UNIVERSITE HASSIBA BENBOUALI DE CHLEF

Faculté de Technologie

Département d'Electronique

MEMOIRE DE MASTER

Domaine : SCIENCES ET TECHNOLOGIES

Filière : : TELECOMMUNICATIONS

Spécialité : SYSTEMES DES TELECOMMUNICATIONS

ESTIMATION DU BRUIT DANS LE SIGNAL DE PAROLE : ALGORITHMES ET PERFORMANCES

Par

Djemia BENNOUA

Encadreur :

M. KADDAI Abdellah

Maître de Conférences «B» à l'UHBC

Chlef, Juin 2024

Dédicace

Grace à l'aide et succès de Allah tout -puisant,

J'ai achevé ce travail et avec toute mon

Modestie et mon respect

Je le dédie à :

Mes deux personnes les plus les chères et les plus aimées

Ma mère DAKOUR Omelkhire et mon père BENNOUA

Boulame, qui ont été ma source de force

A mon frère Abdelkader et mes sœurs surtout ma sœur aînée

Et à son mari qui m'ont beaucoup soutenu

A tous mes chères amis et collègues

Djamia

Remerciements

Tout d'abord, je tiens à exprimer mes sincères remerciements à Allah Tout Puissant, qui m'a été d'une aide et m'a donné patience, force et succès.

J'adresse également mes remerciements à mon encadrant pour sa compréhension, sa coopération et sa gentillesse à mon égard.

Je remercie également ma famille qui m'a toujours encouragé, notamment ma mère et mon père.

J'adresse également mes remerciements à tous mes enseignants et professeurs du parcours académique ainsi qu'à tous mes collègues et amis.

Merci à tous ceux qui m'ont aidé et encouragé et qui est prié Allah pour moi.

الملخص

يتسبب تلف الكلام بسبب وجود ضوضاء إضافية في الخلفية في حدوث صعوبات خطيرة في بيئات الاتصال المختلفة. تناولت هذه المذكرة مسألة تقدير الضوضاء لتحسين الكلام أحادي القناة. في معظم أساليب تحسين الكلام، من المفترض أن يتوفر تقدير لطيف الضوضاء. يعد هذا التقدير ضروريًا لأداء خوارزميات تحسين الكلام. يمكن أن يكون لتقدير الضوضاء تأثير كبير على جودة الإشارة المحسنة. إذا كان تقدير الضوضاء منخفضًا جدًا، فستكون الضوضاء المتبقية مزعجة للسمع، بينما إذا كان تقدير الضوضاء مرتفعًا جدًا، فسيتم تشويه الكلام، مما قد يؤدي إلى فقدان الوضوح. إن أبسط طريقة هي تقدير وتحديث طيف الضوضاء أثناء المقاطع الصامتة (على سبيل المثال أثناء التوقف المؤقت) للإشارة باستخدام خوارزمية الكشف عن النشاط الصوتي. على الرغم من أن هذا النهج قد يعمل بشكل مرض في الضوضاء الثابتة (على سبيل المثال، الضوضاء البيضاء)، فإنه لن يعمل بشكل جيد في بيئات أكثر واقعية (على سبيل المثال، في مطعم) حيث قد تتغير الخصائص الطيفية للضوضاء باستمرار. ولذلك فمن الضروري تحديث طيف الضوضاء بشكل مستمر مع مرور الوقت، وهو ما يمكن تحقيقه باستخدام خوارزميات تقدير الضوضاء. تمت دراسة فئتين من هذه الخوارزميات في هذه المذكرة. أظهرت نتائج المحاكاة فعالية الخوارزميات في فئة المتوسط المعاود.

Résumé

La corruption de la parole due à la présence d'un bruit de fond additif entraîne de graves difficultés dans divers environnements de communication. Ce mémoire a abordé la question de l'estimation du bruit pour le rehaussement de la parole bruitée à canal unique. Dans la plupart des méthodes de rehaussement de la parole, on suppose qu'une estimation du spectre du bruit est disponible. Une telle estimation est essentielle et nécessaire pour les performances des algorithmes de rehaussement de la parole. L'estimation du bruit peut avoir un impact majeur sur la qualité du signal amélioré. Si l'estimation du bruit est trop faible, un bruit résiduel gênant sera audible, tandis que si l'estimation du bruit est trop élevée, la parole sera déformée, entraînant éventuellement une perte d'intelligibilité. L'approche la plus simple consiste à estimer et à mettre à jour le spectre de bruit pendant les segments silencieux (par exemple pendant les pauses) du signal à l'aide d'un algorithme de détection d'activité vocale. Bien qu'une telle approche puisse fonctionner de manière satisfaisante dans un bruit stationnaire (par exemple, le bruit blanc), elle ne fonctionnera pas bien dans des environnements plus réalistes (par exemple, dans un restaurant) où les caractéristiques spectrales du bruit peuvent changer constamment. Il est donc nécessaire de mettre à jour le spectre du bruit en permanence au fil du temps, ce qui peut être réalisé à l'aide d'algorithmes d'estimation du bruit. Deux catégories de ces algorithmes ont été étudiées dans ce mémoire. Les résultats de simulations ont montré l'efficacité des algorithmes de la catégorie à moyenne récursive.

Abstract

Speech corruption due to the presence of additive background noise causes serious difficulties in various communication environments. This master thesis addressed the issue of noise estimation for single-channel noisy speech enhancement. In most speech enhancement methods, it is assumed that an estimate of the noise spectrum is available. Such estimation is essential and necessary for the performance of speech enhancement algorithms. Noise estimation can have a major impact on the quality of the enhanced signal. If the noise estimate is too low, annoying residual noise will be audible, while if the noise estimate is too high, speech will be distorted, possibly leading to loss of intelligibility. The simplest approach is to estimate and update the noise spectrum during silent segments (e.g. during pauses) of the signal using a voice activity detection algorithm. Although such an approach may work satisfactorily in stationary noise (e.g., white noise), it will not work well in more realistic environments (e.g., in a restaurant) where the spectral characteristics of the noise may change constantly. It is therefore necessary to update the noise spectrum continuously over time, which can be achieved using noise estimation algorithms. Two categories of these algorithms were studied in this master thesis. Simulation results showed the effectiveness of algorithms in the recursive average category.

Table des matières

INTRODUCTION GENERALE 1

CHAPITRE 1 : GENERALITES SUR LE SIGNAL DE PAROLE 4

1.1 NIVEAU PHONETIQUE **5**

1.2 NIVEAU ACOUSTIQUE **7**

1.3 PROPRIETES STATISTIQUES D'UN SIGNAL DE PAROLE **8**

1.4 MODELE NUMERIQUE DE PRODUCTION DE LA PAROLE **11**

1.5 GENERALITES SUR LE BRUIT **13**

1.5.1 DEFINITION DU BRUIT 13
1.5.2 CLASSIFICATION DES TYPES DE BRUIT EN FONCTION DE LEUR NATURE 13
1.5.2.1 Bruit physique 13
1.5.2.2 Bruit Psychologique 13
1.5.2.3 Distorsions de canal, écho et évanouissement 14
1.5.3 CLASSIFICATION SELON SA FREQUENCE OU SON TEMPS CARACTERISTIQUES 14
1.5.3.1 Bruit blanc 14
1.5.3.2 Bruits colorés 14
1.5.3.3 Bruit impulsif 15

1.6 TECHNIQUES D'ANALYSE **16**

1.6.1 ANALYSE TEMPORELLE A COURT TERME 16
1.6.2 ANALYSE FREQUENTIELLE A COURT TERME 16
1.6.3 FONCTIONS DE FENETRAGE 16
1.6.4 ANALYSE DE FREQUENCE 18
1.6.5 TRANSFORMEE DE FOURIER (TF) 18
1.6.6 TRANSFORMEE DE FOURIER A TEMPS DISCRET (TFTD) 18
1.6.7 TRANSFORMEE DE FOURIER A COURT TERME 19
1.7 CONCLUSION 19

CHAPITRE 2 : ALGORITHMES D'ESTIMATION DU BRUIT 20

2.1 SOUSTRACTION SPECTRALE 21

2.2 ALGORITHMES D'ESTIMATION DE BRUIT 26

2.2.1 ALGORITHMES DE SUIVI DES MINIMA 27
2.2.1.1 Algorithme de statistiques minimales (MS) 28
2.2.1.2 Algorithme de suivi continu des minima spectraux 31
2.2.2 ALGORITHMES D'ESTIMATION DU BRUIT A MOYENNE RECURSIVE 31
2.2.2.1 Moyenne récursive contrôlée des minima (MCRA) 32
2.2.2.2 Algorithme MCRA modifiée (MCRA 2) 35

CHAPITRE 3 : SIMULATIONS ET RESULATS 38

3.1 MESURES OBJECTIVES D'EVALUATIONS 39

3.1.1 RAPPORT SIGNAL SUR BRUIT GLOBAL ET SEGMENTAL 40
3.1.2 PESQ (PERCEPTUAL EVALUATION OF SPEECH QUALITY) 41
3.1.3 MESURE D'ITAKURA-SAITO IS 42
3.1.4 LIKELIHOOD LINEAR REGRESSION (LLR) 43
3.2 BASE DONNEES 43

3.3 RESULTATS DES TESTS 44

3.4 CONCLUSION 52

CONCLUSION GENERALE 53

BIBLIOGRAPHIE 55

Liste des Figures et des Tableaux

FIGURE 1.1 : APPAREIL PHONATOIRE HUMAIN. 5
FIGURE 1.2 : EXEMPLE D'UN SIGNAL DE PAROLE VOISE ET NON VOISE. 7
FIGURE 1.3 : DENSITE SPECTRALE DE PUISSANCE EN ECHELLE LOGARITHMIQUE D'UN SON VOISE. .. 9
FIGURE 1.4 : DENSITE SPECTRALE DE PUISSANCE EN ECHELLE LOGARITHMIQUE D'UN SON NON VOISE. 10
FIGURE 1.5 : AUTOCORRELATION D'UN SON VOISE (A GAUCHE) ET NON VOISE (A DROITE). 10
FIGURE 1.6 : LA FREQUENCE FONDAMENTALE [1]. 11
FIGURE 1.7 : MODELE NUMERIQUE DE PRODUCTION DE LA PAROLE. 12
FIGURE 1.8 : SPECTRE D'UN BRUIT BLANC. 14
FIGURE 1.9 : EXEMPLE DES BRUITS COLORES. 15
FIGURE 1.10 : EXEMPLE DE BRUIT IMPULSIF. 15
FIGURE 1.11 : FENETRE DE HANNING. 17
FIGURE 2.1 : PRINCIPE DE LA SOUSTRACTION SPECTRALE. 23
FIGURE 2.2 : VALEURS DE α A EN FONCTION DU SNR. 25
FIGURE 3.1 : SYSTEME PESQ POUR L'EVALUATION DES PERFORMANCES D'UN REDUCTEUR DE BRUIT. 42

FIGURE 3.2 : FORMES D'ONDES DU SIGNAL PROPRE, DU SIGNAL BRUITE PAR UN BRUIT « CAR » (RSB = 0 DB) ET DU SIGNAL REHAUSSE EN UTILISANT L'ALGORITHME D'ESTIMATION DU BRUIT « DOBLINGER ».....44
FIGURE 3.3 : FORMES D'ONDES DU SIGNAL PROPRE, DU SIGNAL BRUITE PAR UN BRUIT « CAR » (RSB = 0 DB) ET DU SIGNAL REHAUSSE EN UTILISANT L'ALGORITHME D'ESTIMATION DU BRUIT « MARTIN (MS) ». 45
FIGURE 3.4 : FORMES D'ONDES DU SIGNAL PROPRE, DU SIGNAL BRUITE PAR UN BRUIT « CAR » (RSB = 0 DB) ET DU SIGNAL REHAUSSE EN UTILISANT L'ALGORITHME D'ESTIMATION DU BRUIT « MCRA ».....45
FIGURE 3.5 : FORMES D'ONDES DU SIGNAL PROPRE, DU SIGNAL BRUITE PAR UN BRUIT « CAR » (RSB = 0 DB) ET DU SIGNAL REHAUSSE PAR SOUSTRACTION SPECTRALE EN UTILISANT L'ALGORITHME D'ESTIMATION DU BRUIT « MCRA2 ».....46
FIGURE 3.6 : PERFORMANCES DES ALGORITHMES D'ESTIMATION DU BRUIT ETUDIES EN TERMES DU RSB EN FONCTION DU RSB D'ENTREE POUR UN BRUIT « CAR » DE LA BASE « NOIZEUS »...47
FIGURE 3.7 : PERFORMANCES DES ALGORITHMES D'ESTIMATION DU BRUIT ETUDIES EN TERMES DU RSB EN FONCTION DU RSB D'ENTREE POUR UN BRUIT « TRAIN » DE LA BASE « NOIZEUS ». 47
FIGURE 3.8 : PERFORMANCES DES ALGORITHMES D'ESTIMATION DU BRUIT ETUDIES EN TERMES DU RSB EN FONCTION DU RSB D'ENTREE POUR UN BRUIT « STREET » DE LA BASE « NOIZEUS ». 48
FIGURE 3.9 : PERFORMANCES DES ALGORITHMES D'ESTIMATION DU BRUIT ETUDIES EN TERMES DU RSB SEGMENTAL EN FONCTION DU RSB D'ENTREE POUR UN BRUIT « CAR » DE LA BASE « NOIZEUS ».....48
FIGURE 3.10 : PERFORMANCES DES ALGORITHMES D'ESTIMATION DU BRUIT ETUDIES EN TERMES DU RSB SEGMENTAL EN FONCTION DU RSB D'ENTREE POUR UN BRUIT « TRAIN » DE LA BASE « NOIZEUS ».....49
FIGURE 3.11 : PERFORMANCES DES ALGORITHMES D'ESTIMATION DU BRUIT ETUDIES EN TERMES DU RSB SEGMENTAL EN FONCTION DU RSB D'ENTREE POUR UN BRUIT « STREET » DE LA BASE « NOIZEUS ».....49
FIGURE 3.12 : PERFORMANCES DES ALGORITHMES D'ESTIMATION DU BRUIT ETUDIES EN TERMES DU PESQ EN FONCTION DU RSB D'ENTREE POUR UN BRUIT « CAR » DE LA BASE « NOIZEUS ».50
FIGURE 3.13 : PERFORMANCES DES ALGORITHMES D'ESTIMATION DU BRUIT ETUDIES EN TERMES DU PESQ EN FONCTION DU RSB D'ENTREE POUR UN BRUIT « TRAIN » DE LA BASE « NOIZEUS ». 51

FIGURE 3.14 : PERFORMANCES DES ALGORITHMES D'ESTIMATION DU BRUIT ETUDIES EN TERMES DU PESQ EN FONCTION DU RSB D'ENTREE POUR UN BRUIT « STREET » DE LA BASE « NOIZEUS ». 51

TABLEAU 1.1 : PROPRIETES DU BRUIT. 13

TABLEAU 3.1 : CLASSIFICATION DES CRITERES D'EVALUATION OBJECTIVE LES PLUS COMMUNEMENT UTILISES. 40

Introduction Générale

Une grande partie de l'interaction entre les humains se fait via la communication vocale. C'est pourquoi la recherche en sciences de la parole et de l'audition se poursuit depuis des siècles pour comprendre la dynamique et les processus impliqués dans la production et la perception de la parole. Le domaine du traitement de la parole est essentiellement une application des techniques de traitement du signal aux signaux acoustiques en utilisant les connaissances offertes par les chercheurs dans le domaine des sciences de l'audition. Les avancées explosives de ces dernières années dans le domaine de traitement du signal numérique ont donné un formidable élan au domaine du traitement de la parole. Les techniques de traitement du signal numérique sont plus sophistiquées et avancées que leurs homologues analogiques. La facilité et la rapidité de représentation, de stockage, de récupération et de traitement des données vocales ont contribué au développement de techniques de traitement de la parole efficaces et efficientes pour résoudre les problèmes liés à la parole.

La présence de bruit de fond dans la parole réduit considérablement l'intelligibilité de la parole. La dégradation de la parole affecte gravement la capacité d'une personne, qu'elle soit malentendante ou normale, à comprendre ce que dit l'orateur. Des algorithmes de réduction du bruit ou de rehaussement de la parole sont utilisés pour réduire ce bruit de fond et améliorer la qualité de perception et l'intelligibilité de la parole.

Le rehaussement de la parole a trouvé de nombreuses applications, notamment avec l'essor de la reconnaissance vocale automatique et des communications mobiles. Dans les systèmes de reconnaissance vocale automatique, les performances se dégradent considérablement dans le cas d'environnements défavorables avec un Rapport Signal à Bruit (RSB) très faible. Il a été constaté que le taux de reconnaissance peut être amélioré en appliquant un algorithme de rehaussement à la parole dégradée afin d'en améliorer l'intelligibilité. Egalement dans le cas des communications mobiles, le signal vocal est dégradé par différents types de bruit dans le canal de communication. Il existe donc un besoin pour un système de rehaussement de la parole au niveau du récepteur. De nombreux systèmes de rehaussement de la parole ont été développés sur la base des principes de soustraction spectrale et de filtrage de Wiener. Les caractéristiques communes de toutes ces méthodes sont l'estimation du spectre de puissance d'une parole propre en utilisant le spectre de puissance de la parole bruitée et du bruit. Dans les systèmes de rehaussement de la parole à canal unique, il n'y aura accès qu'à la parole bruitée et les statistiques de bruit doivent donc être estimées à partir de la parole bruitée elle-même.

Habituellement, l'estimation du spectre de bruit est obtenue à partir des premières millisecondes de parole bruitée qui sont des régions de silence. Cette hypothèse est valable pour le cas d'un bruit stationnaire (par exemple, bruit blanc) dans lequel le spectre du bruit ne varie pas beaucoup dans le temps. La mise à jour de l'estimation du bruit dans les détecteurs d'activité vocale traditionnels est limitée aux trames sans parole. Cela n'est pas suffisant dans le cas d'un bruit non stationnaire (environnement plus réaliste) dans lequel le spectre de puissance du bruit varie même pendant l'activité vocale. Il est donc nécessaire de mettre à jour le spectre du bruit en continu au fil du temps, ce qui est réalisé par des algorithmes d'estimation du bruit. Ce mémoire focalise sur des algorithmes adaptés aux environnements de bruit hautement non stationnaires.

