$$

Avec : N est la taille de la trame.

1.3.4. Propriétés statistiques du signal vocal

L'audiogramme ou l'évolution temporelle du signal vocal ne fournit pas directement les traits acoustiques du signal, il est nécessaire de mener un ensemble de calculs statiques. Le signal vocal peut être vu comme un processus aléatoire non stationnaire [1].

1.3.4.1. Densité de probabilité

Si N_ξ représente le nombre d'échantillons $x(n)$ dont l'amplitude est comprise entre $[\xi - \Delta\xi/2, \xi + \Delta\xi/2]$ alors que $n \in [-N, N]$, la densité de probabilité du signal x supposé ergodique et stationnaire est donnée par :

$$P_x(\xi) = lim_{N\to\infty}\left[\frac{N_\xi}{2N+1}\right] \tag{1.8}$$

1.3.4.2. La valeur moyenne et la variance

La valeur moyenne d'un système stationnaire vaut :

$$\mu_x = \int_{-\infty}^{+\infty} \xi P_x(\xi) d\xi = lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x(n) \quad (1.9)$$

Pour le signal vocal cette moyenne est supposé nulle, elle ne contient aucune information utile. La variance est donnée par :

$$\sigma^2{}_x = \int_{-\infty}^{+\infty} \xi^2 P_x(\xi) d\xi = lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x^2(n) \quad (1.10)$$

1.3.4.3. Fonction d'autocorrélation

Soit $x(n)$ un signal numérique obtenu par échantillonnage avec période T_e et $x(n+k)$ le signal lui-même décalé par le temps discret d'indice k. Soit N le nombre de points utilisés pour calculer la moyenne ($T = NT_e$) [19].

La fonction d'auto corrélation d'un signal ergodique et stationnaire s'exprime par :

$$\emptyset_{xx}(k) = lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x(n)x(n+k) \quad (1.11)$$

La fonction d'autocovariance est définie et estimée par des formules identiques après avoir soustrait la moyenne μ_x et comme $\mu_x = 0$ alors, ces deux notions peuvent être confondues. On a aussi : $\sigma_x^2 = \emptyset_{xx}(0)$.

Le coefficient d'autocorrélation est défini par $p_x(k) = \frac{\emptyset_{xx}(k)}{\emptyset_{xx}(0)}$ sa valeur est comprise entre 1 et -1.

1.3.4.4. Densité spectrale de puissance

Densité spectrale de puissance $S_{xx}(\theta)$ est la transformée de fourrier de la fonction d'auto corrélation :

$$S_{xx}(\theta) = \sum_{k=-k} \emptyset_{xx}(k) e^{-jk\theta}, \qquad \theta = \omega T_e \quad (1.12)$$

$$S_{xx}(\theta) = \sum_{k=-k} \emptyset_{xx}(k) w(k) e^{-jk\theta} \quad (1.13)$$

Où $w(k)$ est une fonction fenêtre.

1.3.5. Chaine d'acquisition du signal de parole

La parole apparaît physiquement comme une variation de la pression de l'air causée et émise par le système articulatoire. La phonétique acoustique étudie ce signal en le transformant dans un premier temps en signal électrique grâce au transducteur approprié : le microphone.

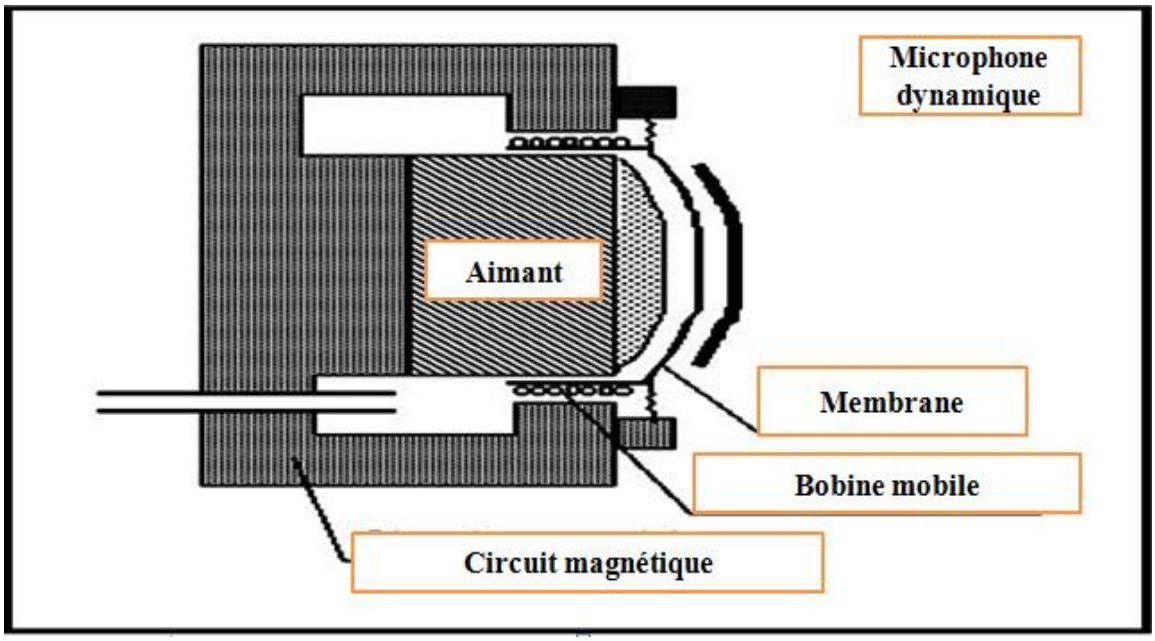


Figure 1.6 Microphone dynamique [20].

Une bobine mobile est solidaire de la membrane. Cette dernière, sous l'effet des variations de pression acoustique fait osciller la bobine dans un champ magnétique annulaire produit par un aimant permanent. La bobine coupe les lignes de force du champ magnétique.

En raison de ces oscillations périodiques de la membrane, et donc de la bobine dans le champ, il y a induction d'un courant électrique dans la bobine mobile - courant qui peut être rendu utilisable par une amplification appropriée [20]. De nos jours, le signal électrique résultant est le plus souvent numérisé. Il peut alors être soumis à un ensemble de traitements statistiques qui visent à en mettre en évidence les traits acoustiques : sa fréquence fondamentale, son énergie, et son spectre. Chaque trait acoustique est lui-même intimement lié à une grandeur perceptuelle : pitch et intensité. L'opération de numérisation, schématisée à la figure (1.7) requiert successivement : un filtrage de garde, un échantillonnage, et une quantification.

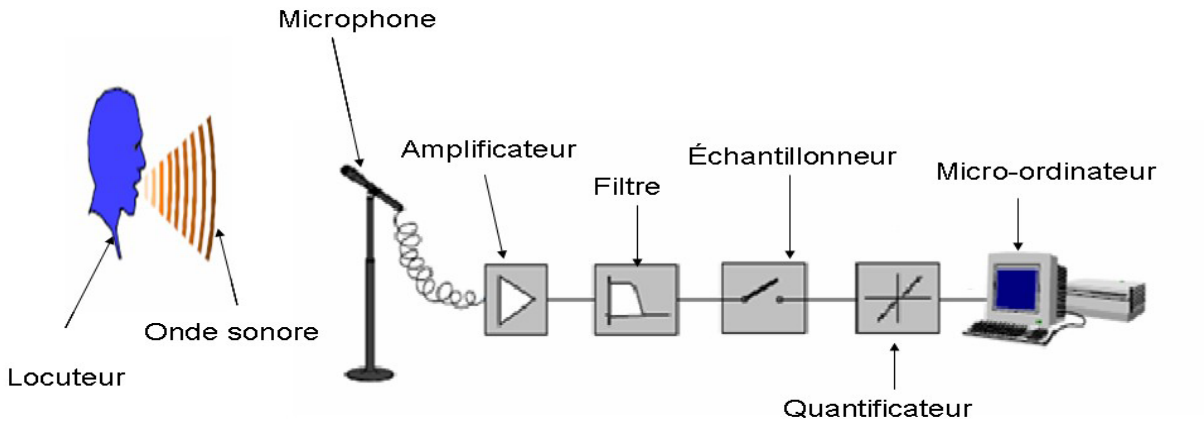


Figure 1.7 Chaine d'acquisition du signal de parole [21].

1.4. Bruit dans le signal de parole [16]

1.4.1. Définition d'un bruit

Le bruit peut être défini comme un "son indésirable" et une énergie acoustique audible qui nuit au bien-être physiologique et / ou psychologique des personnes, ou qui dérange ou altère la commodité ou la paix de quiconque. On peut généraliser en disant que le son devient indésirable lorsqu'il :

• gêne la communication vocale,

• entrave le processus de réflexion,

• interfère avec la concentration,

• gêne les activités (travail ou loisirs),

• présente un risque pour la santé en raison de troubles auditifs.

1.4.2. Différents types de bruit

Les différents bruits affectant un message peuvent être divisés :

- **Selon la nature du bruit**

Elle contient, elle-même, trois catégories :

- ✓ **Bruit à bande large**

Qui touche une bande fréquentielle un peu importante (large bande), à titre d'exemple on trouve le bruit blanc.

Figure 1.8 Spectre d'un bruit blanc [22].

