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 : Télécommunications

Spécialité : Systèmes des Télécommunications

Par

- **Alsalkini Rama**
- **Belkadi Hanane**

Intitulé

Suppression du bruit dans les signaux parole avec les méthodes basées sur des modèles statistiques

Evalué le : ……

Par la commission d'évaluation composée de* :

Nom & Prénom	***Grade***	***Qualité***	***Etablissement***
M.	***MCB***	***Président***	***Univ-BBA***
M. Asbai Nassim	***MCA***	***Encadreur***	***Univ-BBA***
M.	***….***	***Examinateur***	***Univ-BBA***

Année Universitaire 2020/2021

Sommaire

Remerciement

Nous remercions Dieu Tout-Puissant de nous avoir donné le courage de développer cet humble travail, et ce travail a été accompli avec l'aide de nombreuses personnes qui nous sont chères.

Tout d'abord, nous tenons à remercier notre encadreur « Dr. A.S.BAJ » pour nous avoir apporté ses précieux conseils et son soutien lors de la préparation de ce mémoire.

Nous remercions également les membres du jury de leur avoir rendu hommage en acceptant d'examiner et de juger notre travail. Toutes les personnes qui nous ont soutenus jusqu'au bout et qui n'ont jamais cessé de nous donner des conseils, nous vous remercions beaucoup.

Finalement, nous remercions toutes les personnes qui ont participé de près ou de loin à l'élaboration de cet ouvrage avec des conseils ou non.

Dédicace

Tout d'abord je me félicite pour tous les efforts que j'ai fournis pendant 19 ans pour en arriver là où j'en suis aujourd'hui, ça n'a pas été facile mais je l'ai fait.

Je dédie cet ouvrage, A mes très chers parents, source de vie. Mon cher père, la lampe qui ne s'éteint jamais et qui a fait des efforts au fil des années pour que je gravisse les échelons du succès. Ma chère mère, qui m'a donné le paradis sous ses pieds, qui m'a rempli d'amour et de tendresse et m'a fait me sentir heureux et en sécurité, et qui, chaque fois que je me sentais désespéré, m'a donné la volonté.

A mes chers frères Mebarek et Amine et mes sœurs Sara et Nesrine, source de joie et de bonheur.

A tout ma famille, source d'espoir et de motivation, A mes proches, et ceux que me donnent de l'amour et de la vivacité. Aux enfants de la maison, Mohammed et Syed.

A tous mes amis, surtout mon cher ami Rama avant que tu ne sois mon partenaire.

A vous chers lecteur.

À tous ceux qui m'ont encouragé et soutenu même avec le mot. Surtout Zakaria, qui m'a soutenu tout au long de mon parcours universitaire.

Hanane Belkadi

إ هـد ا ء

أخيرا انتهت الحكاية رفعت قبعتي مودعة السنين التي مضت راجية من الله أن يفتح لي طرقا مكللة بالنجاح

الحمد والشكر لله الذي أنار لي درب العلم والمعرفة، الحمد لله الذي بتوفيقه وتسهيل منه جل في علاه أكملت مسيرتي العلمية وأعانني على انهاء رسالة الماجستير أهدي عملي هذا

إلى وطني الجريح الى كل محافظة من محافظات سوريا وإلى كل حي من أحياءك يا حمص...

إلى من تشققت يداه في سبيل رعايتي ... من أدين له بحياتي ... من أكن له مشاعر العرفان والامتنان ... أبي الغالي فراس السلقيني حفظك الله ورعاك.

إلى من لم تدخر نفسا في تربيتي.. من ساندتني في صلاتها ودعاءها.. من تشاركني أفراحي وأماني، من حصدت الأشواك عن دربي لتمهد لي طريق العلم.. أمي الحبيبة وفاء الجرايحي بوركت أياديك.

الى أروع من جسّد الحب بكل معانيه وقدم لي الكثير بصور من الصبر، الحنان، الأمل، من كان لي المرمم، السند، الأمل الملاذ الأمن، وخطى معي خطواتي خطوة بخطوة نبض قلبي وقلمي م. صهيب بهلولي.

إلى تلك العينين البريئتين، ملاكي الصغير، فلذة كبدي ليان بهلولي.

إلى أعز الناس وأقربهم لقلبي أخواتي وإخواني: روى، منير، م.حمزة، زهراتي الصغيرات ريتال ورهف.

إلى الذين ظفرت بهم هدية من القدر عائلتي الثانية عمي الغالي نصر الدين بهلولي وزوجته الحبيبة وإخوتي د. هاجر، د. سارة، أيمن وزوجته، د. صفية، شيماء، منار وأخص بالذكر من كان لي عونا ورافقتني بدعائها طيلة فترة الدراسة والبحث أمي الثانية مليكة كسال ولمن ساعدتني صفية وفقك الله.

إلى من أبعدتنا عنهم المسافات والسنين وما زال في القلب ذكراهم: اقاربي جدي وجدتي حفظكم الله، عماتي وخالاتي جميعهم وأخص بالذكر من قالت لي في يوم من الأيام أن سلاح البنت هو العلم عمتي أ. مطيعة.

إلى صديقاتي الغاليات رفيقات الدرب: حنان بلقاضي، منى حسين، م.منى جحنيط، م.سمراء، م.مروة.

إلى من أبعدتهم عني الغربة ومشاغل الحياة صديقات الطفولة الجميلة: د. أية الحلواني، د. نغم الحلواني، أمان عبد الصمد، راما مندو، لبنى الإخوان، م.نورالسباعي، د. هبة السلقيني.

وأخيرا إلى نفسي التي صبرت وقدمت لتصل إلى هذه اللحظة رغم الكثير من العقبات والحمد لله.

Rama Alsalkini

Liste des tableaux

Liste des figures

Liste des abréviations

DFT Transformée de Fourier Discrète.

FFT Transformée de Fourier Rapide.

TF Transformée de Fourier.

FDP fonction de densité de probabilité.

MAP Maximum à Posteriori**.**

ML Maximum de Vraisemblance**.**

MMSE l'Erreur Quadratique Moyenne Minimale**.**

MSE l'Erreur quadratique moyenne.

STSA Amplitude Spectrale Court Terme.

PESQ Evaluation Perceptive de la Qualité de la Parole

SNR seg Rapport signal sur bruit segmental.

SNR Rapport signal sur bruit.

WSS Pente Spectrale Pondérée.

LLR Logarithme du Rapport de Vraisemblance.

VAD Algorithme de détection d'activité vocale.

Résumé

Dans ce travail, nous avons étudié trois méthodes de rehaussement de la parole (ML, MAP et MMSE) basées sur l'estimation des densités de probabilités des coefficients de la transformée de Fourier du signal bruité qui utilisent le théorème de Bayes pour optimiser le modèle séparation signal-bruit. Ces méthodes intègrent la propriété du système auditif humain pour améliorer la perception et la qualité de la parole rehaussée. D'une part, le théorème de Bayes estime le meilleur modèle probabiliste de signal propre, et réduit donc la distorsion du signal de parole. D'autre part, l'utilisation des propriétés perceptives réduit l'effet du bruit musical. Les résultats de l'évaluation des performances en utilisant diverses mesures objectives et subjectives bien connues (PESQ, SegSNR, SIG, BAK, OVRL, WSS et LLR) confirment l'efficacité du ML par rapport aux autres méthodes. Les résultats obtenus nous amènent à dire que l'incorporation du théorème de Bayes dans le rehaussement perceptif de parole effectue un bon compromis entre la réduction du bruit musical et l'ensemble de qualité du signal vocal rehaussé.

Abstract

In this work, we studied three speech enhancement methods (ML, MAP and MMSE) based on the estimation of the probability densities of the Fourier transform coefficients of the noisy signal which use Bayes theorem to optimize the model. Signal-noise separation. These methods incorporate the property of the human auditory system to improve the perception and quality of enhanced speech. On the one hand, Bayes' theorem estimates the best probabilistic eigen signal model, and therefore reduces the distortion of the speech signal. On the other hand, the use of perceptual properties reduces the effect of musical noise.

The results of performance evaluation using various well-known objective and subjective measures (PESQ, SegSNR, SIG, BAK, OVRL, WSS and LLR) confirm the effectiveness of ML compared to other methods. The results obtained lead us to say that the incorporation of Bayes' theorem in perceptual speech enhancement makes a good compromise between the reduction of musical noise and the overall quality of the enhanced speech signal.

ملخص

في هذا العمل، درسنا ثلاث طرق لتحسين الكلام (ML و MAPو MMSE) بناءً على تقدير الكثافة الاحتمالية لمعاملات تحويل Fourier لإشارة الضوضاء التي تستخدم نظرية Bayes لتحسين نموذج فصل الإشارة والضوضاء. تدمج هذه الطرق خاصية النظام السمعي للإنسان لتحسين إدراك وفهم الكلام المحسن وجودته. من ناحية أخرى ، تقدر نظرية بايز أفضل نموذج إشارة احتمالية، وبالتالي تقلل من تشويه إشارة الكلام. من ناحية أخرى ، فإن استخدام الخصائص الحسية يقلل من تأثير الضوضاء الموسيقية.

تؤكد نتائج تقييم الأداء باستخدام العديد من المقاييس الموضوعية والذاتية المعروفة (PESQ و SegSNRو SIGو BAK و OVRLو WSSو LLR) فعالية ML مقارنة بالطرق الأخرى. تقودنا النتائج التي تم الحصول عليها إلى القول إن دمج نظرية Bayes في تحسين الكلام المفهوم يجعله حل وسط جيد بين تقليل الضوضاء الموسيقية والجودة الإجمالية لإشارة الكلام المحسّنة.

Introduction générale

Introduction général

Le rehaussement de la parole pourrait améliorer la qualité de la parole noyée dans des bruits d'environnement, ce qui se traduit par un large éventail d'applications, telles que la communication vocale mobile, la reconnaissance robuste de la parole, aides pour les malentendants, etc. Par conséquent, le rehaussement de la parole a largement attiré les recherches, et un grand nombre d'algorithmes du rehaussement de la parole, par exemple, méthode de soustraction spectrale (SS) [1], méthode de débruitage par ondelettes [2], méthode des sous-espaces [3], amélioration de la parole basée sur le modèle de la perception auditive humaine [4].

Dans ce travail, nous continuons toujours, dans le même axe de recherche en étudiant trois méthodes de soustraction du bruit de la parole, à savoir ; Maximum de Vraisemblance (ML) [5], l'Erreur Quadratique Moyenne Minimale (MMSE) [5], et Maximum à Posteriori (MAP*)* [4]. Ces méthodes sont basées sur le calcul des probabilités des amplitudes spectraux dans le domaine de la transformée de Fourier discrète (DFT), c'est-à-dire que la parole améliorée est obtenue en estimant Coefficients DFT de la parole propre à partir de la parole bruitée.

Les objectifs de ce travail peuvent être résumés en quatre points importants :

1. déterminer une approche prometteuse pour améliorer l'intelligibilité de signaux de parole fortement bruités

2. utiliser les outils algorithmiques originaux nécessaires au bon fonctionnement du système d'amélioration de l'intelligibilité.

3. implémenter et simuler sur Matlab les trois méthodes suscitées, afin d'avoir des résultats expérimentaux qui permettent une comparaison scientifique fructueuse entre elles.

4. valider le bon fonctionnement de ces méthodes.

Ce travail est structuré comme suit :

Le chapitre 1 présente les bases théoriques générales du traitement du signal audio. Et les différentes techniques d'analyse ainsi que la définition et les caractéristiques du bruit sont étudiées.

Dans le chapitre 2 nous avons présenté les méthodes de rehaussement de la parole de l'état de l'art, basées sur le principe la théorie de l'estimation statistique.

Le chapitre 3 contient les résultats de simulation et discussions, Les mesures objectives utilisées dans le cadre de ce travail, sont concerné également l'évaluation du niveau de distorsion entre le signal propre et le signal rehaussé

L'évaluation de ces trois méthodes a été menée par un ensemble de mesures objectives et subjectives dont les résultats obtenus, confirmés par les spectrogrammes des signaux propres, bruités et rehaussés.

Chapitre 1 : Généralités sur le signal de parole et le bruit

1.1 Introduction

Ce chapitre présente les bases théoriques générales du traitement du signal audio [6], et les différentes techniques d'analyse qui lui sont appliquées ainsi que la définition et l'étude de différentes caractéristiques temporelles et fréquentielles des bruits d'environnements.

1.2. Signale audio numérique

Un signal audio numérique est une représentation de signaux sonores via un flux de données binaires, le signal numérique est constitué de la séquence de codage de bits affectés à chaque valeur analogique discrète. Un signal numérique est un signal discret dont l'amplitude a été quantifiée [7].

La qualité de ce processus dépend principalement de deux valeurs:

1.2.1. Profondeur de bits

Le terme profondeur de bits décrit le nombre de bits avec lesquels chaque échantillon est enregistré. Il correspond à la résolution à laquelle chaque échantillon est quantifié.