Ce mémoire est présenté en trois chapitres :

- Dans le ***premier chapitre***, nous exposons des généralités sur le signal de parole ainsi que sur le rehaussement de la parole, nous aborderons également les variantes de la technique de soustraction spectrale.
- Au ***deuxième chapitre***, certains des algorithmes d'estimation du bruit existants sont expliqués en détail. Les avantages et les inconvénients de ces algorithmes sont discutés à la fin du chapitre.

- Le ***troisième chapitre*** fera l'objet de la valorisation des algorithmes d'estimation du bruit par l'évaluation de ses performances en utilisant plusieurs critères d'évaluation objectifs pour différents types et niveaux de bruit.
- Les principaux résultats obtenus sont résumés dans une conclusion générale.

Chapitre 1

Généralités sur le signal de parole

Au cours des dernières décennies, l'évolution technologique et scientifique a permis le développement considérable de nouveaux outils pour le traitement de la parole. De manière générale, le domaine des applications pratiques utilisant les résultats des chercheurs est de plus en plus large tels que la reconnaissance de la parole, la reconnaissance du locuteur, les systèmes de dialogue, la traduction automatique, le rehaussement de la parole connaissent un essor conséquent.

La recherche dans le domaine de rehaussement de la parole s'est concentrée sur la suppression du bruit de fond additif. Du point de vue du traitement du signal, le bruit additif est plus facile à gérer que le bruit convolutif ou les perturbations non linéaires. Le but ultime de l'amélioration de la parole est d'éliminer le bruit additif présent dans le signal vocal et de restaurer le signal vocal dans sa forme originale. Plusieurs méthodes ont été développées à la suite de ces efforts de recherche. La plupart de ces méthodes ont été développées avec certaines ou d'autres contraintes auditives, perceptuelles ou statistiques placées sur les signaux de parole et de bruit. Cependant, dans des situations réelles, il est très difficile de prédire de manière fiable les caractéristiques du signal de bruit ou les caractéristiques exactes de la forme d'onde vocale.

Par conséquent, en réalité, les méthodes de rehaussement de la parole ne sont pas optimales et ne peuvent réduire la quantité de bruit dans le signal que dans une certaine mesure. En raison de la nature sous-optimale de ces méthodes, une partie du signal vocal peut être déformée au cours du processus. Il existe donc un compromis entre les distorsions de la parole traitée et la quantité de bruit supprimée. L'efficacité du système de rehaussement de la parole peut donc être mesurée en fonction de ses performances à la lumière de ce compromis.

Ce chapitre présente une revue des notions fondamentales de la production de la parole, les caractéristique du ce signal, les technique d'analyse utilisés ainsi les différents types du bruit.

1.1 Niveau phonétique

Le signal de parole est complexe et possède de nombreuses caractéristiques qui permettent de le modéliser correctement. C'est un vecteur de communication naturelle aux êtres humains. Les principaux organes composant l'appareil phonatoire sont [1] : les poumons, la trachée artère, le larynx, le pharynx, les cordes vocales, la glotte et se termine par les cavités buccales et nasales (voir figure 1.1).

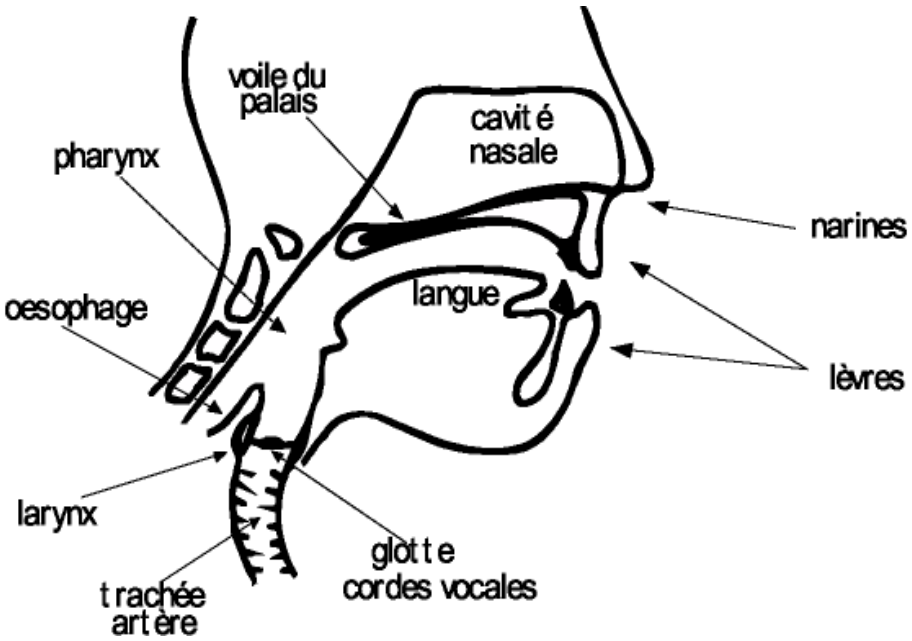

Figure 1.1 : Appareil phonatoire humain.

La parole peut être décrite comme le résultat de l'action volontaire et coordonnée d'un certain nombre de muscles. Cette action se déroule sous le contrôle du système nerveux central qui reçoit en permanence des informations par rétroaction auditive et par les sensations kinesthésiques.

Comme il présenté à la figure 1.1. L'appareil respiratoire fournit l'énergie nécessaire à la production de sons, en poussant de l'air à travers la trachée-artère. Au sommet de celle-ci se trouve le larynx ou la pression de l'air est modulée avant d'être appliquée au conduit vocal. Le larynx est un ensemble de muscles et de cartilages mobiles qui entourent une cavité située à la partie supérieure de la trachée .Les cordes vocales sont en fait deux lèvres symétriques placées en travers du larynx. Ces lèvres peuvent fermer complètement le larynx et, en s'écartant progressivement, déterminer une ouverture triangulaire appelée glotte. L'air y passe librement pendant la respiration et la voix chuchotée, ainsi que pendant la phonation des sons non-voisés [1].

Les sons voisés résultent au contraire d'une vibration périodique des cordes vocales. Le larynx est d'abord complètement fermé, ce qui accroît la pression en amont des cordes vocales, et les force à s'ouvrir, ce qui fait tomber la pression, et permet aux cordes vocales de se refermer; des impulsions périodiques de pression sont ainsi appliquées au conduit vocal, composé des cavités pharyngienne et buccale pour la plupart des sons. Lorsque la luette est en position basse, la cavité nasale vient s'y ajouter en dérivation. Notons pour terminer le rôle prépondérant de la langue dans le processus phonatoire. Sa hauteur détermine la hauteur du pharynx : plus la langue est basse, plus le pharynx est court. Elle détermine aussi le lieu d'articulation, région de rétrécissement maximal du canal buccal, ainsi que l'aperture, écartement des organes au point d'articulation. Dans la figure (1.2), on présente le signal correspondant au mot (bonjour). On y voit le graphe du signal complet, et deux zones du son (bonjour). La première correspond au son non voisé (j), alors que la deuxième illustre le phonème voisé (ou).

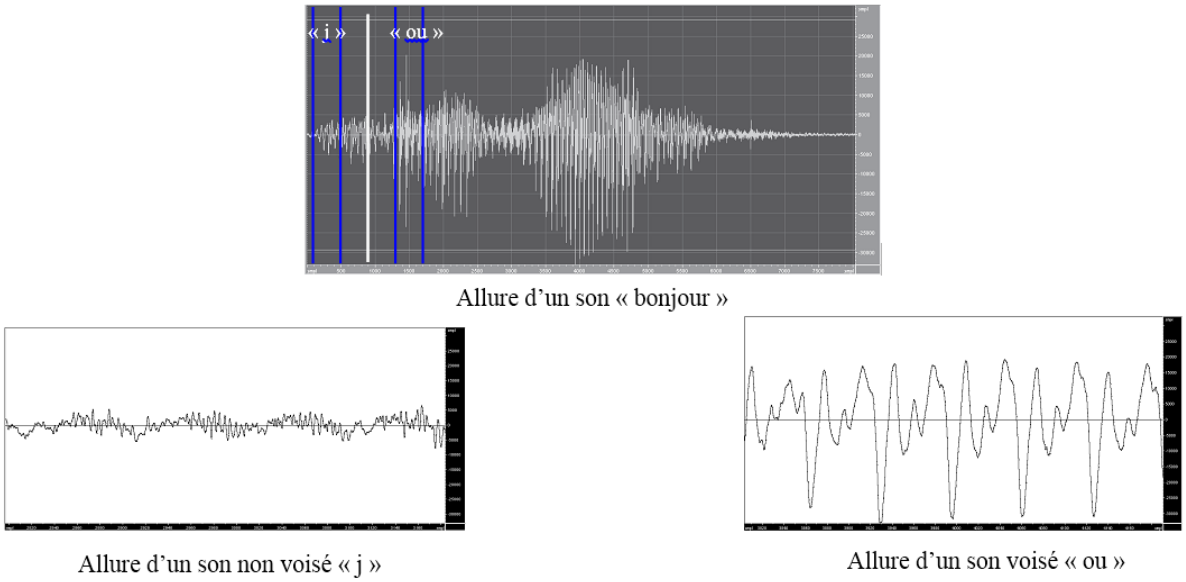

Figure 1.2 : Exemple d'un signal de parole voisé et non voisé.

1.2 Niveau acoustique

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 (lui-même associé à un préamplificateur). 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.

L'opération de numérisation, requiert successivement : un filtrage de garde, un échantillonnage, et une quantification.

L'échantillonnage transforme le signal à temps continu *x(t)* en signal à temps discret *x(nTe)* défini aux instants d'échantillonnage, multiples entiers de la période d'échantillonnage *Te*; celle-ci est elle-même l'inverse de la fréquence d'échantillonnage *fe*. Pour ce qui concerne le signal vocal, le choix de *fe* résulte d'un compromis. Son spectre peut s'étendre jusque 12 kHz. Il faut donc en principe choisir une fréquence *fe* égale à 24 kHz au moins pour satisfaire raisonnablement au théorème de Shannon4. Cependant, le coût d'un traitement numérique, filtrage, transmission, ou simplement enregistrement peut être réduit d'une façon notable si l'on

accepte une limitation du spectre par un filtrage préalable. C'est le rôle du filtre de garde, dont la fréquence de coupure *fc* est choisie en fonction de la fréquence d'échantillonnage retenue. Pour la téléphonie, on estime que le signal garde une qualité suffisante lorsque son spectre est limité à 3400 Hz et l'on choisit *fe* = 8000 Hz [1].

1.3 Propriétés statistiques d'un signal de parole

Rappelons que, si le signal est échantillonné à une fréquence compatible avec le théorème de Shannon, l'estimation de sa statistique peut être faite sur ses échantillons. Le signal vocal peut être vu comme une réalisation particulière d'un processus aléatoire non stationnaire ; ses propriétés moyennes doivent donc être estimées sur des intervalles de temps très importants, par exemple sur plusieurs dizaines de secondes, et moyennées pour plusieurs locuteurs. On parle dans ce cas de statistique à long terme.

On définit également une statistique à court terme, estimée sur des tranches temporelles de durée de 10 à 30 ms, pendant lesquelles le signal peut être considéré comme quasi stationnaire étant donné l'inertie propre aux muscles articulatoires [2].

Les principales caractéristiques des signaux de parole sont :

- On constate expérimentalement (à partir d'une estimation faite sur des segments de parole de l'ordre de 50 s)que la densité de probabilité à long terme de la parole est très proche de la distribution Gamma et relativement proche de la distribution laplacienne.
- Taux de passage par zéro pour un signal échantillonné, il y a passage par zéro lorsque deux échantillons successifs sont de signe opposé qui présente une répartition gaussienne avec une moyenne de l'ordre de 49/10ms pour les sons non voisés et de 14/10ms pour les sons voisés ;ces deux répartitions se recouvrent partiellement.la mesure du taux de passage par zéro contribue donc, tout comme la mesure de l'énergie du signal ,à la détection voisé/non voisé [4].
- La densité spectrale de puissance à court terme (transformée de Fourrier de la fonction d'auto- corrélation) qui présente, lorsque la tranche est voisée, une structure périodique fine qui correspond aux harmoniques de l'excitation glottique. Les maximums de l'enveloppe, de ce spectre, sont les formants, ils correspondent aux résonances du conduit vocal. Par contre, le spectre d'un signal non voisé ne présente

aucune structure particulière, sauf une accentuation vers les hautes fréquences (figures 1.3 et 1.4) [4].

- L'autocorrélation dans un son voisé est plus élevée par rapport au son non voisé (Figure 1.5).

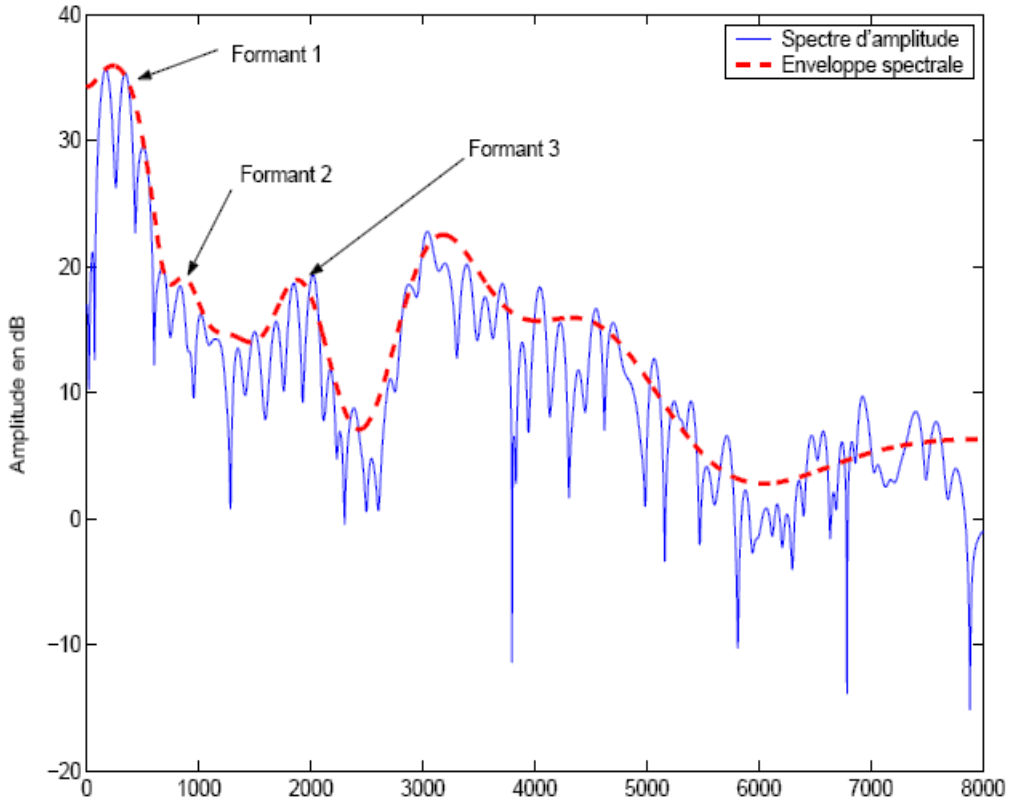

Figure 1.3 : Densité spectrale de puissance en échelle logarithmique d'un son voisé.

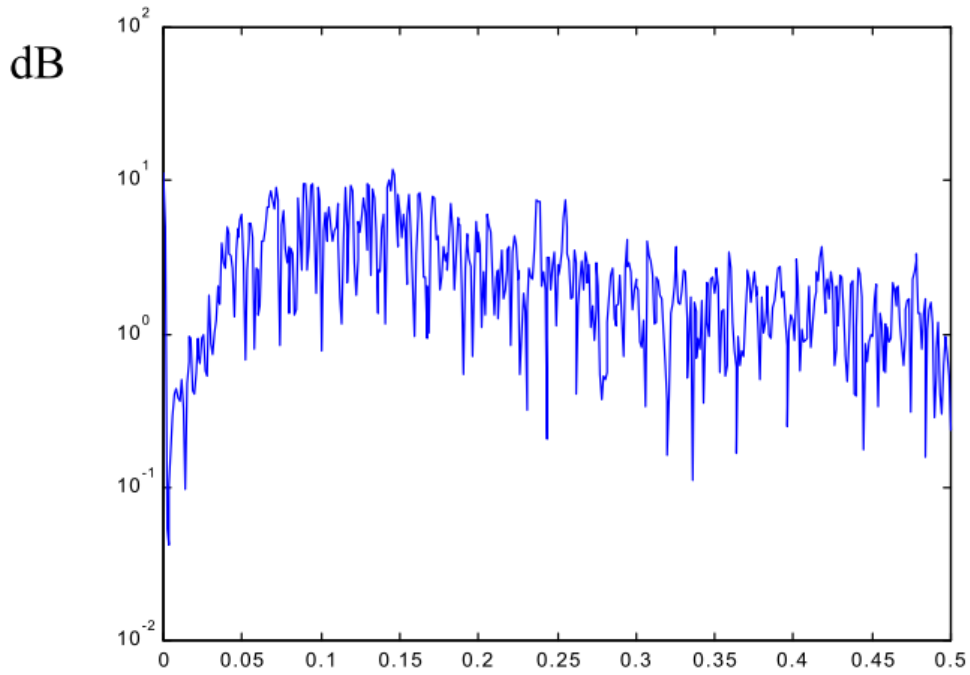

Figure 1.4 : Densité spectrale de puissance en échelle logarithmique d'un son non voisé.

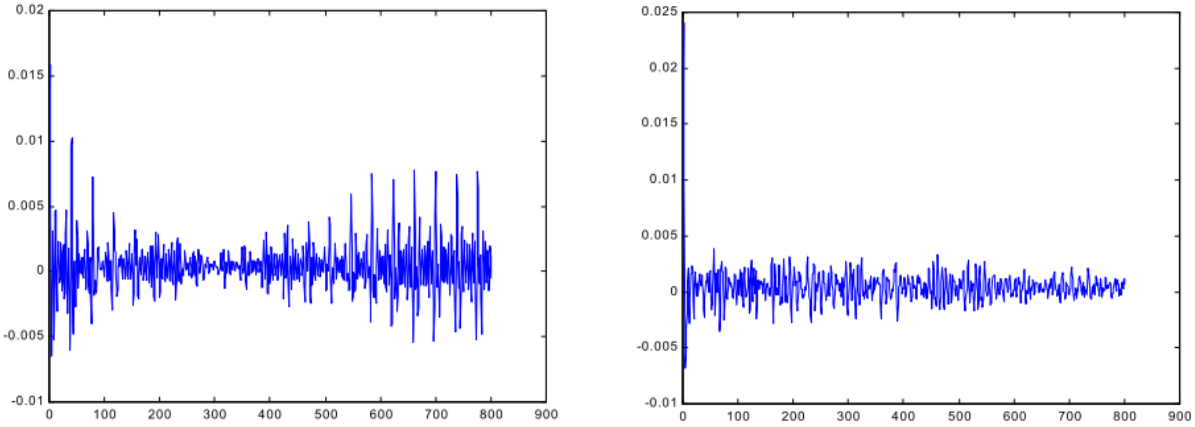

Figure 1.5 : Autocorrélation d'un son voisé (à gauche) et non voisé (à droite).