- ✓ **Bruit à bande étroite**

Qui touche une bande fréquentielle limitée par un intervalle (qui a un spectre de raies bien défini), dans ce cas on trouve par exemple le bruit des outils de travail des ouvriers tel que le bruit du petit matériel électrique. 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 intermittent dans le temps (bruit impulsionnel)**

C'est un signal qui s'annule dans certains intervalles de temps. Ce bruit nous aide à l'extraction du signal de parole pendant ces intervalles.

- ➢ **Selon les caractéristiques de stationnarité du bruit**

Elle contient aussi trois classes : bruits stationnaires ou quasi-stationnaires, bruits rythmiques et bruits aléatoires.

- ✓ **Bruits stationnaires ou quasi-stationnaires**

On trouve ce type dans les organes moteurs fonctionnant moins vite. Ce type de bruit affecte beaucoup plus les voyelles dans un signal de parole.

- ✓ **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.

✓ **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 bruyant. Un exemple des bruits aléatoire est le bruit de soudures qui correspond à des raies verticales à des instants aléatoires dans le temps.

➢ **Selon le mode d'interférence**

C'est la classification la plus importante, qui nous intéresse. Il y a deux types principaux : les bruits additifs et les bruits convolutionnels (multiplicatifs).

✓ **Les bruits additifs**

Le bruit additif est important lorsque le signal est formé par l'addition de parole et de bruit propres. Ainsi, la réduction de bruit réalise la tâche de séparer ces deux signaux de la manière la plus optimale.

Les bruits additifs sont dus à la multiplicité des systèmes de communication dans un même environnement. Plusieurs émetteurs et plusieurs récepteurs pouvant être confinés dans un même espace, les messages de tous les émetteurs peuvent donc se trouver en concurrence sur une même voie sans que les récepteurs possèdent un mécanisme infaillible pour isoler le message qui leur est destiné. L'émetteur et le récepteur peuvent aussi se trouver en présence d'un ou de plusieurs équipements générant un bruit de fond de puissance variable. Les bruits additifs peuvent être subdivisés en trois groupes en fonction des lieux où ils peuvent être rencontrés :

a) Bruits des systèmes industriels

Ils peuvent être très intenses et sont, par nature, non stationnaires. Ils correspondent aux bruits émis par des machines possédant une faible isolation phonique.

b) Bruits des moyens de transport

Ils correspondent aux bruits qui peuvent être observés dans divers véhicules tels que les voitures, les trains et les avions…etc.

c) Bruits des milieux administratifs et urbains

Ce sont les bruits présents dans les bureaux, les domiciles ou dans les concentrations urbaines. Ces bruits peuvent être très variés (climatisation, bruit de parole) mais sont peu intenses.

Il est également possible de classer, dans ce type des bruits produits par les systèmes industriels. Ce sont les bruits qui peuvent être rencontré dans les systèmes de ventilation, les

machines à écrire ou les ordinateurs. À cette liste peuvent être ajoutés les bruits de mobiles par rapports à l'auditeur tels que les voitures.

- ✓ **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.

- ➢ **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 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 le plus remarquable 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 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.

Comme notre but est essentiellement le débruitage de parole, dans notre mémoire nous utiliserons des bruits additifs décorrélé avec la parole, tels que le bruit de conversation appelé bavardage (babble), le bruit de voiture (car), et le bruit blanc gaussien.

1.5. Conclusion

Dans ce chapitre, nous avons présenté des notions fondamentales sur le signal de parole, leurs propriétés ainsi que les différents types de bruits qui peuvent affecter ce signal.

Nous avons aussi remarqué que le signal vocal est très complexe, du fait de sa grande variabilité, ce qui rend toute tentative de le modéliser ou de reconnaître très délicate.

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

Chapitre 2

Méthodes de rehaussement de la parole

2.1. Introduction

L'amélioration de la parole est un problème important dans de nombreuses applications de traitement de la parole, telles que les communications mobiles. L'objectif principal de rehaussement de la parole est d'améliorer la qualité et l'intelligibilité du signal. Au cours des dernières décennies, diverses approches ont été proposées pour résoudre ce problème, telles que la méthode de soustraction spectrale. Nous présentons dans ce chapitre quelque méthode de l'état de l'art les plus utilisées, notamment celles relatives à la Soustraction Spectrale, les différentes méthodes dérivées seront également présentées dans ce chapitre.

2.2. Méthode de Soustraction Spectrale (SS)

L'algorithme spectral soustractif est historiquement l'un des premiers algorithmes proposé pour la réduction du bruit. Plus d'articles ont été écrits décrivant les variations de cet algorithme que tout autre algorithme. Il repose sur un principe simple. En supposant un bruit additif, on peut obtenir une estimation du spectre du signal propre en soustrayant une estimation du spectre de bruit du spectre de parole bruitée. Le spectre de bruit peut être estimé et mis à jour, pendant les périodes où le signal est absent. L'hypothèse faite est que le bruit est stationnaire ou est un processus à variation lente et que le spectre de bruit ne change pas de manière significative entre la mise à jour de périodes. Le signal amélioré est obtenu en calculant la Transformée de Fourier discrète inverse du spectre de signal estimé en utilisant la phase du signal bruit. L'algorithme est simple du point de vue du calcul car il n'implique qu'un avant et un une transformée de Fourier inverse.

Le simple traitement de soustraction a un prix. Le processus de soustraction a besoin à faire soigneusement pour éviter toute distorsion de la parole. Si trop est soustrait, alors certains les informations vocales peuvent être supprimées, alors que si trop peu est soustrait, alors beaucoup du bruit parasite demeure. De nombreuses méthodes ont été proposées pour atténuer, et dans certains cas, éliminer la plupart de la distorsion de la parole introduite par la soustraction spectrale traitée. Ce chapitre présente l'algorithme de soustraction spectrale de base [2] et décrit certaines des variantes de l'algorithme visant à réduire la parole et distorsion du bruit.

Principe

Les différentes techniques de rehaussement présentées dans ce mémoire se basent sur un signal de parole corrompu par un bruit additif et non corrélé tel que :

$$y(n) = x(n) + d(n) \tag{2.1}$$

Où n représente l'index temporel discrète $y(n)$, $x(n)$ et $d(n)$ représentent respectivement la parole bruitée, le signal de parole propre et le bruit.

Nous supposons que le signal de parole est quasi-stationnaire sur des fenêtres d'analyse de 20 à 30ms. La Transformée de Fourier à Court Terme (TFCT) de $y(n)$ (2.1) est donnée par :

$$Y(w) = X(w) + D(w) \tag{2.2}$$

Où $Y(w)$, $X(w)$ et $D(w)$ sont les transformées de Fourier du signal $y(n)$, $x(n)$ et $d(n)$ respectivement, et w est le Variable de fréquence. En soustraction spectrale, le signal entrant $x(n)$ est tamponné et divisé en segments de N échantillons de longueur. Chaque segment est fenêtré, en utilisant une fenêtre Hanning ou Hamming, puis transformé par transformée de Fourier discrète (DFT) en N échantillons spectraux. Les fenêtres atténuent les effets des discontinuités aux extrémités de chaque segment.

Le signal fenêtré est donné par

$$\begin{aligned} y_w(n) &= w(n)y(n) \\ &= w(n)[x(n) + d(n)] \\ &= x_w(n) + d_w(n) \end{aligned} \tag{2.3}$$

L'opération de fenêtrage peut être exprimée dans le domaine fréquentiel comme

$$\begin{aligned} Y_W(w) &= W(w) * Y(w) \\ &= X_W(w) + D_W(w) \end{aligned} \tag{2.4}$$

Où l'opérateur $*$ désigne la convolution. Tout au long de ce chapitre, il est supposé que les signaux sont fenêtrés, et par conséquent, pour simplifier, nous supprimons l'utilisation de l'indice w pour les signaux fenêtrés.

$$Y(w) = X(w) + D(w) \tag{2.5}$$

$Y(w)$ Et $D(w)$ peuvent être exprimés en forme polaire comme suit

$$Y(w) = |Y(w)|\, e^{j\Phi_y(w)} \tag{2.6}$$

$$D(w) = |D(w)|\, e^{j\Phi_d(w)} \tag{2.7}$$

Ou $|Y(w)|$ et $|D(w)|$ sont les spectres d'amplitude du signal de parole bruité et du bruit additif, respectivement, et $\Phi_y(w)$ et $\Phi_d(w)$ sont leurs phases.

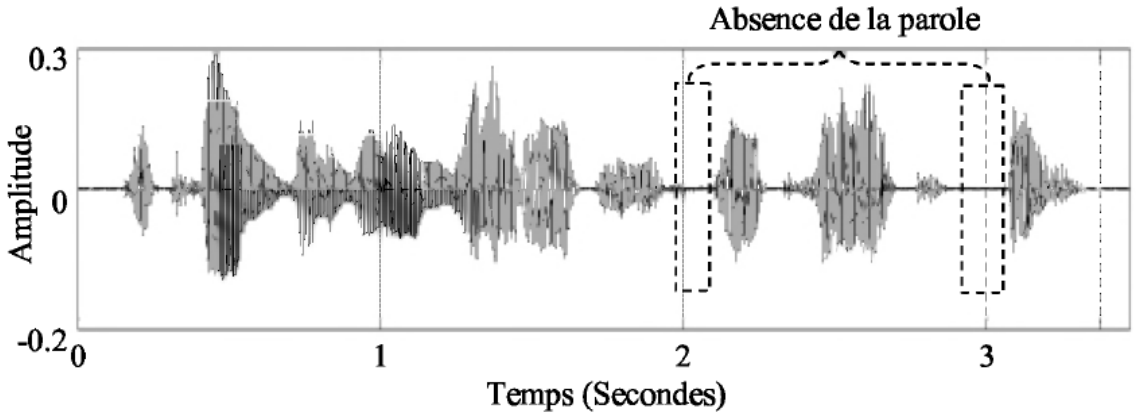


Figure 2.1 Allure temporelle d'un signal de parole non bruité.