Lorsque la valeur des valeurs de profondeur de bits est plus élevée, le résultat obtenu serait plus proche de la réalité. Cette perte d'information est appelée erreur de quantification, et c'est la différence entre la valeur réelle et celle assignée [8].

1.2.2. Fréquence d'échantillonnage

La fréquence d'échantillonnage est le nombre d'échantillons par unité de temps pris à partir d'un signal continu pour produire un signal discret. Il est exprimé en Hz par unité de temps. Le théorème de Nyquist-Shannon stipule que pour qu'un signal soit correctement échantillonné, la fréquence d'échantillonnage doit être au moins supérieure à deux fois la fréquence la plus élevée à échantillonner [9].

Tableau 1 : Exemple d'utilisation de la fréquence d'échantillonnage et de la profondeur de bits.

Application	la profondeur de bits	fréquence d'échantillonnage
Téléphone	8 KHZ	8
Disque compact (CD)	44.1 KHZ	16
Audio DVD	96 KHZ	24

1.3. Les techniques d'analyse

L'analyse de la parole est une étape indispensable à toute application de synthèse, de codage, ou de reconnaissance.

1.3.1. Fonctions de la fenêtre

Les fonctions mathématiques appelées «fenêtres» sont utilisées dans l'analyse et le traitement du signal pour atténuer le problème dans lequel le signal audio n'est pas stationnaire. Le fenêtrage permet l'analyse d'étages stationnaires du signal audio, et consiste à regrouper un certain nombre d'échantillons consécutifs dans un segment, de manière à traiter récemment chaque segment individuellement.

Il existe de nombreux types de fenêtrage, par exemple: fenêtrage rectangulaire, Hanning, Hamming, Gauss, Triangulaire, etc [10].

Leur équation mathématique :

$h(t) = \{1\ si\ t\ \in [T_1, T_2]\ 0 \qquad sinon$ (1.1)

h(t) : signal.

T : duré de temps.

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

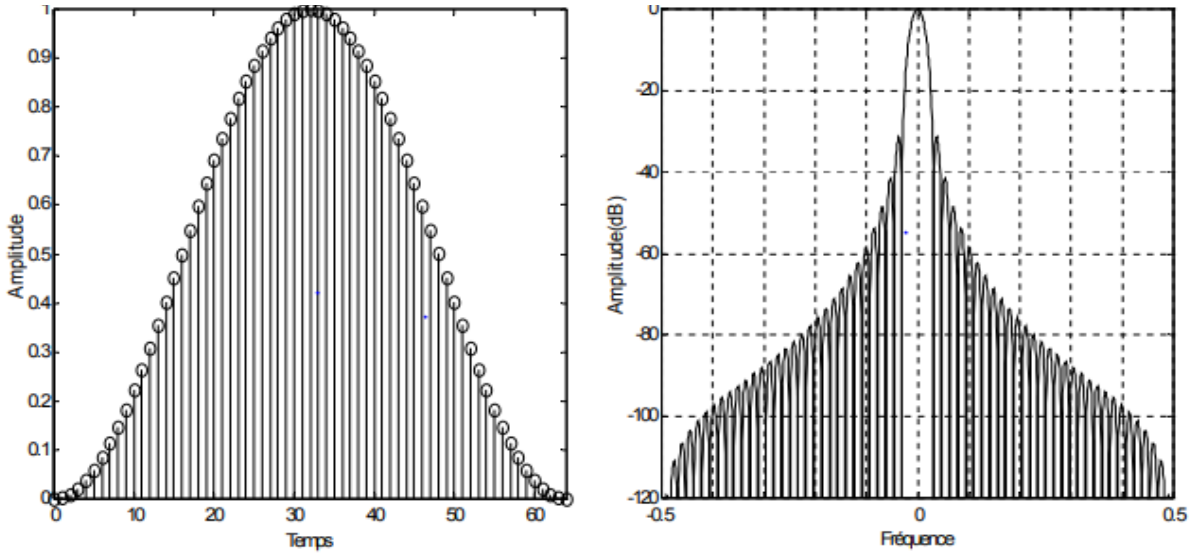


Figure 1 : exemple de fenêtrage Hanning [10].

1.3.2. Analyse de fréquence

Le résultat de la dégradation du signal dans les composantes sinusoïdales, le signal est converti dans le domaine fréquentiel et à cet effet, deux outils mathématiques sont utilisés, selon la nature du signal continu ou discret.

Pour les signaux continus, une transformée de Fourier rapide (FFT) est utilisée, tandis que pour les signaux discrets, l'outil doit être une transformée de Fourier discrète (DFT).

1.3.2.1. La Transformée de Fourier (TF)

La Transformée de Fourier est un outil mathématique qui permet de passer de la représentation temporelle à la représentation fréquentielle d'un signal. Ainsi qu'elle est une qui permet de représenter en fréquence (développement sur une base d'exponentielles) des signaux qui ne sont pas périodiques [11]. Son expression est la suivant :

$$TF[X(t)] = X(f) = \int_{-\infty}^{+\infty} x(t)e^{-j2\pi ft}\,dt \qquad (1.2)$$

$X(f)$= la transformé de fourrier du signal x (t).

$$x(t) = \int_{-\infty}^{+\infty} X(f)e^{-j2\pi ft}df \qquad (1.3)$$

1.3.2.2. Transformée de Fourier discrète (DFT)

La transformée de Fourier discrète est une méthode qui permet de décrire un signal discret en fonction de la fréquence. S'applique aux signaux discrets périodiques. Produit un spectre discret [11].

$$X_{DFT}[k] = \sum_{n=0}^{n-1} x[n] e^{-j2\pi kn/N} \quad (1.4)$$

1.4. La parole « vecteur de communication »

La parole est un système structuré qui permet aux humains de communiquer entre eux. Les informations du message vocal sont transmises par la fluctuation de la pression atmosphérique émise par l'appareil vocal, qui est le signal vocal. Le signal est analysé par l'oreille et l'information générée est transmise au cerveau qui l'interprète. Au sens strict, le contenu d'un signal de parole n'est représenté que par son intelligibilité. La parole montre comme une porteuse d'information [12], en analogie avec le principe de porteuse en transmission radio.

1.4.1. Paramètres du signal de parole

Dans plusieurs domaines d'analyse et de traitement du signal vocal, on définit par paramètres prosodiques :

- la fréquence fondamentale (vibration des cordes vocales).
- l'intensité de la voix (ou énergie).
- la durée successive des segments syllabiques.

Ces paramètres prosodiques prennent une importance particulière pour donner aux systèmes de synthèse une meilleure intelligibilité tout en permettant aux systèmes de reconnaissance d'effectuer une analyse ou segmentation par ordre d'unité phonétique. La variation dans le temps de ces paramètres (intonation) véhicule divers indices caractéristiques de l'individu que ce soit au niveau de son état physique (âge, sexe, physiologie), de son état émotionnel ou de son accent régional. Cependant, les paramètres prosodiques ne sont utilisés en général que pour faire rehausser légèrement les performances de ces systèmes [13].

1.4.2. La fréquence fondamentale

Notée f0, celle-ci constitue une caractéristique très importante de la voix. Elle correspond à la fréquence de vibration des cordes vocales lors de la production de voyelles ou de consonnes voisées. Elle génère des variations prosodiques, c'est à dire de mélodie et d'intonation, qui contribuent à l'identification du sexe, de l'âge, de l'identité du locuteur, ainsi que la signification du message prononcé [14].

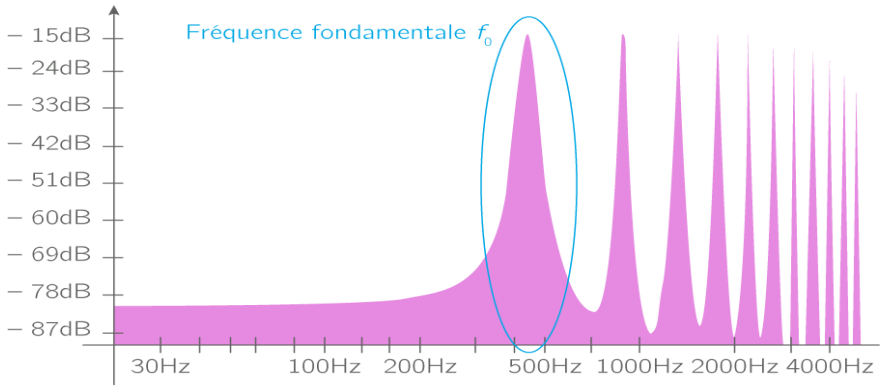


Figure 2 : fréquence fondamentale [15].

1.4.3. Spectre fréquentiel

Le spectre fréquentiel d'un domaine temporel d'un signal est la représentation de ce signal dans le domaine fréquentiel.

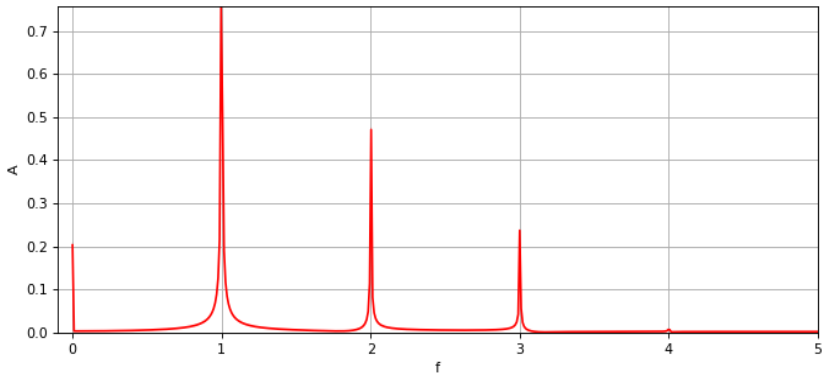


Figure 3 : Spectre fréquentiel d'un signal [16].

1.4.4. Energie

Soit x(t) un signal quelconque (fonction complexe) [17].

L´énergie totale est définie par :

$$W_x = \int_{-\infty}^{+\infty} |x(t)|^2 dt \qquad (1.5)$$

1.5. Le bruit

Le bruit est une oscillation de l'air qui, frappant le tympan, est interprété par l'oreille et le cerveau. On parle en général de bruit pour les sons non nécessaires et qui déplaisent.

Le son ou bruit est caractérisé par sa fréquence, sa vitesse de propagation et son amplitude. Si l'on se réfère à la physiologie humaine pour définir le bruit, il correspondra à un « phénomène acoustique produisant une sensation auditive considérée comme désagréable ou gênante » [18].

1.5.1. Le bruit peut être classé en

1.5.1.1. Bruit additif

Le bruit additif peut être vu comme le bruit provenant de diverses sources qui coexistent dans le même environnement acoustique.

Lorsque le signal est formé en ajoutant une substance propre parole et bruit. L'additif de bruit est pris en compte, donc la réduction du bruit effectue la tâche de séparer ces deux signaux de la meilleure façon.

La première suggestion est de traiter l'élimination du bruit comme une estimation des paramètres du problème, moyennant quoi une estimation optimale de la parole propre peut être faite selon le critère de certains facteurs. Par exemple, le SNR (Rapport signal sur bruit) pour estimer une parole nette par rapport au son d'origine serait un exemple [19].

1.5.1.2. Signaux parasites

Dans le cas d'un signal vocal, les signaux d'interférence proviennent d'autres haut-parleurs que ceux d'intérêt. Les signaux parasites sont créés par des variations rapides de tension ou de courant à travers des conducteurs. Ils peuvent venir se superposer à d'autres signaux et les brouiller [20].

1.5.1.3. Réverbération

Cet effet est produit par la propagation par trajets multiples. Il se produit dans des environnements acoustiques clos ou semi clos et c'est une forme de distorsion.

1.5.1.4. Écho

Il est produit par le couplage entre les microphones et les haut-parleurs. C'est une autre forme de distorsion.

1.5.2. Les effets psychologiques et physiologiques du bruit

Le bruit interfère avec les activités telles que les études, le travail, le sommeil ou même les loisirs. Il provoque de la fatigue, et le bruit peut donc être très perturbant dans la vie de tous les jours.

Cela peut provoquer des irritations et des maux de tête. Les bruits avec de nombreux décibels peuvent provoquer une surdité temporaire ou permanente.

Psychologiquement, il a des effets négatifs sur la productivité et l'efficacité des travailleurs. [21].

1.5.3. Classification des types de bruit en fonction de leur nature

Selon sa nature, le bruit peut être classé en différents types

1.5.3.1. Bruit physique

Le bruit physique est externe au locuteur et à l'auditeur. Il comprend des choses telles que les bruits de la construction d'une route à l'extérieur de votre fenêtre qui rendent difficile d'entendre ce qui est dit.

1.5.3.2. Bruit Psychologique

Le bruit psychologique est une interférence mentale qui vous empêche d'écouter. Si votre esprit erre lorsque quelqu'un vous parle, le bruit dans votre tête empêche la communication.

1.5.3.3. Distorsions de canal, écho et évanouissement

Ce bruit est le résultat de caractéristiques non idéales des canaux de communication.

Les communications par téléphone mobile sont particulièrement sensibles à la propagation caractéristique du canal [22].