- La fréquence fondamentale est par définition l'inverse de la période de vibration des cordes vocales. Ses variations définissent le pitch qui constitue la perception de la hauteur, où les sons s'ordonnent de grave à aigu. Seuls les sons voisés engendrent une sensation de hauteur tonale bien définie.

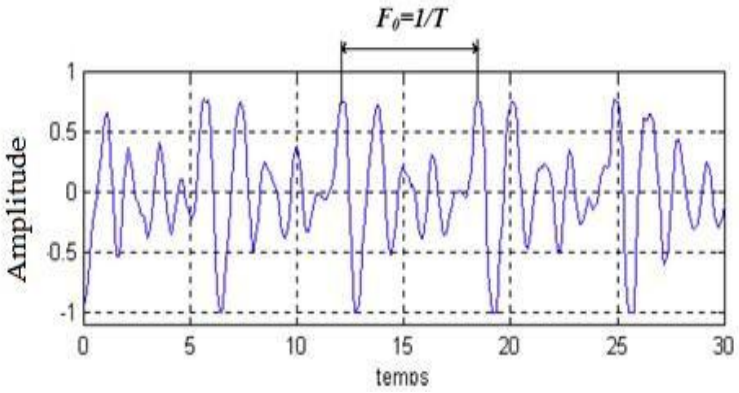

Figure 1.6 : La fréquence fondamentale [1].

L'extraction de la fréquence fondamentale n'est pas une tâche facile pour les trois raisons suivantes :

- La périodicité de vibration des cordes vocales n'est pas nécessairement parfaite ;
- Il est difficile des éparer la fréquence fondamentale des effets du trait vocal ;
- La plage de dynamique de la fréquence fondamentale est très grande.

Une analyse d'un signal de parole n'est pas complète tant qu'on n'a pas mesuré l'évolution temporelle de la fréquence fondamentale.

Il a été constaté qu'à l'intérieur des zones voisées la fréquence fondamentale évoluent lentement dans le temps. Elle s'étend approximativement de 70 à 250 Hz chez les hommes, de 150 à 400 Hz chez les femmes, et de 200 à 600 Hz chez les enfants.

1.4 Modèle numérique de production de la parole

Une modélisation exhaustive de la production de la parole est très difficile et, pour des raisons pratiques, inefficace. L'idée de base dans la modélisation numérique est d'arriver à un modèle linéaire qui produit en sortie un signal équivalent au signal vocal. Le modèle est correct dans la mesure où sa sortie s'approche du signal vocal sans modéliser les phénomènes physiques intrinsèques à la production du signal vocal [5]. La figure 1.7 présente un tel modèle général qui est utilisé dans le traitement numérique de la parole.

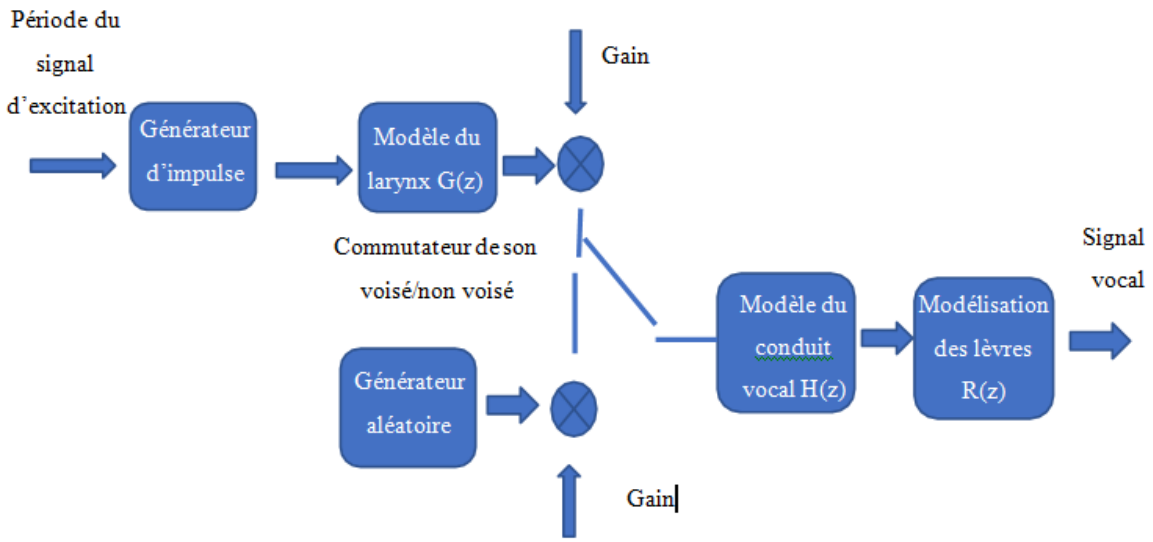

Figure 1.7 : Modèle numérique de production de la parole.

Dans ce modèle général, on utilise deux sources d'excitation. Pour les sons non voisés, la source d'excitation est un bruit blanc. Pour la production des sons voisés, la source d'excitation est un train périodique d'impulsions qui traverse un filtre passe bas d'ordre 2 de fonction de transfert $G(z)$ donnée par [5] :

$$G(z) = \frac{\mathrm{A}}{(1+\mathrm{b}_1 \mathrm{z}^{-1})(1+\mathrm{b}_2 \mathrm{z}^{-2})} \tag{1.1}$$

Le conduit vocal peut être modélisé par une succession de tubes acoustiques élémentaires. Chaque tube ou résonateur mécanique est assimilé à un filtre numérique d'ordre 2. La transmittance globale du modèle est de la forme [5] :

$$H(z) = \frac{B}{\prod_{k=1}^{n}(1+b_{1k}z^{-1})(1+b_{2k}z^{-2})} \tag{1.2}$$

Au bout du conduit vocal le son passe à travers l'ouverture des lèvres. Celles-ci sont vues comme une composante qui transforme le débit volumique dans une onde de pression à une certaine distance. Dans le domaine spectral, le rayonnement des lèvres a l'effet d'un filtrage passe haut. Le plus simple filtre numérique qui a cette propriété est [5] :

$$R(z) = C(1 - z_0 z^{-1}) \tag{1.3}$$

En conclusion, la fonction de transfert globale est de la forme [4] :

- Pour les sons non voisés : $T(z) = H(z)R(z)$
- Pour les sons voisés : $T(z) = G(z)H(z)R(z)$

1.5 Généralités sur le bruit

1.5.1 Définition du bruit

Le bruit est défini, au sens large, comme étant tout signal perturbateur entachant à un degré ou un autre l'intégrité d'un signal utile (dans notre cas le signal de parole). En physique, acoustique et en traitement de signal, bien que le bruit soit par nature aléatoire, il possède certaines caractéristiques statistiques, spectrale sous partiales [6]. Le tableau suivant représente des exemples des différentes classes auxquelles un bruit peut appartenir.

Tableau 1.1 : Propriétés du bruit.

Propriétés	Types
Structure	Continu/Intermittent/Impulsif
Type d'interaction	Additif/Convolutif
Comportement temporel	Stationnaire/Non-stationnaire
Bande de fréquence	Etroite/Large
Dépendance	Corrélé/Décorrélé
Propriétés spatiales	Cohérent/Incohérent

1.5.2 Classification des types de bruit en fonction de leur nature

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

1.5.2.1 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.

1.5.2.2 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.

1.5.2.3 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 aux caractéristiques du canal [6].

1.5.3 Classification selon sa fréquence ou son temps caractéristiques

1.5.3.1 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.

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

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

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

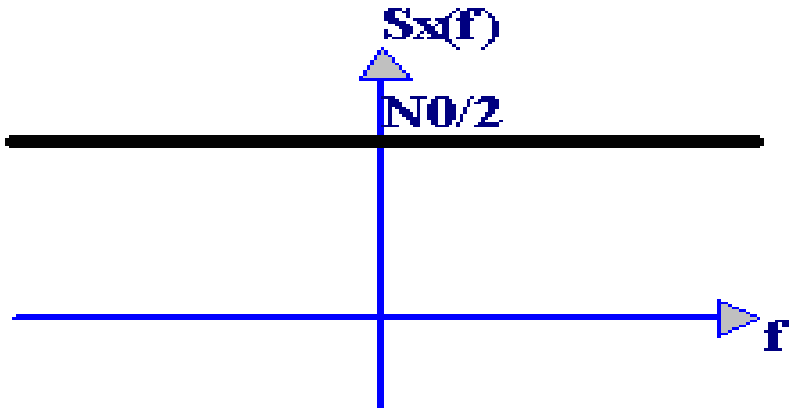

Figure 1.8 : Spectre d'un bruit blanc.

1.5.3.2 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 largeur 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).

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

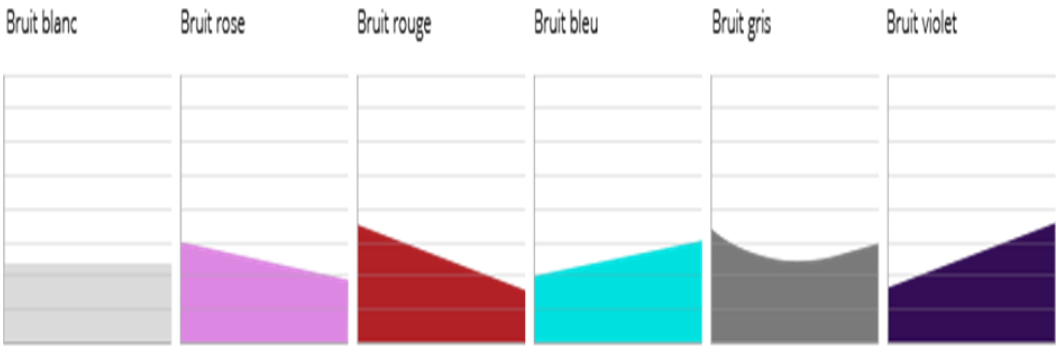

Figure 1.9 : Exemple des bruits colorés.

1.5.3.3 Bruit impulsif

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

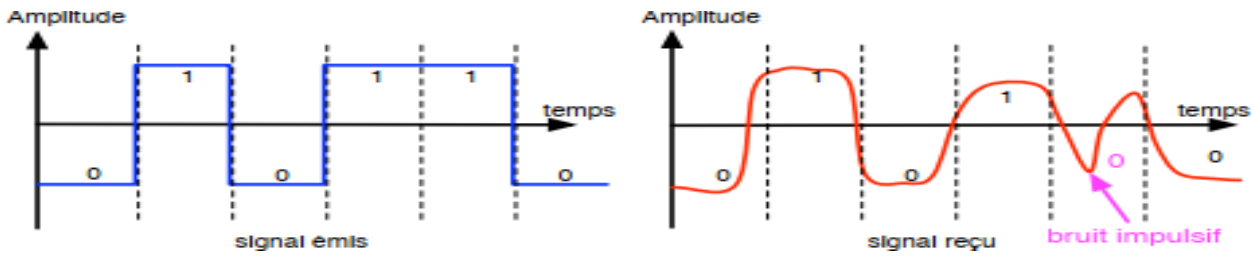

Figure 1.10 : Exemple de bruit impulsif.

1.6 Techniques d'analyse

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

1.6.1 Analyse temporelle à court terme

La représentation à court terme du signal dans le domaine temporel consiste à extraire la représentation du signal trame par trame, l'idée de base étant de travailler sur chaque trame. Dans certains cas, le signal est supposé stationnaire dans une trame donnée. La taille de trame est un paramètre important du signal de parole et une fois que cette taille est fixée, elle est choisie pour contenir au moins 2-3 périodes de signal dans tous les cas. Pour les signaux de parole continus de, il est raisonnable de supposer que les voix masculines s'éteignent à 50 Hz et les voix féminines à 100 Hz, ce qui donne des tailles de fenêtre d'environ 30 ms pour les hommes et 20 ms pour les femmes [4].

1.6.2 Analyse fréquentielle à court terme

En effet, les signaux étudiés ne sont pas périodiques, la décomposition en série de Fourier doit être appliquée aux signaux non périodiques, ce qui donne cependant des harmoniques.

D'un point de vue pratique, la seule analyse réalisable est une analyse à court terme dans le domaine fréquentiel, cette analyse nécessite l'utilisation d'une fenêtre sur le signal, ce qui modifie la décomposition en harmoniques.

Cependant, avec une fenêtre suffisamment grande, l'analyse fréquentielle est généralement capable de séparer les harmoniques du signal et ainsi d'estimer leurs propriétés [4].

1.6.3 Fonctions de fenêtrage

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 [4].

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 :

$$W_{Hann} = 0.5 + 0.5 \cdot \cos(\frac{2n\pi}{N}) \quad (1.4)$$

Où :

N est la longueur de fenêtre (nombre d'échantillons).

n est 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 [4].

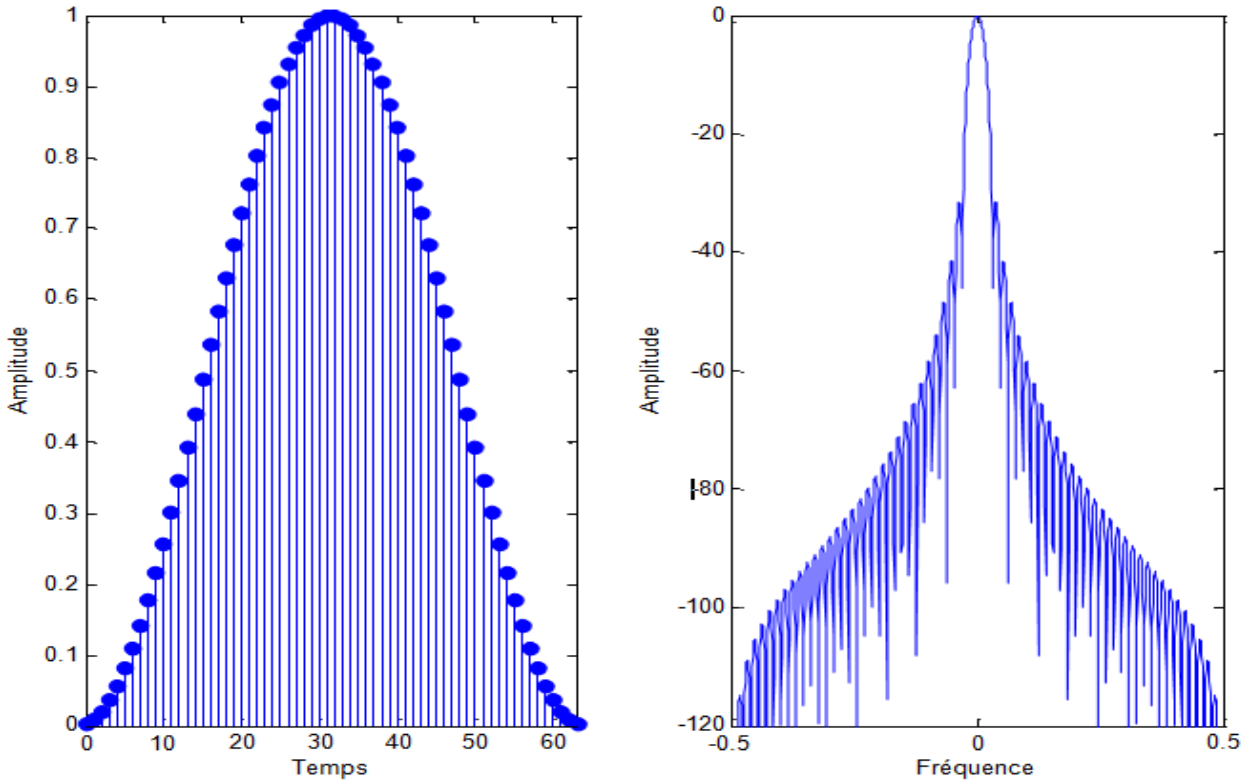

Figure 1.11 : Fenêtre de Hanning.

1.6.4 Analyse de fréquence

Pour convertir le signal dans le domaine fréquentiel, Il y a deux outils mathématiques sont utilisés, selon la nature du signal continu ou discret.

Pour les signaux continus, la transformation de Fourier (TF) est utilisée, tandis que pour les signaux discrets ou les séquences, l'outil doit être la transformation de Fourier à temps discret (TFTD). Ce projet se concentre sur la TFTD, car les signaux audio numériques sont des signaux discrets ou des séquences [5].

1.6.5 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 permet de représenter en fréquence (développement sur une base d'exponentielles) des signaux qui ne sont pas périodiques [5]. Son expression est la suivante :

$$TF(x(t)) = X(f) = \int_{-\infty}^{+\infty} x(t)e^{-j2\pi ft}\,dt \tag{1.5}$$

1.6.6 Transformée de Fourier à temps discret (TFTD)

La transformée de Fourier à temps discret, ou transformée de Fourier d'une suite $x[n]$, est une fonction $X(e^{j\omega})$, continue et périodique, de période 2π. Elle peut être calculée avec l'expression suivante :

$$X(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} x[n]e^{-j\omega n} \tag{1.6}$$

La transformée de Fourier à temps discret inverse (TFTDI) renverra la séquence d'origine, étant son expression appelée synthèse :

$$x[n] = \frac{1}{2\pi}\int_{-\pi}^{+\pi} X(e^{j\omega})e^{j\omega n}\,d\omega \tag{1.7}$$

Comme $X(e^{j\omega})$ une fonction complexe avec des composants réels et imaginaires, elle peut être représentée comme :

$$X(e^{j\omega}) = X_R(e^{j\omega}) + jX_I(e^{j\omega}) \tag{1.8}$$

ou, sous forme polaire (module et phase) :

$$X(e^{j\omega}) = \left|X(e^{j\omega})\right|.e^{j(argX(e^{j\omega}))} \tag{1.9}$$

où $|X(e^{j\omega})|$ est le module et $arg\left[X(e^{j\omega})\right]$ est la phase.