Le spectre d'amplitude $|D(w)|$ du bruit additif est estimé par le calcul de sa valeur moyenne pendant l'absence de la parole bruitée comme est montrée dans la figure 2.1 la phase $\Phi_d(w)$ est remplacée par la phase de la parole bruitée $\Phi_y(w)$, car la phase n'affecte pas l'intelligibilité de la parole, mais peut affecter sa qualité dans un certain sens [23]. Le spectre estimé du signal original est donné par

$$\hat{X}(w) = [|Y(w)| - |\widehat{D}(w)|]e^{j\Phi_y(w)} \quad (2.8)$$

Ou $|\widehat{D}(w)|$ 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 $\hat{X}(w)$. On peut remarquer à partir de (2.8) que $|\hat{X}(w)| = |Y(w)| - |\widehat{D}(w)|$ peut prendre des valeurs négatives en raison des erreurs d'estimation du bruit. Plusieurs solutions sont envisageables pour remédier à ce problème [2,24]. Parmi ces techniques on peut citer la rectification demi-onde qui permet de mettre à zéro ces valeurs négatives, comme suit :

$$\begin{cases} |\hat{X}\,(\mathrm{w})\,| = |Y(w)| - |\widehat{D}(w)| & \text{si } |Y(w)| > |\widehat{D}(w)| \\ 0 & ailleurs \end{cases} \quad (2.9)$$

L'algorithme de soustraction spectrale en amplitude (2.8) peut être étendu à la soustraction spectrale de puissance en multipliant $Y(w)$ (2.5) par sa conjuguée $Y^*(w)$ comme suit :

$$|Y(w)|^2 = |X(w)|^2 + |D(w)|^2 + X(w).D^*(w) + X^*(w).D(w)$$

$$=|X(w)|^2 + |D(w)|^2 + 2.\mathrm{Re}\,\{X(w).D^*(w)\} \quad (2.10)$$

Les termes $|D(w)|^2$, $X(w).D^*(w)$ et $X^*(w).D(w)$ ne peuvent pas être obtenus directement, mais approximés par $E\{|D(w)|^2\}$, $E\{X(w).D^*(w)\}\}$ et $E\{X^*(w).D(w)\}$, où $E\{.\}$ désigne l'opérateur espérance mathématique. Généralement $E\{|\widehat{D}(w)|^2\}$ est estimée durant la période d'absence d'activité vocale est désignée par $|\widehat{D}(w)|^2$.

Si nous supposons que $d(n)$ et $x(n)$ est à moyenne nulle et non corrélés avec $x(n)$, les termes $E\{X(w).D^*(w)\}$ et $E\{X^*(w).D(w)\}$, deviennent nuls. La valeur estimée du spectre de puissance du signal propre devient :

$$|\hat{X}(w)|^2 = |Y(w)|^2 - |\widehat{D}(w)|^2 \quad (2.11)$$

L'équation (2.11) décrit l'algorithme de soustraction spectrale en puissance et peut être exprimée par :

$$\left|\hat{X}(w)\right|^2 = H^2(w)\ |Y(w)|^2 \tag{2.12}$$

$$H(w) = \sqrt{1 - \frac{|\hat{D}(w)|^2}{|Y(w)|^2}} \tag{2.13}$$

$H(w)$ Est connu comme la fonction du gain ou la fonction de suppression. Dans la théorie des systèmes linéaires, elle est connue comme la fonction de transfert. Cette fonction est réelle et positive et prenne des valeurs dans l'intervalle [0,1]. Les valeurs dans cet intervalle, représentent la quantité de la suppression ou l'atténuation qui doit être appliquée au spectre de puissance du signal de la parole bruitée y_k.

L'algorithme de la soustraction spectrale peut être présenté sous une forme plus générale comme

$$\left|\hat{X}(w)\right|^p = |Y(w)|^p - \left|\hat{D}(w)\right|^p \tag{2.14}$$

Où p est l'exposant de puissance. Si $p = 1$, alors l'équation (2.14) correspondra à l'algorithme de la soustraction spectrale d'amplitude originale [2]. Si $p = 2$, alors l'équation (2.14) correspondra à l'algorithme de la soustraction spectrale de puissance. La forme générale de l'algorithme de la soustraction spectrale est montrée dans la figure 2.2.

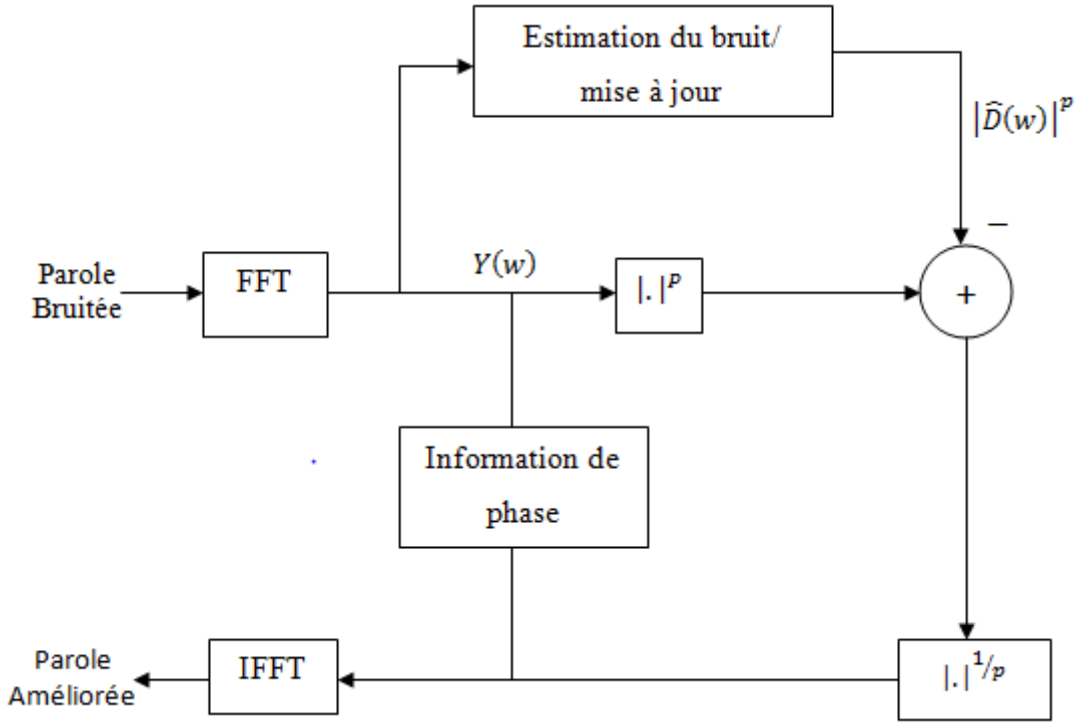


Figure 2.2 Schéma synoptique d'un système de débruitage de la parole par la soustraction spectrale [25].

La soustraction spectrale est efficace du point de vue complexité de calcule, elle est caractérisée par un simple mécanisme de contrôle du compromis entre la distorsion du signal de parole améliorée et le bruit résiduel dans ce dernier. L'inconvénient majeur de cette technique est qu'elle introduise dans le signal amélioré un bruit musical gênant.

2.2.1. Variantes de la soustraction spectrale

Plusieurs variantes de la soustraction spectrale ont été proposées dans la littérature (voir par exemple [26, 27]). Une approche générale permet d'écrire :

$$|\hat{X}(w)| = \begin{cases} \left[|Y(w)|^{\gamma 1} - \alpha|\widehat{D}(w)|^{\gamma 1}\right]^{\gamma 2} & si\ |Y(w)|^{\gamma 1} > \alpha|\widehat{D}(w)|^{\gamma 1} \\ \beta|\widehat{D}(w)|^{\gamma 1 \gamma 2} & sinon \end{cases} \quad (2.15)$$

Ou α est un paramètre de sur-estimation du bruit [2], β est un paramètre de contrôle de niveau du bruit résiduels [24] et les exposants γ_1 et γ_2 sont des termes qui ont un effet sur l'intelligibilité de la parole débrutiée [26].

Le cas ou γ_1 = 2 et γ_2 = 0.5 correspond à la soustraction spectrale de puissance alors que le nom soustraction spectrale d'amplitude est réservé au cas $\gamma_1 = \gamma_2 = 1$.

Grâce aux différents paramètres de contrôle, cette généralisation de la soustraction spectrale introduit une certaine flexibilité. On peut en profiter pour réaliser un compromis entre une bonne réduction du bruit, un faible bruit résiduel et une faible distorsion du signal utile.

