République Tunisienne
Ministère de l'Enseignement Supérieur
et de la Recherche Scientifique

Université de Sfax
École Nationale d'Ingénieurs de Sfax

Ecole Doctorale
Sciences et Technologies

Thèse de *DOCTORAT*

Génie Electrique
N° d'ordre: 50-2011

THESE

Présentée à

L'École Nationale d'Ingénieurs de Sfax

En vue de l'obtention du

DOCTORAT

Dans la discipline *Génie Electrique*

Par

Fathi KALLEL

Assistant à l'Institut Supérieur d'Informatique
et du Multimédia de Gabès

Algorithmes de Réduction du Bruit en Vue d'une Amélioration de l'Intelligibilité de la Parole: Cas de la Prothèse Cochléaire

Soutenue le 13 Décembre 2011, devant le jury composé de :

M.	**Lotfi KAMMOUN**	Professeur à l'ENIS	*Président*
M.	**Noureddine ELLOUZE**	Professeur à l'ENIT	*Rapporteur*
M.	**Jaime LOPEZ-KRAHE**	Professeur à l'Université Paris 8	*Rapporteur*
M.	**Mounir SAMET**	Professeur à l'ENIS	*Examinateur*
M.	**Ahmed BEN HAMIDA**	Professeur à l'ENIS	*Directeur*
M.	**Christian BERGER-VACHON**	Professeur à l'UCBL	*Directeur*
M.	**Dan GNANSIA**	Responsable de recherche à Neurelec	*Invité Industriel*
M.	**Abdelmoneem GHORBEL**	Professeur au CHU de Sfax	*Invité Médical*

Année 2011

THESE DE L'UNIVERSITE DE LYON

délivrée par

L'UNIVERSITE CLAUDE BERNARD LYON 1

et préparée en cotutelle avec

L'ECOLE NATIONALE D'INGENIEURS DE SFAX

ECOLE DOCTORALE
Electronique Electrotechnique Automatique

DIPLOME DE DOCTORAT

(arrêté du 7 août 2006 / arrêté du 6 janvier 2005)

soutenue publiquement le 13 Décembre 2011

par

M Fathi KALLEL

Algorithmes de Réduction du Bruit en Vue d'une Amélioration de l'Intelligibilité de la Parole: Cas de la Prothèse Cochléaire

Directeurs de thèse:
Pr Ahmed BEN HAMIDA, Professeur
Pr Christian BERGER-VACHON, Professeur

JURY: M. Lotfi KAMMOUN, Professeur
M. Noureddine ELLOUZE, Professeur
M. Jaime LOPEZ-KRAHE, Professeur
M. Mounir SAMET, Professeur

A la mémoire de mes grands parents...
A la mémoire de mon père...
A ma mère...
A ma femme et ma fille...
A ma soeur...
A mes frères et leurs femmes...
A mes beaux parents...
A ma belle soeur...
A mon beau frère...
A mes ami(e)s...

Remerciements

Je tiens tout d'abord à exprimer toute ma gratitude et mes vifs remerciements à mes professeurs Ahmed BEN HAMIDA et Christian BERGER-VACHON de m'avoir accordé leur confiance en me proposant ce sujet de thèse et d'avoir dirigé mon travail. Leur savoir faire et leurs conseils m'étaient précieux pour mener à bien ce travail. Je les prie de trouver ici l'expression de ma profonde reconnaissance.

Je remercie cordialement le professeur Lotfi KAMMOUN de l'Ecole Nationale d'Ingénieurs de Sfax de m'avoir honoré en acceptant de présider le jury de cette soutenance de thèse. J'adresse également mes remerciements à mes deux rapporteurs, Pr Noureddine ELLOUZE de l'Ecole Nationale d'Ingénieurs de Tunis et Pr Jaime LOPEZ-KRAHE de l'université Paris 8, pour leur lecture attentive de la thèse et les remarques constructives qu'ils m'ont faites. Je suis aussi très reconnaissant au Pr Mounir SAMET de l'Ecole Nationale d'Ingénieurs de Sfax pour avoir examiné cette thèse. Un sincère remerciement aux Pr Abdelmoneem GHORBEL, chef du service ORL à l'hôpital Habib BOURGUIBA de Sfax et Dr Dan GNANSIA, responsable de recherche à Neurelec d'avoir accepté de participer au jury de ma thèse.

Je tiens à remercier tous les membres du laboratoire d'Electronique et des Technologies de l'Information et de l'unité de recherche Advanced Technologies for Medecine and Signals de l'Ecole Nationale d'Ingénieurs de Sfax. Je remercie également tous les membres du groupe Parole, Audiologie, Communication & Santé du Centre de Recherche en Neurosciences de Lyon pour leur accueil et les moments agréables passés durant ma thèse.

Je voudrais aussi exprimer mes plus profonds remerciements à tous les membres de ma famille KALLEL et ma belle famille KAMMOUN. Plus particulièrement à, ma mère Saloua, ma belle mère Mongia, mon beau père Abdelwaheb pour leurs encouragements, ma femme Amal pour son soutien illimité pour atteindre ce but, sa compréhension mais aussi sa patience. Je n'oublierai pas évidement mon adorable fille Emna qui par sa présence et son amour, m'a sans cesse revitalisé et donné du courage pour atteindre cet objectif. Je finirais par une pensée pour mon père Adel qui nous a quitté un jour triste de Janvier 2002. Il m'a instruit durant les années scolaires et sans cesse aidé financièrement et moralement durant ma carrière académique sans pour autant eu l'occasion de voir cette thèse achevée.

J'aimerais enfin remercier tous ceux que j'ai côtoyés au cours de ces années à l'université de Sfax, à l'université de Lyon 1 et à l'université de Gabès, et qui m'ont aidé de près ou de loin dans la réalisation de ce travail.

Il est toujours difficile de remercier l'ensemble des personnes qui ont contribué à l'aboutissement de ce travail de recherche. Que ceux qui ne sont pas mentionnés ne m'en tiennent pas rigueur.

Table des matières

Remerciements **ii**

Table des figures **ix**

Liste des algorithmes **xii**

Liste des tableaux **1**

Introduction Générale **2**

1 Implant Cochléaire et Effet du Bruit **6**

1 Introduction 7

2 Système auditif humain et mécanisme de l'audition humaine 7

2.1 Physiologie du système auditif humain 7

2.2 Mécanisme de l'audition humaine 9

3 Réhabilitation de la surdité 10

3.1 Aides auditives 10

3.2 Implants cochléaires 11

4 Implant cochléaire : Principe de fonctionnement 12

4.1 Outils de traitement du signal de parole pour l'implant cochléaire . 13

4.1.1 Analyse et filtrage du signal de parole 13

4.1.2 Détection d'enveloppe 14

4.1.3 Vitesse de stimulation 16

4.1.4 Fonction de compression 17

4.2 Stratégie de codage pour implant cochléaire 18

4.2.1 Stratégies à base d'extraction des formants 18

4.2.1.a Multi-Peak ou MPEAK 18

4.2.1.b Spectral Maxima Sound Processor ou SMSP 18

4.2.2 Stratégie favorisant la représentation fréquentielle 18
4.2.2.a Spectral Peak Strategy ou SPS 18
4.2.2.b Adaptative Stimulation Rate ou ASR 19
4.2.3 Stratégies favorisant la représentation temporelle 19
4.2.3.a Compressed Analog ou CA 19
4.2.3.b Continuous Interleaved Sampling ou CIS 19
4.2.3.c Simultaneus Analogue Stimulation ou SAS 19
4.2.4 Stratégie hybride 19
4.2.4.a Spectral Peak Strategy ou SPEAK 19
4.2.4.b Advanced Combination Encoders ou ACE 20
5 Différents types d'appareillages 20
6 Nature et caractéristiques du bruit 21
6.1 Différentes classes du bruit 21
6.2 Bruit musical 21
7 Implant cochléaire et effet du bruit 22
7.1 Effet du bruit 'speech-shaped noise' sur la reconnaissance des consonnes et de la parole avec la stratégie de codage CIS 22
7.2 Effet du bruit SSN sur la reconnaissance des consonnes et des voyelles avec la stratégie de codage SPEAK 23
7.3 Effet du bruit SSN sur la reconnaissance des consonnes, des voyelles et de la parole continue avec les stratégies de codage SPEAK, CIS et SAS 23
7.4 Effet du bruit de type 'Babble' multilocuteurs sur la reconnaissance de la parole continue avec les stratégies de codage SPEAK, CIS et SAS 24
8 Conclusion 24

2 Techniques de base pour la réduction de Bruit 25
1 Introduction 26
2 Techniques mono-voie pour la réduction de bruit 26
2.1 Importance de la phase 27
2.2 Rapport signal à bruit : RSB a priori et RSB à postériori 27
2.3 Algorithme d'estimation du bruit à base de DAV 28
2.4 Estimation de $P_{xmin}(f,k)$ basée sur l'utilisation d'une fenêtre de recherche 30
2.5 Estimation de $P_{xmin}(f,k)$ suivant une méthode récursive 32

2.6 Algorithme d'estimation du signal débruité 33
2.6.1 Estimateurs basés sur les statistiques moyennes du bruit . 33
2.6.1.a Soustraction spectrale 33
2.6.1.b Soustraction spectrale multi-bandes 35
2.6.1.c Filtre de Wiener 35
2.6.2 Estimateurs avec modèle de parole 36
2.6.2.a Filtre de MacAulay et Malpass 37
2.6.2.b Filtre d'Ephraim et Malah 37
2.7 Algorithmes de réduction de bruit mono-voie dédiés pour implant cochléaire 38
2.7.1 Méthode de la soustraction spectrale non linéaire 38
2.7.2 Méthode à sous-espace signal 39
2.7.3 Exploitation de la fonction sigmoïdale pour la réduction du bruit 40
2.7.4 Exploitation de la fonction de compression pour le débruitage du signal de parole 41
3 Aperçu général sur les techniques de réduction du bruit bi-voie 41
3.1 Hypothèse des bruits corrélés 42
3.2 Hypothèse de bruits décorrélés 43
3.2.1 Cohérence 43
3.2.2 Formation de voie 45
3.2.2.a Formation de voie fixe 45
3.2.2.b Formation de voie adaptative 46
3.2.2.c Algorithme de la soustraction spectrale bi-voie . . . 47
3.3 Hypothèse de bruits diffus ou peu corrélés 49
4 Conclusion 50

3 Effet de la distribution fréquentielle sur l'intelligibilité de la parole en milieux silencieux et bruité : Cas de l'implant cochléaire bilatéral 52

1 Introduction 53
2 Stimulation cochléaire bilatérale vs unilatérale 53
3 Simulateur d'implant 55
4 Comparaison des performances d'une stimulation bilatérale symétrique et décalée 61
4.1 Matériel et méthode 62

4.1.1 Sujets 62
4.1.2 Matériel 62
4.1.3 Protocole expérimental 62
4.2 Résultats 64
4.2.1 Représentation des signaux 64
4.2.2 Evaluation subjective 66
4.2.2.a Environnement silencieux 68
4.2.2.b RSB=6 dB 68
4.2.2.c RSB=0 dB 68
4.2.2.d RSB=-6 dB 69
4.2.3 Discussion 69
4.2.3.a Stimulation bilatérale vs Unilatérale 69
4.2.3.b Vitesse d'analye 69
5 Conclusion 70

4 Exploitation de la technique de soustraction spectrale bi-voie pour la réduction du bruit : Cas de l'implant cochléaire bilatéral 71

1 Introduction 72
2 Algorithmes de la soustraction spectrale bi-voie pour la réduction du bruit 72
2.1 Estimation des dsp : un point crucial 73
2.2 Estimation bi-voie de la dsp du bruit 74
3 Algorithmes d'estimation des signaux de parole rehaussés 77
3.1 Algorithme de la soustarction spectrale non linéaire bi-voie 78
3.2 Algorithme de la soustraction spectrale multi-bande bi-voie 78
4 Etude comparative et évaluation des performances 80
4.1 Matériel phonétique 81
4.2 Evaluation objective 82
4.3 Evaluation subjective 85
4.3.1 Sujets 85
4.3.2 Procédure 85
4.3.3 Résultats 86
4.3.3.a Résultats en simulation avec des sujets normoentendants 86
4.3.3.b Résultats avec des sujets bilatéralement implantés . 88

4.3.4 Discussion 90
4.3.4.a Algorithme de réduction du bruit vs RSB 90
4.3.4.b Algorithme de réduction du bruit vs configuration des sources de bruit 90
5 Conclusion 91

5 Exploitation de la technique de soustraction interspectrale pour la réduction du bruit : Cas de l'implant cochléaire bilatéral 92
1 Introduction 93
2 Présentation 93
3 Algorithme de rehaussement 94
3.1 Soustraction interspectrale 95
3.2 Soustraction interspectrale et bruit musical 95
3.3 Estimation des dsp et de la dip des signaux bruités 96
4 Estimation du bruit 97
4.1 Estimation à base d'une détection d'activité vocale 97
4.1.1 Critère de détection 98
4.1.2 Seuil de détection 99
4.2 Estimation permanente de la dip des bruits 100
4.2.1 Estimation récursive 101
4.2.2 Estimation basée sur les informations de phase 101
4.2.3 Estimation en ligne de la dip des bruits 102
4.2.4 Estimation basée sur la technique des statistiques minimales 103
4.2.5 Algorithme proposé pour l'estimation de la dip des bruits 104
5 Evaluation et résultats 106
5.1 Etude comparative préliminaire 106
5.2 Evaluation de l'intelligibilité 107
5.2.1 Matériel phonétique 107
5.2.2 Evaluation objective 109
5.2.3 Evaluation subjective 109
5.2.3.a Sujets 110
5.2.3.b Procédure 110
5.2.3.c Résultats 110
5.3 Discussion 113

5.3.1 Signaux rehaussés vs signaux bruités 113
5.3.2 Algorithme de réduction du bruit vs RSB 113
5.3.3 Algorithme de réduction du bruit vs configuration des sources de bruit 114
5.3.4 Consistance des résultats 114
6 Conclusion 116

Conclusion Générale **117**

Annexe **120**

Bibliographie **122**

Liste des Publications **129**

Table des figures

1.1 Représentation schématique du système auditif humain 7

1.2 Représentation schématique de la cochlée humaine 8

1.3 Représentation schématique de l'implant cochléaire 12

1.4 Représentation schématique des différentes étapes de traitement du signal de parole 13

1.5 Représentation temporelle du signal de parole (rouge) et des signaux d'enveloppes estimés (bleu) 16

2.1 Comportement du DAV en milieux bruité pour des traitements avec et sans chevauchement à différents RSB 29

2.2 Haut : dsp des signaux de parole et de bruit, Bas : dsp du signal bruité pour $\lambda = 0.6$ et $\lambda = 0.9$ 30

2.3 psd estimée du signal bruité pour $\lambda = 0.9$ et $\hat{\lambda}(f, k)$ pour $f = 15$ 31

2.4 psd réelle et estimée du bruit suivant les algorithmes de Doblinger et Farsi pour $f = 15$ 32

2.5 psd réelle et estimée du bruit suivant la méthode de Rangachari pour $f = 15$ 33

2.6 Annulation de bruit avec référence de bruit seul 43

2.7 Formation de voie fixe de type retard et soustraction 46

2.8 Représentation du Beamformer adaptatif 47

2.9 MSC théorique et réelle dans le cas d'un bruit de type Babble 49

3.1 Exemple de représentation fréquentielle du banc de filtres d'analyse (Nb=12) 56

3.2 Répartition en Hz et en Barks des différentes bandes d'analyse dans le cas du mode SBD 59

3.3 Représentation schématique des différents blocs du simulateur d'implant cochléaire considéré 60

3.4 Principe de simulation de la vitesse d'analyse 61

3.5 Protocole expérimental 63

3.6 Représentation temporelle des signaux de parole synthétisés pour chaque canal de stimulation 64
3.7 Représentation temporelle du signal enveloppe estimé 64
3.8 Représentation temps-fréquence du mot 'agis' pour les différents modes de stimulation 65
3.9 Performances obtenues dans le cas des trois modes de stimulation pour différents RSB à 250 Hz 66
3.10 Performances obtenues dans le cas des trois modes de stimulation pour différents RSB à 500 Hz 66