1.5.4. Classification selon sa fréquence ou son temps caractéristiques

1.5.4.1. Bruit blanc

Un bruit blanc est une réalisation d'un processus aléatoire dans lequel la densité spectrale de puissance est la même pour toutes les fréquences.

Le bruit blanc est composé de l'ensemble des fréquences audibles. Il constitue, sur une bande passante de largeur infinie, le bruit de fond thermique de la matière. Il est obtenu en amplifiant le bruit de fond de composants électroniques [23].

On parle souvent de bruit blanc gaussien, il s'agit un bruit blanc qui suit une loi normale de moyenne et variance donnes.

En synthèse et traitement du son, on ne considère que les fréquences comprises entre 20Hz et 20kHz puisque l'oreille humaine n'est sensible qu'à cette bande de fréquences (en fait plutôt 25Hz-19kHz). L'impression obtenue est celle d'un souffle.

Le bruit blanc contient théoriquement toutes les fréquences avec la même intensité.

Figure 4 : spectre d'un bruit blanc.

1.5.4.2. Bruits colorés

Les bruits colorés sont des bruits larges bandes, neutres et continus. Ce sont des signaux aléatoires à propriétés statiques caractéristiques. Il en existe plusieurs standardisés, que l'on différencie en fonction de leur densité spectrale de puissance. Ils ont été créés à partir de bruits blancs (bruit possédant le même niveau sonore par bandes de fréquences à largeurs égales) auxquels on applique un filtre spectral.

Un bruit coloré « haute fréquence » [2000 – 30000 Hz] : bruit identifié comme étant du charriage de petites particules (diamètre ~ 1 cm) [24].

Il existe le bruit rose (bruit possédant le même niveau sonore par bandes d'octaves), rouge (ayant une puissance sonore qui décroit de 6dB par octave), bleu (ayant une puissance sonore qui augmente de 3 dB par octave, gris (bruit rose soumis à une courbe de sonie, bruit violet (ayant une puissance sonore qui augmente de 6 dB par octave).

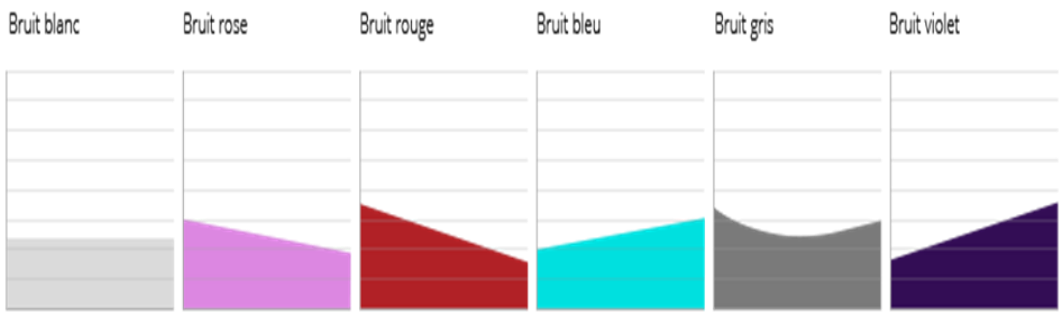


Figure 5 : exemple des bruits colorés [25].

1.5.4.3. Bruit impulsif

Il se compose d'impulsions de courte durée d'amplitude aléatoire et de durée aléatoire.

Se compose d'impulsions de courte durée, dues à une variété de sources telles que comme les commutateurs de bruit, les rainures ou la dégradation de surface des enregistrements audio, les « clics » de claviers d'ordinateur, etc.

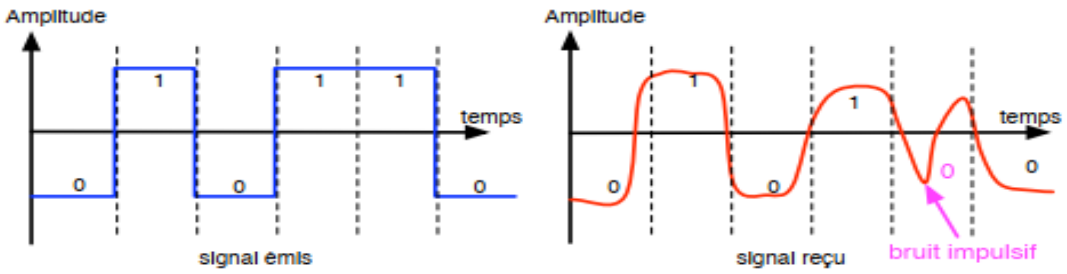


Figure 6 : exemple de bruit impulsif [26].

1.6. Conclusion

Ce chapitre est consacré à l'étude des généralités sur le signal de parole. Une étude de ses propriétés fréquentielles et temporelles est aussi présentée. En addition de son analyse, nous avons aussi étudié différents types de bruits d'environnement qui peuvent affecter ce signal.

Nous avons présenté également la classification des bruits en fonction de leurs natures et aussi selon leurs fréquences.

Chapitre 2 : les méthodes statistiques

2.1 Introduction

Ce chapitre est consacré essentiellement pour étudier les méthodes statistiques, l'objectif étant d'arriver à exposer quelques éléments de leurs théories, permettant ainsi de mieux comprendre leurs utilisations dans le traitement du signal en général et dans le rehaussement du signal de parole en particulier [27].

2.2.1. Maximum de vraisemblance (ML)

L'approche ML est peut-être l'approche la plus populaire en théorie de l'estimation statistique pour dériver des estimateurs pratiques et souvent utilisée même pour les problèmes d'estimation les plus compliqués. Elle a d'abord été appliquée à l'amélioration de la parole par Macaulay et Malpass [28].

Supposons qu'on nous donne un ensemble de données à N points $y = \{y(0), y(1), y(N-1)\}$ qui dépend d'un paramètre inconnu θ. Dans l'amélioration de la parole, y (l'ensemble de données observé) peut être le spectre d'amplitude de la parole bruyante, et le paramètre d'intérêt, peut être le spectre d'amplitude de la parole propre. De plus, supposons que l'on connaisse la PDF (probability density function) de$y$, que l'on note$P(y;\theta)$. La PDF de y est paramétrée par le paramètre inconnu, et nous la désignons par le point-virgule. Comme le paramètre θ affecte la probabilité de y, nous devrions pouvoir déduire les valeurs de à partir des valeurs observées de y ; c'est-à-dire que nous pouvons poser la question, quelle valeur de a le plus probablement produit les données observées y?

Mathématiquement, on peut chercher la valeur de qui maximiser $P(y;\theta)$, c'est-à-dire

$$\hat{\theta}_{ML} = \arg\max p(y;\theta) \tag{2.1}$$

Soit $y(n) = x(n) + d(n)$ le signal de parole bruitée échantillonné constitué du signal propre $x(n)$ et du signal de bruit $d(n)$. Dans le domaine fréquentiel, on a:

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

Pour $w = 2\pi f$ et $K = 0, 1, 2 \ldots . N-1$

N : la langueur de trame en échantillons. L'équation précédente peut également être exprimée sous forme polaire comme :

$$Y_k e^{j\theta_y(k)} = X_k e^{j\theta_x(k)} + D_k e^{j\theta_d(k)} \quad (2.3)$$

Où $\{Y_k, X_k, D_k\}$ désignent les amplitudes et $\{\theta_x(k), \theta_d(k), \theta_y(k)\}$ désignent les phases au niveau du bin de k fréquences de la parole bruitée, de la parole propre et du bruit, respectivement. La fréquence k de la parole bruitée de la parole propre et du bruit, respectivement. Dans l'approche ML, proposée par Macaulay et Malpass [28], les spectres d'amplitude et de phase du signal propre, de la parole propre et du bruit, respectivement. Phase du signal propre, c'est-à-dire $X_k$et $\theta_x(k)$, sont supposés être inconnus mais déterministes. La fdp des coefficients de la transformée de Fourier du bruit D (ω_k) est supposée être une gaussienne complexe à moyenne nulle. Les parties réelles et imaginaires de $D(w_k)$ sont supposées avoir des variances $\lambda_d(k)/2$. Sur la base de ces deux hypothèses, nous pouvons former La densité de probabilité des coefficients DFT de la parole bruitée observée $y(\omega_k)$. La densité de probabilité de $Y(\omega_k)$ est également gaussienne de variance λ_d (k) et de moyenne $X_k e^{j\theta_x(k)}$:

$$p\big(Y(w_k); X_k, \theta_x(k)\big) = \frac{1}{\pi\lambda_d(k)} \exp\left[-\frac{|Y(w_k) - X_k e^{j\theta_x(k)}|^2}{\lambda_d(k)}\right] \quad (2.4)$$

$$p\big(Y(w_k); X_k, \theta_x(k)\big) = \frac{1}{\pi\lambda_d(k)} \exp[-\frac{Y_k^2 - 2X_k Re\{e^{-j\theta_x(k)} Y(w_k)\} + X_k^2}{\lambda_d(k)}] \quad (2.5)$$

Pour obtenir l'estimation ML de $X_k$nous devons calculer le maximum de $p\big(Y(w_k); X_k, \theta_x(k)\big)$ par rapport à X_k. Ce n'est cependant pas simple car $p\big(Y(w_k); X_k, \theta_x(k)\big)$ est fonction de deux paramètres inconnus : l'amplitude et la phase. Le paramètre de phase est considéré comme un paramètre de nuisance (c'est-à-dire un paramètre indésirable), qui peut être facilement éliminé en « l'intégrant ». Plus précisément, nous pouvons éliminer le paramètre de phase en maximisant à la place la fonction de vraisemblance moyenne suivante:

$$P_L(Y(w_k); X_k) = \int_0^{2\pi} p(Y(w_k); X_k, \theta_x p(\theta_x) d\theta_x \quad (2.6)$$

Pour plus de simplicité nous supprimons l'indice k de la phase. en supposant une distribution uniforme sur $p(\theta_x) = \frac{1}{2\pi}$, $\theta_x = [0, 2\pi]$, la fonction de vraisemblance devient :

$$P_L(Y(w_k); X_k) = \frac{1}{\pi\lambda_d(k)} \exp\left[-\frac{Y_k^2 + X_k^2}{\lambda_d(k)}\right] \frac{1}{2\pi} \int_0^{2\pi} \exp\left[\frac{2X_k Re(e^{-j\theta_x} Y(w_k)}{\lambda_d(k)}\right] d\theta_x \quad (2.7)$$

L'intégrale dans l'équation précédente est connue sous le nom de fonction de Bassel modifiée du premier type et est donnée par :

$$I_0(|x|) = \frac{1}{2\pi}\int_0^{2\pi} \exp\left[\mathrm{Re}\left(xe^{-j\theta_x}\right)\right] d\theta_x \quad (2.8)$$

$$I_0(|x|) = \frac{1}{\sqrt{2\pi|x|}}\exp(|x|) \quad (2.9)$$

Et la fonction de vraisemblance dans l'équation 2.9 se simplifie en :

$$P_L(Y(\omega_{\mathbf{k}}); X_k) = \frac{1}{\pi\lambda_d(k)}\frac{1}{\sqrt{2\pi\frac{2X_kY_k}{\lambda_d(k)}}}\exp[-\frac{{X_k}^2+{Y_k}^2-2X_kY_k}{\lambda_d(k)}] \quad (2.10)$$

Après avoir différencié la fonction de log-vraisemblance $\log P_L(Y(w_k); X_k)$ par rapport à l'inconnu (paramètre) X_k, et mis la dérivée à zéro, nous obtenons l'estimation ML du spectre de l'amplitude :

$$\widehat{X}_k = \frac{1}{2}[Y_k + \sqrt{Y_k^2 - \lambda_d}\,] \quad (2.11)$$

En utilisant la phase bruitée θ_y, au lieu deθ_x, nous pouvons exprimer l'estimation du spectre du signal propre (complexe) comme

$$\widehat{X}(w_k) = \widehat{X}_k e^{j\theta_y} = \widehat{X}_k \frac{Y(\omega_k)}{\omega_k}$$

$$= [\frac{1}{2} + \frac{1}{2}\sqrt{\frac{Y_k^2-\lambda_d(k)}{Y_k^2}}]\, Y(\omega_k) \quad (2.12)$$

En laissant $\gamma_k \triangleq \frac{{Y_k}^2}{\lambda_d(k)}$ désigner le rapport signal sur bruit (SNR) a posteriori au mesuré sur la base des données observée, l'équation précédente peut également s'écrire sous la forme :

$$X(\omega_k) = [\frac{1}{2} + \frac{1}{2}\sqrt{\frac{\gamma_{k-1}}{\gamma_k}}]\, Y(\omega_k) \quad (2.13)$$

$$= G_{ML}(\gamma_k)Y(\omega_k) \quad (2.14)$$

Si les coefficients DFT de la parole sont modélisés comme des processus aléatoires gaussiens indépendants de moyenne nulle, mais que c'est la variance du signal $\lambda_X(k)$, qui est inconnue et déterministe, nous obtenons une fonction de vraisemblance différente. Comme le signal et le bruit sont supposés être indépendants, la variance de$Y(w_k)$, désignée par $\lambda_Y(k)$, est donnée

par $\lambda_Y(k) = \lambda_X(k) + \lambda_Y(k)$. Par conséquent, la densité de probabilité de Y (w_k) est donnée par :

$$p\big(Y(\omega_k);\lambda_x(k)\big) = \frac{1}{\pi[\lambda_x(k)+\lambda_d(k)]}\exp[-\frac{Y_k^2}{\lambda_x(k)+\lambda_d(k)}] \quad (2.15)$$