2.3. Soustraction spectrale non linéaire

Dans [24**],** il a été supposé que le bruit affecte de la même façon toutes les composantes spectrales. Pour cela, un seul facteur de soustraction α est utilisé pour soustraire une surestimation du bruit sur tout le spectre. Par ailleurs, cette supposition n'est plus opportune dans le cas des bruits environnementaux tel que les bruits de car, de restaurant, de rue, etc. Certains bruits affectent plus la région des basses fréquences que la région des hautes fréquences, d'où la nécessité de favoriser l'utilisation d'un facteur de soustraction dépendant la fréquence.

La méthode NSS-SB (Non linear Spectral Subtraction) [28] peut être considérée comme une modification de la méthode de Berouti, été ce, par l'utilisation d'un facteur de soustraction variable associé à un processus de soustraction non linéaire.

La règle de soustraction utilisée dans l'algorithme de soustraction non linéaire a la forme suivant :

$$|\hat{X}(w)| = \begin{cases} |\bar{Y}(w)| - \alpha(w)N(w) & si \ \ |\bar{Y}(w)| > \alpha(w)N(w) + \beta.|\bar{D}(w)| \\ \beta|\bar{Y}(w)| & sinon \end{cases} \qquad (2.16)$$

Où

β est le plancher spectral (fixé à 0.1 dans [28])

$|\bar{Y}(w)|$ et $|\bar{D}(w)|$ sont les estimations lissées de la parole bruitée et du bruit, respectivement.

$\alpha(w)$ est un facteur de soustraction dépendant de la fréquence.

$N(w)$ est une fonction non linéaire du spectre de bruit.

Les estimations lissées de la parole bruitée (notées$|\bar{Y}(w)|$) et le bruit (noté comme$|\bar{D}(w)|$) s'obtiennent comme suit :

$$|\bar{Y}_i(w)|=\mu_y|\bar{Y}_{i-1}(w)|+(1-\mu_y)|Y_i(w)|$$
$$|\bar{D}_i(w)|=\mu_d|\bar{D}_{i-1}(w)|+(1-\mu_d)|\widehat{D}_i(w)| \quad (2.17)$$

Ou

$|\bar{Y}_i(w)|$ est le spectre d'amplitude de la parole bruitée obtenue dans la i^{ème} trame.

$\left|\widehat{D}_i(w)\right|$ est l'estimation du spectre d'amplitude du bruit pour la i^{ème} trame.

Les constantes μ_y, μ_d prennent des valeurs comprises entre $0.1 \leq \mu_y \geq 0.5$ et $0.5 \leq \mu_d \geq 0.9$

Le filtre de suppression est donné par :

$$H_{NSS}(w) = \frac{|\bar{Y}(w)|-\alpha(w)N(w)}{|\bar{Y}(w)|} \quad (2.18)$$

$\alpha(w)$ est le facteur de soustraction dépendant de la fréquence et du RSB a posteriori :

$$\alpha(w) = \frac{1}{1+\gamma p(w)} \quad (2.19)$$

γ est un scalaire et $p(w)$ la racine carrée du RSB a posteriori :

$$p(w) = \frac{|\bar{Y}(w)|}{|\bar{D}(w)|} \quad (2.20)$$

$N(w)$ est obtenue à partir de la valeur maximale du spectre d'amplitude du bruit $\left|\widehat{D}_i(w)\right|$ sur les 40 trames précédentes :

$$N(w) = \max_{i-40\leq j\leq i}\left(\left|\widehat{D}_j(w)\right|\right) \quad (2.21)$$

2.4. Technique de Soustraction Spectrale basée sur le Moyennage Adaptative du Gain (AGA)

Nous savons que les deux facteurs responsables de la présence du bruit musical après soustraction spectrale sont la grande variance de l'estimation spectrale et la variabilité de la fonction de gain. Pour cela, Gustafsson et al [29] ont proposé une méthode de Soustraction Spectrale dite AGA (Adaptative Gain Averaging) qui consiste à diviser la trame d'analyse

courante en sous trames en vue d'obtenir un spectre à faible résolution. Le spectre d'amplitude de chaque sous trame est moyenné pour obtenir un spectre à faible variance suivi d'un lissage temporel de la fonction de suppression par le biais d'un moyennage exponentiel adaptatif.

La figure 2.3 montre le schéma fonctionnel de l'algorithme de soustraction spectrale proposé [29]. Le signal d'entrée est divisé en trames d'échantillons L et subdivisé en sous-trames constituées chacune de $M(M < L)$ échantillons. Les spectres calculés en chaque sous-trame sont moyennés pour produire une estimation de spectre d'amplitude à faible variance, noté $\left|\bar{Y}_i^{(M)}(w)\right|$, où l'exposant indique le nombre de spectres composants (c'est-à-dire la taille de la FFT), et l'indice i indique le numéro de la trame. De $\bar{Y}_i^{(M)}(w)$, une fonction de gain est donnée par :

$$G_i^{(M)}(w) = 1 - k\frac{\left|\hat{D}_i^{(M)}(w)\right|}{\left|\bar{Y}_i^{(M)}(w)\right|} \tag{2.22}$$

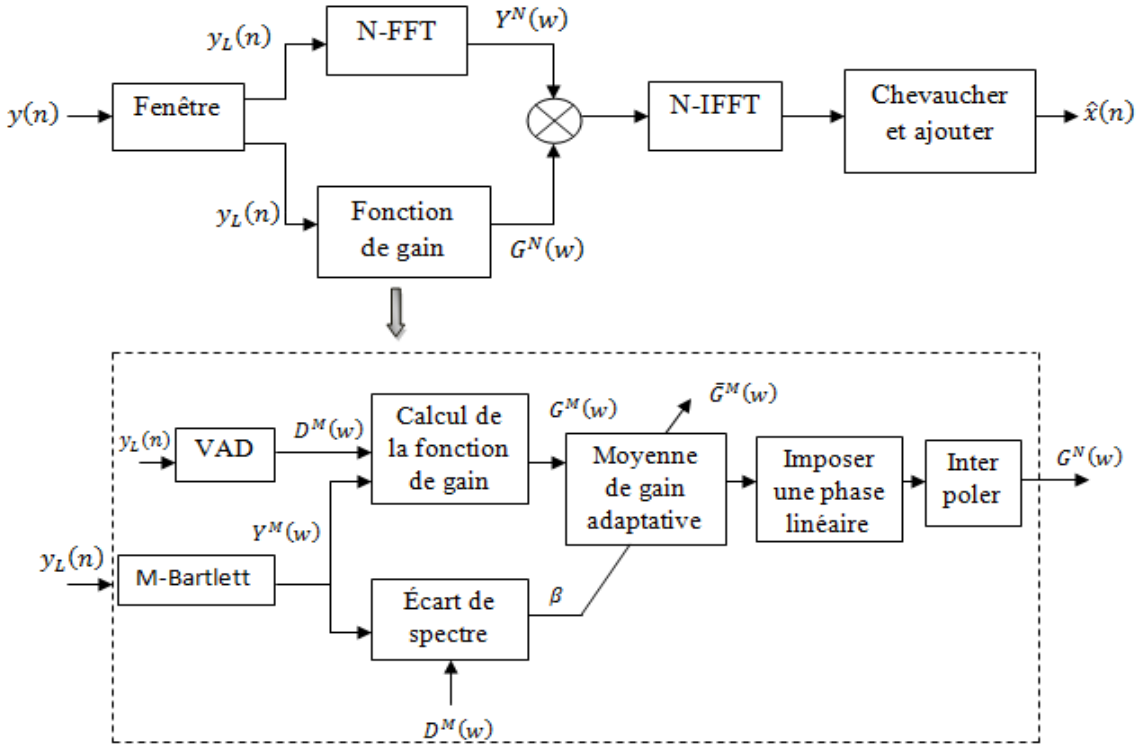


Figure 2.3 Schéma fonctionnel de l'algorithme de soustraction spectrale avec gain adaptatif moyen.

Où

k est le facteur de soustraction (défini sur $k = 0.7$ dans [29]), et $\left|\widehat{D}_i^{(M)}(w)\right|$ est la valeur estimée du spectre d'amplitude du bruit, mis à jour durant les segments d'absence d'activité vocale. Pour réduire la variabilité de la fonction de gain, $G_i^{(M)}(w)$ est moyennée selon la relation suivante :

$$\bar{G}_i^{(M)}(w) = \alpha_i \bar{G}_{i-1}^{(M)}(w) + (1-\alpha_i) G_i^{(M)}(w) \tag{2.23}$$

Où $\bar{G}_i^{(M)}(w)$ désigne la fonction de gain lissée dans la trame i, et α_i le paramètre de lissage adaptatif. Le paramètre de moyennage adaptatif α_i est dérivé d'une mesure de discordance β_i définie comme suit :

$$\beta_i = min\left\{\frac{\sum_{w_k=0}^{M-1}\left|\left|\bar{Y}_i^{(M)}(w_k)\right|-\left|\widehat{D}_i^{(M)}(w_k)\right|\right|}{\sum_{w_k=0}^{M-1}\left|\widehat{D}_i^{(M)}(w_k)\right|}\right\} \tag{2.24}$$

La mesure de discordance précédente évalue de manière grossière le changement spectral relatif au bruit de fond. Une petite valeur d'écart suggérerait une valeur relativement stationnaire, situation de bruit de fond, alors qu'une valeur d'écart importante suggérerait des situations dans lesquelles la parole ou un bruit de fond très variable est présent. Le paramètre de moyennage adaptatif α_i est calculé à partir de β_i comme suit :

$$\alpha_i = \begin{cases} \gamma\alpha_{i-1} + (1-\gamma)(1-\beta_i) & if\ \ \alpha_{i-1} < 1-\beta_i \\ 1-\beta_i & sinon \end{cases} \tag{2.25}$$

Où γ est un paramètre de lissage ($\gamma = 0.8$ dans [29]). Le paramètre adaptatif α_i est permis de diminuer rapidement, permettant à la fonction de gain de s'adapter rapidement au nouveau signal d'entrée, mais ne peut augmenter que lentement.