4.1 Schéma global du système de réduction de bruit considéré 73
4.2 Effet du coefficient d'oubli sur l'estimation de la MSC, pour $f = 300Hz$ . . 74
4.3 Variation temporelle de la dsp réelle du bruit au niveau des voies droite et gauche 75
4.4 Représentation temporelle des dsp réelle et estimée du bruit au niveau des voies droite et gauche 77
4.5 RSBS pour différentes bandes fréquentielles en fonction des trames 79
4.6 Protocole expérimental 81
4.7 SSMB-BV : Scores PESQ pour différent nombre de bandes fréquentielles . 83
4.8 SSMB-BV : Distances IS pour différent nombre de bandes fréquentielles . . 83
4.9 Représentation temps-fréquence des différents signaux de parole 84
4.10 Performances des algorithmes SSNL-BV et SSMB-BV en présence d'une seule source de bruit à différents niveaux du RSB 87
4.11 Performances des algorithmes SSNL-BV et SSMB-BV en présence de trois sources de bruit à différents niveaux du RSB 87
4.12 Pourcentage de reconnaissance des phonèmes par sujet pour les algorithmes SSNL-BV et SSMB-BV à différents RSB : Cas d'une seule source de bruit . 89
4.13 Pourcentage de reconnaissance des phonèmes par sujet pour les algorithmes SSNL-BV et SSMB-BV à différents RSB : Cas de trois sources de bruit . . 89

5.1 Représentation schématique du système de réduction de bruit 94
5.2 Représentation temps-fréquence des bruit résiduels dans le cas de l'algorithme de la (a)soustraction spectrale (b) soustraction interspectrale 96
5.3 Caractéristiques du bruit diffus 98
5.4 Evaluation des performances du DAV considérée dans le cas d'un bruit Babble à un RSB=0dB 100
5.5 Schéma bloc de l'estimateur de la dip des bruits proposé 105
5.6 dip réelle et estimées des bruits en utilisant les approches DAV, SM et SMA 106

5.7 Protocole expérimental 108
5.8 Performances de l'algorithme de la SIS dans le cas de chacun des estimateurs de la dip des bruits considérés en présence d'une seule source de bruit 111
5.9 Performances de l'algorithme de la SIS dans le cas de chacun des estimateurs de la dip des bruits considérés en présence de trois sources de bruit . 111
5.10 Comparaison des performances des approches SBD, SSMB-BV et SIS . . . 115

Liste des algorithmes

1 Répartition fréquentielle des différentes bandes d'analyse 57
2 Différentes étapes nécessaires pour l'estimation des signaux rehaussés . . . 105

Liste des tableaux

1.1 *Principaux implants cochléaires unilatéral commercialisés* 20

1.2 *Différentes classes du bruit* 21

3.1 *Configuration des différentes bandes d'analyse : mode SU* 58

3.2 *Configuration des différentes bandes d'analyse : mode SBS* 58

3.3 *Configuration des différentes bandes d'analyse : mode SBD* 59

3.4 *Pourcentages de reconnaissance moyens en milieux silencieux et bruités aux différents niveaux du RSB* 67

4.1 *Nombre de raies spectrales pour chaque bande de fréquences et pour différents nombres de bandes d'analyse* 82

4.2 *Données biographiques des patients implantés* 85

5.1 *Temps de traitement pour l'estimation de la dip des bruits avec les méthodes DAV, SM et SMA* 107

5.2 *Scores PESQ des signaux bruité et rehaussés* 109

5.3 *Comparaison des performances des algorithmes de la soustraction interspectrale et la soustraction spectrale* 115

Introduction Générale

La surdité ou bien le dysfonctionnement du système auditif est un handicap qui peut être parfois grave pour l'être humain. Cette surdité conduit le malentendant à vivre dans un monde de silence, isolé de toute vie sociale. Les avancées technologiques en ingénierie biomédicale associées aux développements de la médecine ont permis une amélioration globale des conditions de vie et une augmentation sensible de la longévité de l'homme en donnant naissance à de nouveaux appareillages biomédicaux tel que la prothèse cochléaire. En effet, la prothèse cochléaire est un appareillage biomédical, implantable au niveau de l'oreille humaine, qui permet de faire bénéficier certaines personnes atteintes d'une surdité profonde ou totale bilatérale d'un niveau d'audition inaccessible avec les prothèses auditives traditionnelles.

La prothèse cochléaire assure la stimulation directe des neurones cochléaires. La pose d'électrodes dans des zones bien définies dans la cochlée permet de stimuler sélectivement les cellules sensorielles suivant différentes fréquences et différentes intensités électriques qui sont générées suite à un traitement spécifique du signal de parole. La stimulation électrique apportée par les électrodes permet un niveau de compréhension intéressant tout en notant une adaptation progressive du patient à son appareillage suite à une phase de rééducation.

Différents travaux de recherche ont été établis afin d'évaluer l'intelligibilité de la parole chez les sujets implantés en environnements silencieux et bruité. Les résultats ont montré une bonne intelligibilité de la parole variant entre 80% et 90% en milieu silencieux. Toutefois, les capacités de perception de la parole par les patients implantés se dégradent en environnement bruité. Afin d'améliorer l'intelligibilité de la parole en milieu bruité, différents algorithmes de réduction de bruit, appelés aussi algorithmes de débruitage, ont été développés dans le cas de l'implant cochléaire. Ces algorithmes peuvent être classés principalement en deux catégories. Pour la première catégorie, les algorithmes de débruitage sont intégrés en totalité au niveau de la stratégie de codage adopté pour l'analyse du signal de parole au niveau de l'implant en modifiant certaine fonction tel que la fonction de compression. Pour la deuxième catégorie, une étape de prétraitement basés sur un algorithme de débruitage spécifique est tout d'abord adoptée. Le signal de parole rehaussé ainsi obtenu à la suite de cette étape de prétraitement est ensuite traité par l'algorithme de codage utilisé par l'implant. Dans ce cas, différents algorithmes de débruitage initialement développés pour des sujets normoentendants ont été adoptés dans le cas de l'implant cochléaire.

Le développement progressif de la technologie a rendu possible l'implantation des patients d'une manière bilatérale. L'implantation cochléaire bilatérale consiste à poser un implant cochléaire au niveau de chaque oreille. Elle permet de plus en plus l'accès à l'audition humaine naturelle qui est une audition binaurale. Ceci est assuré grâce à la restauration des capacités de localisation spatiale, l'amélioration de la sélectivité fréquentielle et de la reconnaissance de la parole dans le bruit. De nombreux travaux de recherche ont mis en évidence les performances d'une stimulation bilatérale par rapport à une stimulation unilatérale en milieux silencieux et bruité. Les résultats ont montré qu'une stimulation bilatérale permet d'améliorer l'intelligibilité de la parole surtout en milieu bruité par rapport à une stimulation unilatérale. D'autre part, le patient implanté se trouve très souvent dans des environnements où le niveau de bruit est assez élevé. Dans ce cas, même avec une stimulation bilatérale, l'intelligibilité de la parole reste insuffisante et la perception sonore est difficile. Afin d'améliorer davantage l'intelligibilité en milieu bruité, différents algorithmes de débruitage ont été développés dans le cas de l'implant cochléaire bilatéral. Ces algorithmes, basés sur un traitement multi-microphones, présentent l'inconvénient du coût assez élevé et de la complexité. Un traitement à base de deux microphones (chaque implant est équipé d'un seul microphone) pourrait être considéré comme étant le meilleur compromis entre un traitement mono-microphone et un traitement multi-microphones.

Dans ce travail de thèse, nous traitons le problème de l'intelligibilité de la parole dans le cas de l'implant cochléaire, particulièrement l'implant cochléaire bilatéral, en proposant de nouvelles approches de traitement du signal pour le débruitage du signal de parole.

Le présent rapport est structuré en cinq chapitres qui se résument par ce qui suit :

Dans le premier chapitre, nous commençons par donner un aperçu général sur la physiologie de l'audition humaine et les différents types d'appareillages biomédicaux dédiés à la réhabilitation à la surdité. Nous nous intéressons par la suite à décrire le principe de fonctionnement de l'implant cochléaire en détaillant les différents outils de traitement du signal exploités pour l'analyse et le codage du signal de parole ainsi que leur effet sur l'intelligibilité de la parole chez les patients implantés. Nous introduisons enfin la notion de bruit et son influence sur l'intelligibilité de la parole chez les sujets implantés.

Le deuxième chapitre est consacré à présenter une étude bibliographique des principaux algorithmes proposés dans la littérature pour la réduction du bruit. Ces algorithmes sont classés en deux catégories : les algorithmes mono-voie et les algorithmes bi-voie.
Dans le cas des algorithmes mono-voie, une seule observation bruitée issue d'un seul microphone est exploité pour la réduction du bruit et l'estimation du signal rehaussé. Dans une première partie, nous nous intéressons à l'estimation de la dsp du bruit : une composante principale dans tout système de débruitage. En effet, la qualité du signal de parole rehaussé dépend énormément de la qualité de l'estimateur du bruit considéré. Dans une deuxième partie, nous présentons différents algorithmes de débruitage pour l'estimation des signaux rehaussés. Nous nous intéressons en particulier aux algorithmes de la soustraction spectrale, soustraction spectrale multi-bande et le filtrage de Wiener. Nous présentons aussi quelques algorithmes de débruitage mono-voie proposés dans le cas de l'implant cochléaire tel que le filtrage à base de la fonction sigmoïdale.

Dans le cas des algorithmes bi-voie, deux observations bruitées issues de deux microphones sont exploitées pour le rehaussement des signaux bruités. Une présentation de différentes techniques de débruitage bi-voie sera présentée suivant l'hypothèse sur les signaux de bruit (corrélé, décorrélé ou diffus). Dans le cas de l'implant cochléaire bilatéral, les travaux traitant le problème de réduction du bruit sont minimes et la majorité est basée sur la technique du Beamformer adaptatif.

Dans le troisième chapitre, nous présentons une première approche de traitement du signal de parole visant l'amélioration de l'intelligibilité de la parole dans le cas de l'implant cochléaire. Le protocle expérimental adopté pour la comparaison des performances de trois modes de stimulation cochléaire est alors présenté. Les performances de ces trois modes de stimulation ont été comparées dans le cas de deux vitesses d'analyse, en milieux silencieux et bruité à différents niveaux du RSB. Le premier mode est la stimulation unilatérale. Le deuxième mode est la stimulation bilatérale symétrique où les bancs de filtres utilisés pour l'analyse des signaux de parole au niveau des voies droite et gauche sont identiques. Le troisième mode est la stimulation bilatérale décalée où les deux oreilles sont stimulées par des signaux différents (le banc de filtres utilisé pour l'analyse du signal de parole au niveau de la voie droite est fréquentiellement décalé par rapport à celui de la voie droite). Les performances de ces trois modes de stimulation ont été comparées d'une manière subjective avec une population de cinquante sujets normoentendants.

Dans le quatrième chapitre, les performances de deux algorithmes bi-voie pour la réduction du bruit dédiés pour implant cochléaire bilatéral sont présentées. Ces algorithmes sont basés sur deux étapes de traitement : Une étape d'estimation de la densité spectrale de puissance (dsp) du bruit au niveau de chaque voie suivi d'une étape d'estimation des signaux rehaussés. Le principe d'un algorithme bi-voie pour l'estimation de la dsp du bruit est d'abord présenté. Cet algorithme a été basé sur le calcul des dsp et de la densité interspectrale (dip) des signaux bruités et il a été développé sous l'hypothèse des bruits parfaitement décorrélés. L'estimation des signaux rehaussés a été déterminée en se basant sur la technique de la soustraction spectrale. Le principe des algorithmes de la soustraction spectrale non linéaire bi-voie et la soustraction spectrale multi-bande bi-voie sont alors détaillés. Les performances de ces deux algorithmes ont été évaluées et comparées d'une manière subjective dans le cas de l'implant cochléaire bilatéral, en simulation en premier lieu avec des sujets normoentendants puis avec des sujets implantés portant l'implant cochléaire Digisonic SP binaural de Neurelec. Cette étude comparative est établie à différents niveaux du RSB et dans le cas de deux configurations spatiales des sources de bruit (d'abord en présence d'une seule source de bruit, puis en présence de trois sources de bruit).

Dans le dernier chapitre, un deuxième algorithme de réduction de bruit dédié pour l'implant cochléaire bilatéral est proposé. Cet algorithme est basé sur la méthode de la soustraction interspectrale qui est développé sous l'hypothèse des bruits peu corrélés ou diffus. L'exploitation de l'algorithme de la soustraction interspectrale repose principalement sur l'estimation de la dip des bruits. Dans ce travail, une nouvelle approche basée sur l'algorithme des statistiques minimales améliorées a été proposée pour l'estimation de la dip des bruits. Deux autres approches présentées dans la littérature basées sur la

technique de détection d'activité vocale (Guérin, 2002) et la technique des statistiques minimales (Rahmani *et al.*, 2009b) ont été aussi considérées. Les performances de l'algorithme de la soustraction interspectrale sont évaluées dans le cadre de l'implant cochléaire bilatéral en simulation dans le cas de chacun des estimateurs de la dip des bruits considérés avec des sujets normo-entendants. Cette étude comparative est établie dans les mêmes conditions expérimentales que l'expérience précédente (différents niveaux du RSB et deux configurations des sources de bruit).

Chapitre 1

Implant Cochléaire et Effet du Bruit

Sommaire

1	Introduction	7
2	Système auditif humain et mécanisme de l'audition humaine	7
3	Réhabilitation de la surdité	10
4	Implant cochléaire : Principe de fonctionnement	12
5	Différents types d'appareillages	20
6	Nature et caractéristiques du bruit	21
7	Implant cochléaire et effet du bruit	22
8	Conclusion	24

1 Introduction

La perception des sons sous forme de parole et de musique est l'une des activités quotidiennes de notre vie humaine. D'autre part, l'audition est un sens fragile qui subit l'attaque du temps de manière irréversible conduisant souvent à une surdité. En effet, pour un malentendant, le déficit auditif peut représenter un handicap majeur tant du point de vue physique que du point de vue social. Le progrès réalisé ces dernières années dans les domaines de la neurophysiologie, de l'électronique et aussi de la chirurgie a permit de restaurer, chez des personnes atteintes d'un certain niveau de surdité, une audition inaccessible avec les prothèses conventionnelles grâce aux prothèses cochléaires. La prothèse cochléaire comporte principalement deux modules : un implant fabriqué autour d'une puce dédiée utilisant des technologies électroniques de pointe assurant la stimulation des terminaisons nerveuses auditives et un processeur externe permettant le traitement et l'analyse des signaux de parole captés de l'environnement extérieur. Le contrôle et l'alimentation de l'unité implantée est assurée via un lien inductif.

2 Système auditif humain et mécanisme de l'audition humaine

2.1 Physiologie du système auditif humain

Le système auditif humain possède des capacités remarquables. Sa sensibilité s'étend sur une plage de fréquences audibles allant de 20 Hz à 20 kHz et couvre un ensemble de pressions acoustiques de 20 μPa à 20 Pa. En outre, les capacités de discrimination aussi bien en fréquence qu'en intensité sont excellentes. Par contre, nous remarquons une complexité notable de fonctionnement de la chaîne de traitement que constitue l'oreille. En effet, l'appareil auditif est divisé en trois éléments de nature et de fonctions différentes mais complémentaires : l'oreille externe, l'oreille moyenne et l'oreille interne. Une représentation schématique du système auditif humain est donnée par la figure 1.1.

FIGURE 1.1 – Représentation schématique du système auditif humain

L'oreille externe est le premier maillon de la chaîne que constitue l'appareil auditif. Elle reçoit les vibrations acoustiques aériennes et les transmet à l'oreille moyenne. L'oreille externe est formée d'une structure cartilagineuse appelée pavillon, du canal auditif et du tympan. L'oreille moyenne débute par la membrane tympanique ou le tympan qui a une forme d'entonnoir et possède un diamètre d'environ 10 mm. Le tympan vibre sous l'effet

de la pression acoustique qui est recueillie et focalisée par l'oreille externe. Cette vibration sera transmise à une chaîne d'osselets qui est formée par le marteau, l'enclume et un étrier fortement relié au tympan. Ces trois osselets sont maintenus ensembles en forme d'un pont flexible reliant le tympan à la fenêtre ovale. Le dernier élément de la chaîne auditive est l'oreille interne qui est l'organe principal de l'audition. L'oreille interne est responsable de la transduction du signal acoustique en message nerveux. Son anatomie et sa physiologie très complexes sont à l'origine des capacités auditives très performantes. Elle regroupe deux organes sensoriels distincts : le vestibule, organe de l'équilibre et la cochlée, organe de l'audition. La cochlée est un canal de forme spirale, enroulé un certain nombre de tours, d'une longueur d'environ 35 mm et d'un diamètre de l'ordre de 1.3 mm à 2.2 mm. Une représentation schématique de la cochlée humaine est donnée par la figure 1.2.