En maximisant la fonction de vraisemblance $p\big(Y(w_k);\lambda_x(k)\big)$ par rapport à$\lambda_x(k)$

$$\hat{\lambda}_x(k) = Y_k^2 - \lambda_d(k) \quad (2.16)$$

En supposant que $X_k^2 = \lambda_x(k)$et $D_k^2 = \lambda_d(k)$ et $(Y_k^2 - \lambda_d(k))>0$ nous obtenons une estimation du spectre d'amplitude du signal :

$$\widehat{X}_k = \sqrt{Y_k^2 - D_k^2} \quad (2.17)$$

Notez que cet estimateur de X_k, n'est rien d'autre que l'estimateur par soustraction du spectre de puissance. Par conséquent, l'approche originale de soustraction du spectre de puissance peut être dérivée en utilisant principes de ML en supposant que les coefficients de la transformée de Fourier du signal et du bruit sont modélisés comme des processus aléatoires gaussiens indépendants et la variance du signal $\lambda_x(k)$, est inconnue mais déterministe. Comme dans l'équation (2.14), nous pouvons calculer l'estimation du spectre de signal propre obtenu par soustraction du spectre de puissance comme :

$$\widehat{X}_k(\omega_k) = \widehat{X}_{k_k} e^{j\theta_y} = \widehat{X}_{k_k}\frac{Y(\omega_k)}{Y_k}$$

$$= \sqrt{\frac{Y_k^2-\lambda_d(k)}{Y_k^2}}]\, Y(\omega_k) \quad (2.18)$$

En terme de Y, l'équation précédente peut être écrite comme :

$$X(\omega_k) = \sqrt{\frac{\gamma_{k-1}}{\gamma_k}}\, Y(\omega_k) \quad = G_{ps}(\gamma_k)Y(\omega_k) \quad (2.19)$$

Où GPS (γ_k) est la fonction de gain de la méthode de soustraction du spectre de puissance. Enfin, il est intéressant de noter que si nous substituons l'estimation ML de $\lambda_x(k)$ (équation 2.16) dans l'équation du filtre de Wiener :

$$\widehat{X}(\omega_k) = \frac{\lambda_x(k)}{\lambda_x(k)+\lambda_d(k)}Y(\omega_k) \quad (2.20)$$

$$\widehat{X}(\omega_k) = \frac{Y_k^2 - \lambda_d(k)}{Y_k^2} Y(\omega_k) = \frac{\gamma_{k-1}}{\gamma_k} Y(\omega_k) \quad (2.21)$$

$$= G_{ps}^2(\gamma_k) Y(\omega_k) \quad (2.22)$$

En comparant l'équation 2.19 avec l'équation 2.22, nous voyons que l'estimateur de Wiener est le carré de l'estimateur par soustraction du spectre de puissance. Par conséquent, l'estimateur de Wiener fournit une plus grande atténuation spectrale que l'estimateur par soustraction du spectre de puissance, pour une valeur fixe deγ_k. Enfin, il convient de souligner que la règle de suppression ML n'est jamais utilisée seule, car elle ne fournit pas suffisamment d'informations sur la qualité du spectre. Dans [28], il a été utilisé en conjonction avec un modèle de parole à deux états qui incorporait l'incertitude de présence du signal.

L'estimateur du maximum de vraisemblance est très souvent utilisé [29].

Les problèmes auxquels il se heurte :

• difficulté à résoudre le problème de maximisation, par exemple due à la présence de nombreux maxima locaux.

• coût calculatoire parfois élevé dû à une maximisation compliquée [29].

2.2.1. Filtrage de Wiener Paramétrique

Les techniques de rehaussement de la parole les plus couramment utilisées sont des généralisations de techniques comme le filtrage de Wiener à court-terme [30] et les méthodes de soustraction spectrale. Ces méthodes ont été largement étudiées et employées dans plusieurs applications. Cette famille d'algorithmes peut être rassemblée sous une même catégorie appelée filtrage de Wiener paramétrique ou soustraction spectrale paramétrique pouvant être exprimés par une formulation paramétrique générale. Le gain spectral correspondant à cette para métrisation est alors défini par :

$$H(\omega) = [1 - (\frac{\left[\widehat{N}(\omega)\right]^2}{\lceil X(\omega) \rceil^2})^{\frac{\delta}{2}}]^{\rho}$$

2.3. L'erreur quadratique moyenne MMSE

L'estimateur de Wiener peut être dérivé en minimisant l'erreur entre un modèle linéaire du spectre propre et le spectre vrai. L'estimateur de Wiener est considéré comme l'estimateur spectral complexe optimal (au sens de l'erreur quadratique moyenne), mais n'est pas l'estimateur d'amplitude spectrale optimal. Reconnaissant l'importance de l'amplitude spectrale à court terme (STSA : Short Term Spectral Amplitude) sur l'intelligibilité et la qualité de la parole, plusieurs auteurs ont proposé des méthodes optimales pour obtenir les amplitudes spectrales à partir d'observations bruitées. En particulier, des estimateurs optimaux ont été recherchés qui a minimisé l'erreur quadratique moyenne entre les grandeurs estimées et réelles :

$$e = E\{(\widehat{X}_k - X_k)^2\} \tag{2.23}$$

Où

$\widehat{X}_k$: L'amplitude spectrale estimée à la fréquence ω_K

X_k : La véritable amplitude du signal propre

La minimisation de l'équation 2.23 peut être effectuée de deux manières, selon la façon dont nous réalisons l'attente. Dans l'approche MSE (mean square error) classique, l'espérance est faite par rapport à $p(Y; X_k)$, où Y désigne le spectre de parole bruitée observé $Y = [Y(\omega 0)\ Y(\omega 1) \ldots\ Y(\omega_{N-1})]$. Dans l'approche bayésienne MSE, l'attente est faite par rapport à la pdf jointe $p(Y; X_k)$, et la MSE bayésienne est donnée par :

$$BMSE(\widehat{X}_k) = \iint (X_k - \widehat{X}_k)^2\ p(Y, X_k) dY\ dX_k \tag{2.24}$$

La minimisation de la MSE bayésienne par rapport à $\widehat{X}_k$ conduit à l'estimateur optimal de la MMSE donné par :

$$\widehat{X}_k = \int X_k\ p\left(\frac{X_k}{Y}\right)\ dX_k \tag{2.25}$$

$$= E[X_k/Y]$$

$$= E\left[\frac{X_k}{Y(\omega_0)Y(\omega_1)\ldots\ldots Y(\omega_{N-1})}\right] \tag{2.26}$$

Qui est la moyenne de la PDF postérieur des amplitudes spectrales propres, c'est –à-dire $p(X_k, Y)$, est la PDF des amplitudes après toutes les données (c'est-à-dire les spectres

complexes de parole bruyante, Y, dans notre cas) aient été observées. En revanche, la pdf a priori deX_k, c'est-à-dire$p(X_k)$, fait référence à la pdf des amplitudes propres avant que les données ne soient observées.

Notez qu'il existe deux différences fondamentales entre l'estimateur de Wiener et l'estimateur MMSE donné dans l'équation (2.27).

Premièrement, dans la dérivation du filtre de Wiener, nous avons supposé que $\widehat{X}_k = H_k \{Y(\omega_k)\}$ pour un filtre inconnuH_k, c'est-à-dire que nous avons supposé qu'il existe une relation linéaire entre les deux. Nous avons supposé qu'il existe une relation linéaire entre $Y_k(\omega_k)$ (les données observées) et $\widehat{X}_k(\omega_k)$.

Deuxièmement, le filtre de Wiener est obtenu en évaluant la moyenne de la FDP postérieure de $X(\omega_k)$ plutôt que x_k ; c'est-à-dire qu'il est donné par $E[X(\omega_k)/\ Y_k(\omega_k)]$. Le filtre de Wiener est donc l'estimateur de spectre complexe optimal (au sens MMSE) et non l'estimateur de spectre d'amplitude optimal selon le modèle supposé.

L'estimateur MMSE donné dans l'équation 2.26, contrairement à l'estimateur de Wiener, ne suppose pas l'existence d'une relation linéaire entre les données observées et l'estimateur, mais il nécessite une connaissance des distributions de probabilité des coefficients DFT de la parole et du bruit. En supposant que nous ayons des connaissances préalables sur les distributions des coefficients DFT de parole et de bruit, nous pouvons évaluer la moyenne de la PDF postérieure de X_k, c'est-à-dire la moyenne de $p(X_k /(Y))$.

Cependant, mesurer les vraies distributions de probabilité des coefficients de transformée de Fourier de la parole a été difficile, en grande partie parce que le signal de parole (et parfois le bruit) n'est ni un processus stationnaire ni un processus ergodique. Plusieurs ont tenté de mesurer les distributions de probabilité en examinant le comportement à long terme des processus. Cependant, on peut se demander si les histogrammes des coefficients de Fourier, obtenus à l'aide d'une grande quantité de données, mesurent la fréquence relative des coefficients de la transformée de Fourier plutôt que la véritable densité de probabilité des coefficients de la transformée de Fourier.

Pour contourner ces problèmes, Ephraim et Malah [31] ont proposé un modèle statistique qui utilise les propriétés statistiques asymptotiques des coefficients de la transformée de Fourier. Ce modèle repose sur deux hypothèses :

1. Les coefficients de la transformée de Fourier (parties réelle et imaginaire) ont une distribution de probabilité gaussienne. La moyenne des coefficients est nulle et les variances des coefficients varient dans le temps en raison de la non-stationnarité de la parole.

2. Les coefficients de la transformée de Fourier sont statistiquement indépendants et, par conséquent, non corrélés.

L'hypothèse gaussienne est motivée par le théorème central limite, car les coefficients de la transformée de Fourier sont calculés comme une somme de N variables aléatoires. Considérons, par exemple, le calcul des coefficients de transformée de Fourier de la parole bruitée, $Y(\omega_k)$:

$$Y(\omega_k) = \sum_{n=0}^{N-1} y(n) e^{-j\omega_k n} = y(0) + a_1 y(1) + a_2 y(2) + \cdots a_{N-1}\, y(N-1) \qquad (2.27)$$

Où $a_{m=e^{-j\omega_k m}}$ sont des constantes.

Y(n) Sont les échantillons dans le domaine temporel du signal de parole bruitée.

2.3.1. Estimateur MMSE d'ordre *p*

Sachant que l'estimateur MMSE d'amplitude est exprimée par :

$$\hat{S}_k = \frac{\sqrt{\pi}}{2}\frac{\sqrt{v_k}}{\gamma_k}\exp\left(-\frac{v_k}{2}\right)\left[(1+v_k)I_0\left(\frac{v_k}{2}\right) + v_k I_1\left(\frac{v_k}{2}\right)\right]X_k \qquad (2.28)$$

Dans l'estimateur MMSE d'ordre p [5], la fonction de suppression est calculée en fonction d'un facteur en puissance p. La valeur de $p = 1$, nous permet d'obtenir l'estimateur MMSE d'origine du spectre d'amplitude de l'équation (2.28). Il a été remarqué que si *p* est grand, il résulte de l'apparition d'un niveau de bruit résiduel assez élevé. En d'autres termes si *p* est petit, il en résulte une distorsion du signal. Pour cela, il a été procédé à l'adaptation du facteur *p* par rapport à chaque segment du signal.

2.3.2. Estimateurs MMSE basés sur des distributions non gaussiennes

Pour ces estimateurs [32-33] et [34], on suppose que les coefficients TFD du spectre d'amplitude du signal bruité et de la phase peuvent être modélisés par une loi Gamma ou de Laplace, ce qui a permet d'avoir des résultats améliorés par rapport à la méthode MMSE [35].