2.5. Conclusion

Dans ce chapitre nous avons étudié la soustraction spectrale qui est historiquement l'un des premiers algorithmes proposés pour l'amélioration de la parole bruitée. Dans ce procédé, le spectre de bruit est estimé pendant les pauses de la parole, et il est soustrait du spectre de la parole bruitée pour estimer la parole propre. Ceci est également réalisé en multipliant le spectre de la parole bruitée avec une fonction de gain et en le combinant plus tard avec la phase de la parole bruitée. L'inconvénient de cette méthode est la présence de distorsions de traitement, appelées bruit résiduel. Des variantes de la méthode ont été développées au cours de ce chapitre pour remédier à cet inconvénient. Ces variantes forment une famille d'algorithmes de type soustractif spectral. Le but de ce chapitre est de fournir une étude comparative, par un développement mathématique des différentes formes d'algorithmes de type soustraction à savoir ; 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).

Chapitre 3

Résultats et discussions

3.1. Introduction

L'évaluation de la qualité de la parole est l'un des problèmes les plus importants dans le domaine de communication, pour évaluer les différents systèmes de communication.

Ce chapitre est consacré à l'évaluation des performances des méthodes de rehaussement de la parole étudiées dans notre travail à savoir ; SS, SS-NSS et SS-AGA en termes de l'intangibilité et la perception de la parole rehaussée, ainsi sa qualité après le débruitage.

Dans ce chapitre, plusieurs mesures de qualité subjective et objective, et des RSB (Rapport Signal sur Bruit) qui déterminent la qualité de la parole rehaussée seront présentées et calculées pour chaque méthode par des données du monde bruité simulé (environnement bruité). Ainsi des discussions et des comparaisons des résultats de simulation seront accomplies pour analyser les points forts et faibles de chaque méthode.

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 [30] 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 [31] 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 [32]. 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 [33].

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) [7],
- Evaluation Perceptive de la Qualité de la Parole (PESQ : Perceptual Evaluation of Speech Quality) [8],
- Pente Spectrale Pondérée (WSS : Weighted Slop Spectral) [9],
- Logarithme du Rapport de Vraisemblance (LLR : Log Likelihood Ratio) [10].

3.3. Résultats et discussions

À partir de la base de données NOIZEUS, nous avons extrait 15 phrases mixtes avec des bruits d'aéroport (airport noise), de chahut (babble noise), de véhicule (car noise), de salle d'exposition (exhibition_hall noise), de restaurant (restaurant noise), de rue (street noise), de gare de train (train noise) et bruit blanc gaussien (white noise) 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 SS, SS-NSS et SS-AGA étudié dans notre travail a été effectuée par les mesures objectives et subjectives et SNR de signal rehaussée 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 SS donne de meilleurs résultats en terme BAK, segSNR et SNR de sortie par rapport aux deux autres méthodes SS-NSS et SS-AGA.
- Aussi nous remarquons, que la méthode de rehaussement de la parole SS-NSS donne de meilleurs résultats en terme SIG, OVL et WSS par rapport aux deux autres méthodes SS et SS-AGA a l'exception des cas suivantes :
    - Pour le bruit d'aéroport à 10 dB, la valeur du SIG de la méthode SS-AGA est plus grande,

- Pour les bruits salle d'exposition et train à 10 dB, les valeurs du SIG et OVL de la méthode SS-AGA sont meilleurs,
- Pour le bruit restaurant à 5 dB, les valeurs du SIG et OVL de la méthode SS-AGA sont plus grandes.
- Pour le bruit blanc, les valeurs du WSS de la méthode SS-AGA sont meilleures.

Parmi les huit bruits utilisés dans notre travail, nous choisissons deux bruits, le bruit le plus couramment utilisé dans notre domaine d'étude (télécommunication) et le bruit le plus utilisé dans notre vie en générale : le bruit blanc gaussienne et le bavardage des locuteurs (bruit Babble),

À partir du tableau 2, pour le bruit Babble

- En terme de PESQ, nous remarquons que la méthode SS-NSS présente les meilleures valeurs par rapport aux autres méthodes SS et SS-AGA, sauf le cas du SNR d'entrée 10 dB, la valeur du PESQ de la méthode SS-AGA est grande,
- En terme LLR, nous remarquons que la méthode SS donne de meilleures valeurs par rapport aux autres méthodes SS-NSS et SS-AGA, sauf le cas du SNR d'entrée 0 dB, la méthode SS-NSS présente de meilleures valeurs.

À partir du tableau 5, pour le bruit blanc

- En terme PESQ, nous remarquons que la méthode SS donne des meilleures valeurs par rapport aux autres,
- En terme LLR, nous remarquons que la méthode SS-NSS donne de meilleures résultats par rapport aux autres méthodes SS et SS-AGA.

D'après les résultats de ces quatre tableaux, nous constatons que la méthode d'amélioration du signal de parole SS est la meilleure en terme de suppression du bruit par rapport aux autres méthodes, en raison que la méthode SS utilise un facteur de soustraction fixe pour traiter chaque trame du signal bruitée : quel que soit le niveau de bruit dans chaque trame du signal bruitée (trame bruitée beaucoup ou peu), la méthode SS soustrait de chaque spectre du trame bruitée la même spectre de bruit estimé. Dans le cas du bruit blanc où nous remarquons que la méthode SS donne les meilleurs résultats du SNR de sortie par rapport aux autres cas des

bruits en raison de la forme du spectre de ce bruit (fixe pour toutes les fréquences). Tandis que la méthode d'amélioration SS-NSS est la meilleure en terme de l'intelligibilité du parole rehaussée par rapport aux autres méthodes, par ce que c'est une méthode qui utilise un facteur de soustraction variable pour traiter les trames du signal bruitée : la méthode SS-NSS soustraire de chaque spectre d'amplitude du trame bruitée un surestimation du bruit selon le SNR du trame, la quantité de soustraction est plus petite pour les trames qui ont SNR élevées et augmente pour les trames qui ont SNR faible, donc les composantes spectrales de la parole pas perdent d'une manière qui rend la parole peu claire lorsqu'elle est rehaussée par cette méthode. Par contre la méthode d'amélioration SS-AGA donne des résultats mauvais, n'encouragent pas à son utilisation dans les cas des bruits qui nous avons étudiées.

À partir de la figure 1 :

Nous remarquons, que la méthode de rehaussement de la parole SS a récupéré les niveaux d'énergie de la parole rehaussé dans les hautes et basses fréquences, par contre les méthodes SS-AGA et SS-NSS ont récupéré les niveaux d'énergie de la parole rehaussée presque dans les basses fréquences seulement, avec un bruit restant dans le signal parole rehaussée par ces deux méthodes.

À partir de 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 SS donne une meilleure amélioration du SNR.

Tableau 3.2 Évaluation objective, subjective et SNR de sortie des méthodes SS, SS-NSS et SS-AGA 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	SS	SS - NSS	SS - AGA
SIG [1 à 5]	**0**	1.6226	**1.9957**	1.5302	1.4454	**1.9060**	1.4903
	5	2.3364	**2.4501**	2.3251	2.1825	**2.4685**	2.2183
	10	2.9770	2.9754	**3.0151**	2.8910	**2.9387**	2.7752
BAK [1 à 5]	**0**	**1.4162**	1.3001	0.8852	**1.3465**	1.2567	0.8413
	5	**1.9060**	1.5606	1.3607	**1.8337**	1.5591	1.3687
	10	**2.4102**	1.8456	1.8404	**2.3662**	1.8336	1.7757
OVL [1 à 5]	**0**	1.3644	**1.7149**	1.1604	1.2383	**1.6395**	1.1646
	5	1.9708	**2.1478**	1.9553	1.8618	**2.1575**	1.8394
	10	2.5359	**2.6509**	2.6461	2.4761	**2.6132**	2.4083
PESQ [0.5 à 4.5]	**0**	1.6023	**1.7791**	1.6527	1.5347	**1.7297**	1.6840
	5	1.9823	2.1073	**2.1489**	1.9399	**2.1105**	2.1096
	10	2.3629	2.5034	**2.5351**	2.3391	2.4712	**2.5249**
segSNR	**0**	**-2.0823**	-8.6012	-6.6530	**-2.5120**	-8.6567	-7.1987
	5	**0.1923**	-8.7635	-7.5229	**-0.1471**	-8.7650	-7.3210
	10	**2.9204**	-9.0827	-7.7209	**2.6206**	-8.8876	-7.8517
WSS	**0**	121.7839	**91.7764**	106.8134	123.2636	**94.0946**	109.7937
	5	98.2352	**75.5172**	88.9444	102.6180	**75.9355**	92.7764
	10	76.7513	**58.9810**	73.5976	78.7201	**60.2389**	74.4015
LLR	**0**	**1.3027**	1.3063	1.3787	1.4224	**1.3442**	1.4308
	5	**1.0377**	1.1991	1.2507	**1.1241**	1.1795	1.2612
	10	**0.8261**	1.0654	1.1110	**0.8786**	1.0712	1.1477
SNR sortie	**0**	**3.7272**	-9.1668	-6.9282	**3.3614**	-8.9580	-7.3273
	5	**7.3216**	-9.0962	-8.0025	**7.1899**	-9.0615	-7.7269
	10	**10.6976**	-9.4068	-8.4524	**10.8377**	-9.1076	-8.3302