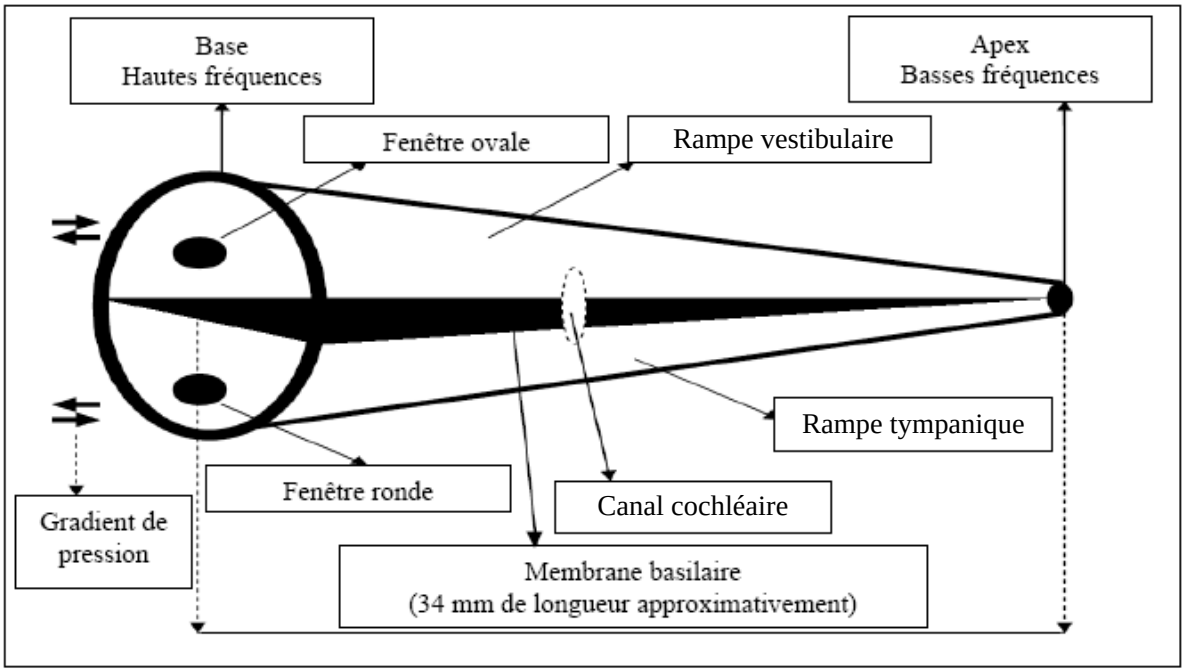


FIGURE 1.2 – Représentation schématique de la cochlée humaine

La cochlée permet de convertir les sons transmis par le tympan et la chaîne des osselets en signaux électriques acheminés au cerveau à travers le nerf auditif. En effet, le système auditif humain agit comme un ensemble de filtres passe-bande chevauchés afin d'effectuer une analyse spectrale du signal de parole. Ces filtres passe-bande sont connus sous le nom des bandes critiques ou des filtres auditifs. La cochlée est constituée de deux rampes qui se rejoignent au niveau de son apex. La première est la rampe tympanique qui aboutit à la fenêtre ronde. La seconde est la rampe vestibulaire qui s'ouvre dans l'appareil vestibulaire. Entre ces deux rampes se trouve le canal cochléaire qui est séparé de la rampe tympanique par la membrane basilaire et séparé de la rampe vestibulaire par la membrane de Reissner. Sur la membrane basilaire se trouvent des cellules ciliées qui transmettent des impulsions électriques aux nerfs auditifs dès l'oscillation de la membrane basilaire. Ces cellules ciliées sont stimulées plus ou moins fortement en fonction du volume sonore.

2.2 Mécanisme de l'audition humaine

Le pavillon capte les vibrations acoustiques et les concentre dans le canal auditif pour les diriger vers le tympan qui se met alors à vibrer. Le canal auditif assure une première amplification des sons collectés d'environ 5 dB à 10 dB pour les fréquences variant autour de 3500 Hz. A partir du tympan, la méthode de conduction devient mécanique en faisant mouvoir la chaîne d'osselets. Une amplification de l'onde sonore d'environ 27 dB est notée au niveau de l'oreille moyenne. Les ondes sonores se transforment par le système de levier formé par les osselets en vibrations mécaniques. Ces vibrations se transforment au marteau fixé par son manche au tympan. Le marteau transmet le mouvement à l'enclume qui fait mouvoir à son tour l'étrier. La platine de l'étrier, fixé sur la fenêtre ovale, induit le déplacement des fluides cochléaires dans la rampe vestibulaire. Les mouvements de ces liquides provoquent des déplacements de la membrane basilaire. A partir de ce moment commence la neurophysiologie de l'audition : Les cellules ciliées internes qui baignent dans ces liquides vont créer un influx suite à un phénomène bioélectrique engendré par leurs cils flottants. L'influx nerveux se transmet par le nerf auditif aux aires cérébrales pour la reconnaissance et la traduction. Sans ce phénomène d'adaptation d'impédance entre le milieu externe (l'air) et le milieu interne (les fluides cochléaires), la majeure partie de l'énergie acoustique serait réfléchie ou perdue. Enfin, la trompe d'Eustache permet de compenser les différences de pression qui s'établissent dans l'oreille moyenne lors des oscillations assurant ainsi un certain équilibre pour éviter certaines confusions.
D'autre part, l'audition humaine est une audition naturellement binaurale, c'est-à-dire une audition stéréophonique assurée par deux oreilles : l'oreille droite et l'oreille gauche. Cette audition binaurale ne permet pas seulement une meilleure intelligibilité de la parole surtout en milieu bruité, mais permet aussi l'identification et la localisation de la position d'une source sonore dans l'espace. Cette particularité est liée notamment à différents phénomènes appelés indices binauraux. Rayleigh (Rayleigh, 1907) a proposé la théorie du duplex qui indique que la localisation d'un signal sonore dans l'espace est déterminée principalement grâce à deux caractéristiques du signal sonore stéréophonique : la différence de temps interaurale et la différence d'intensité interaurale. En effet, l'être humain n'est jamais parfaitement immobile (mouvements de la tête), et ce sont ces petits mouvements qui impliquent de petites variations sur lesquelles le cerveau se base pour son analyse de localisation.

- La différence de temps interaurale (ITD) : Elle correspond à la différence de temps d'arrivée du signal acoustique aux deux oreilles. En effet, la diffraction du son dans l'air implique que l'effet de l'ITD soit plus élevé en basses fréquences qu'en hautes fréquences. Pour la tête humaine, l'ITD varie entre -700 µs et 700 µs (Kuhn, 1977; Wightman et Kistler, 1992) avec une ITD nulle lorsque le signal sonore provient d'une source positionnée à un angle d'incidence de 0°.

- La différence d'intensité interaurale (IID) : Elle correspond à la différence d'intensité du signal acoustique qui existent entre les signaux arrivant sur les deux oreilles. Cette différence d'intensité est utilisée par le système auditif pour les fréquences supérieures à 3000 Hz. Lorsqu'une source sonore est placée du côté de l'oreille latérale, la tête produit un effet d'ombrage appelé ombrage de la tête au niveau de l'oreille controlatérale entrainant ainsi une différence d'amplitude entre les deux oreilles. La

différence d'intensité IID peut varier, en valeur absolue, de 20 dB à 25 dB suivant l'angle d'incidence et la fréquence du signal sonore (Hartmann, 1997).

3 Réhabilitation de la surdité

La description du fonctionnement de l'oreille qui vient d'être faite suppose un fonctionnement normal de toutes ses composantes. Une lésion d'une ou de certaines structures de l'oreille peut engendrer une déficience auditive. Pour certains, elle peut être une baisse ou perte de l'acuité auditive, due principalement au mauvais transfert de l'énergie acoustique dans le système mécanique de conduction. Pour d'autres, le problème est plutôt d'origine nerveuse. Trois types de surdité peuvent être distingués : les surdités de transmission, les surdités de perception et les surdités mixtes rassemblant les deux premières :

- Les surdités de transmission (mécanique) sont dues à une atteinte du tympan et des osselets des éléments anatomiques qui assurent la transmission des sons dans l'oreille moyenne.

- Les surdités de perception (neurosensorielles) sont dues à une atteinte du fonctionnement neurosensoriel de l'audition, c'est à dire, à une atteinte des structures de l'oreille interne ou des voies auditives.

- Les surdités mixtes : C'est la coexistence des deux types d'atteintes précédentes.

Afin de remédier à ce problème de déficit auditif, différents appareillages peuvent être utilisés. Quand la défaillance auditive est partielle, il est encore possible de maintenir l'ouïe en amplifiant simplement les sons avec une prothèse auditive. Lorsque la défaillance est totale, les prothèses auditives seront inefficaces. Dans ce cas, nous aurons recours à une stimulation électrique reflétant l'environnement sonore extérieur par l'intermédiaire d'une ou de plusieurs électrodes insérées dans la cochlée. Cette fonction est réalisée par la prothèse cochléaire qui est basée sur l'idée de remplacer le mécanisme normal d'audition (oreille externe, moyenne, et une partie de l'oreille interne comprenant les cellules ciliées) par une stimulation directe des restes auditifs afin d'initier un influx nerveux sur les fibres du nerf auditif provoquant une sensation sonore

3.1 Aides auditives

Les aides auditives ou prothèses auditives sont des systèmes universels pour l'aide à l'audition applicables à toutes les surdités avec une efficacité reconnue mais variable en fonction du type de surdité, de la personne appareillée et des conditions acoustiques ambiantes. Les aides auditives ne traitent pas l'oreille mais le son. Elles n'ont aucun effet thérapeutique mais constituent une aide qui permet d'améliorer la communication. Les aides auditives agissent par amplification du son, mais celle-ci est plus complexe que dans un amplificateur courant. Elle dépend de la fréquence et du niveau du son, de l'ambiance sonore, de la position de la source. Elle varie aussi selon que le son soit du bruit ou de la

parole. Depuis quelques années, les aides auditives bénéficient du traitement numérique du signal de parole. Les prothèses numériques sont plus fiables, plus facilement adaptables et surtout plus confortables que les appareils anciens. Elles permettent une bonne perception de l'environnement sonore et une amélioration de la compréhension de la parole, mais l'utilisation en milieu bruyant pose encore des problèmes. Il existe deux formes d'appareils : les contours d'oreille et les intraauriculaires dont les performances sont à peu près équivalentes avec cependant une plus grande facilité d'usage pour les contours

- Les contours sont portés sur le pavillon et reliés au conduit auditif par un tube terminé par un "embout" qui obture ce dernier, l'embout assure le maintien du tube et l'étanchéité nécessaire pour éviter les sifflements produits par l'effet Larsen. Depuis peu, sont apparus des appareils à embout ouvert, dont le port est plus confortable.

- Les intra-auriculaires sont enfoncés dans le conduit auditif, et invisibles de l'extérieur. Pour diverses raisons, ces appareils ne peuvent pas être utilisés par tous. L'appareillage binaural et stéréophonique est obligatoire pour les enfants et recommandé aux adultes dont la surdité est symétrique. Au fur et à mesure qu'augmente la dissymétrie, les bienfaits de l'appareillage binaural s'estompent. Pour les personnes âgées, le choix de la stéréophonie dépend aussi de l'habileté manuelle du patient qui doit être suffisante pour gérer deux appareils.

3.2 Implants cochléaires

Les implants cochléaires permettent de rétablir une perception sonore chez les personnes souffrant d'une surdité bilatérale, profonde ou totale, et qui ne tirent aucun bénéfice d'une prothèse auditive conventionnelle. A l'intérieur de la cochlée, les cellules ciliées transforment les vibrations sonores en impulsions électriques. Ensuite ces impulsions électriques sont transportées par le nerf auditif jusqu'au cerveau. Si les cellules ciliées ne sont plus présentes ou insuffisamment nombreuses, il n'y a plus de codage naturel des vibrations sonores en impulsions électriques nerveuses. Le rôle d'un implant cochléaire est de remplacer l'ensemble de l'oreille (externe, moyenne et cochlée) pour transformer les vibrations sonores en impulsions électriques. Ces signaux transmis par le nerf auditif jusqu'au cerveau produisent des sensations auditives. Le cerveau doit alors apprendre, ou réapprendre à décoder ce nouveau message sonore.
L'implant cochléaire est composé d'une partie externe à base d'un processeur de traitement vocal permettant le traitement des signaux de parole, le contrôle et l'alimentation de l'unité implantée, et d'une partie interne appelée implant ou récepteur. L'étage de stimulation est un dispositif électronique qui permet d'appliquer un courant électrique aux terminaisons du nerf auditif à l'intérieur de la cochlée. Un étage de transmission permet de véhiculer l'information sonore codée issue de l'étage de traitement à l'étage de stimulation. L'étage de transmission assure aussi la liaison entre l'étage de traitement et l'étage de stimulation à travers des liens inductifs opérant par émission radiofréquence ou bien à travers une émission optique. Les différentes parties de la prothèse cochléaire sont données par la figure 1.3.

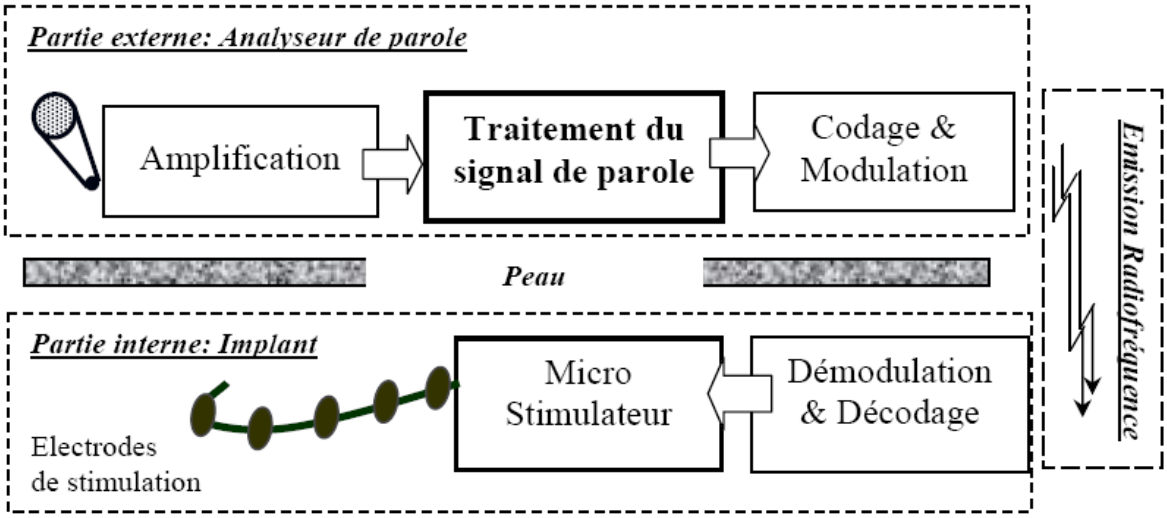


FIGURE 1.3 – Représentation schématique de l'implant cochléaire

Le microphone permet de capter le signal acoustique de l'environnement extérieur et le convertit en un signal électrique. Un processeur spécifique de traitement du signal permet de traiter ce signal moyennant différentes technique de traitement du signal afin de générer les stimuli adéquats. Le traitement usuel est basé sur un vocodeur à canaux utilisant un banc de filtres passe-bande afin de découper le signal en différentes régions fréquentielles correspondantes aux positions fréquentielles des différentes électrodes au niveau de la cochlée. Le signal de stimulation est envoyé à l'implant par l'antenne à travers la peau. Cette antenne est maintenue sur le cuir chevelu en regard du récepteur implanté grâce à un aimant. L'implant, qui est un petit boîtier en céramique placé chirurgicalement sous la peau, reçoit les informations du processeur externe et les répartit sur les différentes électrodes placées dans la cochlée. Les terminaisons du nerf auditif en contact avec les électrodes transmettent les impulsions électriques jusqu'au cerveau qui les interprète comme des signaux sonores. La réhabilitation des surdités de perception profondes bilatérales est aujourd'hui possible grâce à l'implant cochléaire bilatéral. Un implant cochléaire binaural en cours de développement permet non seulement la restitution de l'audition bilatérale mais aussi la capacité de localisation des sources sonores dans l'espace. Bref, un tel appareillage binaural permet d'apparenter le fonctionnement naturel du système auditif.

4 Implant cochléaire : Principe de fonctionnement

La partie externe de l'implant cochléaire est conçue autour d'un processeur spécifique de traitement du signal. L'élaboration d'une stratégie de stimulation convenable pour la prothèse cochléaire consiste à extraire les composantes essentielles de la parole et de les transmettre jusqu'à l'oreille interne. Le but est de faire percevoir le malentendant, suite à une stratégie de stimulation, les différentes variantes de la parole. Les impulsions de stimulation sont des valeurs de courant électrique qui correspondent à des quantités de charges électriques injectées dans le tissu biologique.

4.1 Outils de traitement du signal de parole pour l'implant cochléaire

Le signal de parole capté du milieu extérieur est traité selon un algorithme bien spécifique afin d'extraire tout les paramètres nécessaires à la stimulation électrique tel que l'énergie et la fréquence fondamentale. Le traitement du signal comporte principalement les quatre étapes suivantes :

- Prétraitement du signal de parole.
- Filtrage du signal de parole suivant un découpage spécifique moyennant un banc de filtres. Chaque bande de fréquences est associée à une électrode de stimulation.
- Détection d'enveloppe pour chaque canal.
- Calcul et extraction des paramètres nécessaires à la stimulation.