2.3.3. MMSE d'amplitude

Pour déterminer l'estimateur MMSE nous devons d'abord calculer la pdf postérieure de X_k, c'est –à-dire p $[X(\omega_k)/\ Y_k(\omega_k)\ ]$. Nous pouvons utiliser la règle de Bayes pour le déterminer comme :

$$p(X_k/y(\omega_K)) = p(Y(\omega_K)/X_k)\ p(X_k)\ /p(Y(\omega_k)) = p\left(\frac{Y(\omega_K)}{X_k}\right)\ p(X_k)\ /\int_0^\infty p(Y(\omega_K)/X_k)p(X_k)\,dx_k \qquad (2.29)$$

Où x_k, est une réalisation de la variable aléatoire X_k. Notez que $p(Y(\omega_K)$ est un facteur de normalisation requis pour s'assurer que $p(\ X_k/IY(\omega_K))$ s'intègre à 1. En supposant une indépendance statistique entre les coefficients de transformation, c'est-à-dire :

$$E[(\frac{X_k}{Y(\omega_0)Y(\omega_1)\ldots\ldots (Y\omega_{n-1})}] \ = E\ (X_k/y(\omega_K)) \qquad (2.30)$$

Et en utilisant l'expression précédente pour P $(X_k/y(\omega_K))$ L'estimateur de l'équation 2.26 se simplifie en :

$$\widehat{X}_k = E[x_k/Y(\omega_k)] = \int_0^\infty x_k\ p(\frac{x_k}{Y(\omega_k)}))d\ x_k = \int_0^\infty x_k\ p(Y(\omega_k)\ p(x_k)dx_k/\int_0^\infty p(Y(\omega_k)p(x_k)dx_k \qquad (2.31)$$

Depuis

$$p(Y(\omega_k)\backslash p(X_k) = \int_0^{2\pi} p(Y(\omega_k)\backslash x_k,\theta_x)p(x_k,\theta_x)d\theta_x \qquad (2.32)$$

Où θ_x, est la réalisation de la variable aléatoire de phase de $X(\omega_k)$ (pour plus de clarté on laisse désormais tomber l'indice k en θ_x), on obtient :

$$\hat{x}_k = \frac{\int_0^\infty \int_0^{2\pi} x_k\ p(Y(\omega_k)\backslash x_k,\theta_k)p(x_k,\theta_x)d\theta_x dx_k}{\int_0^\infty \int_0^{2\pi} p(Y(\omega_k)\backslash x_k,\theta_k)p(x_k,\theta_x)d\theta_x dx_k} \qquad (2.33)$$

Ensuite, nous devons estimer $p(Y(\omega_k)\backslash(x_k),\theta_x)$ et $p\ (x_k,\theta_x)$. D'après le modèle statistique supposé, nous savons que $Y\ (\omega_k)$ est la somme de deux variables aléatoires gaussiennes complexes de moyenne nulle. Par conséquent, la Fdp conditionnelle $p(Y(\omega_k)\backslash(x_k),\theta_x)$ sera également gaussienne :

$$P(Y((\omega_k)\backslash x_k,\theta_k) = p_D\ Y((\omega_k) - X((\omega_k)) \tag{2. 34}$$

Où p_D est la fdp des coefficients de la transformée de Fourier du bruit $D\ (\omega_k)$. L'équation précédente devient alors :

$$P(Y\ ((\omega_k)\backslash x_k,\theta_k) = \frac{1}{\pi\,\lambda_d(K)}\exp\left\{-\frac{1}{\lambda_d(K)}|Y((\omega_k) - X((\omega_k)|^2\right\} \tag{2.35}$$

Où $\lambda_d(K) = E\{|D(\omega_k)|\}^2$ est la variance de la kième composante spectrale du bruit. Pour les variables aléatoires gaussiennes complexes, nous savons que l'amplitude (X_k) et la phase $\theta_X(k)$ des variables aléatoires de $X\ (\omega_k)$ sont indépendantes et nous pouvons donc évaluer la fdp conjointe $P\ (x_k,\theta_x)$ comme le produit des fdp individuelles, c'est-à-dire $P(x_k,\theta_x) = p\ (x_k)\ p(\theta_x)$. La Fdp de x_k est de Rayleigh puisque $x_k = \sqrt{r(k)^2} + \sqrt{i(k)^2}$ où $r(k) = Re\{X\ (\omega_k)\}$ et $i(k) = Im\{X(\omega_k)\}$ sont des variables aléatoires gaussiennes. La fdp de $\theta_x(k)$ est uniforme en $(-\pi,\pi)$ et donc la probabilité conjointe $P\ (x_k,\theta_x)$ est donnée par [36] :

$$P\ (x_k,\theta_k) = \frac{x_k}{\pi\,\lambda_d(K)}\exp\left\{-\frac{{x_k}^2}{\lambda_x(k)}\right\} \tag{2.36}$$

Où $\lambda_x(K)$ - $E\ \{|X(\omega_k)|\}^2\}$ est la variance de la kième composante spectrale du signal propre. En remplaçant les équations 2.35 et 2.36 dans l'équation 2.31, nous obtenons finalement l'estimateur de magnitude MMSE optimal

$$\hat{x}_k = \sqrt{\lambda_k}\Gamma(1.5)\Phi(-0.5,1;-v_k)Y_k \tag{2.37}$$

Où Γ désigne la fonction gamma, ϕ (a, b; c) désigne la fonction hypergéométrique confluente, λ_k est donné par :

$$\lambda_k = \frac{\lambda_x\,(k)\lambda_d\,(k)}{\lambda_x\,(k)+\lambda_d\,(k)} = \frac{\lambda_x\,(k)}{1+\xi_k} \tag{2.38}$$

Et V_k est définie par:

$$V_k = \frac{\xi_k}{1+\xi_k}\gamma_k \quad (2.39)$$

Où γ_K et ξ_k est définis par:

$$\gamma_K = \frac{Y_k{}^2}{\lambda_D\,(k)} \quad (2.40)$$

$$\xi_k = \frac{\lambda_x(k)}{\lambda_d(k)} \quad (2.41)$$

Les termes ξ_k et γ_K sont appelés respectivement SNR a priori et a posteriori. Le SNR a priori peut être considéré comme le vrai SNRξ_k de la kième composante spectrale, tandis que le SNRγ_K a posteriori peut être considéré comme le SNR observé ou mesuré de la kième composante spectrale après ajout du bruit.

2.4. Estimateurs de maximum a Posteriori (MAP)

Cet estimateur est basé sur la maximisation de la FDP a posteriori [34-37].

Bien que l'approche MMSE est axée sur le calcul de la moyenne de la FDP a posteriori (Sk|(ωk)) (c.à.d. E{Sk|X(ωk)}). L'estimateur MAP, par contre, vise à trouver le maximum

De (Sk|(ωk)).

Jusqu'à présent, nous avons décrit des algorithmes d'amélioration de la parole pour une estimation d'amplitude spectrale optimale basée sur les principes ML et MMSE. Des estimateurs d'amplitude spectrale basés sur la maximisation de la pdf a posteriori (MAP) ont également été proposés. [34-38]

Bien que l'approche MMSE vise à trouver la moyenne des pdf a posteriori $p(x_k|Y(w_k))$, c'est-à-dire $E(x_k|Y(w_k))$, l'approche MAP vise à trouver le max de $p(x_k|Y(w_k))$. De toute évidence, si la fdp a posteriori est symétrique alors les estimateurs MMSE et MAP sont identiques. Les algorithmes MAP sont souvent utilisés comme alternative aux algorithmes MMSE dans des circonstances où il est extrêmement difficile de dériver des estimateurs MAP. Dans certains cas, il est plus facile de maximiser la PDF a posteriori $p(x_k|Y(w_k))$ plutôt que d'évaluer la moyenne de $p(x_k|Y(w_k))$.

À la section 2.3, nous avons dérivé un estimateur MMSE pour l'amplitude spectrale et un estimateur MMSE pour la phase. Les deux estimateurs ont été dérivés indépendamment et non conjointement. En utilisant le critère MAP, nous pouvons calculer un maximum conjoint a posteriori d'un estimateur spectral de magnitude et de phase. Plus précisément, nous pouvons trouver le maximum du PDF a posteriori $p(x_k, \theta_X | Y(w_k))$. Les estimateurs MAP de l'amplitude et de la phase peuvent être dérivés comme la solution de :

$$(\hat{x}_k, \hat{\theta}_k) = \arg\max_{x_k, \theta_x} p(x_k, \theta_k | Y(w_k)) \quad (2.42)$$

En utilisant la règle de bays, nous pouvons exprimer $p(x_k, \theta_k | Y(w_k))$ comme :

$$p(x_k, \theta_k | Y(w_k)) = \frac{p(Y(w_k) | x_k, \theta_k) p(x_k, \theta_k)}{p(Y(w_k))} \quad (2.43)$$

Puisque p $(Y(w_k))$ n'est pas une fonction de x_k ou θ_k, on peut maximiser p $(Y(w_k)/ x_k, \theta_k)$ p(x_k, θ_k).

Les estimateurs MAP de x_k et θ_k, peuvent alors être obtenus comme la solution de

$$(\hat{X}_k, \hat{\theta}_k) = \quad \arg\max p\,(Y(w_k) | x_k, \theta_k) p(x_k, \theta_k) \quad (2.44)$$

En supposant le modèle statistique gaussien et en utilisant les équations 2.30 et 2.31, nous avons :

$$p(Y(w_k) | x_k, \theta_k) p(x_k, \theta_k) = \frac{x_k}{\pi^2 \lambda_x(k) \lambda_d(k)} \exp\left(-\frac{|Y(w_k) - x_k e^{\theta_x j}|^2}{\lambda_d(k)} - \frac{x_k^2}{\lambda_x(k)}\right) \quad (2.45)$$

Comme la fonction log est une fonction monotone croissante, nous pouvons alternativement maximiser le logarithme de l'équation précédente, c'est-à-dire

$$J_i = \ln[p(Y(w_k) | x_k, \theta_k) p(x_k, \theta_k)] = -\frac{|Y(w_k) - x_k e^{\theta_x}|^2}{\lambda_d(k)} - \frac{x_k^2}{\lambda_x(k)}) + \ln x_k + \text{constant} \quad (2.46)$$

En différenciant J_i par rapport à l'amplitude θ_x et en fixant la dérivée égale à zéro, nous obtenons l'estimateur de magnitude MAP [38] :

$$\frac{\partial J_i}{\partial \theta_x} = 2 \sin(\theta_y - \theta_x) = 0 \quad (2.47)$$

Et donc

$$\theta_x = \theta_y \quad (2.48)$$

L'estimation de phase MAP est donc la phase bruitée, qui se trouve être également l'estimation de phase MMSE.

Maintenant, en différenciant par rapport à l'amplitude x_k et en fixant la dérivée égale à zéro, nous obtenons l'estimateur de magnitude MAP :

$$\hat{X}_k = \frac{\varepsilon_k + \sqrt{\varepsilon_k^2 + 2(1+\varepsilon_k)\varepsilon_k/\gamma_k}}{2(1+\varepsilon_k)} Y_k \tag{2.49}$$

Les équations 2.48 et 2.49 donnent les estimateurs MAP pour l'amplitude et la phase de $\hat{X}_k(w_k)$. Ensuite, nous dérivons l'estimateur MAP pour la magnitude x_k uniquement ; c'est-à-dire que nous cherchons la solution pour

$$\hat{X}_k = \arg\max p\big(x_k \big| Y(w_k)\big) \tag{2.50}$$

En utilisant la règle de Bayes, nous pouvons exprimer $p\big(x_k \big| Y(w_k)\big)$ comme

$$p\big(x_k \big| Y(w_k)\big) = \frac{p(Y(\omega_k)/x_k)}{p(Y(\omega_k)} \; p(x_k) \tag{2.51}$$

Comme $p(Y(w_k))$ n'est pas une fonction de X_k, on peut maximiser $p(Y(\omega_k)/x_k)\, p(x_k)$ qui est :

$$\hat{X}_k = \text{argmax}_{xk}\; p\left(\frac{Y(\omega_k)}{x_k}\right) p(x_k) \tag{2.52}$$

Nous pouvons utiliser les équations 2.26 et 2.7 pour évaluer $p(Y(\omega_k)/x_k)\, p(x_k)$:

$$p(Y(\omega_k)/x_k)\, p(x_k) = \frac{x_k}{{\sigma_k}^2} \exp\left(\frac{-{x_k}^2 + {s_k}^2}{2\,{\sigma_k}^2}\right) I_0\left(\frac{x_k s_k}{{\sigma_k}^2}\right) \tag{2.53}$$

Où

$${\sigma_k}^2 \triangleq \frac{\lambda_k}{2}, {s_k}^2 \triangleq \lambda_k v_k \tag{2.54}$$

Et v_k et λ_k ont été définis dans les équations 2.33 et 2.34, respectivement. Notez que l'équation précédente a la forme du PDF de Rican. Substituer dans l'équation 2.53 l'approximation de la fonction de Bessel

$$I_0(|x|) \approx \frac{1}{\sqrt{2\pi|x|}} \exp(|x|) \tag{2.55}$$

On a

$$p\left(\frac{Y(\omega_k)}{x_k}\right) p(x_k) \approx \frac{1}{\sqrt{2\pi{\sigma_k}^2}} \sqrt{\frac{x_k}{s_k}} \exp\left(-\frac{1}{2}\left[\frac{x_k - s_k}{\sigma_k}\right]^2\right) \quad (2.56)$$

En différenciant le log de $p(Y(\omega_k)/x_k)\, p(x_k)$ par rapport à x_k et en fixant la dérivée à zéro, nous obtenons l'estimateur d'amplitude MAP optimal

$$\hat{X}_k = \frac{\xi_k + \sqrt{{\xi_k}^2 + (1+\xi_k)\xi_k/\gamma_K}}{2(1+\xi_k)} Y_k \quad (2.57)$$

Qui est différent de l'estimateur MAP ponctuel dérivé précédemment (Equation 2.49), par seulement un facteur de 2 à l'intérieur de la racine carrée.