1.6.7 Transformée de Fourier à court terme

La transformée de Fourier à court terme ou la STFT[1] est un outil puissant pour le traitement du signal vocal [5], 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 [5] :

$$X_m(\omega) = \sum_{n=-\infty}^{+\infty} x(n) \cdot w(n - mR) \cdot \mathrm{e}^{-\mathrm{j}\omega \mathrm{n}} \qquad (1.10)$$

où

$x(n)$ est le signal d'entrée l'instant n,

n est l'indice de temps discret,

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

$X_m(\omega)$ est TFTD de signal fenêtrée centré sur le temps mR

R est la taille de saut en échantillons, entre les TFTD successives.

1.7 Conclusion

Dans ce chapitre, nous avons passé en revue des notions fondamentales du signal de parole, ses caractéristiques ainsi que les différents types de bruits peuvent affecter ce signal, en addition de son analyse.

Pour rehausser le signal de parole, plusieurs techniques existent. Nous présentons donc la technique de la soustraction spectrale ainsi que quelques méthodes d'estimation de bruit dans le prochain chapitre.

[1] **S**hort **T**ime **F**ourier **T**ransform.

Chapitre 2
Algorithmes d’estimation du bruit

Si la liberté et la flexibilité offertes par la technologie mobile ont permis de communiquer en dehors des environnements contrôlés, elles ont également introduit de nouveaux défis. Les utilisateurs mobiles communiquent dans différents environnements avec différents niveaux et types de bruit de fond tels que le bruit de la circulation, le bruit des moteurs de voiture, le babillage de plusieurs interlocuteurs comme dans les cafétérias, etc. La suppression du bruit de fond acoustique est un problème pertinent et difficile. Outre la réduction de la fatigue de l'auditeur et l'amélioration de la qualité et de l'intelligibilité de la parole, la réduction du bruit est également cruciale pour obtenir de bonnes performances des algorithmes de codage vocal qui rendent possible la communication mobile.

Les systèmes de rehaussement de la parole à canal unique obtiennent le signal d'entrée en utilisant un seul microphone. Ceci contraste avec les systèmes multicanaux où la présence de deux microphones ou plus permet un traitement à la fois spatial et temporel. Les approches monocanal

sont pertinentes en raison de facteurs de coût et de taille. Ils parviennent à réduire le bruit en exploitant la diversité spectrale entre les signaux de parole et de bruit. Étant donné que les spectres de fréquences de la parole et du bruit se chevauchent souvent, les méthodes monocanal permettent généralement de réduire le bruit au détriment de la distorsion de la parole. La soustraction spectrale est sans doute la plus connue et la plus utilisée des méthodes de rehaussement de parole monocanal.

L'estimation des statistiques du bruit de fond est une caractéristique essentielle des algorithmes de réduction de bruit monocanal. Il s'agit d'une tâche difficile car les estimations doivent être obtenues à partir du signal vocal bruité. Une approche courante consiste à utiliser un détecteur d'activité vocale (VAD) pour identifier les segments temporels du signal où la parole est absente et où le signal est donc constitué uniquement de bruit de fond. Les estimations des statistiques de bruit sont mises à jour pendant ces pauses de parole. Bien que les schémas d'estimation du bruit basés sur VAD présentent l'avantage d'une faible complexité de calcul, ils souffrent de deux problèmes. Premièrement, avec la diminution des rapports signal/bruit, la détection des pauses vocales n'est plus une tâche triviale. Deuxièmement, même si la méthode fonctionne raisonnablement bien dans des environnements à bruit stationnaire, les performances se dégradent dans des environnements où les statistiques de bruit changent continuellement, ce qui est souvent le cas en pratique.

Pour remédier aux défauts des VAD binaires (il n'y a que deux états, présence et absence de parole), des algorithmes d'estimation de bruit (VAD à décision soft) ont été proposés, attribuant une probabilité de présence de parole à chaque segment. Ainsi, il est possible de mettre à jour les statistiques de bruit en continu, en fonction de la probabilité de présence de parole.

Ce chapitre présente la technique de soustraction spectrale ainsi que quelques algorithmes d'estimation de bruit (VAD à décision soft).

2.1 Soustraction spectrale

La soustraction spectrale permet d'estimer l'amplitude spectrale d'un signal stationnaire $s(n)$ dégradé par un bruit additif $b(n)$ non corrélé avec $s(n)$. Soit $y(n)$ le signal observé :

$$y(n) = s(n) + b(n) \tag{2.1}$$

On a alors :

$$S_{yy}(f) = S_{ss}(f) + S_{bb}(f) \tag{2.2}$$

Où $S_{yy}(f), S_{ss}(f)$ et $S_{bb}(f)$ sont les densités spectrales de puissance respectives de $y(n)$, $s(n)$ et $b(n)$. On peut donc en théorie obtenir $S_{ss}(f)$ par simple soustraction de $S_{bb}(f)$ à la densité spectrale de puissance de l'observation :

$$S_{ss}(f) = S_{yy}(f) - S_{bb}(f) \tag{2.3}$$

En pratique, le signal $s(n)$ est considéré comme stationnaire sur une durée de 20 à 40 ms. Cet aspect court-terme nous permet d'écrire une première formule approchée :

$$\left|\hat{S}(f)\right|^2 = \left|Y(f)\right|^2 - E\left\{\left|B(f)\right|^2\right\} \tag{2.4}$$

Où $\left|\hat{S}(f)\right|$ et $\left|Y(f)\right|$ sont les amplitudes des transformations de Fourier court-terme des signaux $s(n)$ et $y(n)$ et $E\left\{\left|B(f)\right|^2\right\}$ représente l'estimée de la densité spectrale de puissance du bruit sur une durée à moyen terme.

Le spectre du bruit est soit calculé sur les périodes de silence, après utilisation d'un système de détection d'activité vocale, soit obtenu à partir des propriétés statistiques de $b(n)$, lorsqu'elles sont connues. Dans les deux cas, l'estimée de densité spectrale de puissance du signal utile de l'équation (2.4) doit toujours rester positive ou nulle. Cette contrainte est respectée en prenant le module de la différence calculée ou en fixant un seuil nul. C'est cette dernière méthode qui est généralement choisie.

Après le calcul de $\left|\hat{S}(f)\right|$, l'estimée temporelle du signal utile est obtenue, en conservant pour $\hat{S}(f)$ la phase de l'observation, par transformée de Fourier inverse et en utilisant la règle de « Chevauchement-addition ». On suppose effectivement que l'oreille est insensible aux légères variations de phase du signal vocal. Cette technique permet, en l'absence de modifications spectrales, la reconstruction parfaite du signal par sommation des signaux découpés par les fenêtres d'analyse.

On a donc :

$$\hat{s}(n) = TF^{-1}\left\{\left|\hat{S}(f)\right| \cdot \exp\left[j \cdot Arg\left(Y(f)\right)\right]\right\} \tag{2.5}$$

Il est important de rappeler que les méthodes de soustraction spectrale ne permettent pas la reconstruction de la phase du signal de parole, le traitement n'agit que sur le module ou le module au carré du spectre du signal d'observation.

Le principe de la soustraction spectrale est donné par la figure 2.1[1].

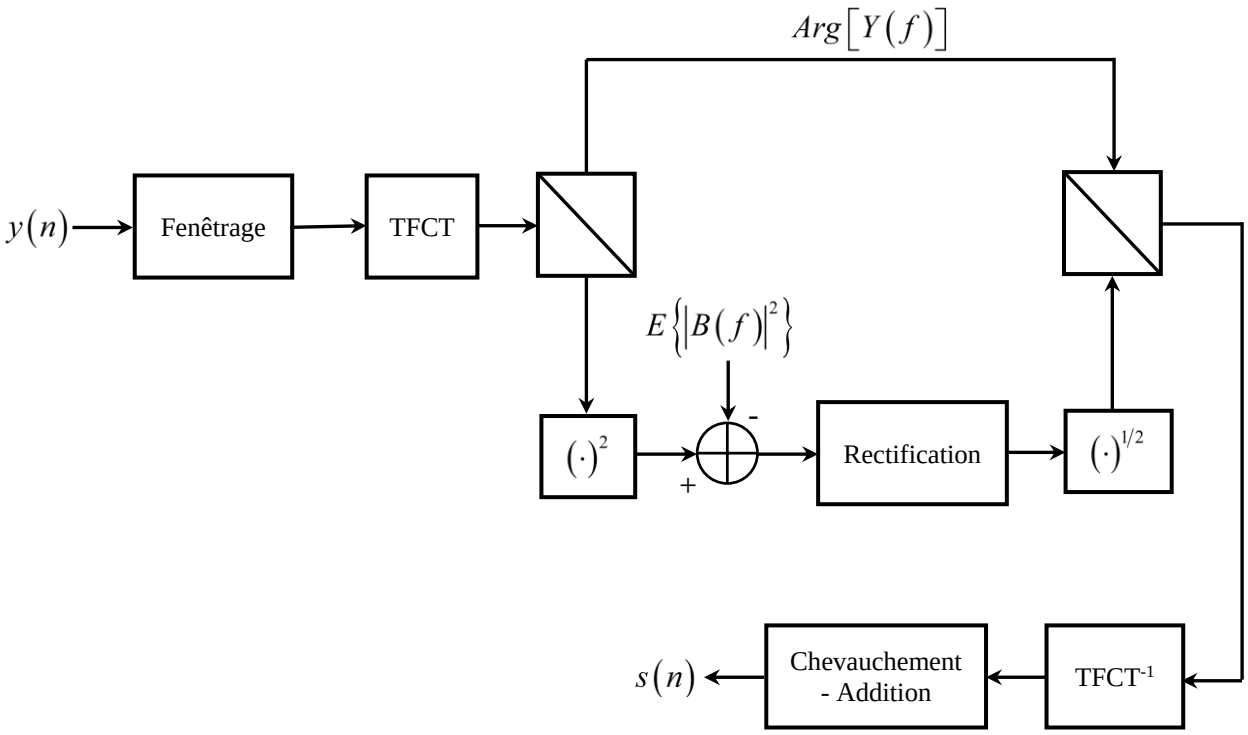

Figure 2.1 : Principe de la soustraction spectrale.

De nombreuses modifications ont été apportées à la soustraction spectrale classique, dite « soustraction spectrale de puissance » donnée par l'équation 2.4. En effet, cette technique engendre un bruit résiduel très gênant, appelé bruit « musical ». Cet inconvénient majeur est dû au fait que l'estimée du bruit $E\left\{\left|B(f)\right|^2\right\}$ correspond à un moyennage à moyen terme qui fait disparaître les variations sur chaque trame de bruit original. Des raies apparaissent alors

[1] L'abréviation TFCT indique la transformée de Fourier à Court-Terme.

aléatoirement dans le spectre estimé et donnent à l'écoute un aspect musical. Parmi les méthodes développées qui essaie de limiter ce phénomène en introduisant des modifications, on trouve la technique développée par Berouti [7]. Elle consiste à soustraire l'estimée de la densité spectrale de puissance du bruit multipliée par un facteur α à la densité spectrale de puissance de l'observation :

$$\left|\hat{S}(f)\right|^2 = \left|Y(f)\right|^2 - \alpha \cdot E\left\{\left|B(f)\right|^2\right\} \tag{2.6}$$

avec $\alpha \geq 0$.

Les excursions spectrales sont alors réduites et un grand nombre de raies du signal utile vont être éliminés. Plus le rapport signal sur bruit sera faible et plus cette méthode éliminera des raies spectrales, ce qui entrainera des brusques variations dans le spectre estimé. Berouti propose alors comme solution complémentaire d'introduire un plancher spectrale minimal.

Afin d'éviter les vallées trop importantes dans le spectre du signal estimé, Berouti propose de les combler par une portion du spectre du bruit. Ce niveau minimum couvrira également les raies indésirables restantes, caractéristiques du bruit musical. On obtient ainsi une première forme de soustraction spectrale modifiée :

$$\left|\hat{S}(f)\right|^2 = \begin{cases} \left|Y(f)\right|^2 - \alpha \cdot E\left\{\left|B(f)\right|^2\right\} & \text{si } \left|Y(f)\right|^2 - \alpha \cdot E\left\{\left|B(f)\right|^2\right\} > \beta \cdot E\left\{\left|B(f)\right|^2\right\} \\ \beta \cdot E\left\{\left|B(f)\right|^2\right\} & \text{sinon} \end{cases} \tag{2.7}$$

Il faudra alors faire un compromis sur les valeurs α (supérieur à 1) et β (qui est très inférieur à 1) afin d'optimiser le traitement en fonction des situations rencontrées.

Le signal de la parole rehaussée, qui résulte de l'application de cette méthode est affecté par deux types de bruits :

- Bruit à large bande, connu sous le nom de bruit résiduel.
- Bruit à bande étroite, connu sous le nom de bruit musical.

Ces deux bruits apparaissent sous forme de pics et de vallées dans le spectre du signal rehaussé. Ils ont une distribution aléatoire et changent aléatoirement en fréquence et en amplitude d'une trame à l'autre. La réduction des pics spectraux du bruit résiduel est assurée par le facteur de soustraction qui prend toujours une valeur supérieure à l'unité $(\alpha > 1)$. Si une valeur élevée de α

est prise, la réduction du bruit à large bande se fait, mais cela provoque une distorsion du signal de la parole.

Afin d'obtenir des valeurs optimales du facteur de soustraction, il faut prendre en compte que α est une fonction du rapport signal sur bruit segmental, dont sa valeur réelle est donnée par la relation suivante :

$$\alpha = \begin{cases} 5 & SNR < -5dB \\ \alpha_0 - (SNR/s) & -5dB < SNR < 20dB \\ 1 & SNR > 20dB \end{cases} \qquad (2.8)$$

avec :

α_0 est la valeur de α pour un $SNR = 0$ (Dans la pratique $3 < \alpha_0 < 6$).

1/s est la pente de la droite dans la figure 2.2.

SNR : le rapport signal sur bruit segmental estimé.

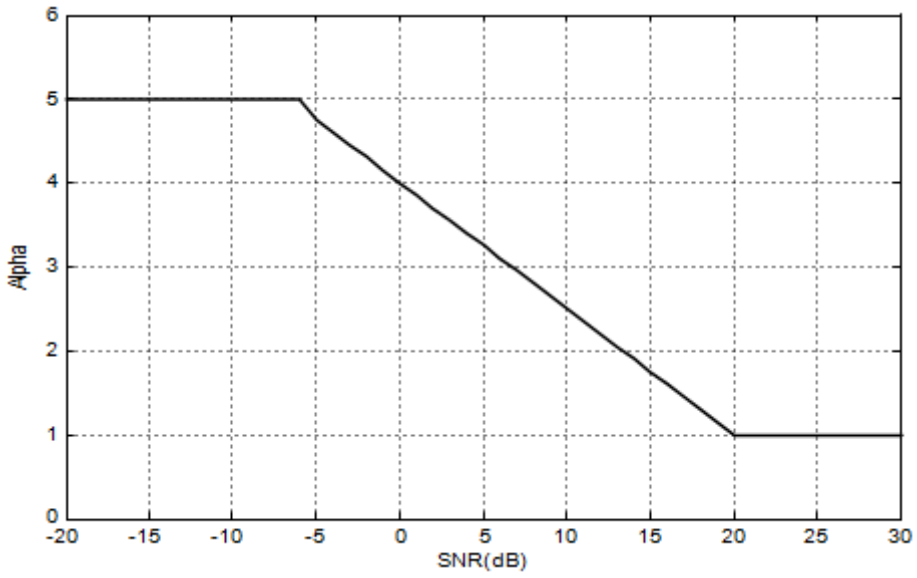

Figure 2.2 : Valeurs de α en fonction du SNR.

Outre la réduction des pics, il y a le problème du remplissage des vallées d'où la réduction du bruit musical. Cela est effectué par le facteur de lissage β qui prend des valeurs dans l'intervalle $0 < \beta << 1$.

- $\beta > 0$: les pics du bruit résiduel sont masqués par les composantes spectrales voisines.
- $\beta << 1$: le bruit à large bande est plus bas par rapport à celui obtenu dans le cas où $\beta = 0$.

Le choix de β a donc une importance majeure pour le spectre de puissance du signal propre estimé $\left|\hat{S}(f)\right|^2$, Il est constaté aussi que :

- Si β est faible : le bruit résiduel sera réduit, mais le bruit musical sera audible.
- Si β est grande : le bruit musical n'est pas audible mais le bruit résiduel reste présent.

L'inconvénient majeur de cette technique est l'apparition d'un bruit musical. Toutefois, malgré cet inconvénient du bruit musical, la méthode de la soustraction spectrale reste performante en termes d'atténuation du bruit.

2.2 Algorithmes d'estimation de bruit

Dans cette section, certains des algorithmes d'estimation du bruit existants, basés sur le suivi du bruit en utilisant le spectre de puissance de la parole bruitée, sont discutés. La plupart de ces algorithmes peuvent être globalement classés en deux classes, à savoir :

- ***Algorithmes d'estimation du bruit à moyenne récursive :*** Le bruit a un effet non uniforme sur le spectre de la parole. En conséquence, différentes bandes de fréquences du spectre auront effectivement des RSB différents. Si, par exemple, le bruit est de nature passe-bas (par exemple, le bruit d'une voiture), alors la région haute fréquence du spectre sera la moins affectée, tandis que la région basse fréquence sera la plus affectée. Par conséquent, le spectre de bruit peut être estimé et mis à jour de manière plus fiable sur la base d'informations extraites de la région haute fréquence du spectre de la parole bruitée plutôt que de la région basse fréquence. Plus généralement, pour tout type de bruit, nous pouvons estimer et mettre à jour des bandes de fréquences individuelles du spectre du bruit chaque fois que la probabilité d'absence de parole sur une bande de fréquences particulière est élevée ou lorsque le RSB effectif sur une bande de fréquences particulière est extrêmement faible.

- ***Algorithmes de suivi des minima*** : La puissance du signal de parole bruitée dans les bandes de fréquences individuelles diminue souvent jusqu'au niveau de puissance du bruit, même pendant l'activité vocale. Une estimation du niveau de bruit dans des bandes

de fréquences individuelles peut être obtenue en suivant le minimum, dans une courte fenêtre (0.4 à 1 s), du spectre de la parole bruitée dans chaque bande de fréquences.