La figure 1.4 illustre schématiquement ces différentes étapes de traitement :

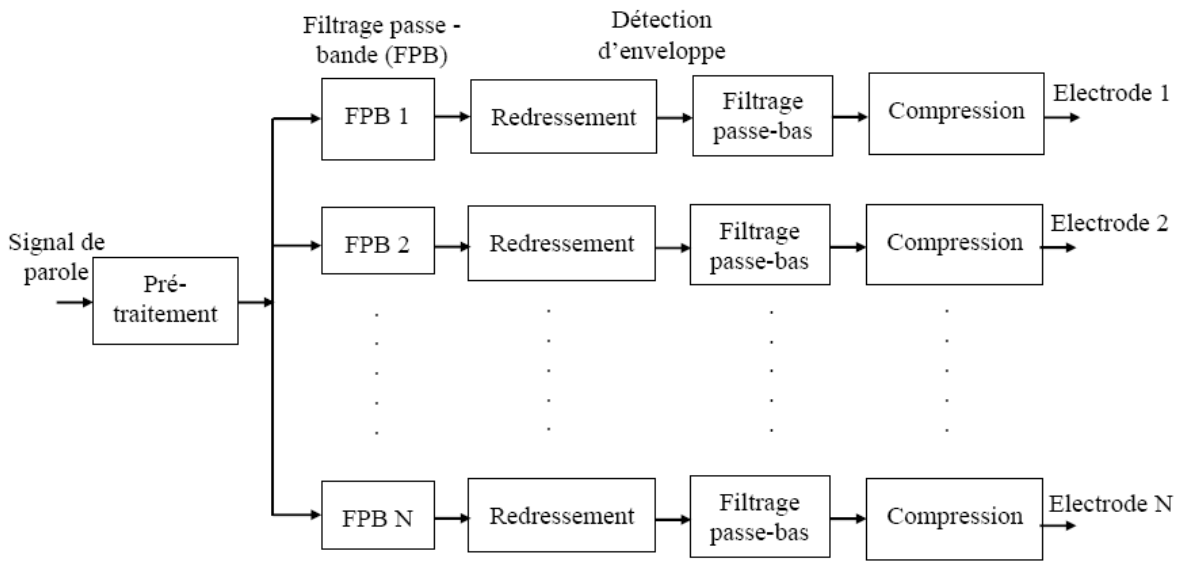


FIGURE 1.4 – Représentation schématique des différentes étapes de traitement du signal de parole

4.1.1 Analyse et filtrage du signal de parole

La première phase est une phase de prétraitement du signal de parole (amplification, filtrage anti-repliement...). Le signal est ensuite traité par un banc de filtres passe-bande. Ce banc de filtres est conçue suivant un découpage spectral spécifique. En effet, il existe plusieurs types de découpages du spectre de fréquences audible limité par les fréquences

F_{min} (la fréquence minimale) et F_{max} (la fréquence maximale) tels que le découpage linéaire, logarithmique... Dans le cas d'un découpage linéaire, toutes les bandes sont fréquentiellement équidistantes. La largeur de chaque bande est donnée par l'équation 1.1.

$$\frac{F_{max} - F_{min}}{N} \tag{1.1}$$

où N est le nombre de bandes.
Un autre type de découpage du spectre audible souvent utilisé dans certains appareillages est basé sur un modèle logarithmique. D'autres appareillages utilisent un découpage selon les bandes critiques qui est un découpage non linéaire. Différents modèles acoustiques non linéaires ont été élaborés afin de déterminer les limites des différentes bandes fréquentielles (Tranmüller, 1990). Un modèle approximatif très souvent utilisé permettant le passage du domaine de représentation fréquentiel f en Hz vers le domaine de représentation en bandes critiques B en barks est donnée par l'équation 1.2.

$$B(f) \quad = \quad 6.7 Argsh(\frac{f-20}{600}) \tag{1.2}$$

L'effet de la largeur des filtres passe-bande sur la reconnaissance de la parole a été étudié par Skinner (Skinner *et al.*, 1995). Trois types de découpage du spectre fréquentiel suivant une échelle logarithmique, une échelle de Mel (Fant, 1973) et les bandes critiques (Zwicker et Fastl, 1999) ont été étudiés. Les résultats ont montré que la majorité des sujets présentent de meilleures performances avec un découpage suivant les bandes critiques par rapport à un découpage logarithmique. Les performances obtenues avec l'échelle de Mel sont les moins bonnes par rapport aux deux autres types de découpage. Ceci peut être expliqué par le nombre de bandes de fréquences attribuées autour des formants F_1 et F_2 du signal de parole. En effet, pour la plupart des voyelles, le formant F_1 est situé dans la bande de fréquences 0-1 kHz et le formant F_2 est situé dans la bande de fréquences 1-3 kHz. Le découpage suivant l'échelle de Mel présente le plus petit nombre des bandes de fréquences (4 bandes) autour du formant F_1, par contre avec le découpage suivant les bandes critiques et l'échelle logarithmique, 6 bandes sont considérées autour du formant F_1. Les meilleures performances obtenues avec le découpage suivant les bandes critiques par rapport au découpage suivant l'échelle logarithmique sont dues au fait que le nombre de bandes autour du formant F_2 est plus important avec le premier type de découpage (7 bandes) par rapport aux deuxième type où seulement 6 bandes de fréquences sont considérées.

4.1.2 Détection d'enveloppe

Deux méthodes différentes peuvent être utilisées afin d'extraire l'enveloppe du signal de parole filtré relatif à chaque bande fréquencielle. La première méthode comporte une étape de redressement suivie d'une étape de filtrage passe-bas caractérisé par une fréquence de coupure qui peut varier entre 150 Hz et 400 Hz. Cette méthode est caractérisée par sa simplicité et son faible temps de traitement. La vitesse de stimulation au niveau de l'implant doit être deux fois plus importante que la fréquence de coupure du filtre passe-bas afin de respecter le théorème de Shannon. La fréquence de coupure du filtre passe-bas contrôle la profondeur de modulation des enveloppes. Plus la fréquence de coupure

est faible, plus la profondeur de modulation du signal filtrée est faible. Des études en simulation (Shannon *et al.*, 1995) n'ont montré aucun effet significatif de la fréquence de coupure sur l'intelligibilité de la parole par des sujets normo-entendant. Ceci a été également confirmé par des études traitant des patients portant des implants cochléaires (Lobo *et al.*, 2002).
La deuxième technique qui permet l'extraction de l'enveloppe du signal de parole est basée sur la Transformation d'Hilbert. En effet, si nous considèrons un signal de parole x(t), sa transformée d'Hilbert y(t) correspond à la réponse de ce signal x(t) à un filtre de quadrature dont la réponse impulsionnelle est donnée par l'équation 1.3 (Rosen, 1992).

$$h(t) = \frac{1}{\pi.t} \tag{1.3}$$

Le signal y(t) peut s'écrire dans le domaine temporel suivant l'équation 1.4.

$$y(t) = h(t) * x(t) \tag{1.4}$$

où l'opérateur * désigne le produit de convolution.
Cette expression s'écrit dans le domaine spectral suivant l'équation 1.5.

$$Y(f) = H(f).X(f) \tag{1.5}$$

Avec $Y(f)$, $X(f)$ et $H(f)$ correspondent respectivement aux Transformées de Fourier (TF) des signaux $y(t)$, $x(t)$ et $h(t)$.
La TF de la réponse impulsionnelle $h(t)$ du filtre est donnée par l'équation 1.6.

$$H(f) = -j.sign(f) \tag{1.6}$$

tel que $j^2 = -1$ et le signal $sign(f)$ est donné par l'équation 1.7.

$$sign(f) = \begin{cases} 1 & \text{si } f > 0 \\ 0 & \text{si } f = 0 \\ -1 & \text{si } f < 0 \end{cases} \tag{1.7}$$

La détermination de l'enveloppe du signal $x(t)$ se fait suite au calcul du signal analytique $z(t)$.

$$z(t) = x(t) + j.y(t) \tag{1.8}$$

L'enveloppe $e(t)$ du signal $x(t)$ est donnée par le module du signal analytique $z(t)$.

$$e(t) = |z(t)| = \sqrt{x(t)^2 + y(t)^2} \tag{1.9}$$

La figure 1.5 présente l'enveloppe d'un signal de parole obtenu par redressement et filtrage passe-bas à 200 Hz et 400 Hz ainsi qu'avec la méthode basée sur la Transformée d'Hilbert.

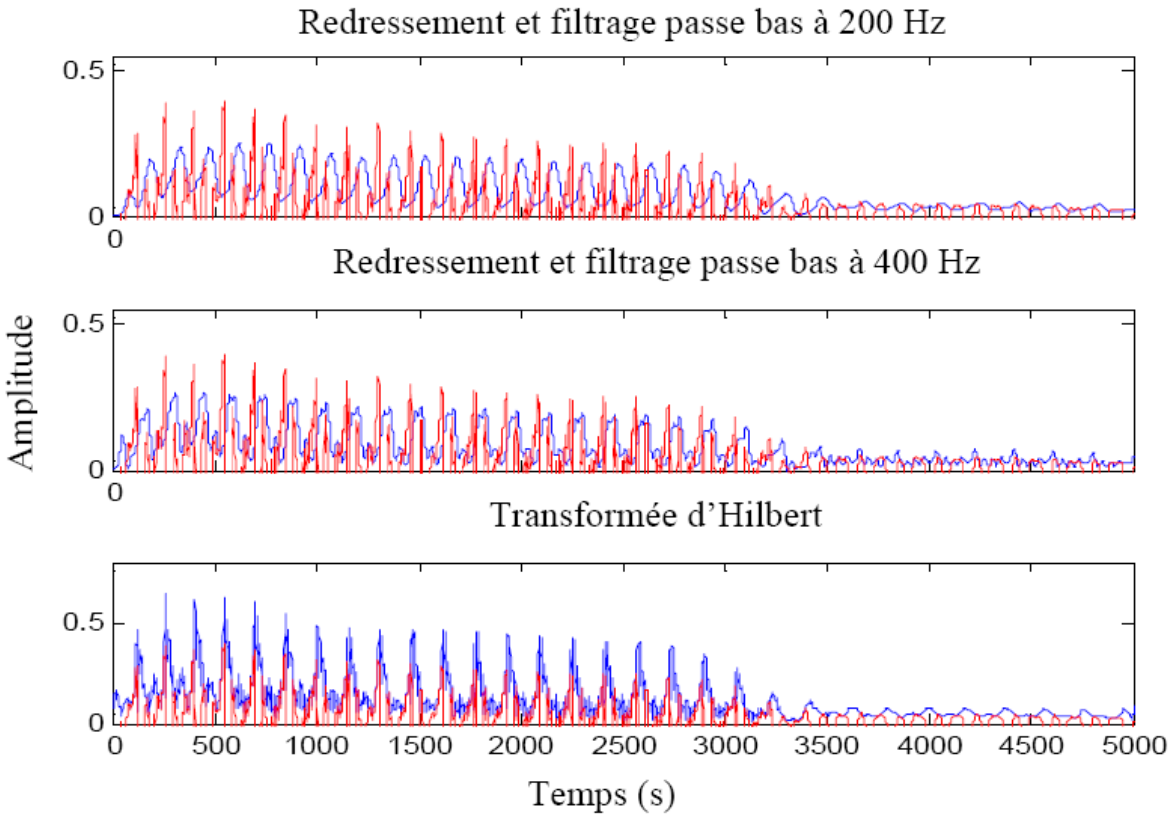


FIGURE 1.5 – Représentation temporelle du signal de parole (rouge) et des signaux d'enveloppes estimés (bleu)

4.1.3 Vitesse de stimulation

La vitesse de stimulation correspond au nombre d'impulsions délivrées pour chaque électrode par secondes (pps). La transmission de l'enveloppe des signaux acoustiques est principalement limitée par une telle vitesse de stimulation. Le choix de la vitesse de stimulation est variable suivant le système d'implant cochléaire commercialisé et il varie des plus faibles valeurs où la vitesse est inférieure à 500pps jusqu'aux hautes valeurs correspondantes à des vitesses supérieures à 3500pps. Il serait raisonnable de s'attendre à ce que les performances sont de plus en plus meilleures que la vitesse de stimulation est importante puisque les hautes vitesses de stimulation permettent une meilleure représentation temporelle du signal. Les avantages de l'utilisation d'une vitesse de stimulation élevée sont limités par l'interaction entre les électrodes de stimulation. Le degré d'interaction est de plus en plus important que la distance entre deux électrodes adjacentes est faible. En effet, le nombre d'électrodes ainsi que la distance qui sépare deux électrodes adjacentes sont des paramètres variables d'un fabriquant à un autre, ce qui explique la variation de l'effet de la vitesse de stimulation sur l'intelligibilité de la parole suivant le type de l'appareillage implanté. Différents travaux de recherche ont étudié l'effet de la vitesse de stimulation sur l'intelligibilité de la parole pour différents types d'appareillages en milieux silencieux et bruité. Lorsque la fréquence de stimulation est supérieure à 300pps, aucune amélioration significative de l'intelligibilité de la parole n'est observé (Komal *et al.*, 2009; Verschuur, 2005). Toutefois, d'autres études ont démontré que cette amélioration n'est

plus significative à partir des hautes vitesses de stimulation (Friesen *et al.*, 2001; Galvin et Fu, 2009).
L'effet de la vitesse de stimulation a été aussi testé pour différentes stratégie de codage et pour différent nombres d'électrodes de stimulation. Dans le travail de Friesen et al., (Friesen *et al.*, 2001), les auteurs ont étudié l'effet de la vitesse de stimulation sur la perception des mots par des patients portant trois types d'appareillages utilisant la stratégie de codage CIS (Nucleus24, Clarion C1 et Clarion C2) pour différents nombres d'électrodes de stimulation (4, 8, 12 et 16 électrodes). Les résultats ont montré que la vitesse de stimulation n'a pas d'effet significatif lorsque 8, 12 ou 16 électrodes sont utilisées. D'autres travaux de recherche ont montré de meilleures performances avec des hautes et moyennes vitesses de stimulation lorsque la stratégie CIS est considérée comme stratégie de codage (Verschuur, 2005; Loizou *et al.*, 2000; Nie *et al.*, 2006; Plant *et al.*, 2007). L'effet des vitesses de stimulation 500pps, 1200pps et 3500pps sur la perception de la parole a été aussi étudié lorsque la stratégie ACE (Advanced Combinaison Encoder) est considérée (Weber *et al.*, 2007). Les résultats n'ont montré aucun effet significatif global de la vitesse de stimulation, mais une certaine variabilité est observé en milieu bruité.

4.1.4 Fonction de compression

La dynamique acoustique du signal de parole est largement plus importante que la dynamique électrique du système auditif humain. C'est pour cela que la fonction de compression constitue une composante importante de l'implant cochléaire vu qu'elle assure la transformation des amplitudes acoustiques en amplitudes électriques. La fonction logarithmique est couramment utilisée pour la compression car elle permet d'adapter la stimulation acoustique à la stimulation électrique et restaurer l'amplitude normale du signal de parole audible. Cette fonction de compression est donnée par l'équation 1.10.

$$E^i = \begin{cases} E^i_{min} & \text{si } A^i < A_{min} \\ E^i_{min} + k^i(A^i - A_{min})^p & \text{si } A_{min} < A^i < A_{max} \\ E^i_{max} & \text{si } A^i > A_{max} \end{cases} \tag{1.10}$$

où E^i représente l'amplitude électrique compressée correspondante à l'amplitude acoustique non compressée A^i, i est l'indice de la bande considérée.
La dynamique acoustique est comprise entre les valeurs A_{min} et A_{max}. Le facteur p est un facteur de puissance. La constante k^i est choisie de telle manière que l'amplitude électrique compressée est égale à E^i_{max} lorsque l'amplitude acoustique est maximale (A^i=A^i_{max}). Fu et Shannon (Fu et Shannon, 1999) ont étudié l'effet de la variation du facteur de puissance sur la reconnaissance des voyelles et des consonnes avec trois patients porteurs de l'implant cochléaire Nucleus22 et utilisant la stratégie de codage SPEAK. Les valeurs du facteur de puissance p choisies sont les suivantes : 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5 et 0.75. Le matériel est formé d'un ensemble de 16 voyelles et 16 consonnes enregistrées avec différents locuteurs (Hillenbrand *et al.*, 1995). Les résultats de l'expérience ont montré que le pourcentage de reconnaissance moyen des voyelles est de 36.8% pour un facteur de puissance p=0.005. Le pourcentage de reconnaissance moyen des voyelles est de 50% pour tous les facteurs de puissance compris entre les valeurs 0.2 et 0.5. Le facteur de puissance p=0.75 engendre une diminution du pourcentage de reconnaissance à la valeur 40.7%.