Tableau 3.3 Évaluation objective, subjective et SNR de sortie des méthodes SS, SS-NSS et SS-AGA 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 Objective subjective	Entrée SNR (dB)	Bruit véhicule			Bruit salle d'exposition		
		SS	SS - NSS	SS - AGA	SS	SS - NSS	SS - AGA
SIG [1 à 5]	**0**	1.4081	**2.1718**	2.0256	1.1994	**2.0454**	1.6192
	5	2.2437	**2.5905**	2.4690	2.0180	**2.5668**	2.4463
	10	2.8910	**2.9387**	2.7752	2.6114	2.9961	**3.0618**
BAK [1 à 5]	**0**	**1.4903**	1.3493	1.2836	**1.3021**	1.2779	1.1004
	5	**1.9813**	1.6052	1.6465	**1.8499**	1.5827	1.5442
	10	**2.3662**	1.8336	1.7757	**2.3108**	1.8331	1.9605
OVL [1 à 5]	**0**	1.2716	**1.8083**	1.5523	1.0629	**1.6794**	1.2210
	5	1.9467	**2.2193**	2.0639	1.7633	**2.1941**	2.0352

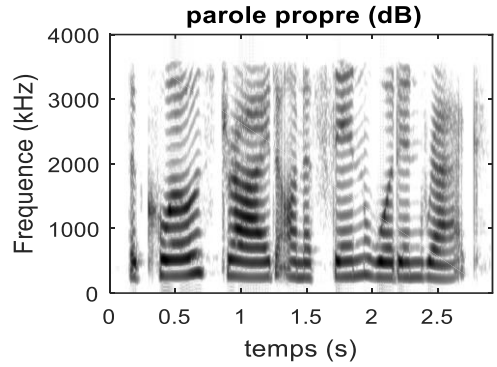
	10	2.4761	**2.6132**	2.4083		2.2960	2.6246	**2.7363**
PESQ	**0**	1.6293	1.7597	**1.7648**		1.4648	**1.6546**	1.6397
[0.5 à 4.5]	**5**	2.0290	2.0895	**2.1847**		1.9223	2.0845	**2.1592**
	10	2.3391	2.4712	**2.5249**		2.2996	2.4435	**2.5495**
segSNR	**0**	**-1.1260**	-8.3561	-4.8554		**-1.9046**	-8.1212	-5.7689
	5	**1.1041**	-8.3890	-5.9824		**0.5858**	-8.2265	-7.1353
	10	**2.6206**	-8.8876	-7.8517		**2.9678**	-8.5470	-7.4266
WSS	**0**	121.6559	**85.6348**	88.3062		130.2949	**90.7637**	95.7393
	5	98.8676	**71.2945**	75.7447		105.6971	**75.6353**	87.2965
	10	78.7201	**60.2389**	74.4015		87.0512	**61.4946**	73.7427
LLR	**0**	1.5282	**1.1774**	1.1845		1.5590	**1.1938**	1.2530
	5	1.1496	**1.0893**	1.1378		1.2467	**1.0713**	1.1569
	10	**0.8786**	1.0712	1.1477		1.0542	**0.9882**	1.0993
SNR sortie	**0**	**5.4893**	-7.7741	-4.8439		**3.5877**	-7.7252	-6.0551
	5	**8.7768**	-8.2569	-7.0553		**7.5498**	-8.3000	-7.9946
	10	**10.8377**	-9.1076	-8.3302		**10.8747**	-8.9168	-8.5844

Tableau 3.4 Évaluation objective, subjective et SNR de sortie des méthodes SS, SS-NSS et SS-AGA, 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.

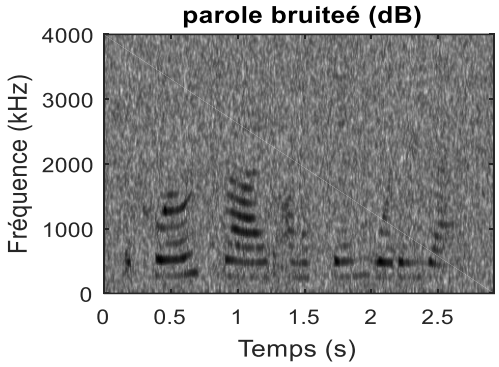
Mesures Objective Subjective	**Entrée SNR (dB)**	**Bruit restaurant**				**Bruit rue**		
		SS	**SS - NSS**	**SS - AGA**		**SS**	**SS - NSS**	**SS - AGA**
SIG	**0**	1.6732	**2.0715**	2.0414		1.3273	**2.1320**	2.0114
[1 à 5]	**5**	2.3620	2.5050	**2.6182**		2.0594	**2.6370**	2.4374
	10	2.9474	**3.0094**	2.9940		2.7495	**3.0676**	2.9289
BAK	**0**	**1.4264**	1.2954	1.1926		**1.4242**	1.3499	1.2663
[1 à 5]	**5**	**1.8857**	1.5334	1.5182		**1.9255**	1.5898	1.5892
	10	**2.3827**	1.8457	1.8481		**2.4006**	1.8719	1.8695
OVL	**0**	1.3828	**1.7486**	1.6303		1.2010	**1.7715**	1.5485
[1 à 5]	**5**	1.9718	2.1604	**2.1893**		1.8165	**2.2250**	2.0056
	10	2.5038	**2.6583**	2.5894		2.4122	**2.6920**	2.6500
PESQ	**0**	1.5555	**1.7612**	1.7084		1.5678	1.7166	**1.7246**
[0.5 à 4.5]	**5**	1.9533	2.0800	**2.1180**		1.9421	2.0471	**2.0765**
	10	2.3227	2.4878	**2.5361**		2.3622	2.4882	**2.5121**
segSNR	**0**	**-2.3075**	-8.7594	-7.4768		**-1.7389**	-8.2322	-6.1378
	5	**-0.0367**	-8.9476	-7.9728		**0.6373**	-8.4872	-7.2021
	10	**2.6608**	-8.8891	-7.8432		**3.2203**	-8.6778	-7.2922
WSS	**0**	115.1062	**89.7996**	104.3053		121.3805	**83.7245**	93.0055
	5	97.0930	**75.8726**	88.8884		96.7079	**69.7255**	78.4552
	10	75.5907	**59.6269**	71.8888		80.7809	**57.8295**	70.4546
LLR	**0**	1.2845	**1.2393**	1.3479		1.5731	**1.2076**	1.2549
	5	**1.0059**	1.1267	1.1695		1.2968	**1.0330**	1.0885
	10	**0.8414**	1.0176	1.0979		1.0115	**0.9770**	1.0757
SNR sortie	**0**	**2.6117**	-9.8208	-8.5657		**3.8138**	-8.4406	-6.9270
	5	**6.6136**	-9.5604	-8.4924		**7.8416**	-8.6460	-7.8480
	10	**10.2792**	-9.0707	-8.4786		**10.9168**	-9.1236	-8.2857

Tableau 3.5 Évaluation objective, subjective et SNR de sortie des méthodes SS, SS-NSS et SS-AGA, 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.