Tous les algorithmes fonctionnent de la manière suivante. Tout d'abord, le signal est analysé à l'aide de spectres à court terme calculés à partir de courtes trames d'analyse qui se chevauchent (généralement des fenêtres de 20 à 30 ms avec un chevauchement de 50 % entre les trames adjacentes). Ensuite, plusieurs trames consécutives, formant ce que nous appelons un segment d'analyse, sont utilisées dans le calcul du spectre de bruit. La durée typique de ce segment peut aller de 400 ms à 1 s. Les algorithmes d'estimation du bruit sont basés sur les hypothèses suivantes :

- Le segment d'analyse est suffisamment long pour contenir des pauses vocales et des segments de signaux de faible énergie.
- Le bruit présent dans le segment d'analyse est plus stationnaire que la parole. Autrement dit, le bruit change à un rythme relativement plus lent que la parole.

Les hypothèses précédentes imposent des restrictions contradictoires sur la durée du segment d'analyse. Le segment d'analyse doit être suffisamment long pour englober les pauses vocales et les segments de faible énergie, mais il doit également être suffisamment court pour suivre les changements rapides du niveau de bruit. De ce fait, la durée choisie du segment d'analyse résultera d'un compromis entre ces deux contraintes.

2.2.1 Algorithmes de suivi des minima

Comme mentionné précédemment, les algorithmes de suivi des minima reposent sur l'hypothèse que la puissance du signal de parole bruitée dans les bandes de fréquences individuelles décroît souvent jusqu'au niveau de puissance du bruit, même pendant l'activité vocale. Par conséquent, en suivant le minimum de puissance de la parole bruitée dans chaque bande de fréquences, on peut obtenir une estimation approximative du niveau de bruit dans cette bande. Deux algorithmes différents ont été proposés pour l'estimation du bruit. Le premier algorithme, appelé algorithme de statistiques minimales (MS^2) [8], suit le minimum du spectre de puissance vocale bruyante dans une fenêtre finie (segment d'analyse), tandis que le deuxième algorithme [9]

[2] En anglais : **M**inimum-**S**tatistics Algorithm

suit le minimum d'une façon continue sans exiger l'utilisation d'une fenêtre. Ce dernier est appelé l'algorithme de Doblinger.

2.2.1.1 Algorithme de statistiques minimales (MS)

La méthode de Martin [8] est basée sur des statistiques minimales et un lissage optimal de la densité spectrale de puissance de la parole bruitée. Cette méthode repose sur deux observations majeures. La première observation est l'indépendance de la parole propre et du bruit, ce qui implique que le spectre de puissance de la parole bruitée est respectivement égal à la somme du spectre de puissance de la parole propre et du bruit. C'est-à-dire :

$$\left|Y(\lambda,k)\right|^2 = \left|X(\lambda,k)\right|^2 + \left|D(\lambda,k)\right|^2 \tag{2.9}$$

Où $\left|Y(\lambda,k)\right|^2$, $\left|X(\lambda,k)\right|^2$ et $\left|D(\lambda,k)\right|^2$ sont respectivement le spectre de puissance de la parole bruitée, de la parole propre et du bruit et λ et k dénotent respectivement l'indice de temps et l'indice de fréquence. La deuxième observation est que le spectre de puissance de la parole bruitée devient souvent égal au spectre de puissance du bruit. Cela se produit pendant les segments de pause de la parole et également entre les mots et les syllabes. Par conséquent, l'estimation de la densité spectrale de puissance du bruit est obtenue en suivant séparément le minimum de parole bruitée dans chaque bande de fréquences. De plus, comme le minimum est biaisé vers des valeurs inférieures, une estimation non biaisée a été obtenue en multipliant par un facteur de biais qui est dérivé des statistiques du minimum local.

a) Principes

Étant donné que la densité spectrale de puissance de la parole bruitée est égale à la somme de la puissance du bruit et de la puissance de la parole propre, la variance du bruit est estimée en suivant la densité spectrale de puissance minimale de la parole bruitée sur une longueur de fenêtre fixe. Cette longueur de fenêtre est choisie suffisamment large pour combler le pic le plus large de n'importe quel signal de parole. Il est trouvé expérimentalement [10] que des longueurs de fenêtre d'environ 0.8 à 1.4 s donnent de bons résultats.

Pour rechercher le minimum, une version récursive du premier ordre de la densité spectrale de puissance de la parole bruitée est utilisée :

$$P(\lambda,k) = \alpha \cdot P(\lambda-1,k) + (1-\alpha)\cdot\left|Y(\lambda,k)\right|^2 \tag{2.10}$$

où α est la constante de lissage. Pour améliorer les performances de la procédure de suivi du minimum, les modifications suivantes ont été apportées :

1. Remplacement du facteur de lissage constant dans l'équation. (2.10) par le facteur de lissage dépendant du temps et de la fréquence.
2. Dériver un facteur de biais pour l'estimation du bruit puisque le suivi du minimum est biaisé vers des valeurs plus faibles.
3. Amélioration de la vitesse de suivi de l'algorithme pour les niveaux élevés de bruit.

b) Détermination du facteur de lissage optimal

Le paramètre de lissage utilisé dans l'équation (2.10) doit être très faible pour suivre la non-stationnarité du signal de parole. D'un autre côté, il doit être proche de un (1) pour que la variance du suivi du minimum soit aussi petite que possible. Il est donc nécessaire d'utiliser des facteurs de lissage dépendant du temps et de la fréquence au lieu d'un facteur de lissage fixe.

La condition exigée est que le spectre de puissance lissé $P(\lambda,k)$ soit égal à la variance du bruit $\sigma_D^2(\lambda,k)$ pendant les segments de pauses de parole. Par conséquent, le paramètre de lissage est dérivé en minimisant l'erreur quadratique moyenne conditionnelle entre $P(\lambda,k)$ et $\sigma_D^2(\lambda,k)$ comme suit :

$$E\left\{\left(P(\lambda,k)-\sigma_D^2(\lambda,k)\right)^2 \middle| P(\lambda-1,k)\right\} \tag{2.11}$$

où :

$$P(\lambda,k)=\alpha(\lambda,k)\cdot P(\lambda-1,k)+\left(1-\alpha(\lambda,k)\right)\cdot\left|Y(\lambda,k)\right|^2 \tag{2.12}$$

Notez que dans l'équation (2.12), le facteur de lissage dépendant du temps et de la fréquence $\alpha(\lambda,k)$ a été utilisé au lieu du α fixe tel que défini dans (2.10). La substitution de (2.12) dans (2.11) et la mise à zéro de la dérivée première ont donné la valeur optimale pour $\alpha(\lambda,k)$:

$$\alpha_{opt}(\lambda,k)=\frac{1}{1+\left(P(\lambda-1,k)/\sigma_D^2(\lambda,k)-1\right)^2} \tag{2.13}$$

Mais dans la mise en œuvre en temps réel, la variance vraie du bruit $\sigma_D^2(\lambda,k)$ est remplacée par la dernière valeur estimée de la variance du bruit $\hat{\sigma}_D^2(\lambda-1,k)$. Un facteur de correction $\alpha_c(\lambda)$ a

été calculé en utilisant le rapport entre le périodogramme lissé moyen et la variance de bruit estimée [8]. Le paramètre de lissage final, après l'ajout du facteur de correction, est donné comme suit :

$$\hat{\alpha}_{opt}(\lambda,k)=\frac{\alpha_{\max}\cdot\alpha_c(\lambda)}{1+\left(P(\lambda-1,k)/\hat{\sigma}_D^2(\lambda-1,k)-1\right)^2} \tag{2.14}$$

où : $\alpha_{\max}=0.96$.

c) Détermination du facteur de biais

Comme le minimum est biaisé vers des valeurs inférieures, le facteur de biais pour compenser le minimum du spectre de puissance de la parole bruitée a été dérivé en utilisant les statistiques du minimum des estimations corrélées de la densité spectrale de puissance de la parole bruitée. Il a été indiqué que puisque la fonction de densité de probabilité (fdp) de $P(\lambda,k)$ est pondérée par $\sigma_D^2(\lambda,k)$, les statistiques minimales des estimations à court terme $P_{\min}(\lambda,k)$ sont également pondérées par $\sigma_D^2(\lambda,k)$. Ainsi, le terme de biais a été dérivé en trouvant la moyenne de la densité spectrale de puissance minimale pour $\sigma_D^2(\lambda,k)=1$ qui, après simplification, donne :

$$B_{\min}(\lambda,k)\approx 1+(L-1)\frac{2}{\tilde{Q}_{eq}(\lambda,k)} \tag{2.15}$$

où L est la longueur de la fenêtre sur laquelle le minimum a été trouvé et $\tilde{Q}_{eq}(\lambda,k)$ appelé « degrés de liberté équivalents », est une fonction du périodogramme lissé et de la variance précédente du bruit. L'estimation non biaisée du bruit est donc obtenue comme suit :

$$\hat{\sigma}_D^2(\lambda,k)=B_{\min}(\lambda,k)\cdot P_{\min}(\lambda,k) \tag{2.16}$$

d) Recherche efficace du minimum

Le minimum pour la parole bruitée a été trouvé sur L trames consécutives. Dans le pire des cas (niveaux élevés de bruit), la recherche du minimum est en retard de $2L$ trames. Pour réduire ce retard, le segment d'analyse de L trames a été divisé en U sous-fenêtres de V échantillons chacune ($L=U\cdot V$). Dans ce cas, le délai maximum a été réduit aux trames $L+V$ par rapport aux $2L$ trames dans le cas précédent.

2.2.1.2 Algorithme de suivi continu des minima spectraux

Dans [9], un algorithme d'estimation du bruit efficace sur le plan calcul a été proposé pour le rehaussement de la parole. Contrairement à l'utilisation d'une fenêtre de temps spécifiée pour suivre le minimum de parole bruitée, l'estimation du bruit a été mise à jour en continu en lissant séparément les spectres de puissance de parole bruitée dans chaque bande de fréquences à l'aide d'une règle de lissage non linéaire.

Pour suivre le minimum du spectre de puissance de la parole bruitée, une version lissée à court terme du périodogramme de la parole bruitée a été donnée comme suit :

$$P(\lambda,k)=\alpha\cdot P(\lambda-1,k)+(1-\alpha)\cdot\left|Y(\lambda,k)\right|^2 \tag{2.17}$$

avec α comme facteur d'oubli compris entre 0.7 et 0.9. La règle non linéaire utilisée pour estimer le spectre de bruit en suivant séparément le minimum du spectre de puissance de la parole bruitée dans chaque bande de fréquences est donnée comme suit :

$$\begin{cases} P_{\min}(\lambda,k)=\gamma\cdot P_{\min}(\lambda-1,k)+\dfrac{1-\gamma}{1-\beta}\left(P(\lambda,k)-\beta\cdot P(\lambda-1,k)\right) & \text{si } P_{\min}(\lambda-1,k)<P(\lambda,k) \\ P_{\min}(\lambda,k)=P(\lambda,k) & \text{ailleurs} \end{cases} \tag{2.18}$$

où $P_{\min}(\lambda,k)$ est l'estimation du bruit et les valeurs des paramètres sont $\alpha=0.7$, $\beta=0.96$ et $\gamma=0.998$.

De plus, le facteur β dans le suivi du minimum peut être ajusté pour faire varier le temps d'adaptation de l'algorithme. Le temps d'adaptation typique était de 0.2 à 0.4 s avec les valeurs mentionnées ci-dessus.

2.2.2 Algorithmes d'estimation du bruit à moyenne récursive

Les algorithmes de moyenne récursive dans le temps exploitent l'observation selon laquelle le signal de bruit a généralement un effet non uniforme sur le spectre de la parole, dans le sens où certaines régions du spectre sont plus affectées par le bruit que d'autres. En d'autres termes, chaque composante spectrale aura généralement un RSB effectif différent. Par conséquent, nous pouvons estimer et mettre à jour des bandes de fréquences individuelles du spectre de bruit chaque fois que le RSB effectif sur une bande de fréquences particulière est extrêmement faible. De manière équivalente, nous pouvons mettre à jour les bandes de fréquences individuelles du spectre de bruit

chaque fois que la probabilité que la parole soit présente sur une bande de fréquences particulière est faible. Cette observation a conduit au type d'algorithmes de moyenne récursive dans lesquels le spectre de bruit est estimé comme une moyenne pondérée des estimations de bruit passées et du spectre de parole bruitée actuel. Les poids changent de manière adaptative en fonction soit du RSB effectif de chaque bande de fréquences, soit de la probabilité de présence vocale.

2.2.2.1 Moyenne récursive contrôlée des minima (MCRA[3])

Dans [11], une nouvelle approche, appelée la moyenne récursive contrôlée des minima (MCRA), a été introduite pour l'estimation du bruit. Cette dernière a été mise à jour en faisant la moyenne des valeurs spectrales passées de la parole bruitée qui a été contrôlée par des facteurs de lissage en temps et en fréquence. Ces facteurs de lissage ont été calculés en fonction de la probabilité de présence du signal dans chaque bande de fréquence séparément. Cette probabilité a été à son tour calculée en utilisant le rapport du spectre de puissance de la parole bruitée à son minimum local calculé sur un temps de fenêtre fixe.

a) Estimation du spectre de bruit

La détermination du spectre de puissance du bruit à partir de la probabilité de présence du signal est basée sur les deux hypothèses suivantes :

$$\begin{cases} H_0(\lambda,k):\ Y(\lambda,k) = D(\lambda,k) \\ H_1(\lambda,k):\ Y(\lambda,k) = X(\lambda,k) + D(\lambda,k) \end{cases} \tag{2.18}$$

où $Y(\lambda,k)$, $X(\lambda,k)$ et $D(\lambda,k)$ représentent la transformée de Fourier à court terme de la parole bruitée, de la parole propre et du bruit respectivement et $H_0(\lambda,k)$ et $H_1(\lambda,k)$ représentent respectivement les hypothèses de parole absente et de parole présente. La variance du bruit est représentée par $\sigma_D^2(\lambda,k) = E\left[\left|D(\lambda,k)\right|^2\right]$. La mise à jour de l'estimation du bruit pour les deux hypothèses ci-dessus peut être écrite comme suit :

$$\begin{cases} H_0'(\lambda,k):\ \hat{\sigma}_D^2(\lambda+1,k) = \alpha_d \cdot \hat{\sigma}_D^2(\lambda,k) + (1-\alpha_d)\left|Y(\lambda,k)\right|^2 \\ H_1'(\lambda,k):\ \hat{\sigma}_D^2(\lambda+1,k) = \hat{\sigma}_D^2(\lambda,k) \end{cases} \tag{2.19}$$

[3] En anglais : **M**inima **C**ontrolled **R**ecursive **A**veraging (MCRA)

où $\hat{\sigma}_D^2(\lambda,k)$ est l'estimation de la variance du bruit et α_d ($0<\alpha_d<1$) est le facteur de lissage. L'équation (2.19) est basée sur le principe selon lequel l'estimation du bruit est mise à jour chaque fois qu'un silence est détecté, sinon elle reste constante. L'estimation globale du bruit a été obtenue sur la base de la probabilité de présence de la parole comme suit :

$$\hat{\sigma}_D^2(\lambda+1,k)=\hat{\sigma}_D^2(\lambda,k)\cdot p'(\lambda,k)+\left\{\alpha_d\cdot\hat{\sigma}_D^2(\lambda,k)+(1-\alpha_d)\left|Y(\lambda,k)\right|^2\right\}\cdot\left(1-p'(\lambda,k)\right) \quad (2.20)$$

où $p'(\lambda,k)=P\left(H_1'(\lambda,k)\middle|Y(\lambda,k)\right)$ dénote la probabilité conditionnelle de présence de la parole. La variance du bruit pour les deux hypothèses définies en (2.19) a été remplacée et simplifiée comme suit :

$$\hat{\sigma}_D^2(\lambda+1,k)=\tilde{\alpha}_d(\lambda,k)\cdot\hat{\sigma}_D^2(\lambda,k)+\left[1-\tilde{\alpha}_d(\lambda,k)\right]\cdot\left|Y(\lambda,k)\right|^2 \quad (2.21)$$

où

$$\tilde{\alpha}_d(\lambda,k)=\alpha_d+(1-\alpha_d)\cdot p'(\lambda,k) \quad (2.22)$$

b) Calcul de la probabilité de présence de la parole

La probabilité de présence vocale $p'(\lambda,k)$ dans chaque bande de fréquences est calculée en utilisant le rapport entre le spectre de puissance de la parole bruitée et son minimum local. Pour trouver le minimum local, un spectre de puissance lissé de parole bruitée utilisant une moyenne récursive de premier ordre a été calculé comme suit :

$$P(\lambda,k)=\alpha_s\cdot P(\lambda-1,k)+(1-\alpha_s)\cdot\left|Y(\lambda,k)\right|^2 \quad (2.23)$$

où α_s ($0<\alpha_s<1$) est le facteur de lissage. Cela a permis de réduire la variance du minimum local du spectre de puissance de la parole bruitée. Ensuite, le minimum local a été trouvé sur une longueur de fenêtre fixe de *L* trames par comparaison par échantillon du spectre de puissance lissé de la parole bruitée et du minimum local de la parole bruitée comme suit :

$$\begin{cases}P_{\min}(\lambda,k)=\min\left\{P_{\min}(\lambda-1,k),P(\lambda,k)\right\}\\ P_{tmp}(\lambda,k)=\min\left\{P_{tmp}(\lambda-1,k),P(\lambda,k)\right\}\end{cases} \quad (2.24)$$

où $P_{\min}(\lambda,k)$ et $P_{tmp}(\lambda,k)$ sont respectivement le minimum local et la variable temporaire. Pour chaque L trames traitées, la variable temporaire a été mise à jour comme suit pour éviter que $P_{\min}(\lambda,k)$ ne soit en retard sur le minimum global :

$$\begin{cases} P_{\min}(\lambda,k) = \min\left\{P_{tmp}(\lambda-1,k), P(\lambda,k)\right\} \\ P_{tmp}(\lambda,k) = P(\lambda,k) \end{cases} \tag{2.25}$$

Ce paramètre L est la longueur de la fenêtre sur laquelle le minimum est mis à jour. Ce L doit être choisi pour couvrir le pic le plus large du spectre de la parole. Ce paramètre contrôle également la mise à jour du spectre de puissance du bruit, en particulier pour les niveaux de bruit croissants. Sur la base des expériences [11] avec différents locuteurs et types de bruit, des longueurs de fenêtre de 0.5 à 1.5 s se sont révélées optimales. Une longueur de fenêtre très petite peut entraîner une surestimation du bruit si la longueur de la fenêtre est inférieure à la largeur du pic de parole lui-même. De plus, une longueur de fenêtre très grande retardera la mise à jour de l'estimation de la variance du bruit, en particulier pour les niveaux de bruit élevés.