Dans le cas des consonnes, le pourcentage moyen de reconnaissance est de 56% lorsque le facteur de puissance p=0.05 est utilisé et il atteint la valeur 70.7% lorsque le facteur de puissance est égal à 0.2. Comme pour les voyelles, le pourcentage moyen de reconnaissance diminue à la valeur 43.5% avec le facteur de puissance p=0.75. Les résultats indiquent que les meilleures performances, pour les voyelles ainsi que pour les consonnes, sont obtenues lorsqu'une fonction de compression non linéaire est utilisée, ce qui contredit la linéarité de la fonction d'audition naturelle.

4.2 Stratégie de codage pour implant cochléaire

Plusieurs stratégies de traitement du signal ont été développées pour les prothèses cochléaires multi-canaux. Elles peuvent être divisés en trois catégories qui diffèrent dans la manière d'analyse du signal de parole et dans sa présentation aux électrodes de stimulation (Ghorbel, 2007).

4.2.1 Stratégies à base d'extraction des formants

4.2.1.a Multi-Peak ou MPEAK

Le processeur réalise une extraction des estimations des formants F_0, F_1, F_2 et utilise par la suite trois filtres passe-bande pour transmettre des informations supplémentaires entre 2000 Hz et 8000 Hz. La fréquence de stimulation est fixe.

4.2.1.b Spectral Maxima Sound Processor ou SMSP

A chaque cycle de traitement, le processeur sélectionne les 6 bandes d'amplitude maximale et adresse l'information aux 6 électrodes correspondantes. Cette stratégie n'est plus utilisée.

4.2.2 Stratégie favorisant la représentation fréquentielle

4.2.2.a Spectral Peak Strategy ou SPS

Le signal de parole est décomposé en 20 bandes fréquentielles. La vitesse de stimulation est fixe. Son principe est de stimuler seulement un certain nombre d'électrodes où les énergies sont les plus élevées. Le nombre réel de maxima (1 à 10) est déterminé par le niveau du signal et de sa composition spectrale. Généralement, 6 maxima sont sélectionnés par le régleur d'implant. La bande passante est dépendante du nombre d'électrodes utilisables. Ainsi, dans le cas où 20 électrodes sont utilisées, la bande passante sera comprise entre 250 Hz et 10800 Hz ; mais si seulement 16 électrodes sont disponibles, la bande passante ne sera comprise qu'entre 250 Hz et 5700 Hz.

4.2.2.b Adaptative Stimulation Rate ou ASR

La vitesse de stimulation est adaptative. C'est une évolution de SPEAK avec une détection automatique des composantes transitoires (consonnes) et stationnaires (voyelles) de la parole par une analyse statistique afin d'adapter la fenêtre d'analyse FFT. Pour une voyelle, la fenêtre d'analyse est large avec un rythme de stimulation lent, aboutissant à une bonne résolution fréquentielle sur 15 canaux. Pour les consonnes, la fenêtre d'analyse est réduite avec un rythme de stimulation rapide pour respecter les caractéristiques temporelles des consonnes. Nous avons alors un codage sur un plus petit nombre de canaux.

4.2.3 Stratégies favorisant la représentation temporelle

4.2.3.a Compressed Analog ou CA

Cette stratégie consiste à stimuler quatre zones de la cochlée par quatre électrodes. L'étage de filtrage est formé par un banc de filtres passe-bande analogiques répartis sur le spectre de fréquence perçu par une oreille saine en autant de bandes qu'il y a d'électrodes. Chaque filtre permet de garder une bande de fréquence bien spécifique et éliminer les autres. L'énergie du signal issu de chaque bande sera convertie en un stimulus électrique pour le site spécifié de la cochlée.

4.2.3.b Continuous Interleaved Sampling ou CIS

C'est la stratégie la plus rapide actuellement. Les électrodes sont stimulées séquentiellement (c'est à dire les unes après les autres) ce qui limite les interférences. Son originalité est que l'information spectrale n'est pas dominante et nous avons un bon profil de la structure temporelle.

4.2.3.c Simultaneus Analogue Stimulation ou SAS

Une oreille en bon état de fonctionnement traite les sons simultanément dans l'ensemble de la cochlée. Pour imiter cette conception naturelle, l'implant peut fournir des informations auditives simultanément à divers endroits de la cochlée. Ainsi, SAS convertit par traitement numérique les sons en ondes analogiques sinusoïdales parvenant simultanément aux diverses électrodes. L'information est délivrée sur 8 canaux et les impulsions sont bipolaires.

4.2.4 Stratégie hybride

4.2.4.a Spectral Peak Strategy ou SPEAK

Le signal de parole est décomposé en 20 bandes fréquentielle. Seulement un certain nombre d'électrodes où les énergies sont les plus élevées sont utilisées pour la stimulation.

Le nombre réel de maxima (1 à 10) est déterminé par le niveau du signal et de sa composition spectrale, ainsi que le réglage propre à chaque patient. La bande passante est dépendante du nombre d'électrodes utilisables. Le rythme de stimulation est déterminé selon la fréquence fondamentale F_0.

4.2.4.b Advanced Combination Encoders ou ACE

ACE est sensée allier le meilleur de CIS et de SPEAK. De CIS, nous avons retenu la vitesse de stimulation constante paramétrable et de SPEAK le système de stimulation sélectionnant les sites d'activation nerveuse de manière dynamique en fonction de l'énergie spectrale.

5 Différents types d'appareillages

Il est classique de distinguer les implants en fonction du nombre d'électrodes, monoélectrode et multiélectrodes, et en fonction du siège des électrodes, intra ou extracochléaires. Actuellement, les implants intracochléaires multiélectrodes sont les plus utilisés. Le tableau 4.1 expose les principaux implants commercialisés, leur type et leur stratégie de codage. Nous notons ici qu'une simulation acoustique de l'implant cochléaire est très sou-

TABLE 1.1 – *Principaux implants cochléaires unilatéral commercialisés*

Constructeur	**Système**	**Stratégie**	**Nombre électrode**
Advanced Bionics	HiRes Bionic Ear System	CIS	16
Cochlear	Nucleus Freedom	SPEAK	22
MedEL	COMBI 40+	CIS	12
Neurelec	Digisonic SP	ACE	20

vent utilisée pour étudier l'intelligibilité de la parole chez les sujets porteurs d'un implant cochléaire en absence de tous facteurs interférents (profondeur d'insertion des électrodes, causes de la surdité, nombre d'années d'expérience avec l'implant ...). Le signal de sortie du simulateur de l'implant cochléaire est un signal de parole synthétisé. Un simulateur d'implant est dit à bruit à bande étroite si un bruit aléatoire est considéré comme signal d'excitation pour tous les segments de parole (Shannon *et al.*, 1995). Si les signaux d'excitation sont des signaux sinusoïdaux dont les fréquences correspondent à la fréquence centrale de chaque bande fréquentielle d'analyse, le simulateur d'implant est dit sinusoïdal (Dorman *et al.*, 1997). Différent travaux de recherche ((Whitmal *et al.*, 2007) par exemple) ont montré que les résultats obtenus avec les simulateurs d'implant cochléaire donnent des résultats similaires avec les résultats donnés par les implants cochléaires.

6 Nature et caractéristiques du bruit

6.1 Différentes classes du bruit

Nous appelons bruit tout signal nuisible qui se superpose au signal utile en un point quelconque d'un système de transmission ou d'une chaîne de mesure. Il constitue donc une gêne dans la compréhension du signal utile, qui est dans notre cas, la parole. En physique, en acoustique et en traitement du signal, bien que le bruit soit, par nature, aléatoire, il possède certaines caractéristiques statistiques, spectrales ou spatiales. Le tableau 1.2, extrait de (Virag, 1999), représente les différentes classes auxquelles un bruit peut appartenir. Comme notre but est essentiellement le débruitage et l'amélioration de

TABLE 1.2 – *Différentes classes du bruit*

Propriétés	Types
Structure	Continu/Impulsif/Périodique
Type d'interaction	Additif/Multiplicatif/Convolutif
Comportement temporel	Stationnaire/Non-stationnaire
Bande de fréquences	étroite/large
Dépendance	Corrélé/ Décorrélé
Propriétés statistiques	Dépendant/Indépendant
Propriétés spatiales	Cohérent/Incohérent

l'intelligibilité de la parole pour les sujets porteurs d'implants cochléaires, nous nous limitons dans notre étude aux bruits additifs, stationnaires ou non stationnaires, et décorrélés avec la parole (indépendance au sens statistique), tels que le bruit de conversation appelé Babble et le bruit blanc gaussien (ce dernier est souvent utilisé mais peu réaliste).

6.2 Bruit musical

Le bruit musical est un bruit résiduel perceptuellement gênant qui apparaît suite au débruitage de la parole par des algorithmes d'atténuation spectrale à court terme tels que la soustraction spectrale ou le filtrage de Wiener. Le spectre du bruit musical est particulièrement tonal, d'où le caractère musical. Son énergie moyenne est plus faible que celle du bruit initial, mais sa dispersion en fréquence est plus grande, ce qui le rend plus gênant que le bruit de départ du point de vue perception. Parmi les raisons d'apparition de ce type de bruit (Cappé, 1994), nous citons :

- Le traitement non linéaire des composantes négatives du signal rehaussé,
- L'estimée non précise de la densité spectrale de bruit,
- L'estimation basée sur des périodogrammes,
- La variabilité de la fonction de gain appliquée au signal bruité...

Nous considèrons le signal de parole bruité $x(n)$ tel que $x(n) = s(n) + b(n)$ où $s(n)$ et $b(n)$ indiquent respectivement le signal propre et le bruit à l'intant discret n. Soient $X(f,k)$, $S(f,k)$ et $B(f,k)$ les transformées de Fourier Discrètes à court terme respectives des signaux $x(n)$, $s(n)$ et $b(n)$ tel que $X(f,k) = S(f,k) + B(f,k)$, où k et f indiquent respectivement l'indice de la trame et la varible discrète fréquentielle.
Dans le cas d'un problème de débruitage linéaire classique où nous cherchons un estimateur $H(f,k)$ tel que $\hat{S}(f,k) = H(f,k)X(f,k)$, l'erreur dûe à ce filtrage est donnée par l'équation suivante :

$$\begin{aligned} \hat{E}(f,k) &= \hat{S}(f,k) - S(f,k) = H(f,k)\left(S(f,k) + B(f,k)\right) - S(f,k) \\ &= \left(H(f,k) - 1\right) S(f,k) + H(f,k)B(f,k) \end{aligned} \tag{1.11}$$

L'expression $\left(H(f,k) - 1\right) S(f,k)$ représente la distorsion du signal tandis que $H(f,k)B(f,k)$ désigne le bruit résiduel contenant le bruit musical. Dans le cas où $0 < H(f,k) < 1$, il est généralement très difficile de réduire le bruit musical sans apporter de distorsion sur le signal de parole. Le besoin d'un compromis entre la distorsion et le bruit musical s'avère donc le meilleur moyen pour augmenter les performances en qualité et intelligibilité.

7 Implant cochléaire et effet du bruit

En milieu silencieux et en absence de bruit, les patients porteurs d'un implant cochléaire bénéficient d'une intelligibilité maximale. Les performances de l'implant se dégradent d'une manière considérable en milieu bruité (lieu de travail, dans la rue, dans les cafétérias...), et cette dégradation est proportionnelle au niveau du bruit. Cette dégradation est variable suivant le type de bruit considéré ainsi que la stratégie de codage utilisée.

7.1 Effet du bruit 'speech-shaped noise' sur la reconnaissance des consonnes et de la parole avec la stratégie de codage CIS

Eddington et al. (Eddington *et al.*, 1997) ont testé l'effet du bruit 'speech-shaped noise' (SSN) sur la reconnaissance des consonnes et de la parole continue avec des patients implantés. Les stimuli sont composés de 24 consonnes et d'un ensemble de listes extraites de la base de données HINT (Hearing in Normal Test) (Nilsson *et al.*, 1994). Les trois patients qui ont participé à cette expérience portent quotidiennement leurs implants. Les deux premiers patients utilisent des processeurs Clarion à huit canaux de stimulation et utilisant la stratégie de codage CIS. Le troisième patient utilise le processeur 'Ineraid' basé sur la stratégie de codage CIS avec six canaux de stimulation. Les résultats ont montré que le pourcentage moyen de reconnaissance des consonnes en milieu silencieux est de 76%, et il chute considérablement pour atteindre environ 35% en milieu bruité à un Rapport Signal à Bruit (RSB) de 0 dB. Dans le cas de la parole continue, le pourcentage moyen de reconnaissance en milieu silencieux est de 90% et décroit à 44% en milieu bruité à un RSB=0 dB.

7.2 Effet du bruit SSN sur la reconnaissance des consonnes et des voyelles avec la stratégie de codage SPEAK

Fu et al. (Fu *et al.*, 1998) ont étudié l'effet du bruit SSN à différents RSB sur la reconnaissance des consonnes et des voyelles avec trois patients utilisateurs d'implants cochléaires. Ces trois patients portaient des implants cochléaires de type Nucleus utilisant la stratégie de codage SPEAK. Dix groupes comportant chacun douze voyelles et six groupes comportant chacun seize consonnes sont considérés (Nilsson *et al.*, 1994). Tous ces stimuli sont noyés dans un bruit de type SSN à des RSB variant de 24 dB à 0 dB avec un pas de 6 dB et de −3 dB à −15 dB avec un pas de 3 dB. Les résultats ont montré que le pourcentage moyen de reconnaissance des voyelles en milieu silencieux est de 66% et il chute considérablement pour atteindre environ 27% en milieu bruité à un RSB=0 dB. Dans le cas des consonnes, le pourcentage moyen de reconnaissance en milieu silencieux est de 70% et il atteint 37% en milieu bruité à un RSB=0 dB.

7.3 Effet du bruit SSN sur la reconnaissance des consonnes, des voyelles et de la parole continue avec les stratégies de codage SPEAK, CIS et SAS

Une étude expérimentale présentée par Friesen et al. (Friesen *et al.*, 2001) avait pour but la comparaison des pourcentages de reconnaissance des voyelles, des consonnes et des mots avec les trois stratégies de codage SPEAK, CIS et SAS en présence d'un bruit de type SSN. Dix patients utilisateurs de l'implant cochléaire Nucleus ainsi que neuf patients utilisant l'implant Clarion ont participé à cette étude. Tous les patients porteurs de l'implant Nucleus utilisaient la stratégie de codage SPEAK. Cinq patients parmi les neufs porteurs de l'implant Clarion utilisaient la stratégie de codage CIS alors que les quatre restants utilisaient la stratégie de codage SAS. Dix groupes de douze voyelles chacun et douze groupes comportant chacun quatorze consonnes sont considérés comme stimuli pour les voyelles et les consonnes respectivement (Hillenbrand *et al.*, 1995). Dix listes comportant chacune cinquante mots, créées par House Ear Institute et Cochlear Corporation, sont considérées comme stimuli dans le cas de la parole continue. Tous ces stimuli sont noyés dans un bruit de type SSN à des RSBs de 15, 10, 5 et 0 dB. Les expériences sont réalisées avec différent nombre de canaux de stimulation. Dans le cas de l'implant Clarion, 2, 3, 4, 6 et 8 électrodes de stimulation sont considérées. Avec l'implant Nucleus, 2, 4, 7, 10 et 20 électrodes de stimulation sont considérées.
Les résultats n'ont pas montré de différence significative entre les processeurs Nucleus et Clarion. Les meilleurs pourcentages de reconnaissance sont obtenus en utilisant le processeur Nucleus avec 20 canaux de stimulation. En effet, Le pourcentage de reconnaissance des voyelles ainsi que celle des consonnes est de 60% en milieux silencieux, alors qu'elle décroit à 43% pour les voyelles et à 30% pour les consonnes en milieux bruité à un RSB=0 dB. Dans le cas des mots isolés, le pourcentage de reconnaissance est de 47% en milieux silencieux et décroit en milieu bruité à 20% et à 5% pour des RSB de 5 dB et 0 dB respectivement. Dans le cas de la parole continue, le pourcentage de reconnaissance est de 85% en milieux silencieux et décroit significativement en présence du bruit à 60%, 40% et 10% à des RSB de 10, 5 et 0 dB respectivement.