Les estimateurs MAP donnés dans les équations 2.49 et 2.56 ont été comparés avec l'estimateur linéaire-MMSE (équation 2.39) en termes de différence dans la quantité de suppression appliquée à différents SNR_s. Les comparaisons ont indiqué que pour des valeurs élevées de ξ_k et γ_K , les estimateurs MAP et MMSE sont presque identiques. La différence entre les deux estimateurs était la plus grande lorsque ξ_k et γ_K étaient très petits. Pour des valeurs très petit ξ_k et γ_K, la différence de gain maximale entre l'estimateur MAP conjoint et l'estimateur linéaire-MMSE était de 5 dB, et la différence de gain maximale entre l'estimateur MAP et l'estimateur linéaire-MMSE était de 2 dB.

2.5. Conclusion

Ce chapitre a été consacré à l'étude des méthodes de rehaussement de la parole de l'état de l'art, basées sur des modèles statistiques pour une estimation optimale de l'amplitude spectrale. Nous nous sommes concentrés sur des estimateurs non linéaires de l'amplitude (c'est-à-dire le module des coefficients DFT) plutôt que sur le spectre complexe du signal (comme le fait le filtre de Wiener), en utilisant divers modèles statistiques et critères d'optimisation. Ces estimateurs non linéaires prennent explicitement en compte la fonction de densité de probabilité (FDP) du bruit et les coefficients DFT de la parole et utilisent, dans certains cas, des distributions a priori non gaussiennes. Ces estimateurs sont souvent combinés avec des modifications de gain à décision douce qui prennent en compte la probabilité de présence de parole.

Chapitre 3 : Résultats de simulation et discussions

3.1. Introduction

Dans ce chapitre, nous allons simuler sous Matlab les trois méthodes statistiques de rehaussement de la parole étudiées dans le chapitre 2. Le corpus de parole que nous avons utilisé est basé sur les données NOIZEUS et NIST [39]. Le choix de ces deux bases de données est motivé par le fait que NOIZEUS nous a permis, d'une part, l'utilisation des mêmes phrases phonétiquement équilibrées, des mêmes types de bruits et des mêmes mesures objectives et subjectives qui ont été déjà utilisées pour évaluer les méthodes de rehaussement de parole proposées dans la littérature.

3.2. Evaluation des performances

3.2.1. 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 [40]. 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

3.2.2. 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 [40].

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 2 : 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

3.2.3. 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)[39],
- Evaluation Perceptive de la Qualité de la Parole (PESQ : Perceptual Evaluation of Speech Quality)[36],
- Pente Spectrale Pondérée (WSS : Weighted Slop Spectral) [40]
- Logarithme du Rapport de Vraisemblance (LLR : Log Likelihood Ratio)[40].

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 ML, MAP et MMSE étudiées dans notre travail a été effectuée par les mesures objectives et subjectives et SNR de signal rehaussée illustrés dans les tableaux 3, 4, 5 et 6, et par les spectrogrammes, les allures temporelles de signaux et la mesure SNR amélioré illustrées dans les figures ci-dessous.

Tableau 3 : Evaluation objective et subjectives en utilisant des données disjointes composées de phrases extraites des bases de données de l'Aéroport et Babble. La méthode ML, MAP et MMSE sont comparées entre elles et la parole est corrompue avec le bruit de chahut et aéroport. Les meilleures performances sont indiquées en Gras.

Mesure objective subjective	**Entrée SNR (dB)**	**Bruit aéroport**			**Bruit Babble**		
		ML	**MAP**	**MMSE**	**ML**	**MAP**	**MMSE**
SIG	**0**	**2.5911**	2.5280	2.0062	**2.5114**	2.4818	1.8772
	5	**3.0715**	2.9942	2.6171	**3.0461**	2.9824	2.4554
	10	**3.5755**	3.4884	3.0893	**3.5529**	3.4677	3.1102
BAK	**0**	**1.7187**	**1.7201**	1.4390	**1.6797**	1.6961	1.3298
	5	**2.1202**	2.0875	1.9457	**2.1079**	2.0823	1.8283
	10	**2.5632**	2.4965	2.3648	**2.5413**	2.4776	2.3818
COVL	**0**	**2.0565**	2.0007	1.5146	**1.9924**	**1.9529**	1.4175
	5	**2.4746**	2.4034	2.0765	**2.4569**	**2.3948**	1.9508
	10	**2.9155**	2.8370	2.5145	**2.8996**	**2.8219**	**2.5320**
WSS	**0**	72.3242	**69.4544**	**111.5998**	71.4111	**66.9383**	110.5484
	5	56.5414	**54.5396**	**96.1641**	56.5752	**54.3479**	96.2712
	10	42.1481	**41.6801**	**78.6838**	42.3676	**41.8706**	78.4880
LIR	**0**	**0.8934**	0.9431	1.1619	**0.9477**	0.9739	1.2389
	5	**0.7261**	0.7748	0.9800	**0.7448**	0.7844	1.0514
	10	**0.5409**	0.5874	0.8098	**0.5561**	0.6007	0.8285
PESQ	**0**	**1.7716**	1.7090	1.6890	**1.7186**	1.6474	1.6695
	5	**2.0474**	1.9725	2.0165	**2.0376**	1.9662	2.0066
	10	**2.3522**	2.2802	2.3252	**2.3440**	2.2714	2.3164
SNR sortie	**0**	0.2057	1.0062	**1.3157**	0.0677	1.0250	**1.2817**
	5	4.8847	**4.9531**	3.4341	4.8650	**4.9767**	3.6088
	10	**9.5637**	8.8580	6.8397	**9.5673**	8.8278	6.3299
SNR seg	**0**	-4.0614	-3.8829	**-2.9000**	-4.3805	-4.0762	**-3.1362**
	5	-1.5340	-1.7082	**-1.3476**	-1.6512	-1.7643	**-1.5087**
	10	**1.5850**	1.0219	1.2193	**1.3243**	0.8079	0.7423

Dans le tableau 3 (bruit aéroport et bruit babble) nous pouvons remarquer que la méthode 1 est la meilleure méthode par rapport autres en terme des SIG, BAK, OVL, PESQ, LLR. En termes de WSS la méthode 2 est la meilleure.

Quand snr est 10, la méthode 3 est la meilleure.

Pour snr sortie 0 db la meilleure méthode est la méthode 3.
5 db la meilleure méthode est la méthode 2.
10 db la meilleure méthode est la méthode 1.

- Le problème : oui il supprime le bruit par contre la qualité, inéligibilité de parole, perception sont dégradée.

Tableau 4 : Evaluation objective en utilisant des données disjointes composées de phrases extraites des bases de données voiture et salle d'exposition. La méthode ML, MAP et MMSE sont comparées entre elles et la parole est corrompue avec le bruit de voiture et salle d'exposition. Les meilleures performances sont indiquées en Gras.

Mesure Objective subjective	**Entrée SNR (dB)**	**Bruit voiture**			**Bruit salle d'exposition**		
		ML	**MAP**	**MMSE**	**ML**	**MAP**	**MMSE**
SIG	**0**	**2.3869**	2.3198	1.7324	**2.1592**	**2.1354**	1.3307
	5	**2.9422**	2.8226	2.6348	**2.7722**	**2.7023**	2.3352
	10	**3.5529**	3.4677	3.1102	**3.2905**	**3.1995**	3.0231
BAK	**0**	**1.6886**	**1.6820**	1.4096	1.6426	**1.6486**	1.2305
	5	**2.0746**	**2.0178**	1.9465	**2.0654**	**2.0282**	1.9114
	10	**2.5413**	**2.4776**	2.3818	**2.4844**	**2.4186**	2.3803
COVL	**0**	**1.9159**	1.8402	1.3474	**1.7508**	**1.7023**	0.9986
	5	**2.3614**	2.2474	2.0938	**2.2545**	**2.1809**	1.8725
	10	**2.8996**	2.8219	2.5320	**2.7014**	**2.6130**	2.4714
WSS	**0**	66.8469	**63.2884**	107.5057	66.3856	**61.3746**	113.9838
	5	53.6318	**52.2632**	92.1369	53.4748	**50.6990**	95.3357
	10	42.3676	**41.8706**	78.4880	42.6970	**41.0009**	79.0730
LIR	**0**	**1.0786**	1.1153	1.2090	**1.2423**	1.2520	**1.3564**
	5	**0.8121**	0.8731	0.9582	**0.9523**	0.9913	**1.0800**
	10	**0.5561**	0.6007	0.8285	**0.7306**	0.7788	**0.9495**
PESQ	**0**	**1.6672**	1.5656	1.6571	**1.5621**	1.4646	1.4564
	5	**1.9362**	1.8214	2.0408	1.8912	1.8004	**1.9239**
	10	**2.3440**	2.2714	2.3164	2.2116	2.1177	**2.2811**
SNR sortie	**0**	0.5563	**1.2426**	1.2147	0.0426	**0.9227**	0.7754
	5	**5.3698**	5.1132	3.7195	4.8619	**4.8704**	4.0306
	10	**9.5673**	8.8278	6.3299	**9.5594**	9.0110	6.4459
SNR seg	**0**	-4.5775	-4.2456	**-2.2248**	-4.5785	-4.2355	**-2.4648**
	5	-1.9702	-2.1058	**-0.5682**	-1.7846	-1.7719	**-0.3148**
	10	**0.9061**	0.4221	0.7498	0.9793	0.5665	**1.7419**

Dans le tableau 4 : (bruit voiture et bruit de la salle d'exposition) nous pouvons remarquer que la méthode 1 est la meilleure par rapport aux autres méthodes en terme de SIG, BAK, OVL, LLR. Par contre dans le tableau (bruit de la salle d'exposition) la méthode qui donne le meilleur résultat en termes d'intelligibilité et la qualité globale de la parole est MAP car il est le plus grand que les autres dans [0 et 5] dB et faible (gênant, mais pas désagréable). En termes de WSS, MAP est aussi la meilleure méthode.

En terme de PESQ, la meilleure méthode est la méthode ML dans toutes les entrée au bruit de voiture, mais le bruit de la salle d'exposition à l'entrée 5 et 10, la meilleure méthode est la méthode MMSE par contre à l'entrée 0 la méthode ML donne un meilleur résultat.

En termes de snr sortie, la meilleure méthode est la méthode MAP pour le bruit de voiture et le bruit de la salle d'exposition à l'entrée 0, la meilleure méthode est la méthode ML à l'entrée 5 et 10.

On peut conclure que quand le SNR sortie est plus important que SNR entré, le bruit diminue, il y a une amélioration, et vice versa.

Tableau 5 : Evaluation objective en utilisant des données disjointes composées de phrases extraites des bases de données de la rue et le restaurant. La méthodes ML,MAP et MMSE sont comparées entre elles et la parole est corrompue avec le bruit de rue et Restaurant. Les meilleures performances sont indiquées en Gras.

Mesure objective subjective	**Entrée SNR (dB)**	**Bruit rue**			**Bruit restaurant**		
		ML	**MAP**	**MMSE**	**ML**	**MAP**	**MMSE**
SIG	0	**2,4748**	2,2052	2,1901	2,5461	**2,6387**	2,2472
	5	2,9586	**2,9685**	2,6600	**3,2047**	3,2062	2,9330
	10	**3,6940**	3,6062	3,1777	**3,6873**	3,6374	3,3509
BAK	0	**1,8303**	1,6638	1,7108	1,7005	**1,8222**	1,6355
	5	2,1401	**2,1467**	2,0063	2,1782	**2,2093**	2,0800
	10	**2,6485**	2,5935	2,3874	**2,6008**	2,5659	2,4749
COVL	0	**2,0191**	1,6638	1,7973	1,9666	**2,1032**	1,8260
	5	2,3864	**2,3924**	2,1963	2,5990	**2,6017**	2,4198
	10	**3,0331**	2,9520	2,6608	**3,0168**	2,9756	2,8117
WSS	0	52,0616	**46,9030**	96,9962	57,1969	**52,3816**	96,9576
	5	44,8167	**41,2977**	81,4404	47,9166	**43,3861**	83,3214
	10	34,7744	**32,6470**	74,0959	37,8304	**35,4160**	66,9894
LIR	0	**1,1480**	1,1835	1,2523	**0,9468**	0,9899	1,1597
	5	**0,8671**	0,8791	0,9735	**0,7143**	0,7416	0,9338
	10	**0,5366**	0,5908	0,8367	**0,5101**	0,5538	0,7903

PESQ	0	**1,7109**	1,2473	1,6603	1,5623	1,7177	**1,7348**
	5	1,9256	1,9101	**1,9781**	**2,1194**	2,1008	1,9521
	10	**2,4314**	2,3466	2,3802	**2,4208**	2,3765	2,3644
SNR sortie	0	0,3231	1,1101	**1,8069**	-0,3390	0,7865	**1,6804**
	5	4,4787	**4,7843**	4,6659	4,3097	4,7536	**4,8759**
	10	**9,6431**	9,2634	6,5790	**9,3210**	8,9418	6,6549
SNR seg	0	-4,0811	-3,8974	**-2,2392**	-4,4424	-4,2252	**-2,5416**
	5	-1,5966	-1,7665	**-1,0678**	-2,1188	-1,9870	**-0,8287**
	10	**1,5185**	1,0531	1,2918	1,1822	0,6958	**1,5376**

Dans le tableau 5 : nous pouvons remarquer que la méthode qui donne le meilleur résultat en termes d'intelligibilité et la qualité globale de la parole est ML car il est le plus grand que les autres, et subjectivement il est moyen (perceptible et un peu gênant). Ainsi en terme LIR elle est la meilleure méthode, car il a le plus petit résultat parmi les autres dans tous les RSB [0, 5 et 10] dB.