Le rapport entre le spectre de puissance vocale bruitée et son minimum local est défini comme suit :

$$S_r(\lambda,k) = P(\lambda,k) / P_{\min}(\lambda,k) \tag{2.26}$$

Une règle de décision pour trouver les régions de présence de la parole a été déterminée en résolvant la fonction de coût minimum de Bayes et est donnée comme suit :

$$\begin{cases} \text{Si } S_r(\lambda,k) > \delta \quad \text{alors} \quad \text{présence de la parole} \quad \rightarrow I(\lambda,k) = 1 \\ \qquad\qquad\quad \text{sinon} \quad \text{absence de la parole} \quad \rightarrow I(\lambda,k) = 0 \end{cases} \tag{2.27}$$

La probabilité de présence de parole est obtenue en utilisant la règle de décision ci-après :

$$\hat{p}'(\lambda,k) = \alpha_p \cdot \hat{p}'(\lambda-1,k) + (1-\alpha_p) \cdot I(\lambda,k) \tag{2.28}$$

où α_p est une constante de lissage et $I(\lambda,k)$ est la fonction indicatrice de la décision de présence de parole définie en (2.27).

Après avoir calculé la probabilité de présence de parole, le facteur de lissage dépendant du temps et de la fréquence est calculé à l'aide de (2.22), puis la variance du bruit est mise à jour à l'aide de (2.21).

2.2.2.2 Algorithme MCRA modifiée (MCRA 2)

Soit le signal vocal bruité dans le domaine temporel donné comme suit :

$$y[n] = x[n] + d[n] \tag{2.29}$$

où $x[n]$ est la parole propre et $d[n]$ est le bruit additif. Le spectre de puissance lissé de la parole bruitée est calculé à l'aide de l'équation récursive du premier ordre suivante :

$$P(\lambda,k) = \eta \cdot P(\lambda-1,k) + (1-\eta)\cdot |Y(\lambda,k)|^2 \tag{2.30}$$

où $P(\lambda,k)$ est le spectre de puissance lissé, λ est l'indice de trame, *k* est l'indice de fréquence, $|Y(\lambda,k)|^2$ est le spectre de puissance à court terme de la parole bruitée et η est une constante de lissage.

a) Suivi du minimum de la parole bruitée

La règle non linéaire de Doblinger [9] est utilisée dans cet algorithme pour suivre le minimum de la parole bruitée en faisant continuellement la moyenne des valeurs spectrales passées :

$$\begin{cases} P_{\min}(\lambda,k) = \gamma \cdot P_{\min}(\lambda-1,k) + \dfrac{1-\gamma}{1-\beta}\left(P(\lambda,k) - \beta \cdot P(\lambda-1,k)\right) & \text{si } P_{\min}(\lambda-1,k) < P(\lambda,k) \\ P_{\min}(\lambda,k) = P(\lambda,k) & \text{ailleurs} \end{cases} \tag{2.31}$$

où $P_{\min}(\lambda,k)$ est le minimum local du spectre de puissance vocale bruitée et β et γ sont des constantes déterminées expérimentalement.

b) Probabilité de présence de la parole

L'approche adoptée pour déterminer la présence de la parole dans chaque bande de fréquences est similaire à l'algorithme précédent [11].

$$S_r(\lambda,k) = P(\lambda,k) / P_{\min}(\lambda,k) \tag{2.32}$$

Ce rapport est comparé à un seuil dépendant de la fréquence, et si le rapport s'avère supérieur au seuil, il est considéré comme une bande de fréquences où la parole est présente, sinon il est considéré comme une bande de fréquences où la parole est absente. Ceci est basé sur le principe selon lequel le spectre de puissance de la parole bruitée sera presque égal à son minimum local

lorsque la parole est absente. Par conséquent, plus le rapport est petit dans (2.32), plus la probabilité qu'il s'agisse d'une région uniquement bruitée est élevée et vice versa. La décision de présence vocale peut être résumée comme suit :

$$\begin{cases} \text{Si } S_r(\lambda,k) > \delta(k) & \text{alors} \quad \text{présence de la parole} \quad \rightarrow I(\lambda,k)=1 \\ & \text{sinon} \quad \text{absence de la parole} \quad \rightarrow I(\lambda,k)=0 \end{cases} \tag{2.33}$$

où $\delta(k)$ est le seuil dépendant de la fréquence déterminé expérimentalement. Notez que dans [11], un seuil fixe a été utilisé à la place de $\delta(k)$ pour toutes les fréquences. À partir de la règle ci-dessus, la probabilité de présence de parole, $p(\lambda,k)$, est mise à jour en utilisant la récursion de premier ordre suivante :

$$p(\lambda,k) = \alpha_p \cdot p(\lambda-1,k) + (1-\alpha_p) \cdot I(\lambda,k) \tag{2.34}$$

où α_p est une constante de lissage. Notez que la récursion ci-dessus exploite implicitement la corrélation pour la présence de parole dans les trames adjacentes.

c) Calcul des constantes de lissage dépendant de la fréquence

En utilisant l'estimation de probabilité de présence vocale ci-dessus, le facteur de lissage dépendant du temps et de la fréquence est calculé comme suit [11] :

$$\alpha_s(\lambda,k) = \alpha_d + (1-\alpha_d) \cdot p(\lambda,k) \tag{2.35}$$

où α_d est une constante. Notez que $\alpha_s(\lambda,k)$ prend ses valeurs dans la plage de :

$$\alpha_d \le \alpha_s(\lambda,k) \le 1 \tag{2.36}$$

d) Mise à jour de l'estimation du spectre de bruit

Enfin, après avoir calculé le facteur de lissage dépendant de la fréquence $\alpha_s(\lambda,k)$ à l'aide de l'équation (2.35), l'estimation du spectre de bruit est mise à jour comme suit :

$$D(\lambda,k) = \alpha_s(\lambda,k) \cdot D(\lambda-1,k) + (1-\alpha_s(\lambda,k)) \cdot |Y(\lambda,k)|^2 \tag{2.37}$$

où $D(\lambda,k)$ est l'estimation du spectre de puissance de bruit. Par conséquent, l'algorithme global peut être résumé comme suit. Après avoir classé les bandes de fréquence en parole

présente/absente à l'aide de l'équation (2.33), nous mettons à jour la probabilité de présence de la parole en utilisant l'équation (2.34) puis utiliser cette probabilité pour mettre à jour le facteur de lissage dépendant du temps-fréquence dans l'équation (2.35). Enfin, l'estimation du spectre de bruit est mise à jour selon l'équation (2.37) en utilisant le facteur de lissage dépendant du temps-fréquence.

Chapitre 3

Simulations et résultats

Il est nécessaire de réaliser des simulations pour vérifier la validité et la faisabilité d'un algorithme avant de pouvoir l'implémenter sur un système temps réel. L'implémentation sur ordinateur permet des modifications et des changements de l'algorithme sans contraintes de temps, de mémoire ou de puissance de calcul. Les simulations de ce projet ont été réalisées à l'aide de Matlab, un logiciel de calcul technique.

Ce chapitre décrit les détails de l'évaluation des performances des différents algorithmes d'estimation du bruit cités dans le chapitre précédent. Ces algorithmes sont implémentés au sein de la méthode de rehaussement de la parole par soustraction spectrale [7]. L'évaluation d'un algorithme de rehaussement de la parole n'est pas simple. Même si les méthodes objectives d'évaluation de la qualité peuvent indiquer une amélioration ou une dégradation de la qualité de la parole sur la base de mesures mathématiques, l'auditeur humain ne croit pas à un simple critère d'erreur mathématique. Par conséquent, des mesures subjectives de l'intelligibilité (compréhension effective de mots ou de phrases) et de la qualité (souffle, bruit, naturel de la voix, clarté, …) sont également nécessaires. La section 3.1 explique les mesures objectives qui

ont été utilisées pour évaluer l'algorithme. La section 3.2 décrit le matériel vocal utilisé pour tester l'algorithme. La section 3.3 traite les résultats obtenus par les simulations.

3.1 Mesures objectives d'évaluations

Lors de l'évaluation des algorithmes de rehaussement de la parole, il est nécessaire d'identifier les similitudes et les différences dans la qualité perçue et l'intelligibilité mesurée subjectivement. La qualité de la parole est un indicateur du « naturel » du signal vocal traité. L'intelligibilité des signaux vocaux est une mesure de la quantité d'informations vocales présentes dans le signal qui sont chargées de transmettre ce que dit le locuteur. La relation entre la parole perçue et l'intelligibilité n'est pas clairement comprise. Même si une parole inintelligible ne peut pas être considérée comme étant de haute qualité, l'inverse n'est peut-être pas vrai. Pour les auditeurs humains, il est important que le signal vocal soit intelligible, même au prix d'une certaine dégradation de la qualité de la parole. Par exemple, les utilisateurs finaux humains pourraient préférer une méthode de rehaussement moins agressive, qui pourrait ne pas supprimer complètement tout le bruit parasite, à un algorithme plus agressif qui pourrait supprimer complètement la composante de bruit mais également réduire l'intelligibilité de la parole. Les tests d'évaluation des performances peuvent être effectués de manière subjective ou objective. Tandis que les mesures subjectives fournissent une large mesure de la performance, puisqu'une grande différence de qualité est nécessaire pour pouvoir être distinguée par l'auditeur. Par conséquent, il devient difficile d'obtenir une mesure fiable des changements dus aux paramètres de l'algorithme. Les mesures objectives, en revanche, fournissent une mesure qui peut être facilement mise en œuvre et reproduite de manière fiable. Les mesures objectives sont basées sur une comparaison mathématique des signaux vocaux originaux et traités. La majorité des mesures objectives de qualité quantifient la qualité de la parole en termes de mesure numérique de distance ou de modèle de perception de la qualité de la parole par le système auditif humain. Il est souhaité que les mesures objectives soient cohérentes avec le jugement de la perception humaine de la parole.

Les mesures objectives de qualité des signaux vocaux les plus communément utilisées sont citées et classées dans le tableau 3.1.

Tableau 3.1 : Classification des critères d'évaluation objective les plus communément utilisés.

Mesures (domaine temporal)	**Mesures (domaine fréquentiel)**	**Mesures (domaine perceptuel)**
RSB[1]	IS[3]	PSQM[5]
RSBseg[2]	LLR[4]	PESQ[6]

Les critères temporels et fréquentiels se basent essentiellement sur l'évaluation de la qualité en termes de comparaison de distorsion de formes entre signal de référence et signal débruité, sans tenir compte de l'aspect perceptif. Certes, c'est une condition nécessaire mais non suffisante dans la mesure où deux signaux pratiquement de même forme peuvent être perçus différemment, d'où l'intérêt d'introduire le facteur psychoacoustique pour tout système ayant pour objectif de conserver la qualité de la parole. Diverses mesures objectives perceptuelles sont élaborées conduisant à de bonnes corrélations avec la perception humaine [12]. Elles sont essentiellement dédiées au codage de la parole, mais trouvent leur application en rehaussement de la parole. A part le fait qu'elles donnent une meilleure corrélation avec la qualité vocale, leur application en rehaussement n'a pas été justifiée jusqu'à présent.

3.1.1 Rapport signal sur bruit global et segmental

Le rapport signal sur bruit, *RSB* (en anglais signal to noise ratio *SNR*), comme son nom l'indique, fournit le rapport entre la puissance moyenne du signal et celle du bruit, c'est le critère le plus couramment utilisé pour désigner la qualité d'une transmission d'information par rapport aux parasites. Il est défini en décibel (dB) par :

$$RSB(dB) = 10 \cdot \log_{10}\left[\frac{\sum_{n=-\infty}^{\infty} x^2(n)}{\sum_{n=-\infty}^{\infty} d^2(n)}\right] \tag{3.1}$$

[1] Rapport Signal sur Bruit global

[2] Rapport Signal sur Bruit segmental

[3] Mesure d'Itakura-Saito

[4] Log-Likelihood Ratio

[5] Perceptual Speech Quality Measure

[6] Perceptual Evaluation of Speech Quality

où $x(n)$ est le signal propre ou rehaussé et $d(n)$ est le bruit.

Le signal de parole étant par nature non stationnaire, certains segments du signal peuvent avoir une énergie plus au moins grande. En supposant que l'énergie du bruit soit à peu près constante, le RSB pourra être soit très faible dans certains segments, soit très élevé dans d'autres segments. Etant donné que la corrélation du RSB global avec la qualité subjective est si médiocre elle est peu intéressante en tant que mesure objective générale de la qualité de la parole. En général, on préfère utiliser le RSB segmental qui est une moyenne des estimations des RSB_{seg} au niveau de chaque trame comme :

$$RSB_{seg} = \frac{10}{M} \cdot \sum_{m=0}^{M-1} \log_{10} \left(\frac{\sum_{i=0}^{N-1} x^2(n)}{\sum_{i=0}^{N-1} (x(n) - \hat{x}(n))^2} \right) \tag{3.2}$$

où N est la longueur de la trame (nombre d'échantillons) et M le nombre de trames dans le signal. La longueur de la trame est liée à la stationnarité du signal de la parole qui varie entre 10ms et 30ms. Comme le logarithme du rapport est calculé avant la moyenne, les trames avec un rapport exceptionnellement grand sont quelque peu moins pesées alors que les trames avec un rapport faible sont un peu plus élevées. On peut observer que ceci correspond bien à la qualité perceptuelle. C'est-à-dire que les trames avec une grande parole et aucun bruit audible ne dominent la qualité perceptive globale, l'existence de trames de bruit seul se détache et conduira à une qualité globale inférieure. Cependant si l'échantillon de parole contient un silence excessif, les valeurs globales du RSB_{seg} diminueront considérablement puisque les trames de silence montrent généralement de grandes valeurs RSB_{seg} négatives. Dans ce cas, les valeurs obtenues durant les trames de silence doivent être exclues de la moyenne en utilisant des détecteurs d'activité vocale. De la même manière l'exclusion des trames ayant des valeurs excessivement grandes ou faibles de la moyenne produit généralement des valeurs RSB_{seg} qui concordent bien avec la qualité. Les valeurs typiques pour la limite supérieure et inférieure sont 35 dB et -10 dB successivement [13].

3.1.2 PESQ (Perceptual Evaluation of Speech Quality)

Plusieurs méthodes de mesure de la qualité existent, la plus principalement utilisée est la note moyenne d'opinion (En anglais MOS : **M**ean **O**pinion **S**core). Le test MOS propose à l'auditeur cinq niveaux d'appréciation possibles {1 : Mauvais ; 2 : Médiocre ; 3 : Passable ; 4 : Bon ; 5 :

Passable). Le moyennage du score sur un nombre important d'auditeurs, donne une note entre 1 et 5 de l'agrément d'écoute.

Les valeurs de MOS sont fiables car elles sont basées sur la perception humaine. Un grand nombre d'auditeurs est requis, de sorte qu'une évaluation raisonnable puisse être faite. Ceci peut être long (demande beaucoup de temps) et cher. Par conséquent, diverses mesures objectives ont été développées et ont comme but de renvoyer la même valeurque celle du test MOS. Parmi eux, on trouve le PESQ (**P**erceptual**E**valuation of **S**peech **Q**uality) normalisé par ITU-T (Union Internationale des Télécommunications – Secteur Télécommunications) en Février 2001. Il est adopté comme la recommandation ITU-T P.862 [14]. Il a été montré que le PESQ peut fournir des résultats fortement corrélés avec les évaluations subjectives du test MOS.

Pour évaluer la qualité d'un signal traité par un réducteur de bruit en utilisant le PESQ, deux entrées sont exigées : le signal traité ou signal à tester, et un signal de référence (*ie.* signal original). La méthode de test est de prendre le signal de parole bruité et on le transmettre à travers le système PESQ et on le compare avec le signal de parole original, comme illustre la figure 3.1.

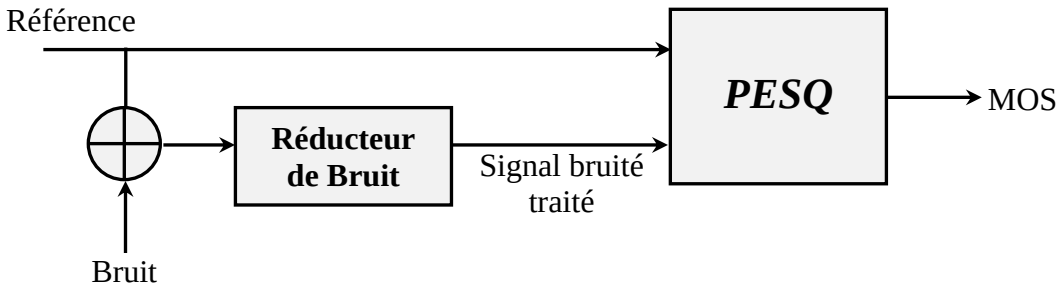

Figure 3.1 : Système PESQ pour l'évaluation des performances d'un réducteur de bruit.

3.1.3 Mesure d'Itakura-Saito IS

La mesure IS est basée sur les similarités et les différences entre le modèle tout-pôles du signal de parole propre et le signal de parole corrompu ou traité. Cette mesure pénalise tout décalage dans les emplacements des formants tandis que les erreurs dans les emplacements des vallées spectrales ne contribuent pas fortement à la distance. Il est calculé comme indiqué dans l'équation suivante [12] :

$$d_{IS}(\vec{a}_p, \vec{a}_c) = \frac{\sigma_c^2}{\sigma_p^2}\left(\frac{\vec{a}_p R_c \vec{a}_p^T}{\vec{a}_c R_p \vec{a}_c^T}\right) + \log_{10}\left(\frac{\sigma_p^2}{\sigma_c^2}\right) - 1 \quad (3.4)$$

où σ_c^2 et σ_p^2 sont les gains du modèle tout-pôles de la parole propre et rehaussée respectivement. $\vec{a}_c$ et $\vec{a}_p$ sont les vecteurs de coefficients de la prédiction linéaire de la parole propre et rehaussée respectivement. R_c et R_p sont les matrices d'autocorrélation de la prédiction linéaire de la parole propre et rehaussée respectivement.

Cette méthode a une corrélation de 0.59 avec les mesures subjectives. Une plage typique de résultats pour la mesure IS est de 1 à 10 telles que les valeurs inférieures indiquant une distance moindre et une meilleure qualité de la parole.