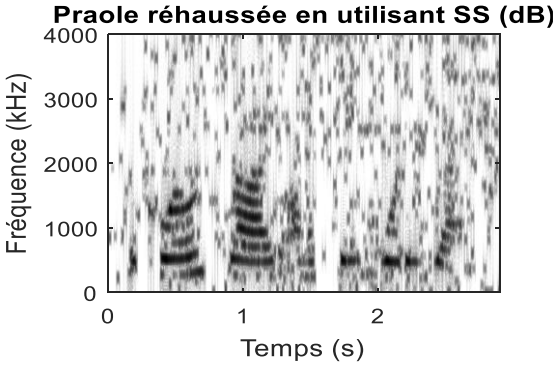
Mesures Objective subjective	**Entrée SNR (dB)**	**Bruit train**				**Bruit blanc**		
		SS	**SS - NSS**	**SS - AGA**		**SS**	**SS - NSS**	**SS -AGA**
SIG [1 à 5]	**0**	1.2575	**2.2869**	1.9857		0.8909	**1.6199**	1.4120
	5	1.9152	**2.7065**	2.6462		1.6360	**2.1365**	1.9035
	10	2.5747	3.1208	**3.2890**		2.2772	**2.6015**	2.4680
BAK [1 à 5]	**0**	**1.4663**	1.3588	1.2171		**1.4825**	1.1091	1.0895
	5	**1.9330**	1.6178	1.6783		**1.9571**	1.4179	1.4038
	10	**2.3870**	1.8642	1.9113		**2.3978**	1.6852	1.7145
OVL [1 à 5]	**0**	1.1340	**1.8626**	1.5248		1.0496	**1.4279**	1.1541
	5	1.7155	**2.2698**	2.2500		1.6641	**1.9070**	1.5662
	10	2.2901	2.6969	**2.7983**		2.2141	**2.3460**	2.1718
PESQ [0.5 à 4.5]	**0**	1.4379	**1.7346**	1.6387		**1.7856**	1.6943	1.6941
	5	1.8476	2.0613	**2.1454**		**2.1521**	2.0574	1.9334
	10	2.2611	2.4354	**2.4906**		**2.5151**	2.3993	2.3714
segSNR	**0**	**-1.5545**	-8.4417	-6.6017		**-0.5155**	-8.4696	-4.0587
	5	**0.6620**	-8.2873	-7.2082		**1.6482**	-8.0233	-4.4182
	10	**3.0729**	-8.6205	-7.4021		**3.7903**	-7.9200	-5.8962
WSS	**0**	108.1498	**81.7864**	90.6276		138.9282	114.4545	**107.2608**
	5	89.4083	**68.4857**	75.2821		115.6370	99.1495	**94.7448**
	10	74.4881	**55.8303**	61.7565		96.7401	85.2416	**83.3017**
LLR	**0**	1.6805	**1.0845**	1.1667		1.9713	**1.4234**	1.4321
	5	1.4453	**0.9845**	1.0865		1.6657	**1.2680**	1.3070
	10	1.1772	**0.9118**	1.0072		1.4205	**1.1381**	1.2082
SNR sortie	**0**	**4.0329**	-8.8489	-7.2255		**7.2255**	-8.1540	-3.2195
	5	**7.8836**	-8.3137	-7.7792		**10.0530**	-7.2727	-5.2123
	10	**11.1626**	-8.9482	-8.3193		**12.2800**	-7.5339	-7.4644



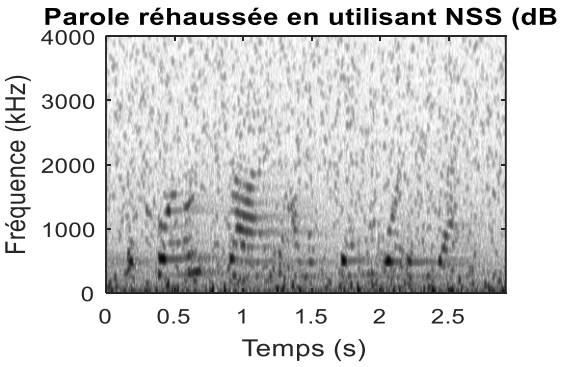
(a)



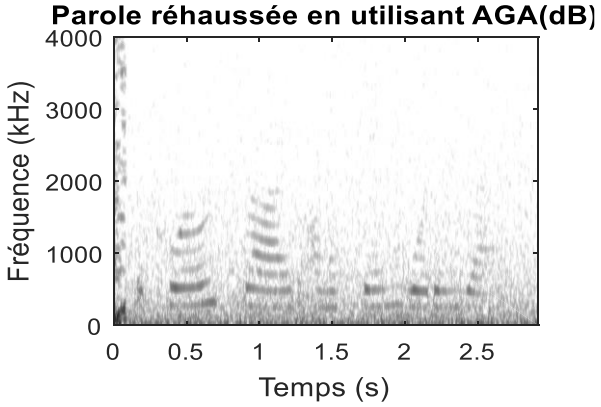
(b)



(c)

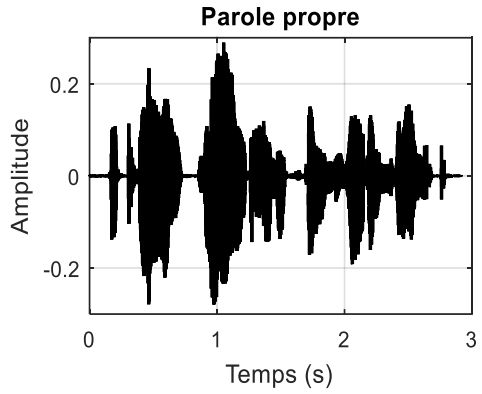


(d)

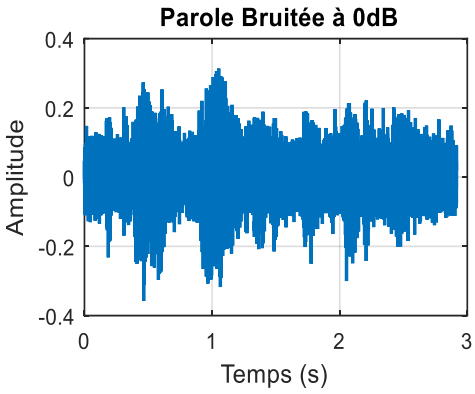


(e)

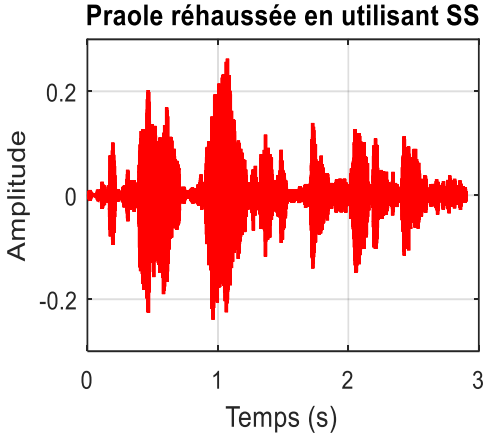
Figure 3.1 Spectrogrammes : (a) parole propre, (b) parole bruitée, (c) parole rehaussée en utilise la méthode SS, (d) parole rehaussée en utilise la méthode SS-NSS, (e) parole rehaussée en utilise la méthode SS-AGA. Cas du bruit blanc 0 dB.



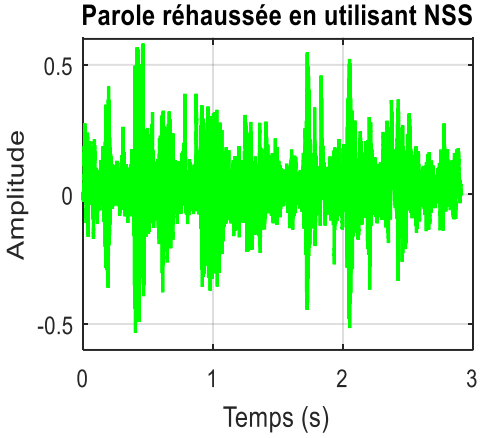
(a)



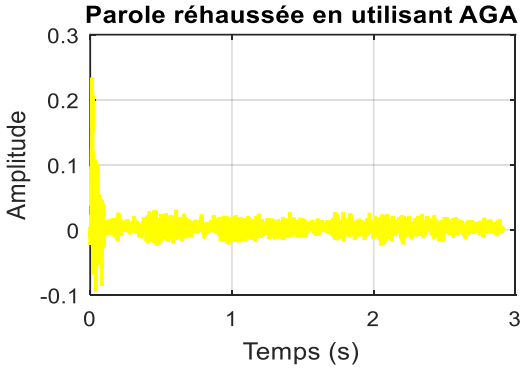
(b)



(c)



(d)



(e)