Par contre WSS la meilleure méthode est le MAP, car il a le plus petit résultat parmi les autres dans tous les RSB [0, 5 et 10] dB.

Nous avons remarqué aussi en termes de signal de parole propre est légèrement naturel et légèrement dégradé, ainsi, il est assez visible dans 10dB et 5dB et très visible dans 0dB.

Nous pouvons remarquer dans le tableau (bruit rue) que la méthode ML a fourni de meilleurs résultats par rapport aux autres méthodes utilisées pour les tests.

Nous pouvons remarquer dans le tableau (bruit restaurant) que la méthode MAP a fourni de meilleurs résultats par rapport aux autres méthodes utilisées pour les tests.

Tableau 6 : Evaluation objective en utilisant des données disjointes composées de phrases extraites des bases de données Train et Blanc. Les méthodes ML, MAP et MMSE sont comparées entre elles et la parole est corrompue avec le bruit de train et bruit blanc. Les meilleures performances sont indiquées en Gras.

Mesure Objective subjective	**Entrée SNR (dB)**	**Bruit train**			**Bruit blanc**		
		ML	**MAP**	**MMSE**	**ML**	**MAP**	**MMSE**

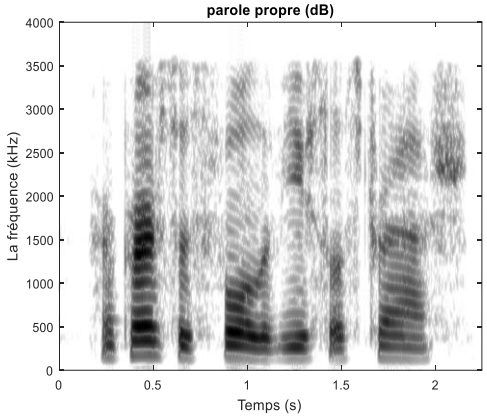
SIG	0	**2,3038**	2,2749	1,9523	1,5693	**1,6168**	1,5157
	5	**2,8325**	2,8157	2,5754	2,1208	2,1766	**2,2162**
	10	**3,378**	3,3076	3,0183	2,6937	**2,7308**	2,7118
BAK	0	1,7752	**1,7795**	1,5846	1,6116	**1,6414**	1,5805
	5	**2,1249**	2,1191	1,9683	1,9760	**1,9882**	1,9812
	10	**2,5113**	2,4838	2,3617	**2,3911**	2,3696	2,3699
COVL	0	**1,8843**	1,8415	1,5449	1,4464	**1,4493**	1,3954
	5	**2,3044**	2,2817	2,1143	1,8971	1,9043	**1,9812**
	10	**2,7372**	2,7112	2,5213	2,3793	2,3764	**2,4263**
WSS	0	51,4073	**47,4527**	89,1859	71,4577	**65,6620**	106,2950
	5	42,1912	**38,9774**	75,1872	63,1299	**57,6049**	93,6826
	10	35,9209	**32,6256**	67,3517	53,0249	**48,4191**	81,3298
LIR	0	**1,2605**	1,2785	1,3386	1,7733	1,7369	**1,6280**
	5	**0,9825**	1,0008	1,0522	1,4906	1,4446	**1,3392**
	10	**0,7539**	0,7742	0,9588	1,2215	**1,1873**	1,1984
PESQ	0	**1,6093**	1,5332	1,5470	1,5659	1,4960	**1,8082**
	5	**1,8743**	1,8298	1,8677	1,8736	1,8051	**2,2183**
	10	2,2120	2,1640	**2,2910**	2,2135	2,1482	**2,5010**
SNR sortie	0	0,1834	1,1475	**2,0714**	1,2777	2,0579	**2,7535**
	5	4,6579	**4,9458**	4,0087	5,9771	**6,0977**	5,9399
	10	**9,5781**	9,1731	6,9055	**10,5951**	9,9156	8,0768
SNR seg	0	-4,2569	-4,0508	**-2,4717**	-4,2965	-3,9376	**1,7398**
	5	-1,7415	-1,8518	**-0,8113**	-1,7719	-1,6729	**0,1910**
	10	1,1329	0,6953	**1,2110**	1,1149	0,7579	**2,0401**

Dans le tableau 6 : nous pouvons remarquer que la méthode qui donne le meilleur résultat en termes d'intelligibilité et la qualité globale de la parole c'est bien MMSE car il est le plus grand que les autre, faible (gênant, mais pas désagréable), En terme LLR elle est aussi la meilleure méthode.

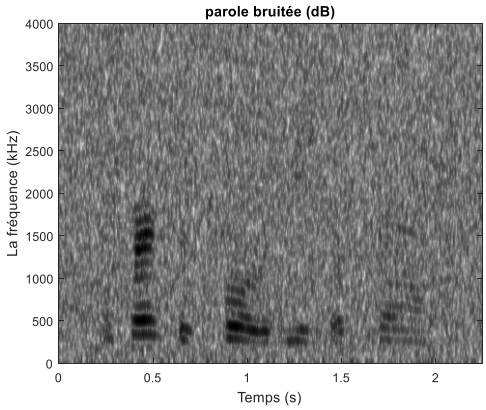
Par contre WSS la meilleure méthode c'est le MAP, car il donne le plus petit résultat parmi les autres dans tous les RSB [0, 5 et 10] dB.

Nous avons remarqué aussi en terme de signal de parole propre est très peu naturel, très peu dégradé dans 0 dB, et assez peu naturel, assez dégradé dans [5 et dB10], pareillement assez peu naturel et dégradé dans 10 dB, ainsi qu'il est très visible dans [0 et 5] dB et visible dans 10 dB.

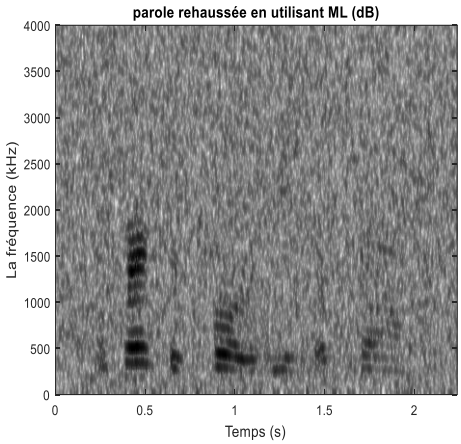
Nous pouvons remarquer dans le tableau que la méthode MAP a fourni de meilleurs résultats par rapport aux autres méthodes utilisées pour les tests.



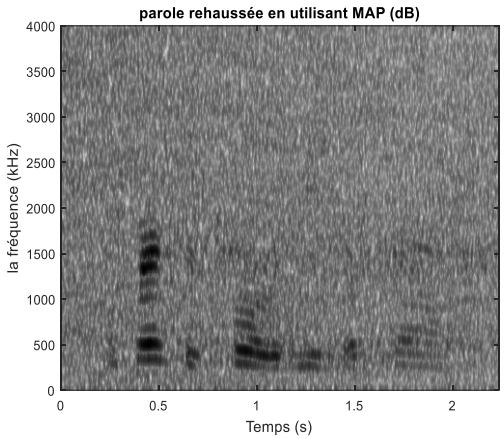
(a)



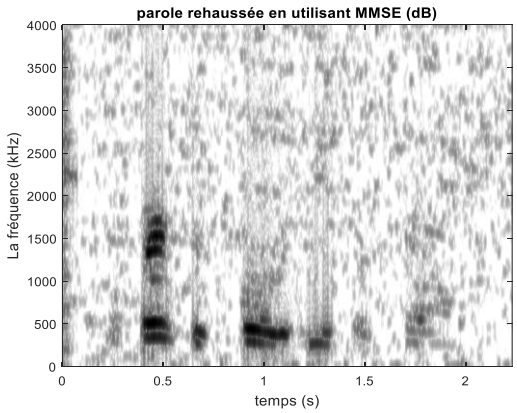
(b)



(c)

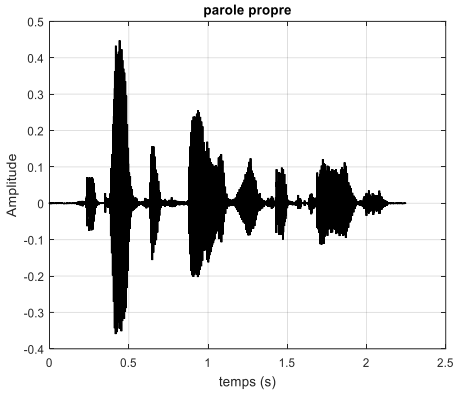


(d)

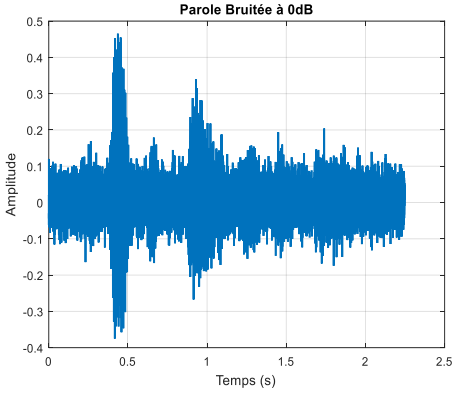


(e)

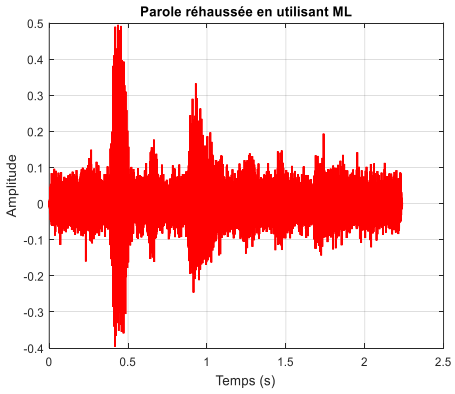
Figure 7 : Spectrogrammes, (a) parole propre, (b) parole bruitée, (c) parole rehaussée en utilisant la méthode ML, (d) parole rehaussée en utilisant la méthode MAP, (e) parole rehaussée en utilisant la méthode MMSE. Cas du bruit blanc 0 dB.



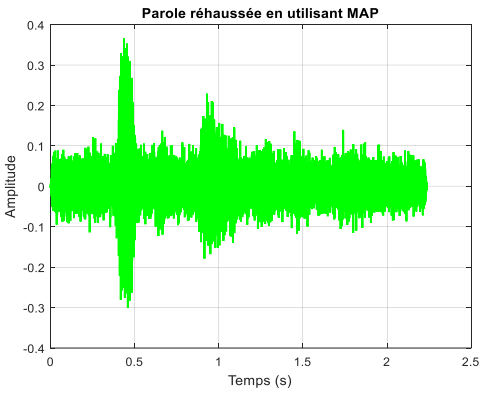
(a)



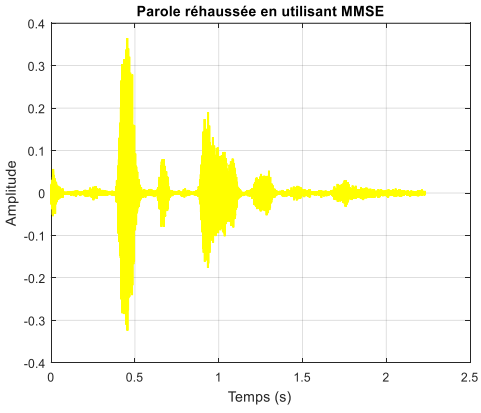
(b)



(c)



(d)



(e)

Figure 8 : (a) signal parole propre, (b) signal parole bruitée, (c) signal parole rehaussée en utilisant la méthode ML, (d) signal parole rehaussée en utilisant la méthode MAP, (e) signal parole rehaussée en utilisant la méthode MSSE. Cas du bruit blanc 0 dB.