La distance IS a été utilisée comme mesure objective pour évaluer les performances des algorithmes étudiés. Les 5 % les plus élevés des valeurs de distance IS ont été rejetés, pour exclure les valeurs de distance spectrale excessivement élevées.

3.1.4 Likelihood Linear Regression (LLR)

Il est bien connu que le processus de production de la parole peut être modélisé de manière efficace avec un modèle de prédiction linéaire (LP). Il y a un certain nombre de mesures objectives qui utilisent la distance entre les deux ensembles de coefficients de prédiction linéaire (LPC) calculée l'original et la parole déformée. Nous ne discuterons que quelques-uns de ceux-ci. Le rapport de vraisemblance (LLR) est une mesure de distance qui peut être directement calculé à partir du vecteur LPC de la parole propre et déformée. La mesure LLR peut être calculée comme suit :

$$d_{LLR}(a_d, a_c) = \log\left(\frac{a_d R_c a_d^T}{a_c R_c a_c^T}\right) \quad (3.5)$$

où a_c est le vecteur LPC pour la parole propre, a_d est le vecteur LPC pour la parole déformée, a^T est la transposée de a, et R_c est la matrice de la prédiction linéaire de la parole propre.

3.2 Base données

Les tests, dans ce travail, ont été réalisés sur cinq (5) fichiers de parole prononcés en langue anglaise. Ils sont choisis parmi les trente phrases de la base de données NOIZEUS [15].

Cette fameuse base des signaux bruités conçu au niveau de l'université de Texas à Dallas par le laboratoire dirigé par Philipos C. Loizou afin de faciliter la comparaison des algorithmes de réduction de bruit. Cette base contient 30 phrases sélectionnées de la base IEEE de façon qu'elles inclure toutes les phonèmes dans la langue anglaise américaine, et sont enregistrées avec une fréquence d'échantillonnage de 25 kHz et après un sous-échantillonnage, cette fréquence devient 8 kHz.

Les bruits qui ont été ajoutés par Loizou *et al.* [15] aux signaux de parole propre sont les suivants : bruit de voiture (noté : **Car**), bruit enregistré dans un train (noté : **Train**), bruit enregistré dans une rue (noté : **Street**).

Les bruits sont ajoutés aux signaux de parole propre à 4 niveaux de RSB d'entrée, à savoir : 0 dB, 5 dB, 10 dB, 15 dB.

3.3 Résultats des tests

Les figures 3.2, 3.3, 3.4 et 3.5 illustrent des exemples de formes d'ondes des signaux propres, bruités et rehaussés de la 6ème phrase de la base NOIZEUS pour un bruit « Car » de la même base.

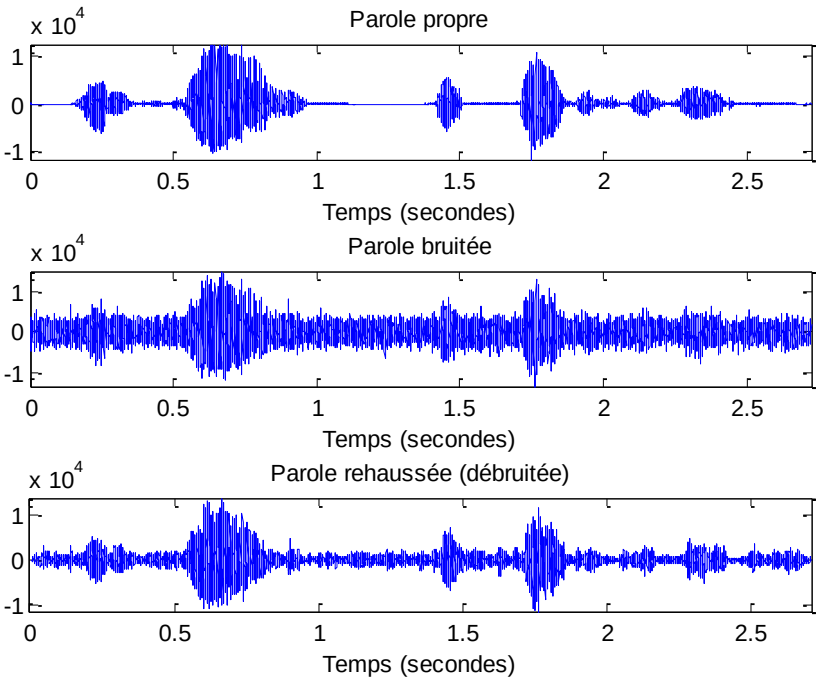

Figure 3.2 : Formes d'ondes du signal propre, du signal bruité par un bruit « **Car** » (RSB = 0 dB) et du signal rehaussé en utilisant l'algorithme d'estimation du bruit « Doblinger ».

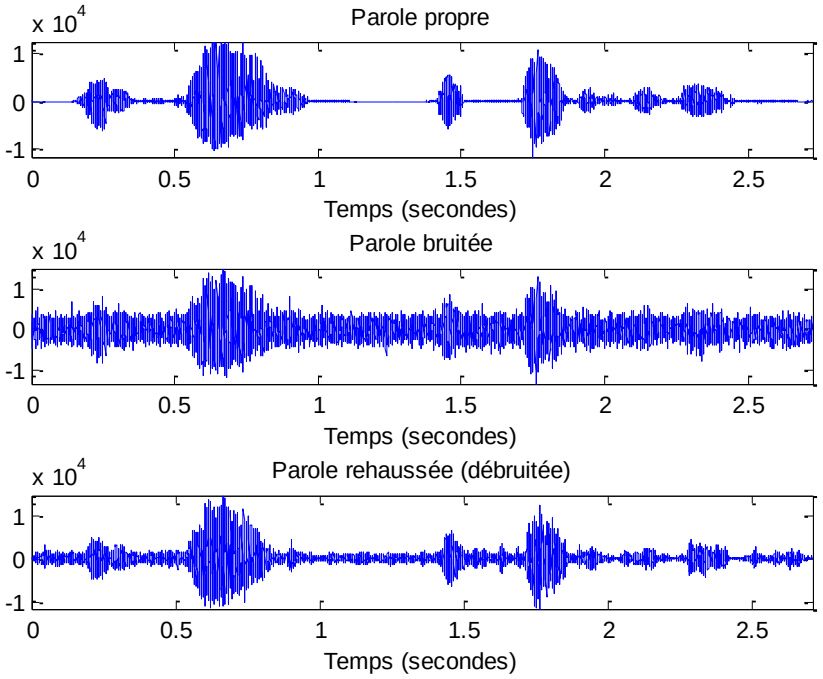

Figure 3.3 : Formes d'ondes du signal propre, du signal bruité par un bruit « **Car** » (RSB = 0 dB) et du signal rehaussé en utilisant l'algorithme d'estimation du bruit « Martin (MS) ».

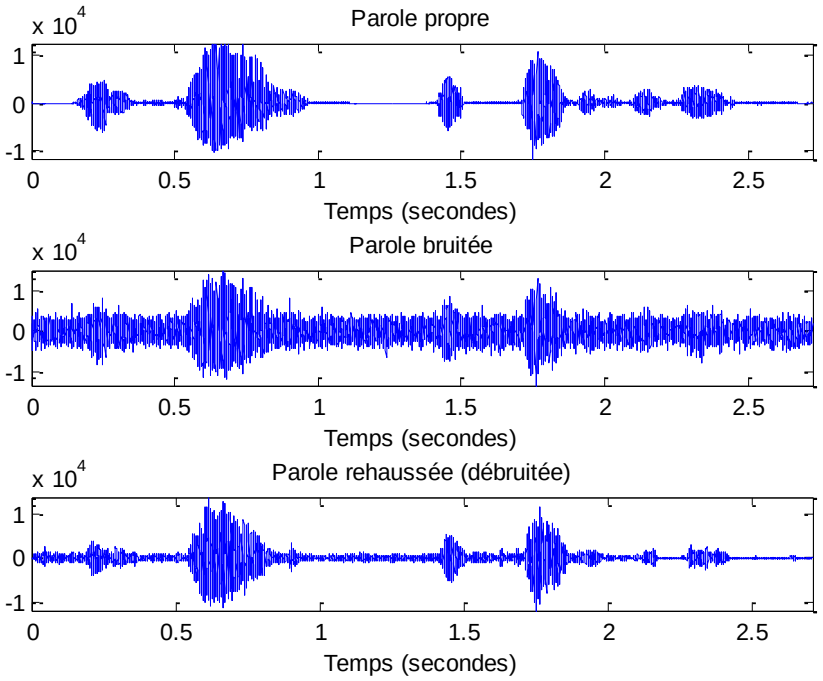

Figure 3.4 : Formes d'ondes du signal propre, du signal bruité par un bruit « **Car** » (RSB = 0 dB) et du signal rehaussé en utilisant l'algorithme d'estimation du bruit « MCRA ».

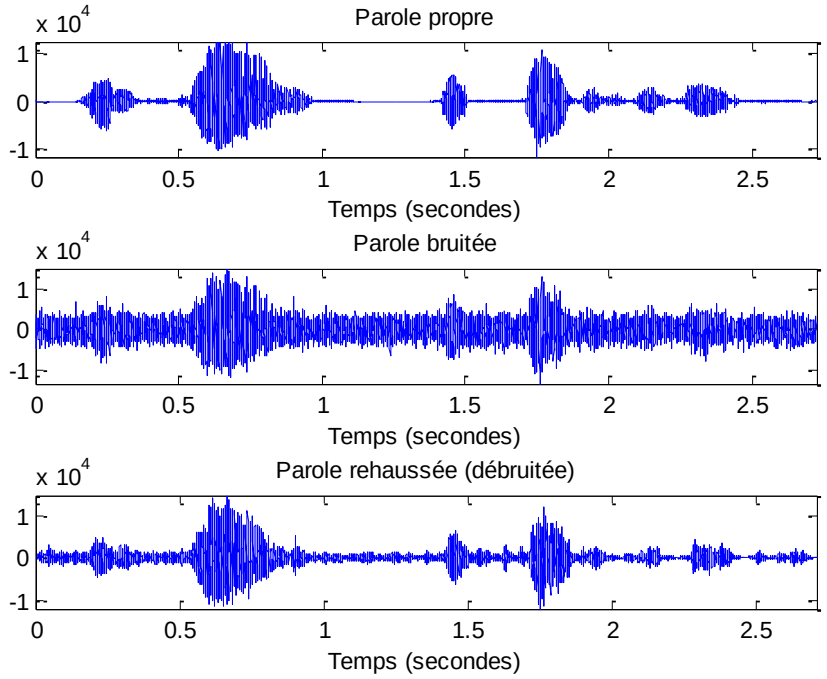

Figure 3.5 : Formes d'ondes du signal propre, du signal bruité par un bruit « **Car** » (RSB = 0 dB) et du signal rehaussé par soustraction spectrale en utilisant l'algorithme d'estimation du bruit « MCRA2 ».

Les performances des algorithmes d'estimation de bruit étudiés en termes de rapport signal sur bruit global et segmental sont données par les figures 3.6, 3.7, 3.8, 3.9, 3.10 et 3.11 pour trois genres de bruit à 4 niveaux. Dans ces figures, noter que l'allure noire « Signal bruité » donne la qualité de la parole bruitée avant l'application de la méthode de rehaussement du bruit.

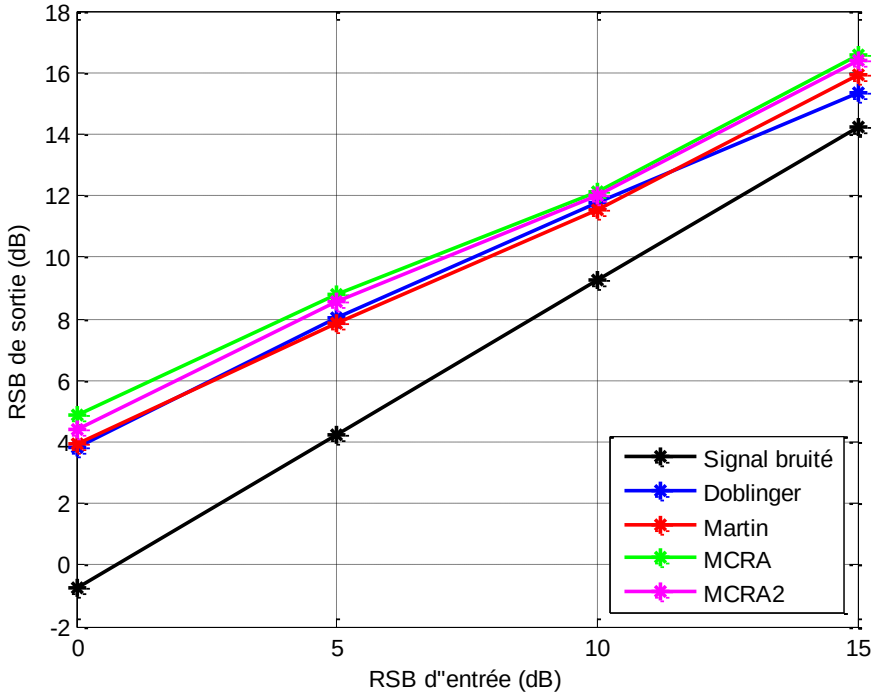

Figure 3.6 : Performances des algorithmes d'estimation du bruit étudiés en termes du RSB en fonction du RSB d'entrée pour un bruit « **Car** » de la base « NOIZEUS ».

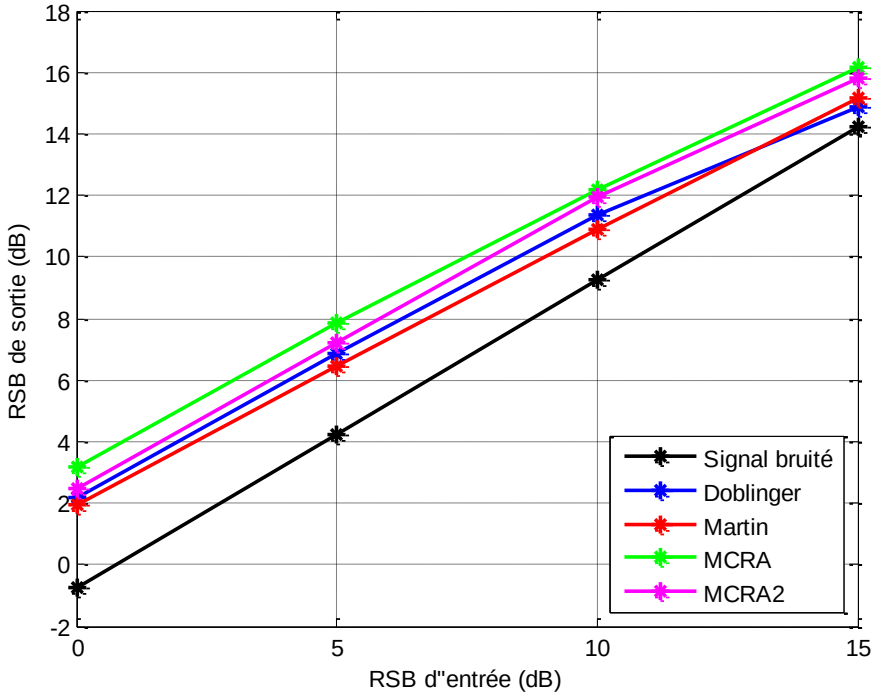

Figure 3.7 : Performances des algorithmes d'estimation du bruit étudiés en termes du RSB en fonction du RSB d'entrée pour un bruit « **Train** » de la base « NOIZEUS ».

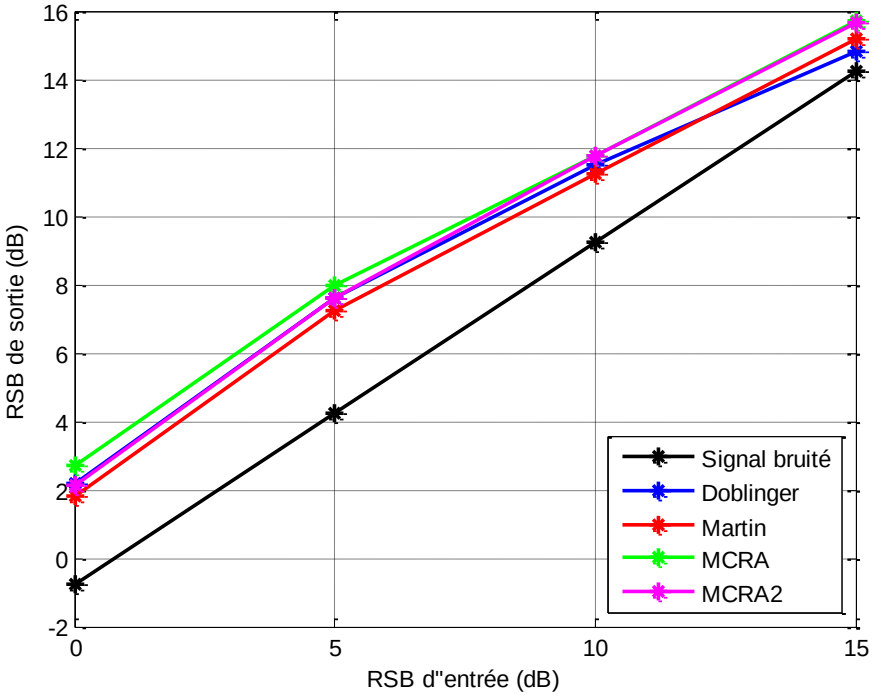

Figure 3.8 : Performances des algorithmes d'estimation du bruit étudiés en termes du RSB en fonction du RSB d'entrée pour un bruit « **Street** » de la base « NOIZEUS ».