7.4 Effet du bruit de type 'Babble' multilocuteurs sur la reconnaissance de la parole continue avec les stratégies de codage SPEAK, CIS et SAS

Fetterman et Domico (Fetterman et Domico, 2002) ont étendu l'étude présentée dans le paragraphe précédent au cas d'un signal de parole noyé dans un bruit de type 'Babble' multilocuteur. Quatre-vingt et seize sujets ont participé à cette expérience. Soixante et six sujets utilisaient l'implant cochléaire Nucleus avec la stratégie de codage SPEAK. Trente sujets portaient l'implant cochléaire Clarion dont vingt et cinq sujets utilisaient la stratégie de codage CIS et cinq sujets utilisaient la stratégie de codage SAS. Les tests subjectifs sont basés sur un matériel de test utilisant des listes développées par l'université de New York (Boothroyd *et al.*, 1985). Le signal de bruit de type Babble avec huit locuteurs est utilisé à des RSBs de 10 dB et 5 dB. Le pourcentage de reconnaissance moyen en milieux silencieux dans le cas de la parole continue est de 82.1%. Ce pourcentage décroit significativement en milieu bruité pour atteindre des valeurs de 73% à un RSB de 10 dB et de 47% à un RSB de 5 dB.

8 Conclusion

L'implant cochléaire est un appareillage biomédical implantable au niveau de l'oreille et destiné à restaurer l'audition chez les personnes atteintes d'une surdité bilatérale, profonde ou totale. Différents paramètres interviennent dans le réglage de cet appareillage.
Les patients implantés bénéficient d'une intelligibilité de la parole comparable à celle obtenue avec des sujets normoentendants en milieu silencieux. En présence du bruit, cette intelligibilité se dégrade d'une manière considérable. Afin de rehausser le signal de parole bruité et améliorer l'intelligibilité de la parole chez les patients implantés, différents algorithmes de réduction du bruit ont été exploités. Cette solution consiste à intégrer au niveau du processeur de traitement du signal un algorithme de débruitage permettant de rehausser le signal avant qu'il soit traité et transmis vers la partie interne de l'implant. Ces algorithmes de débruitage peuvent être classés suivant le nombre d'observation utilisé (nombre de microphones) en algorithmes mono-voie, bi-voie et multi-voie. Le chapitre suivant sera consacré à présenter un état de l'art des différents algorithmes de réduction de bruit mono-voie et bi-voie.

Chapitre 2

Techniques de base pour la réduction de Bruit

Sommaire

1	**Introduction**	**26**
2	**Techniques mono-voie pour la réduction de bruit**	**26**
3	**Aperçu général sur les techniques de réduction du bruit bi-voie**	**41**
4	**Conclusion**	**50**

1 Introduction

Comme mentionné dans le chapitre précédent, l'intelligibilité de la parole chez les patients porteurs d'un implant cochléaire se dégrade d'une manière considérable en milieu bruité. Différent travaux de recherche ont traité ce problème en considérant une étape supplémentaire de prétraitement du signal de parole par un algorithme de débruitage spécifique. Le signal de parole débruité ainsi estimé sera ensuite traité par l'algorithme de codage permettant l'extraction des différents paramètres nécessaires à la stimulation électrique. Ces paramètres seront enfin transmis vers la partie interne. Plusieurs méthodes de débruitage de la parole mono-voie ont été proposées dans la littérature pour le rehaussement de la parole. Ces méthodes ont été initialement destinées pour des sujets normoentendants puis elles sont reconsidérées et testées dans le cas de l'implant cochléaire. Ces algorithmes comportent généralement deux étapes de traitement : Une première étape est destinée à l'estimation du bruit suivie d'une deuxième étape pour estimer le signal de parole débruité.

Les traitements à partir de plusieurs observations de signaux ne sont cependant pas que de simples extensions des traitements mono-voie, car ils permettent d'introduire des informations sur la distribution spatiale des signaux. L'exploitation simple de la localisation de la source du type Global Sidelobe Canceller (GSC) ne peut donner des performances accrues en termes de réduction de bruit du fait du nombre restreint de capteurs. Nous présentons dans la deuxième partie de ce chapitre des méthodes qui permettent de mieux exploiter les caractéristiques spatiales du bruit que ne le fait le GSC, tout en gardant en mémoire que ces traitements doivent être compatibles avec des considérations temps-réel. Ces méthodes sont classées suivant les hypothèses émises a priori concernant les distributions de bruits : bruits corrélés, bruit décorrélés, et finalement le cas de bruits partiellement corrélés, distribution plus en adéquation les environnements de bruit réel.

2 Techniques mono-voie pour la réduction de bruit

Dans ce paragraphe, nous faisons référence aux techniques de réduction de bruit mono-voie qui ont connu un développement spectaculaire ces deux dernières décennies. Il n'est pas question ici de faire un panorama complet de ces techniques, tant la littérature foisonne d'articles sur ce sujet : nous nous attacherons plutôt à décrire les méthodes qui sont à la base du traitement mono-voie (filtre de Wiener, soustraction spectrale). Nous ne pouvons pas parler de réduction de bruit sans aborder les fondements élaborés durant les années 70. Dans le cas mono-voie, l'objectif consiste à estimer un signal utile $s(n)$ noyé dans un bruit additif $b(n)$ supposé indépendant, à partir de la seule observation $x(n)$ qui est la somme de ces deux composantes.

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

Sans hypothèses ou connaissance a priori sur les signaux, le problème est mal posé et ne peut être résolu. Il faut aussi des connaissances sur l'un ou l'autre des signaux, voire sur les deux, afin d'être en mesure de discriminer le signal utile du bruit ambiant.

Les méthodes décrites dans ce paragraphe opèrent dans le domaine de Fourier. En supposant les signaux stationnaires et indépendants, la relation temporelle donnée par l'équation 2.1 s'écrit comme suit :

$$P_x(f) = P_s(f) + P_b(f) \tag{2.2}$$

où $P_x(f)$, $P_s(f)$ et $P_b(f)$ sont les densités spectrales de puissance (dsp) respectives de $x(n)$, $s(n)$ et $b(n)$. La stationnarité est généralement admise pour le bruit, même si cette hypothèse peut s'avérer injustifiée dans certaines conditions. En revanche, la stationnarité de la parole n'est valide que sur un horizon temporel relativement court, de l'ordre de 25 ms, période sur laquelle le signal peut être supposé stationnaire. C'est pour cette raison que nous considèrons plutôt les spectres instantanés, obtenus à l'aide d'une transformée de Fourier à court-terme (TFCT) sur des trames successives. La relation entre les composantes fréquentielles des signaux est donnée par l'équation 2.3.

$$X(f,k) = S(f,k) + B(f,k) \tag{2.3}$$

où f représente la fréquence et k l'indice de la trame courante. $X(f,k)$, $S(f,k)$ et $B(f,k)$ sont les TFCT respectives de $x(n)$, $s(n)$ et $b(n)$.
L'objectif est d'estimer la composante court-terme du signal utile, estimée que nous notons $\hat{S}(f,k)$, en supposant le bruit stationnaire.

2.1 Importance de la phase

Nous devrons plutôt parler de la non importance de la phase, tant ce paramètre a été occulté dans les traitements de rehaussement de la parole. En effet, les algorithmes de rehaussement de la parole implémentés dans le domaine spectral ne proposent qu'un estimateur du module du spectre du signal utile (ou de la puissance) : la phase du signal rehaussé est reconstruite à partir de la phase du signal bruité. Cela se justifie d'un point de vue physiologique par le fait que l'oreille est peu sensible à une modification de phase, tout du moins tant que cette modification reste dans des limites raisonnables. Wang et Lim (Wang et Lim, 1982) ont montré que l'utilisation de la phase du signal pur avec l'amplitude estimée par un algorithme de rehaussement n'améliore pas la qualité du signal rehaussé. Dans un même ordre d'idée, vary (Vary, 1985) a montré qu'une modification de la phase d'un signal pur n'introduit pas de distorsion audible tant que cette modification reste en deçà de $\frac{\pi}{4}$ environ. Les traitements ne s'intéresseront qu'à la modification du module du spectre du signal d'observation, à partir de l'équation 2.4.

$$\hat{S}(f,k) = G(f,k)\ X(f,k) \tag{2.4}$$

où $G(f,k)$ est une loi d'atténuation réelle.

2.2 Rapport signal à bruit : RSB a priori et RSB à postériori

Les différentes lois d'atténuation sont toutes basées sur le raisonnement suivant :

- Un rapport signal à bruit (RSB) élevé correspond très certainement à une forte composante de parole, et le gain $G(f,k)$ doit être quasiment égal à 1.

- Un RSB faible doit mener à une atténuation importante de $X(f,k)$ ($G(f,k)$«1), cette observation ne contenant que du bruit.

Il est généralement possible d'exprimer les lois d'atténuation des filtres en fonction du RSB. Cependant, du fait des différents critères d'estimation choisis pour la dérivation des filtres, nous devons distinguer le RSB a priori et le RSB a posteriori qui sont définis par les équations suivantes :

$$RSB_{prio}(f) = \frac{E\{|S(f)|^2\}}{E\{|B(f)|^2\}} = \frac{P_s(f)}{P_b(f)} \tag{2.5}$$

$$RSB_{post}(f,k) = \frac{|X(f,k)|^2}{E\{|B(f)|^2\}} - 1 = \frac{|X(f,k)|^2}{P_b(f)} - 1 \tag{2.6}$$

où $E\{.\}$ est l'opérateur d'espérance mathématique, $P_s(f)$ est la dsp du signal de parole et $P_b(f)$ est la dsp du bruit. La quantité $RSB_{prio}(f)$, exprime le RSB long-terme en supposant les caractéristiques statistiques du signal utile connues *a priori*, d'où son nom. La grandeur $RSB_{post}(f,k)$ exprime le RSB de la trame courante en considérant le spectre instantané du signal bruité. Dans le cas où $RSB_{post}(f,k)$ devient négatif, il est mis à 0, un RSB négatif n'ayant aucun sens physique. Il faut noter que $RSB_{prio}(f)$ ne dépend *a priori* pas du temps puisqu'il s'agit d'une estimation sur une période importante, *a priori* seulement car les dsp de parole et de bruit sont tout de même amenées à évoluer au cours du temps. Nous considérerons par la suite la notation $RSB_{prio}(f,k)$ avec un indice de trame, tout en gardant en mémoire que ses variations sont plus lentes que celles de $RSB_{post}(f,k)$. La plupart des algorithmes de réduction du bruit qui opèrent dans le domaine spectral nécessitent une estimation précise de la dsp du bruit pour restituer une bonne estimée du signal de parole. La plupart des algorithmes proposés peuvent être classés en deux catégories. La première est basé sur l'estimation du bruit pendant les périodes de silence et la geler pendant les périodes de présence de la parole. Dans ce cas, un détecteur d'activité vocale (DAV) est nécessaire. La deuxième catégorie est basée sur la mise à jour des composantes du bruit d'une manière continue et permanente en se basant sur la dsp du signal bruité.

2.3 Algorithme d'estimation du bruit à base de DAV

Les méthodes classiques d'estimation du psd du bruit sont basées sur un détecteur d'activité vocale. L'idée consiste à estimer le bruit pendant les périodes de silence (présence de bruit seul) puis de le geler pendant les périodes d'activité vocale. Cette méthode classique est relativement répandue, mais elle suppose avoir un DAV fiable. Le DAV est appliqué sur toutes les trames du signal de parole bruité :

- H_0 : période de silence (absence de la parole) : X=B

- H_1 : Présence de la parole : X=S+B.

Le DAV considéré est basé sur un modèle statistique proposé par Sohn et al. (Sohn *et al.*, 1999). Le signal de bruit est estimé durant les périodes de silence, et sa valeur est maintenue constante jusqu'à nouvelle détection de silence. Cette approximation est valable sous le prétexte que nous travaillons avec des trames qui vérifie l'hypothèse de stationnarité du bruit à court terme.
L'objectif du DAV est de pouvoir détecter minutieusement les zones qui seront utilisées pour l'estimation du bruit lorsque nous avons affaire avec un signal de parole bruité. La figure suivante représente le comportement du DAV dans le cas d'un signal bruité (bruit de type Babble à des niveaux de $RSB = 0,\ 5,\ 10$ et 15 dB). Les résultats sont donnés en absence de chevauchement et dans le cas d'un chevauchement de 50%.

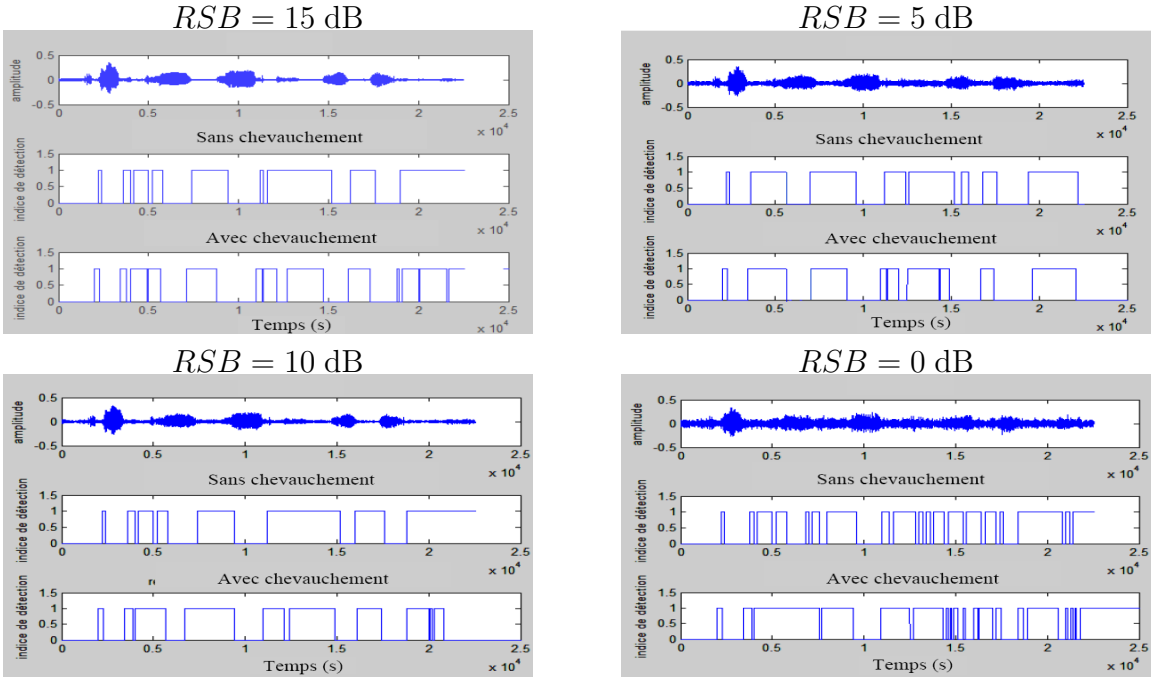


FIGURE 2.1 – Comportement du DAV en milieux bruité pour des traitements avec et sans chevauchement à différents RSB

Les performances de l'algorithme d'estimation du bruit à base de DAV sont acceptables quand il s'agit de bruits stationnaires et quand le RSB est élevé mais deviennent modestes dans les situations inverses. Des techniques alternatives et plus sophistiquées ont donc été élaborées pour pallier ces problèmes. L'une des techniques de base est basée sur la méthode des statistiques minimales (MS, Minimum Statistics) initialement proposée par Martin (Martin, 1994). Cette technique est basée sur le suivi du minimum de la dsp du signal bruité, noté $P_{xmin}(f,k)$, en se basant sur l'hypothèse que pour chaque composante spectrale, la puissance minimale sur une période d'observation donnée correspond à une composante de bruit. Dans ce contexte, différentes méthodes ont été développées pour l'estimation de $P_{xmin}(f,k)$ tel que l'utilisation d'une fenêtre de recherche ou la technique d'estimation récursive. Nous nous intéressons dans ce qui suit à présenter le principe de ces deux algorithmes utilisés pour l'estimation de la dsp du bruit moyennant la technique des statistiques minimales.

2.4 Estimation de $P_{xmin}(f,k)$ basée sur l'utilisation d'une fenêtre de recherche

Martin (Martin, 1994) a proposé un algorithme pour l'estimation du bruit basé sur la méthode des statistiques minimales. Dans ce cas, la dsp du bruit est basée sur le suivi des valeurs minimales de la dsp du signal bruité ($P_x(f)$) le long d'une fenêtre de recherche. Ainsi, l'utilisation de la technique des statistiques minimales élimine le problème de l'utilisation du DAV. La dsp du signal bruité est estimée d'une manière récursive à partir de l'équation 2.7.

$$P_x(f,k) = \lambda P_x(f,k-1) + (1-\lambda)\left|X(f,k)\right|^2 \tag{2.7}$$

où λ est un coefficient de lissage proche de 1.
Ce coefficient de lissage joue un rôle prépondérant dans le sens donné à ces estimations. La figure 2.2 montre les variations de la dsp du signal de parole et du bruit. La dsp estimée du signal bruité $P_x(f,k)$ est aussi représentée pour deux valeurs de $\lambda = 0.6$ et $\lambda = 0.9$.