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

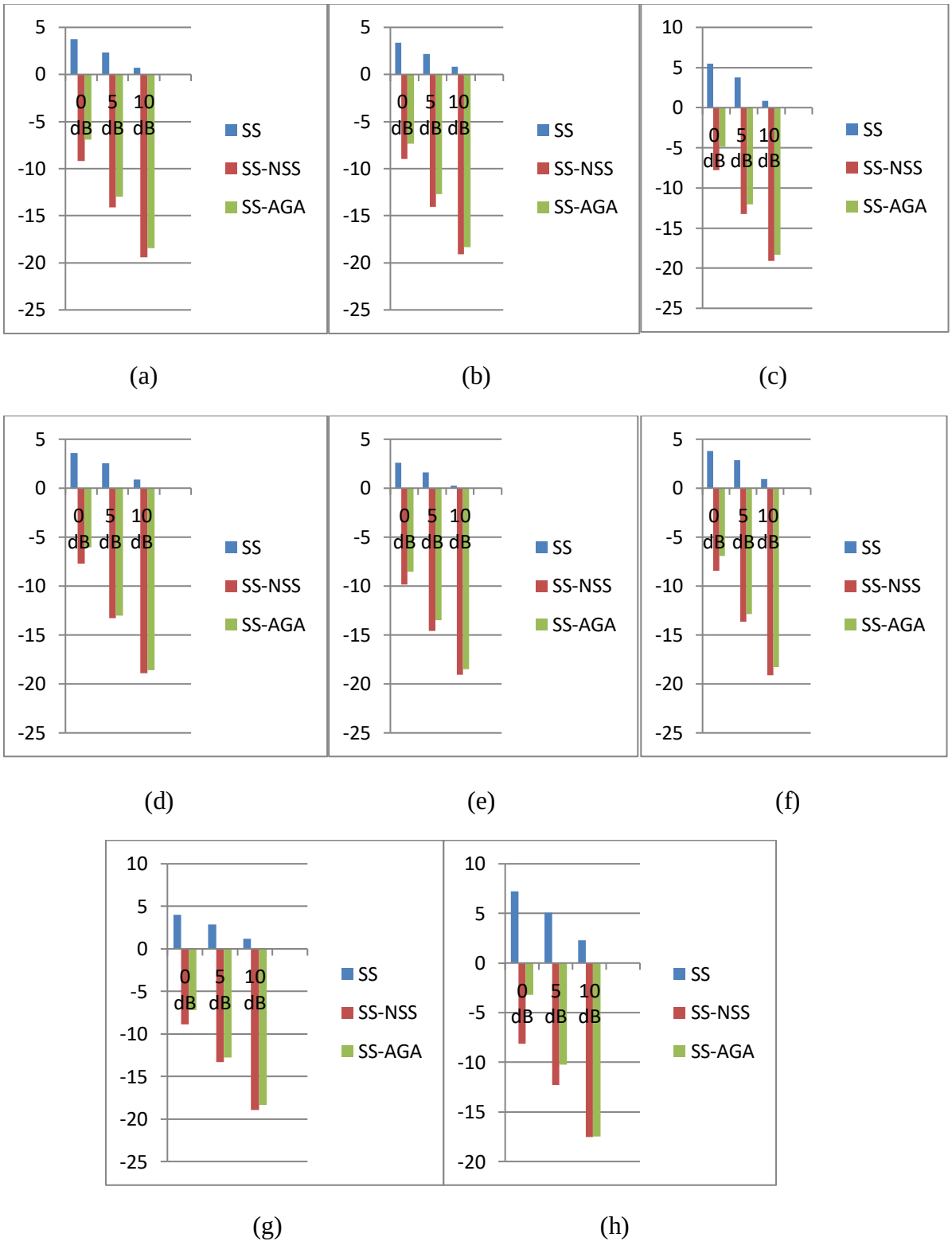


Figure 3.3 Évaluation SNR amélioré pour les méthodes SS, SS-NSS et SS-AGA en utilisant : (a) bruit d'aéroport de NOIZEUS, (b) bruit babble de NOIZEUS, (c) bruit de véhicule de NOIZEUS, (d) bruit salle d'exposition de NOIZEUS, (e) bruit restaurant de NOIZEUS, (f) bruit rue de NOIZEUS, (g) bruit train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

3.4. Conclusion

Dans ce chapitre nous avons testé et évalué trois méthodes différentes de rehaussement de la parole basées sur la soustraction spectrale : SS, SS-NSS et SS-AGA.

Une étude comparative a été réalisée et à montrer la supériorité de la méthode de rehaussement SS en terme de suppression de bruit et amélioration du SNR, et la supériorité de la méthode de rehaussement SS-NSS en terme de l'intelligibilité de parole rehaussé.

Conclusion générale

Des algorithmes capables de réduire le bruit dans le signal de parole, en utilisant la méthode de soustraction spectrale ont été réalisés et testés avec succès dans ce travail. Trois méthodes ont été mises en œuvre à savoir; 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), et leurs performances ont été examinées, pour décider laquelle est la meilleure. Comme mentionné précédemment dans le chapitre 3, toujours de notre point de vue, la mesure de la qualité objective est la plus importante parmi les critères d'évaluation. En conclusion, pour nous, les meilleurs résultats sont obtenus avec l'algorithme de soustraction spectrale non linéaire (SS-NSS). L'algorithme de soustraction spectrale non linéaire a obtenu une valeur d'environ (1.4279) dans l'échelle MOS (pour le cas du bruit blanc à 0 dB), dont la note la plus élevée est 5. Cette valeur est de (0.3583) points au-dessus de la pire méthode (méthode de soustraction spectrale de base (SS)).

Nous pouvons dire que les objectifs initialement proposés ont été sans aucun doute atteints, car l'aspiration principale était d'étudier en profondeur comment fonctionne la soustraction spectrale, et comment trouver une mesure pour l'évaluer. En outre, une étude et une recherche sur la façon dont elle fonctionnerait a été présentée et expliquée.

Bien que les objectifs aient été atteints, ce projet pourrait être poursuivi. La réalisation d'un détecteur d'activité vocale (VAD) pourrait être étudié en profondeur afin d'améliorer les résultats des algorithmes étudiés dans ce travail. Un nouvel objectif pourrait être imposé afin d'obtenir un VAD avec la capacité à ne pas confondre la parole avec le bruit dans tous les cas. Le VAD est une partie du processus, car il n'est pas capable de bien distinguer la voix du bruit à cent pour cent (100%). Comme un résultat, le processus ultérieur n'est pas effectué avec les valeurs optimales, car il dépend de cette partie du processus (VAD) pour estimer le bruit et reconnaître si la voix existe ou pas.

D'un point de vue personnel, la réalisation de ce travail a permis pour nous d'affranchir un nouveau domaine de travail, avec des circonstances et des exigences diverses, pour trouver quelles décisions clés doivent être prises en compte et quelles sont les principales étapes à suivre dans chaque situation. De plus, cette expérience nous a permis à nous nous organiser

ainsi organiser notre emploi du temps, car pour combiner la recherche avec le développement du projet, nous devions faire un effort.

De plus, l'intention d'écrire le projet entièrement en français a été une tâche difficile pour nous, même si nous utilisons cette langue depuis de nombreuses années. Utilisant correctement, tous les mots techniques sur un grand nombre de pages ne sont pas une chose facile pour quelqu'un dont la langue maternelle est assez différente de celle utilisée.

Références Bibliographiques

Références bibliographiques

[1] Boite R, Bourland H, Dutoit T, Hancq J and Leich H. 1999. Traitement de la parole. Presse polytechnique et universitaires romandes.

[2] Boll S.F. 1979. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust Speech Signal Processing. vol 27. pp 113-120.

[3] Paliwal K.K and Basu A. 1987. A speech enhancement method based on Kalman filtering. in Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP).

[4] Eprahim Y. 1992. A bayesian estimation approach for speech enhancement using hidden Markov models. IEEE Trans. Acoust., Speech, Signal Processing, vol. 40, pp. 725-735.

[5] Eprahim Y and Van Trees H.L. 1995. A signal subspace approach for speech enhancement. IEEE Trans. Acoust Speech Signal Processing. vol 3. pp 251-266.

[6] Vasighi S.V. 2000. Advanced digital signal processing and noise reduction, John wiley.

[7] Richards D. 1965. Speech Transmission Performance Of PCM Systems". Electron Letters. P, 40-91.

[8] (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.

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

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

[11] Harris F.J. 1978. On the use of windows for harmonic analysis with the discrete fourier transform. Proceedings of the IEEE. Volume 66. numéro 1. p 51-83.

[12] Allen J.B and Rabiner L.R. nov,1977. A unifed approach to short-time Fourier analysis and synthesis. Proc IEEE. Volume 65. pp 1558-1564.

[13] Ghioum H et Lyzidi R. 2004. Masquage de l'information dans les signaux image et parole.

[14] Damien V. 2007. Analyse et contrôle du signal glottique en synthèse de la parole. L'École Nationale Supérieure des Télécommunications de Bretagne. Bretagne.

[15] Kung M. 1981. Traitement Numérique des signaux. Presse Polytechnique Romandes.

[16] Loucif R et Ayache H. 2004. Débruitage du signal de la parole. Ecole militaire polytechnique. Alger.

[17] Amehraye A. 2009. Débruitage perceptuel de la parole. Thèse de doctorat. L'Ecole Nationale Supérieure des Télécommunications de Bretagne. Bretagne.

[18] https://sites.google.com/site/thuyntranthesisunisa/introduction/the-speech-signal.

[19] https://www.f-legrand.fr/scidoc/docimg/numerique/filtre/autocorrel/autocorrel.html.

[20] http://voyard.free.fr/textes.audio/le_microphone.htm#dynamique.

[21] Bedouhene M et Tabani K. 2008. Les methodes de debruitage du signal de parole. Thèse de master. Universite Mouloud mammeri de Tizi-Ouzou. Tizi-Ouzou.

[22] https://www.probabilitycourse.com/chapter10/10_2_4_white_noise.php.

[23] Paliwal K. and Alseris L. 2005. On the usefulness of STFT phase spectrum in human listening tests. Speech Commun. volume 45 (2). pp 153-170.

[24] Beruoti M, Schwartz R and Makhoul J. 1997. Enhancement of speech corrupted by acoustic noise. in Proc. International Conference on Acoustics. Speech and Signal Processing (ICASSP).

[25] Loizou P.C. 2013. Speech Enhancement Theory and Practice. Taylor & Francis Group. LLC.

[26] Virage N. 1999. Signal channel speech enhancement based on masking properties of the humain auditory system. IEEE Trans. Speech and Audio Processing. volume 7 (2). pp 126-137.

[27] Vary P. 1985. Noise suppression by spectral magnitude estimation-mechanism and theoretical limits. Signal Processing. volume 8. pp 387-400.

[28] Lockwood P and boudy J. 1992. Experiments With A Non-Linear Speech Subtractor (NSS). Hidden Markov Models And Projections For Robust Recognition In Cars. Speech Comm. volume 11 (2-3). pp 215-228.

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

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

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

[32] Fletcher H and Steinberg J.C. 1929. Articulation Testing Methods. Bell Syst. Tech J. volume 8. pp 806-854.

[33] 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-l!!PDF-E&type=items