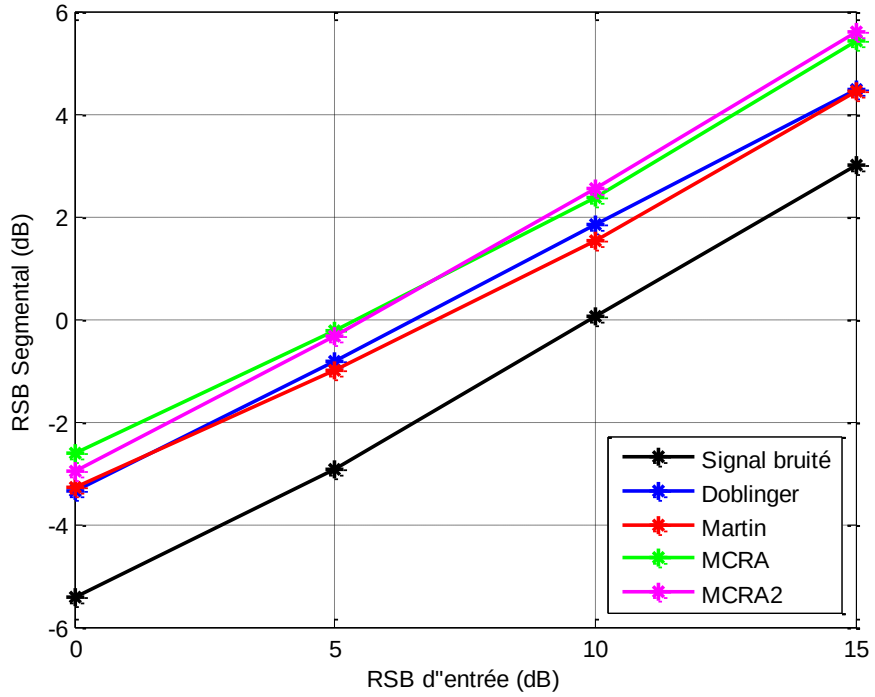

Figure 3.9 : Performances des algorithmes d'estimation du bruit étudiés en termes du RSB segmental en fonction du RSB d'entrée pour un bruit « **Car** » de la base « NOIZEUS ».

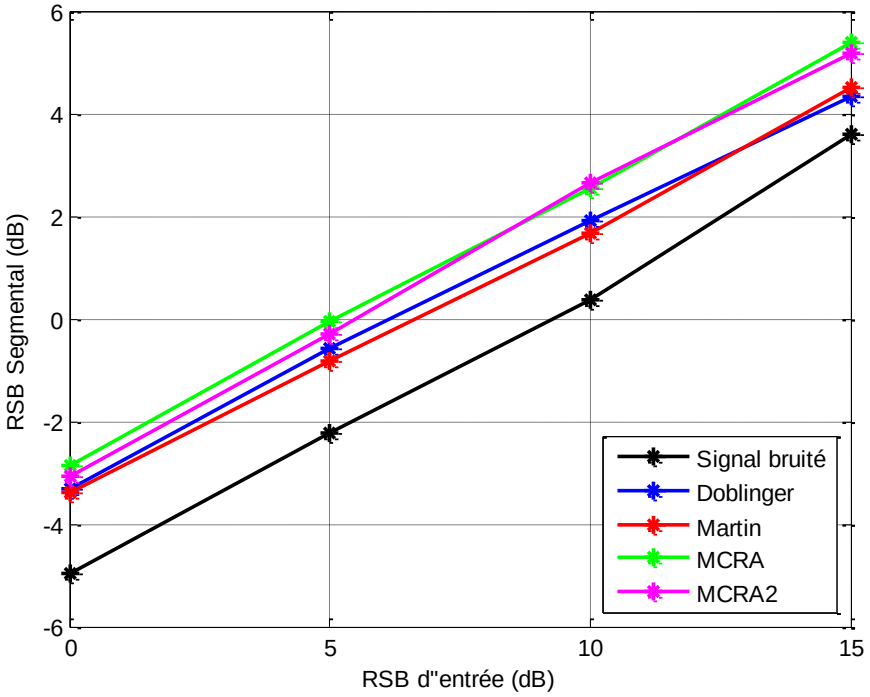

Figure 3.10 : Performances des algorithmes d'estimation du bruit étudiés en termes du RSB segmental en fonction du RSB d'entrée pour un bruit « **Train** » de la base « NOIZEUS ».

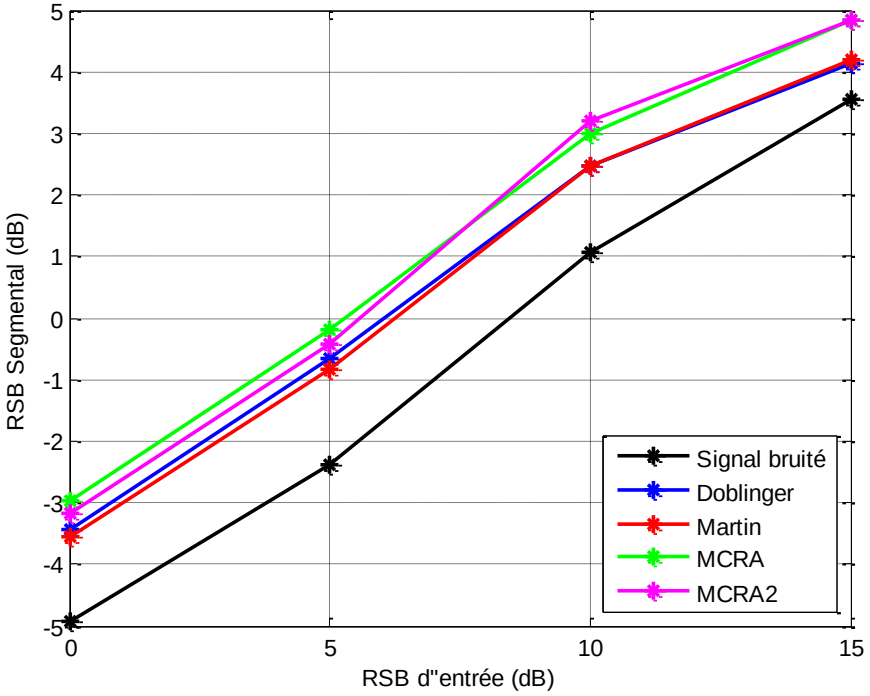

Figure 3.11 : Performances des algorithmes d'estimation du bruit étudiés en termes du RSB segmental en fonction du RSB d'entrée pour un bruit « **Street** » de la base « NOIZEUS ».

Les performances obtenues en termes de rapport signal sur bruit global et segmental montrent la supériorité des algorithmes d'estimation du bruit à moyenne récursive (MCRA et MCRA2) par rapport aux algorithmes de suivi des minima (Doblinger et Martin).

Il est connu que le rapport signal sur bruit global et segmental n'a pas une forte corrélation avec les mesures subjectives. Il est préférable d'utiliser un autre critère objectif afin de bien conclure sur l'algorithme le plus performant.

Une comparaison des performances pour différents types de bruits entre les différents algorithmes d'estimation du bruit étudiés est donnée par les figures 3.12, 3.13 et 3.14 en termes du PESQ.

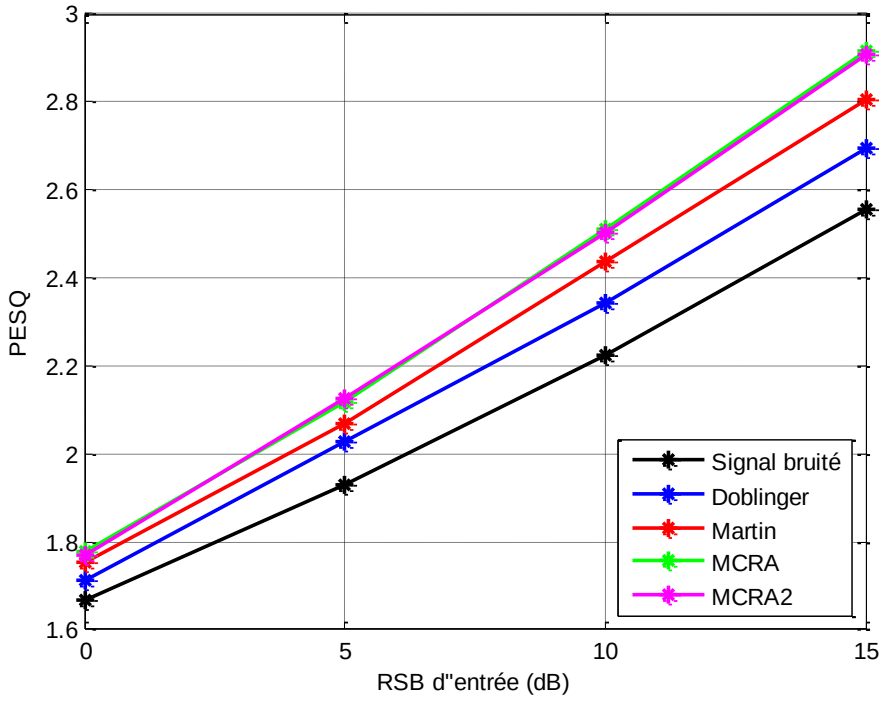

Figure 3.12 : Performances des algorithmes d'estimation du bruit étudiés en termes du PESQ en fonction du RSB d'entrée pour un bruit « **Car** » de la base « NOIZEUS ».

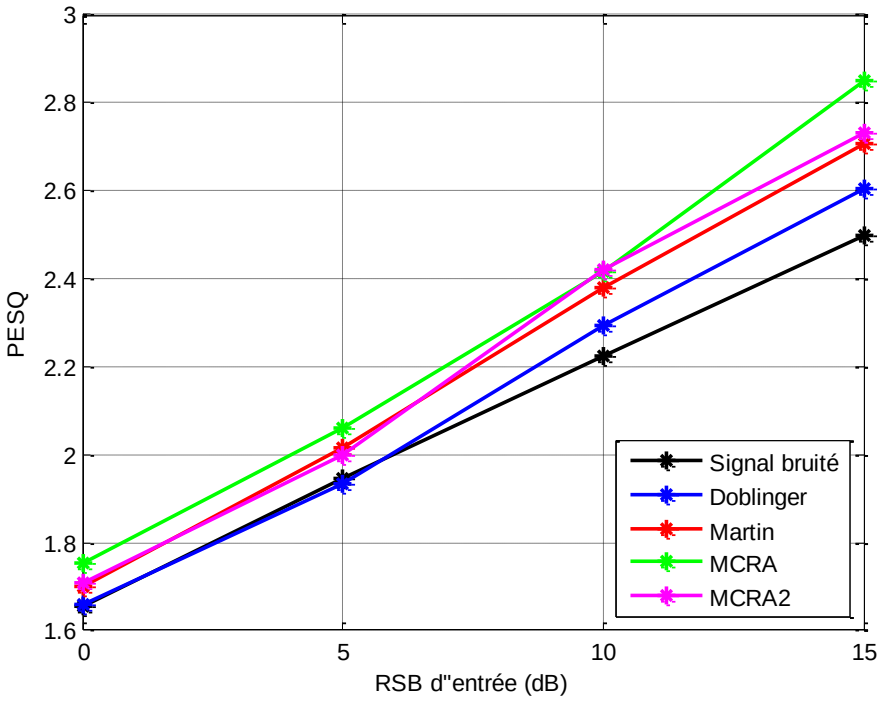

Figure 3.13 : Performances des algorithmes d'estimation du bruit étudiés en termes du PESQ en fonction du RSB d'entrée pour un bruit « **Train** » de la base « NOIZEUS ».

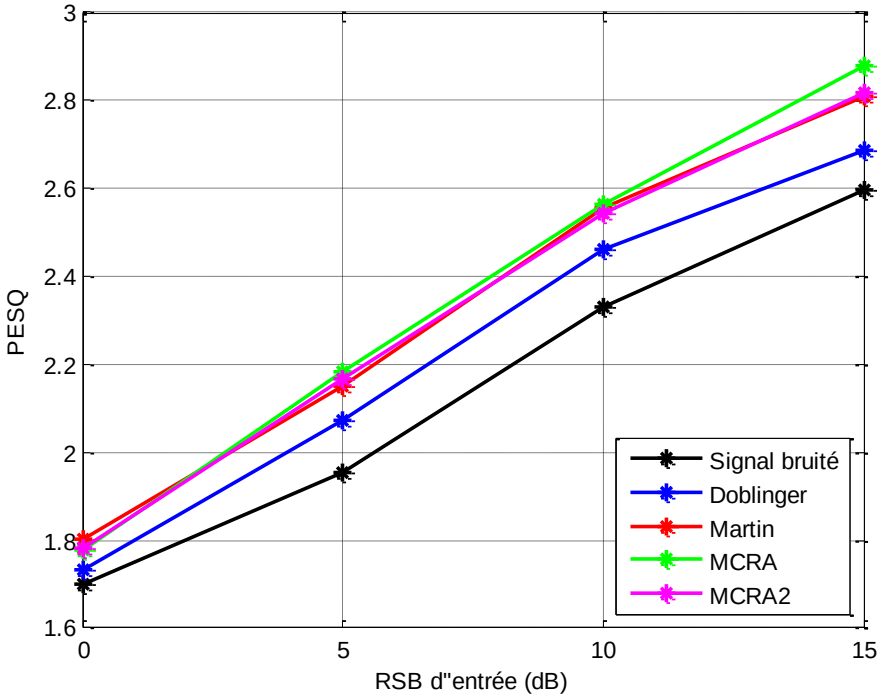

Figure 3.14 : Performances des algorithmes d'estimation du bruit étudiés en termes du PESQ en fonction du RSB d'entrée pour un bruit « **Street** » de la base « NOIZEUS ».

Les performances en termes du PESQ confirment également l'efficacité et la prééminence des algorithmes d'estimation du bruit à moyenne récursive (MCRA et MCRA2).

3.4 Conclusion

Les algorithmes d'estimation du bruit à moyenne récursive (MCRA et MCRA2) offre une nette amélioration par rapport algorithmes de suivi des minima (Doblinger et Martin). La moyenne récursive est donc une procédure intéressante pour estimer le spectre de puissance du bruit. Elle répond plus rapidement aux variations de bruit et, lorsqu'elle est intégrée à un système rehaussement de la parole, produit un RSB segmental plus élevé et un niveau de bruit résiduel musical plus faible.

Conclusion Générale

Lors d’une conversation, l’audition et la parole s’adaptent au bruit de fond dans un environnement bruité. Il est donc possible d'avoir une conversation dans des environnements sonores de fond assez dérangeants. Cependant, lorsque la conversation a lieu au téléphone, les perturbations sont plus gênantes. Les perturbations constituent un problème puisque le cerveau ne recevra pas les informations visuelles supplémentaires et autres informations de base lors de l’interprétation de la parole. Le signal vocal transmis à l'autre partie est capté par un microphone connecté au téléphone. Le signal du microphone contient à la fois de la parole et du bruit selon un certain rapport (rapport signal à bruit, RSB) qui dépend, par exemple, de la distance entre le microphone et la bouche de l'orateur.

Dans les véhicules par exemple, il est courant d'utiliser un accessoire téléphonique mains libres. La principale motivation est de faciliter d’avoir les deux mains sur le volant lors de la conduite. L’inconvénient est que le microphone du téléphone est situé à une plus grande distance (50 cm environ) de la bouche de l’interlocuteur par rapport à la téléphonie portable, dans un environnement très bruité. Pour augmenter le RSB et permettre à l'auditeur d’écouter clairement la parole, une méthode de rehaussement de la parole doit être appliquée.

Dans les systèmes de rehaussement de la parole à canal unique, seule la parole bruitée sera accessible et les statistiques de bruit doivent donc être estimées à partir de la parole bruitée

elle-même. Dans ce travail, nous avons abordé la question de l'estimation du bruit pour le rehaussement de la parole bruitée à canal unique.

Nous avons effectué une évaluation objective de la qualité de quatre différents algorithmes de d'estimation de bruit, à savoir : algorithme de statistiques minimales (MS ou de Martin), algorithme de suivi continu des minima spectraux (de Doblinger), moyenne récursive contrôlée des minima (MCRA) et algorithme MCRA modifiée (MCRA2). Ces algorithmes appartiennent à deux catégories principales, la première appelée « ***algorithmes de suivi des minima*** » englobe l'algorithme de Martin et ce de Doblinger, tandis que les algorithmes MCRA et MCRA2 font partie de la deuxième catégorie appelée « ***algorithmes d'estimation du bruit à moyenne récursive*** ». Ces algorithmes sont implémentés au sein de la méthode de rehaussement de la parole monocanal par soustraction spectrale. D'après les résultats des tests effectués sur des signaux pris de la base de données « NOIZEUS », on peut dire que les algorithmes d'estimation du bruit à moyenne récursive étudiés (MCRA et MCRA2) apportent une amélioration aux performances du signal de la parole et on peut affirmer son efficacité et sa supériorité par rapport aux algorithmes de suivi des minima dans le cas des bruits réels.

Bibliographie

[1] R. Boite, H. Bourlard, T. Dutoit, J. Hancq et H. Leich, *Traitement de la Parole*, Presses Polytechniques Universitaires Romandes, Lausanne, 2000.

[2] R. Boite, M. Kunt, *Traitement de la Parole*, Presses polytechniques romandes, 1987.

[3] L. Calliope, G. Fant, *La parole et son Traitement automatique*, Paris : Masson, 1989.

[4] L. R. Rabiner, *Digital processing of speech signal*, Prentice Hall, 1978.

[5] J. R. Deller, J. H. L. Hansen, J. G. Proakis, *Discrete Time Processing of Speech Signals*, (3e éd) New Jersey: IEEE Press, 1999.

[6] A. Amehraye, *Débruitage perceptuel de la parole*, Thèse de doctorat, Ecole Nationale Supérieure des Télécommunications de Bretagne, Bretagne, 2009

[7] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," *Proc ICASSP*, pp. 208-211, Apr. 1979.

[8] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. on speech and audio processing*, vol. 9, no. 5, pp. 504-512, July 2001.

[9] G. Doblinger, "Computationally efficient speech enhancement by spectral minima tracking

in subbands," in *Proc. 4th Eur. Conf. speech, Communication, and Technology, EUROSPEECH'95*, Madrid, Spain, pp. 1513-1516, Sept. 18-21, 1995.

[10] R. Martin, "Spectral subtraction based on minimum statistics," in *Proc. 7th Eur. Signal Processing Conf. (EUSIPCO'94)*, Edinburgh, U.K., pp. 1182-1185, Sept. 13-16, 1994.

[11] I. Cohen, and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Signal Proc. Letters*, vol. 9, no. 1, pp. 12-15, Jan. 2002.

[12] S. R. Quickenbush. T. P Barnwell and M. A. Clements, *Objective measures of Speech quality*, Prentice-Hall, N.J, USA. 377p. 1988.

[13] J. Hansen and B. Pellom, "An effective quality evaluation protocol for speech enhancement algorithm," in *Proceedings of ICSLP*, Sydney. Australia, pp. 2819-2822, Dec. 1998.

[14] ITU-T, ITU-Recommendation P.862 Perceptual Evaluation of Speech Quality (PESQ) : An Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs, 2001.

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

[16] S. Rangachari, *Noise Estimation Algorithms for Highly Non-Stationary Environments*, Master Thesis, University of Dallas, 2004.