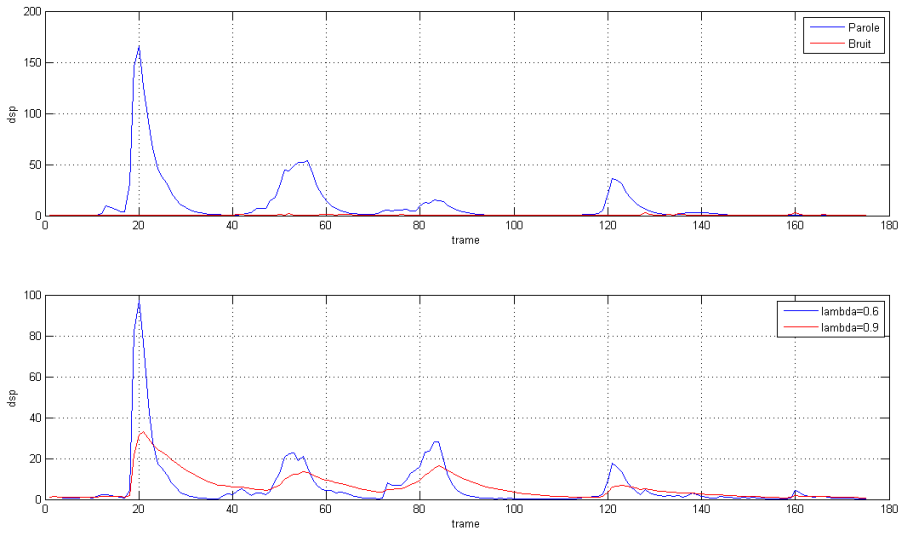


FIGURE 2.2 – Haut : dsp des signaux de parole et de bruit, Bas : dsp du signal bruité pour $\lambda = 0.6$ et $\lambda = 0.9$

Nous remarquons bien que l'estimation de $P_x(f,k)$ est fortement sensible à la valeur du coefficient de lissage considérée. En effet :

- Pour une valeur de λ très proche de 1 (0.9 par exemple), le lissage est important. Ceci entraîne un faible biais dans l'estimation des dsps. En contre partie, un fort coefficient de lissage engendre des temps de montée et de descente importants, et donc des retards sur les fronts montant et descendant de la dsp du signal de bruit sont observés.

- Pour des valeurs plus faibles ($\lambda = 0.6$), les temps de montée et de descente sont évidemment plus faibles, mais l'estimation de $P_x(f,k)$ est plus sensible aux changements dynamiques des signaux.

Afin de pallier ce problème, différent travaux de recherche ont proposé des estimations récursives temps-fréquence du coefficient de lissage. Martin (Martin, 2001) a proposé une estimation temps-fréquence des valeurs de $\hat{\lambda}(f,k)$. Ces valeurs sont estimées afin que $P_x(f,k)$ soit la plus proche de la dsp du signal de bruit $P_b(f,k)$. L'expression de $\lambda_{opt}(f,k)$ est donnée par l'équation 2.8 :

$$\lambda_{opt}(f,k) = \hat{\lambda}(f,k) = \frac{\lambda_{max}\lambda_c(k)}{1 + (\frac{P_x(f,k-1)}{P_b(f,k)} - 1)^2} \tag{2.8}$$

où $\lambda_c(k)$ est un facteur de correction et λ_{max} est une constante égale 0.96.

La figure 2.3 montre les variations de la dsp du signal bruité estimée avec une valeur de $\lambda = 0.9$ et avec les valeurs de $\hat{\lambda}(f,k)$ estimées à partir de l'équation 2.8. Les valeurs de $\hat{\lambda}(f,k)$ sont aussi représentées pour $f = 15$.

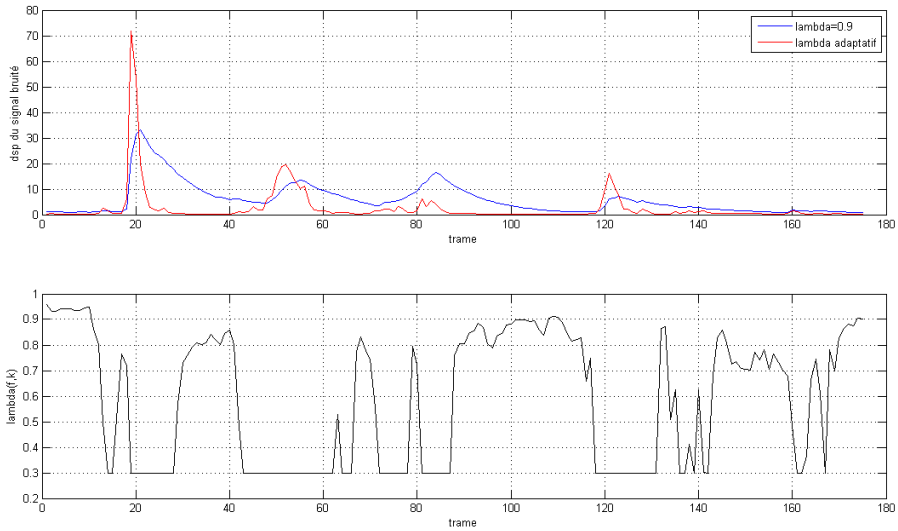


FIGURE 2.3 – psd estimée du signal bruité pour $\lambda = 0.9$ et $\hat{\lambda}(f,k)$ pour $f = 15$

L'estimation de $P_{xmin}(f,k)$ est assurée grâce à l'équation suivante

$$P_{xmin}(f,k) = min\{|P_x(f,m)|, k - L + 1 < m < k\} \tag{2.9}$$

où L est la largeur de la fenêtre de recherche comportant un nombre finie de trames.

Enfin, la dsp du bruit est estimée à partir de l'équation 2.10.

$$\hat{P}_b(f,k) = B.P_{xmin}(f,k) \tag{2.10}$$

où B est un coefficient constant utilisé afin de compenser le biais d'estimation.

2.5 Estimation de $P_{xmin}(f,k)$ suivant une méthode récursive

Dans le paragraphe précédent, les composantes du signal de bruit sont dépendantes de la largeur de la fenêtre utilisée pour l'estimation du minimum $P_{xmin}(f,k)$. En effet, ces minima ne sont pas mis à jour d'une manière continue à chaque trame de parole. Doblinger (Doblinger, 1995) a proposé une autre méthode récursive et non linéaire permettant l'estimation de $P_{xmin}(f,k)$ d'une manière continue à chaque trame de traitement. Cette méthode est décrite par l'équation 2.11.

$$\begin{array}{l} si \quad P_{xmin}(f,k-1) < P_x(f,k) \\ P_{xmin}(f,k) = \gamma P_{xmin}(f,k-1) + \frac{1-\gamma}{1-\beta}(P_x(f,k) - \beta P_x(f,k-1)) \\ sinon \\ P_{xmin}(f,k) = P_x(f,k) \end{array} \tag{2.11}$$

où $\gamma = 0.998$ et $\beta = 0.8$ sont deux constantes déterminées expérimentalement.

Farsi (Farsi, 2010) a proposé un nouvel algorithme pour une meilleure estimation de $P_{xmin}(f,k)$ moyennant l'équation 2.12.

$$\begin{array}{l} si \quad P_{xmin}(f,k-1) < P_x(f,k) \\ P_{xmin}(f,k) = \gamma P_{xmin}(f,k-1) + \frac{1-\gamma}{1-\beta}(P_x(f,k) - \beta P_x(f,k-1)) \\ sinon \\ P_{xmin}(f,k) = \gamma P_{xmin}(f,k-1) - \mu(P_{xmin}(f,k-1) - \zeta P_x(f,k-1)) \end{array} \tag{2.12}$$

où $\mu = 0.001$ et $\zeta = 0.9$ sont deux constantes déterminées expérimentalement.
La dsp du bruit est enfin estimée en utilisant un biais de compensation constant ($B_{min} = 2$). Les variations de $\hat{P}_b(f,k)$ sont données par la figure 2.4.

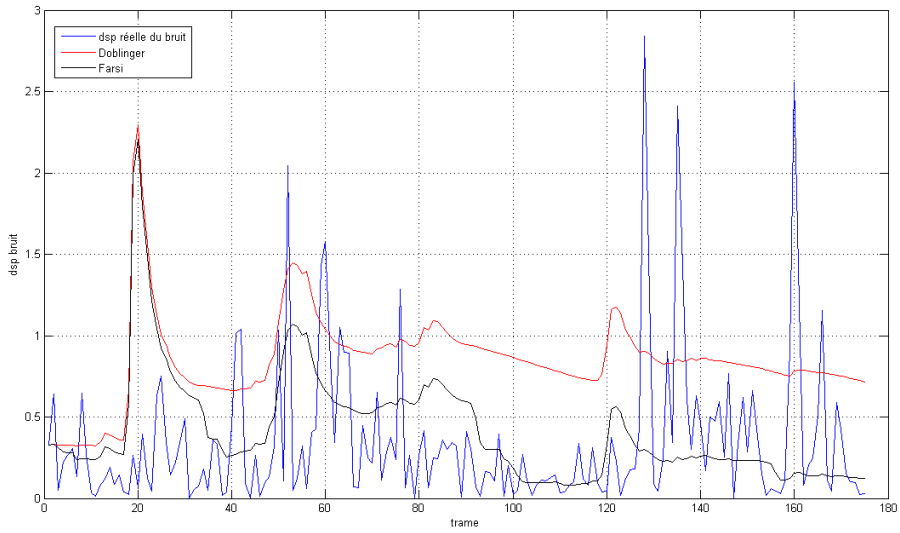


FIGURE 2.4 – psd réelle et estimée du bruit suivant les algorithmes de Doblinger et Farsi pour $f = 15$

Une autre approche récursive pour l'estimation du bruit a été proposé par Rangachari et Loizou (Rangachari et Loizou, 2006). L'estimation de $P_{xmin}(f,k)$ est déterminée de la

même manière que Doblinger (Doblinger, 1995) en utilisant l'équation 2.11. La différence est que la dsp du signal de bruit est estimée d'une manière récursive en utilisant un coefficient de lissage, noté $\alpha_s(f,k)$, déterminé suivant une estimation temps-fréquence suivant l'équation 2.13

$$\alpha_s(f,k) = \alpha_d + (1-\alpha_d)p(f,k) \tag{2.13}$$

où α_d est un coefficient de lissage constant et $p(f,k)$ est la probabilité de présence de la parole.
La dsp du signal de bruit réelle ainsi que celle estimée suivant l'algorithme proposé par Rangachari sont données par la figure 2.5 pour une fréquence discrète $f = 15$.

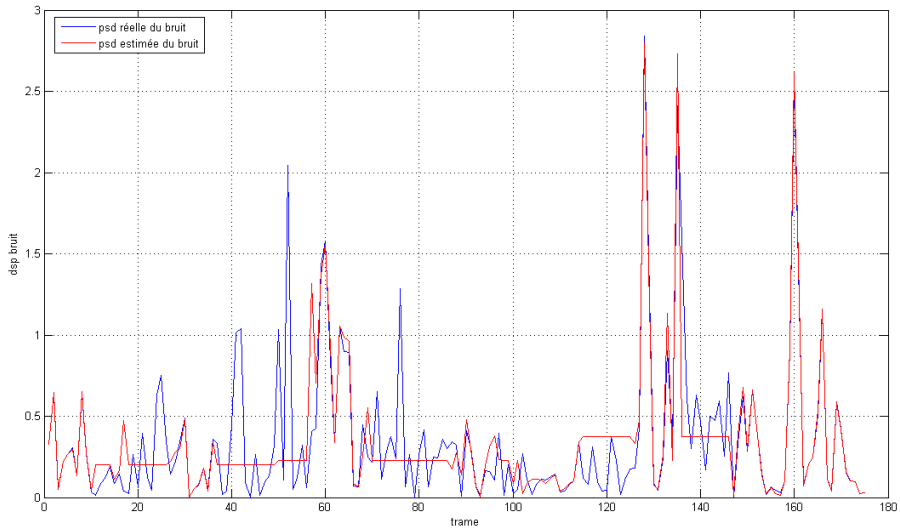


FIGURE 2.5 – psd réelle et estimée du bruit suivant la méthode de Rangachari pour $f = 15$

2.6 Algorithme d'estimation du signal débruité

2.6.1 Estimateurs basés sur les statistiques moyennes du bruit

Les filtres que nous allons décrire ici utilisent des statistiques moyennes dans le but de dériver le filtre optimal pour un critère donné.

2.6.1.a Soustraction spectrale

Dans le domaine de la réduction du bruit, la soustraction spectrale proposée par Boll s'octroie une place de choix (Boll, 1979) : son caractère intuitif et sa simplicité en ont fait un algorithme particulièrement répandu. Elle est basée sur la connaissance *a priori* de la dsp du bruit supposé stationnaire. $P_b(f)$ peut être estimée moyennant l'une des méthodes précédemment décrites. Le module de la TFCT du signal de parole est estimé à partir de

l'équation 2.14.

$$\left|\hat{S}(f,k)\right| = |X(f,k)| - \sqrt{\hat{P}_b(f,k)} \tag{2.14}$$

C'est la soustraction spectrale d'amplitude. Différentes lois issues de la relation 2.14 sont apparues, donnant naissance notamment à la soustraction spectrale de puissance (Vary, 1985). La dsp du signal de parole débruité est estimé relativement à l'équation 2.15.

$$\left|\hat{S}(f,k)\right|^2 = |X(f,k)|^2 - \hat{P}_b(f,k) \tag{2.15}$$

MacAulay (McAulay et Malpass, 1980) a montré que sous hypothèse d'indépendance et de distribution gaussienne des composantes spectrales du bruit et de la parole, l'équation 2.15 correspond à l'estimateur au sens du maximum de vraisemblance de la densité spectrale de puissance du signal utile, ce qui assure théoriquement une distorsion nulle de son spectre. Suite aux erreurs qui peuvent avoir lieu lors de l'estimation du spectre du bruit, des valeurs négatives peuvent apparaitre au niveau du spectre estimé du signal débruité. Ces valeurs sont mises à zéro. Il peut être aussi intéressant d'exprimer l'algorithme de la soustraction spectrale de puissance comme une opération de filtrage donnée par l'équation 2.16.

$$\hat{S}(f,k) = H_{ss}(f,k)X(f,k),\ avec\ H_{ss}(f,k) = \frac{\sqrt{|X(f,k)|^2 - \hat{P}_b(f)}}{|X(f,k)|} \tag{2.16}$$

A partir de cette relation, le filtre s'écrit en utilisant la définition du $RSB_{post}(f,k)$:

$$H_{ss}(f,k) = \sqrt{\frac{RSB_{post}(f,k)}{1 + RSB_{post}(f,k)}} \tag{2.17}$$

Etant donné que le gain $H_{ss}(f,k)$ ne dépend que de la TFCT du signal d'observation et non de statistiques à long-terme, il présente des variations importantes et l'algorithme souffre de bruit résiduel non naturel appelé bruit musical : c'est la conséquence des fluctuations de $B(f,k)^2$ présent dans $|X(f,k)|^2$ autour de sa moyenne $P_b(f)$. De nombreuses études ont été menées afin de réduire cet artefact particulièrement gênant à l'écoute. Nous pouvons citer les aménagements les plus courants suivants :

- Utiliser un spectre lissé $\left|\overline{X(f,k)}\right|$ au lieu de $|X(f,k)|$ permet de réduire ces fluctuations au détriment de la suppression des composantes dynamiques du signal de parole. Le spectre lissé peut être estimé à partir de l'équation 2.18.

$$\overline{X_i(f)} = \frac{1}{2M+1}\sum_{j=-M}^{M} W_j X_{i-j}(f) \tag{2.18}$$

  où i est l'indice de la trame et W_j sont des coefficients de pondération des trames.

- Utiliser un algorithme de la soustraction spectrale où une surestimation volontaire de la dsp du bruit est imposée au détriment des structures fines de la parole (Berouti *et al.*, 1979). Le coefficient de surestimation peut être rendu adaptatif en fonction de la fréquence suivant les types de bruit rencontrés (Bouquin-Jeannés, 1991). Cet algorithme est décrit par les équations 2.19 et 2.20.

$$\left|\hat{S}(f)\right|^2 = |X(f)|^2 - \alpha\hat{P}_b(f) \tag{2.19}$$

$$\left|\hat{S}(f)\right|^2 = \begin{cases} \left|\hat{S}(f)\right|^2 & \text{si } \left|\hat{S}(f)\right|^2 > \beta\hat{P}_b(f) \\ \beta\hat{P}_b(f) & \text{sinon} \end{cases} \tag{2.20}$$

où α est un facteur de surestimation calculé de manière adaptative en fonction du RSB en utilisant l'équation 2.21.

$$\alpha = \begin{cases} 4.75 & \text{si } RSB < -5 \text{ dB} \\ \alpha_0 - \frac{3}{20}RSB & \text{si } -5 \text{ dB} < RSB < 20 \text{ dB} \\ 4.75 & \text{si } RSB > 20 \text{ dB} \end{cases} \tag{2.21}$$

où α_0 est une constante égale à 4 et correspond à la valeur de α désiré pour un RSB=0 dB. Le coefficient β permet de limiter le spectre du signal de parole estimé à la valeur $\beta\hat{P}_b(f)$ et par la suite minimiser le bruit musical.