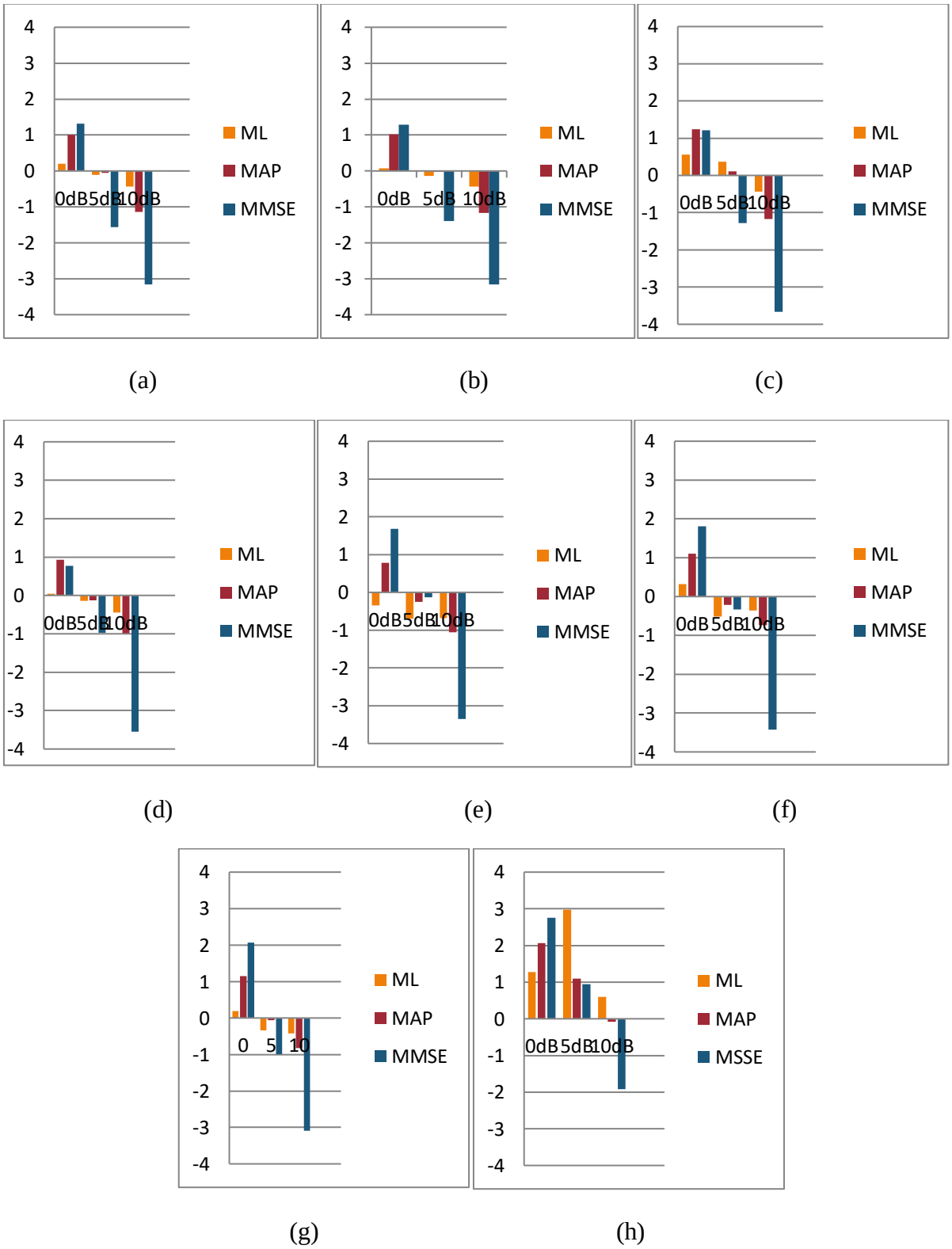


Figure 9 : Évaluation SNR amélioré pour les méthodes ML, MAP et MSSE en utilisant : (a) bruit d'aéroport, (b) bruit babble de NOIZEUS, , (c) bruit de véhicule de NOIZEUS, (d) bruit salle d'exposition de NOIZEUS, (e) bruit restaurant de NOIZEUS, (f) bruit rue de NOIZEUS, (g) bruit train de NOIZEUS, (h) bruit blanc de NOIZEUS.

Cet histogramme représente SNRI (amélioré) ($SNR_{entré} = 0\,,5\ et\ 10$)

$$SNR_I = SNR_{sortie} - SNR_{entré}$$

Si $SNR_{sortie} - SNR_{entré}$ est positive, il y a une amélioration ainsi le bruit est diminué.

Par contre si le résultat est négatif, il n'y a pas d'amélioration, et le bruit est toujours présent voir même augmenté dans certains cas.

3.4. Conclusion

L'une des caractéristiques des méthodes probabilistiques de soustraction de bruit c'est qu'elles travaillent et fonctionnent bien quand le SNR est faible quand le milieu est fortement bruité, donc elles donnent les meilleurs résultats, le bruit est diminué en maximum. Par contre, quand le milieu commence à devenir faiblement bruité, c'est-à-dire quand le snr est augmenté, le bruit est diminué, donc au lieu d'éliminer le bruit elles éliminent la parole .

La meilleure méthode est la methode ML. La robustesse de méthode a été confirmée par les résultats de l'évaluation objective et subjectives, reportés sur les différents tableaux et spectrogrammes en rapport avec les signaux propres, bruités et rehaussés.

Conclusion générale

Conclusion générale

Une application capable de supprimer le bruit dans la parole en utilisant la méthode de soustraction spectrale a été implémentée avec succès dans ce mémoire. Trois méthodes probabilistes (ML, MAP et MMSE) ont été mises en œuvre, et leurs performances ont été examinées, pour décider laquelle est la meilleure. Comme mentionné dans le troisième chapitre, toujours de notre point de vue, la mesure de la qualité objective est la plus importante parmi les mesures d'évaluation. En conclusion, pour nous, les meilleurs résultats sont obtenus avec l'Algorithme de MMSE et mélange de voix avec du bruit blanc. L'algorithme de MMSE a obtenu une valeur d'environ 1.8 dans l'échelle PESR (MOS) pour un SNR=0 dB, dont la note la plus élevée est 2.5 pour SNR=10dB. Cette valeur est de 0.3 points au-dessus de la pire méthode (MAP), et ça sous le bruit blanc.

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 programmer et simuler une application pour l'évaluer. En outre, une étude et une recherche sur la façon dont il 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 pourrait être étudiée en profondeur afin d'améliorer les résultats des algorithmes étudiés. 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. Comme un résultat, le processus ultérieur n'est pas effectué avec les valeurs optimales car elles dépendent de cette partie du processus pour estimer le bruit et reconnaître si la voix est ou pas.

D'un point de vue personnel, la réalisation de ce projet a supposé l'entrée pour nous à un nouveau domaine de travail, avec des circonstances et des exigences diverses, 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 appris à nous organiser dans notre emploi du temps, car pour combiner le recherche et 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. Utiliser correctement, tous les mots techniques sur un grand nombre de pages n’est pas quelque chose facile pour quelqu'un dont la langue maternelle est assez différente de celle utilisée.

Liste des références

[1] SF Boll, Suppression of acoustic noise in speech using spectral subtraction [J]. IEEE Trans. Acoust., Speech Signal Process 27(2), 113–120 (1979)

[2] DL Donoho, De-noising by soft-thresholding [J]. IEEE Trans. Inf. Theory 41, 613–627 (1995)

[3] Y Ephraim, HL Van Trees, A signal subspace approach for speech enhancement. IEEE Trans. Speech Audio Process. 3(4), 251–266 (1995)

[4] N Virag, Single channel speech enhancement based on masking properties of the human auditory system [J]. IEEE Trans. Speech Audio Process. 7(2), 126–137 (1999).

[5] Y Ephraim, D Malah, Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator [J]. IEEE Trans. Acoust. Speech Signal Process 32, 1109–1121 (1984).

[6] Mathieu lagrange. Modélisation long terme de signaux sonores. Machine learning [stat.ML]. Université de nantes, 2019.page 31.

[7] Andrei Doncescu. Université Paul Sabatier. Les bases du traitement des signaux numériques. Page 11.

[8] Marc van. droogenbroeck. 2004-06-15 telecom. ulg.ac.be/teaching /notes /multimedia / node23.

[9] Emilio parrado hernández. Andrea garcía gonzález. Noise cancelling in acoustic voice signals with spectral subtraction. Universidad carols III de Madrid. February 2014. Page 4 et 6.

[10] Avarado, D.: “Efecto del enventanado en la obtención del espectro discrete de una señal”. Universidad técnica particular de loja. Ecuador. 2005.

[11] Gabriel Cormier, Ph.D., ing. Université de moncton. Transformée de fourrier discret. Hiver 2013. Chapitre 7 pages 3 et 27.

[12] Raphael, l. (2011). Speech science primer: physiology, acoustics, and perception of speech. Wolters kluwer health/lippincott williams wilkins, Baltimore,md.

[13] Hassan Ezzaidi. Université du québec à chicoutimi. Discrimination parole/musique et étude de nouveaux paramètres et modèles pour un système d'identification du locuteur dans le contexte de conférences téléphoniques. 4 Octobre 2002. Page 9.

[14] MARIPO tsivery tanjona. Univerite d'antananarivo ecole supérieure polytechnique. Traitement du signal et la reconnaissance de la parole. 23 Avril 2010. Page 14.

[15] https://media.kartable.fr/uploads/finalImages/final_561e71091e7888.76948539.png

[16] https://www.f-legrand.fr/scidoc/figures/numerique/tfd/spectre2/figA.png.

[17] Christophe doignon. Université louis pasteur de strasbourg. Traitement du signal. (2008-2009). Page 21.

[18] Auffret M., Cours magistral de législation, diplôme universitaire nuisances sonores, faculté de pharmacie de nancy, 2014/2015.

[19] Vaseghi, S. V.: "Advanced digital signal processing and noise reduction". John wiley & Sons, LTD. Second edition. 2000.

[20] M. elle amirouch nadia. Université mouloud ammeri de tizi-ouzou. La compatibilité électromagnétique en électronique. « promotion 2017/2018.page 6/7.

[21] Bruno vincent, directeur, docteur en psychologie de l'environnement vincent gissinger, chargé de mission observatoires du bruit. Mai 2011 - révision 2.page 5.

[22] Alexandre boyer. Canaux de transmissions bruits. Institut national des sciences appliquees de toulouse. Septembre 2011.

[23] Ducourneau J., Cours magistral d'acoustique, diplôme universitaire nuisances sonores, faculté de pharmacie de nancy, 2014/2015 : niveaux sonores, analyse spectrale.

[24] Rhône à Pougny – Mesures acoustiques préliminaires. décembre 2014.page 3.

[25] https://www.coopacou.com/fichiers/ACOUPHENE/bruits-colores.png.

[26] http://aplus-development.com/apluseduc/images/upload/1558090065image037.png.

[27] G.Baillargeon, " introduction aux méthode statistique en control de la qualité ", livre d'Applications dans divers secteurs industriels (Traitement de données avec Excel V.2007-2010-2013) pp. 235-236.

[28] McAulay, R. J. and Malpass, M. L. (1980), Speech enhancement using a soft-decision noise suppression filter, IEEE Trans. Acoust. Speech Signal Proces. 28, 137-145.

[29] O. Besson ,"Estimation en traitement du signal", institut supérieur de l'aéronautique et de l'espace (2006).

[30] J. S. Lim, A. V. Oppenheim, Enhancement and Bandwidth Compression of Noisy Speech, In Proc of the IEEE. 67(12) (1979) 1586-1604.

[31] Y. Ephraim, D. Malah, Référence précédente ,1110-1126.

[32] R. Martin, Speech Enhancement Using A MMSE Short Time Spectral Estimation With Gamma Distributed Speech Priors, In Proc. IEEE. ICASSP. (2002) 253-256.

[33] R. Martin, Speech Enhancement Base on MMSE and Super-Gaussian Priors, IEEE. Trans. on SAP. 13(5) 845-856.

[34] T. Lotter, P. Vary, Speech Enhancement by Maximum A Posteriori Spectral Amplitude

Estimation Using A Supergaussian Speech Model, EURASIP, J. Appl. SP 5(7) (2005).

[35] Porter, J. and Boll, S. E (1984), Optimal estimators for spectral restoration of noisy speech, Proceedings of TEEE International Conference on Acoustics, Speech, and Signal Processing, San Diego, CA, pp. 18A.2.1-18A.2.4.

[36] Papoulis, A. and Pillai, S. (2002), Probability, Random Variables and Stochastic Processes, 4th ed., New York: McGraw-Hill P (203).

[37] P. Wolf, S, Godsill, Simple Alternative To Ephraim And Malah Suppression Rule For Speech Enhancement, Proc. 11th IEEE. Workshop. Stat. SP (2001) 201-204.

[38] Walden, A., Percival, B., and McCoy, E. (1998), Spectrum estimation by wavelet thresh-olding of multitaper estimators, IEEE Trans. Signal Process., 46, 3153-3165.

[39] SAADOUNE Adda , " rehaussment de la parole par les méthode PCA –VRE " ,THESE Présentée pour l'obtention du grade de Doctorat En Sciences enElectronique ,Décembre, 2014, page 87.

[40] Y. Hu, P.C. Loizou, NOIZEUS: A Noisy Speech Corpus For Evaluation Of Speech Enhancement Algorithms. Available at: http://www.utdallas.edu/~loizou/speech/noizeus/