2.6.1.b Soustraction spectrale multi-bandes

L'algorithme de la soustraction spectrale multi-bande (Kamath et Loizou, 2002) est un algorithme qui tient compte du fait que le signal de bruit affecte de manière non uniforme le spectre de fréquence du signal de parole. Dans ce cas, Le spectre du signal bruité est découpé en L bandes fréquentielles non chevauchées. L'algorithme de la soustraction spectrale décrit dans le paragraphe précédant est appliqué pour chaque bande. Le spectre du signal de parole estimé pour la $l^{\text{ème}}$ bande est donné par l'équation suivante :

$$\left|\hat{S}(f,k)\right|^2 = |S(f,k)|^2 - \gamma_l\alpha_l\hat{P}_{bl}(f,k) \quad b_l < f < e_l \tag{2.22}$$

où b_l et e_l sont les limites fréquentielles inférieures et supérieures de la $l^{\text{ème}}$ bande (l=1...L). α_l est un coefficient de surestimation calculé à partir de l'équation 2.21 pour chaque bande fréquentielle d'indice l. Le facteur γ_l est un deuxième coefficient qui est réglé pour chaque bande de fréquences et qui permet de mieux personnaliser les propriétés d'élimination du bruit.

2.6.1.c Filtre de Wiener

Le filtre de Wiener est un filtre linéaire optimal au sens du Minimum de l'Erreur Quadratique Moyenne (MEQM). Le signal de sortie est estimé par la relation linéaire suivante :

$$\hat{S}(f,k) = W(f,k).X(f,k) \tag{2.23}$$

où $W(f,k)$ représente les coefficients du filtre de Wiener. L'erreur d'estimation $Er(f)$ est définie comme étant la différence entre le signal désirable $S(f)$ et le signal de parole estimé $\hat{S}(f)$. L'erreur quadratique moyenne est donnée par l'équation suivante :

$$E\{|Er(f)|^2\} = E\{[S(f) - W(f)X(f)][S(f) - W(f)X(f)]^*\} \tag{2.24}$$

Avec $E\{.\}$ est l'opérateur d'espérance mathématique, et le symbole $*$ dénote le conjugué de la quantité complexe. Les coefficients $W(f)$ sont déterminés par minimisation de l'erreur quadratique moyenne :

$$\frac{\partial\{E|Er(f)|^2\}}{\partial W(f)} = 2W(f)P_x(f) - 2P_s(f) = 0 \tag{2.25}$$

Les coefficients du filtre de Wiener peuvent enfin être exprimés suivant l'expression suivante :

$$W(f) = \frac{P_S(f)}{P_x(f)} = \frac{P_S(f)}{P_s(f) + P_b(f)} \tag{2.26}$$

En utilisant la définition du RSB a priori, le filtre de Wiener peut être exprimé sous la forme suivante :

$$W(f) = \frac{RSB_{prio}(f,k)}{1 + RSB_{prio}(f,k)} \tag{2.27}$$

En pratique, la dsp du bruit est apprise sur des périodes de silence, tandis que la dsp du signal n'est pas connue. Elle peut être obtenue par la soustraction spectrale précédente soit à partir du périodogramme lissé ($\hat{P}_S(f,k) = \left|\overline{X(f,k)}\right|^2 - \hat{P}_b(f,k)$), soit directement à partir des observations brutes ($\hat{P}_S(f,k) = |X(f,k)|^2 - \hat{P}_b(f,k)$). Cappé a donné à cette dernière formulation le nom de filtre de pseudo-Wiener (Cappé, 1994). Nous obtenons alors différentes formulations pour le filtre de Wiener. Pour ce dernier cas, nous trouvons la relation entre le filtre de Wiener et la soustraction spectrale de puissance définie par l'équation 2.16 : $W(f,k) = H_{ss}^2(f,k)$.

2.6.2 Estimateurs avec modèle de parole

Le filtre équivalent de la soustraction spectrale de puissance et celui de Wiener sont basés sur des critères de statistiques moyennes. Or, nous avons précisé que la parole ne vérifiait pas l'hypothèse de stationnarité, ce qui limite la portée théorique de ces filtres, bien que, dans les faits, ils soient largement utilisés. Pour cette raison, différentes approches ont été considérées en prenant comme point de départ l'estimation du module de la TFCT du signal de parole, au sens du $MEQM$. Cet estimateur est donné par l'équation suivante :

$$\left|\hat{S}(f,k)\right| = E\{|S(f,k)| \,/\, |X(f,k)|\} \tag{2.28}$$

La forme du filtre obtenu dépend de l'hypothèse faite sur les lois du signal et du bruit.

2.6.2.a Filtre de MacAulay et Malpass

En se basant sur l'observation que la parole présente une structure formantique que nous pouvons approcher par une somme de sinusoïdes à différentes fréquences, Macaulay et Malpass (Macaulay et Malpass, 1980) ont proposé de modéliser les composantes spectrales du signal de parole par une fréquence pure. Le bruit est supposé additif, blanc et gaussien. Le but est alors d'estimer une fréquence et une phase inconnues noyées dans un bruit blanc, en considérant une probabilité de présence de parole due à son intermittence. Les auteurs ont montré que le module du spectre du signal estimé est donné par l'équation suivante.

$$\left|\hat{S}(f,k)\right| = \frac{|X(f,k)| + \sqrt{|X(f,k)|^2 - \hat{P}_b(f,k)}}{2} \frac{\Lambda(f,k)}{1+\Lambda(f,k)} \tag{2.29}$$

Le premier terme dans le membre de droite représente l'estimation du signal utile au sens du maximum de vraisemblance sous hypothèse de présence de la parole. Le second terme représente la probabilité de présence du signal de parole, obtenue à partir de l'hypothèse gaussienne du bruit. La fonction $\Lambda(f,k)$ s'écrit sous la forme suivante.

$$\Lambda(f,k) = \exp\left(-RSB_{prio}(f,k)\right) I_0\left(2\sqrt{RSB_{prio}(f,k)\left(RSB_{post}(f,k)+1\right)}\right) \tag{2.30}$$

Où I_0 est la fonction de Bessel d'ordre 0 et donc difficilement envisageable dans un contexte temps-réel, sauf à considérer une tabulation. Il faut noter qu'elle dépend du $RSB_{prio}(f,k)$ et du $RSB_{post}(f,k)$, c'est-à-dire qu'elle prend en compte à la fois le comportement dynamique de la TFCT du signal d'observation (et donc du signal de parole) ainsi que les informations statistiques des signaux par le biais de $RSB_{prio}(f,k)$.

2.6.2.b Filtre d'Ephraim et Malah

Plutôt que de considérer le signal utile comme une fréquence pure,Ephraim et Malah (Ephraim et Malah, 1984) ont proposé d'approcher les composantes spectrales à court-terme du signal de parole par des composantes gaussiennes aléatoires indépendantes les unes des autres. Les hypothèses de distribution gaussienne du bruit ainsi que d'indépendance entre parole et bruit sont conservées. Ephraim et Malah (Ephraim et Malah, 1984) ont dérivé alors le filtre optimal pour ce modèle statistique dont l'expression fait intervenir des fonctions de Bessel d'ordres 0 et 1. Le module du spectre du signal de sortie est donné par l'équation suivante.

$$\begin{aligned}\left|\hat{S}(f,k)\right| &= \sqrt{\frac{\pi}{2}} \frac{\sqrt{V(f)}}{RSB_{post}(f)+1} \exp\left(-\frac{V(f)}{2}\right) \\ &. \left[(1+V(f))I_O\left(\frac{V(f)}{2}\right) + V(f)I_1\left(\frac{V(f)}{2}\right)\right] \frac{\Lambda(f,k)}{1+\Lambda(f,k)} |X(f,k)|\end{aligned} \tag{2.31}$$

avec les relations suivantes :

$$V(f) = \frac{RSB_{prio}(f)}{1+RSB_{prio}(f)} \left(RSB_{post}(f)+1\right) \tag{2.32}$$

$$\Lambda(f) = \frac{1-q(f)}{q(f)} \frac{\exp(V(f))}{RSB_{prio}(f)+1} \tag{2.33}$$

où q (f) est la probabilité d'absence de la parole à la fréquence f (l'indice temporel k a été omis dans les relations précédentes pour alléger les notations). Ce filtre est généralement trop complexe pour être utilisé tel quel du fait du calcul des fonctions de Bessel. Son intérêt est qu'il dépend de $RSB_{prio}(f,k)$ et de $RSB_{post}(f,k)$. L'estimateur proposé ne tient pas compte du modèle de perception de l'oreille humaine, plus sensible aux variations du logarithme de l'amplitude que de l'amplitude elle-même. Ephraim et Malah (Ephraim et Malah, 1985) ont également proposé un estimateur de la log-amplitude basé sur les mêmes hypothèses gaussiennes de bruit et de parole.

2.7 Algorithmes de réduction de bruit mono-voie dédiés pour implant cochléaire

L'implant cochléaire traite le signal de parole suivant un algorithme de codage spécifique. Différent algorithmes de débruitage mono-voie ont été proposés dans le cas de l'implant cochléaire. Ces algorithmes peuvent être classés en deux catégories. Pour la première catégorie, une étape de prétraitement assurant le rehaussement du signal de parole capté par le microphone est considérée. Le signal rehaussé est ensuite traité suivant la stratégie de codage considérée. Pour la deuxième catégorie, l'algorithme de rehaussement du signal de parole est intégré au niveau de l'algorithme de codage.

2.7.1 Méthode de la soustraction spectrale non linéaire

L'algorithme de la soustraction spectrale non linéaire a été proposé par Yang et Fu (Yang et Fu, 2005) afin d'améliorer la perception du signal de parole par les patients porteurs d'un implant cochléaire en milieux bruités. Cet algorithme comporte principalement deux phases : Une phase de détection d'activité vocale assurant l'estimation du spectre du bruit pendant les périodes de silence et une deuxième phase qui permet l'estimation du spectre du signal de parole débruité $\left|\hat{S}(f,k)\right|$ par application de l'algorithme de la soustraction spectrale tel que $\left|\hat{S}(f,k)\right| = G(f,k)\,|X(f,k)|$. Le gain $G(f,k)$ est estimé d'une manière non linéaire à partir de l'équation suivante :

$$G(f,k) = \left(A - K(f,k)\frac{\left|\bar{B}_M^a(f,k)\right|}{|X_M^a(f,k)|}\right)^{\frac{1}{a}} \tag{2.34}$$

où a est un facteur de puissance et A est une constante.
Afin d'évaluer les performances de l'algorithme de débruitage proposé, une évaluation subjective comportant cinq patients portant l'implant cochléaire Nucleus22, deux autres portant l'implant Clarion et un dernier portant l'implant Med-El est considérée. Les résultats ont montré des améliorations significatives des pourcentages de reconnaissance lorsque l'algorithme de débruitage proposé est considéré. En effet, le pourcentage d'amélioration est de 20% dans le cas du bruit de type speech-shaped noise. Par contre, dans le

cas du bruit de type Babble noise multi-locuteur, cette amélioration est moins importante et elle égale à 7.75% seulement.

2.7.2 Méthode à sous-espace signal

L'une des approches de débruitage de la parole qui a suscité beaucoup d'intérêt est le filtrage à sous-espace signal. Dans cette approche, nous développons un estimateur linéaire non paramétrique du signal de parole propre obtenu par décomposition du signal observé bruité en deux sous-espaces orthogonaux : le sous-espace signal et le sous-espace bruit. La décomposition est achevée soit par valeurs singulières ou par valeurs propres. Un premier algorithme a été proposé par Ephraim et Trees (Ephraim et Trees, 1995) lorsque le signal de bruit interférant est un bruit blanc. Une extension de cet algorithme dans le cas du bruit coloré a été proposée par Hu et Loizou (Hu et Loizou, 2002). Le signal de parole débruité est estimé en éliminant les composantes projetées dans le sous-espace bruit et retenant seulement les composantes projetées dans le sous-espace signal.
Nous considèrons un signal de parole bruité $\bar{x}(n)$ en notation vectorielle tel que $\bar{x}(n) = \bar{s}(n)+\bar{b}(n)$. Le signal de parole débruité, noté $\bar{\hat{s}}(n)$, peut être estimé en utilisant l'équation 2.35.

$$\bar{\hat{s}}(n) = \bar{H}\bar{x}(n) \tag{2.35}$$

où $\bar{H}$ est une matrice de transformation. La suppression de la contribution du bruit dans le sous-espace signal se fait selon un critère dont l'objectif est de trouver les éléments de la matrice de transformation $\bar{H}$. L'erreur associée $\epsilon(n)$ est estimée par la différence entre le signal réel $\bar{s}(n)$ et celui estimé $\bar{\hat{s}}(n)$. Cette erreur est donnée par l'équation suivante.

$$\epsilon = \bar{\hat{s}} - \bar{s} = \bar{H}\bar{x} - \bar{s} = \left(\bar{H}(\bar{s}+\bar{b})\right) - \bar{s} = \left(\bar{H} - \bar{I}\right)\bar{s} + \bar{H}\bar{b} \tag{2.36}$$

où $\bar{I}$ est la matrice identité. Le terme $\epsilon_s^2 = \left(\bar{H} - \bar{I}\right)\bar{s}$ représente la distorsion du signal de parole introduit par cet algorithme et le terme $\epsilon_b^2 = \bar{H}\bar{b}$ représente le bruit résiduel. Le critère d'estimation dans le domaine temporel s'écrit sous forme d'un problème d'optimisation (minimisation de l'énergie de la distorsion du signal) sous contrainte (seuil maximal du bruit résiduel). Ceci est exprimé par l'équation suivante :

$$\begin{aligned} &\min H\epsilon_s^2 \\ &\text{sous contrainte que } \epsilon_b^2 < q.\sigma^2 \end{aligned} \tag{2.37}$$

où σ^2 est la variance du bruit et q est un facteur qui contrôle le niveau admissible du bruit résiduel (0 < q < 1). Ce proplème d'optimisation a été traité et développé par Hu et Loizou (Hu et Loizou, 2002) dans le cas du bruit coloré. Après tout calcul fait, nous aboutissons à l'expression du filtre optimal H_{opt} suivante :

$$\bar{H}_{opt} = \bar{V}^{-T}\bar{\Lambda}\left(\bar{\Lambda} + \mu\bar{I}\right)^{-1}\bar{V}^{T} \tag{2.38}$$

où μ est un paramètre qui prend ses valeurs entre 1 et 20 suivant les valeurs du RSB à court terme estimé. $\bar{V}$ est une matrice diagonale à valeurs propres. La matrice à vecteurs

propres notée $\bar{\Lambda}^T$ permet la projection du signal de parole bruité dans les deux sous-espaces des signaux de parole et du bruit. La matrice diagonale $\bar{\Lambda}\left(\bar{\Lambda}+\mu\bar{I}\right)^{-1}$ permet de multiplier les composantes du signal de parole projetées dans le sous-espace signal par un gain différent de zéro, par contre elle permet d'annuler les composantes du signal projetées dans le sous-espace bruit. Cet algorithme de débruitage est appliqué sur des trames de parole de durée 4 millisecondes avec un chevauchement de 50%. La matrice de covariance du bruit est calculée pendant les premières trames du signal de parole bruité qui sont considérées comme des trames de silence.
Les performances de l'algorithme de débruitage proposé sont évaluées par des tests subjectifs de reconnaissance avec 14 patients porteurs d'un implant cochléaire dont neuf sujets portent l'implant cochléaire Clarion CII et les cinq autres sujets portent l'implant Clarion S-series. L'implant Clarion CII utilise la stratégie de codage CIS, alors que l'implant Clarion S-series utilise la stratégie de codage SAS. Le matériel de test utilisé correspond à des phrases (parole continue) noyé dans un bruit de type speech-shaped noise à un RSB=5 dB. Un premier test avec les signaux bruités est considéré. Dans une deuxième condition, des signaux de parole débruités avec l'algorithme de débruitage proposé sont utilisés. Les pourcentages de reconnaissance moyens obtenus avec la première condition sont de 19%, alors qu'ils augmentent significativement avec la deuxième condition pour atteindre 44%.

2.7.3 Exploitation de la fonction sigmoïdale pour la réduction du bruit

Cet algorithme de réduction de bruit a été proposé par Hu et al. (Hu *et al.*, 2007). Il présente l'avantage de sa simplicité et la possibilité d'intégration au niveau des algorithmes de codage du signal de parole actuellement utilisés. Deux étapes de traitement majeures sont utilisées : Estimation du RSB et calcul d'une fonction poids, ceci étant à chaque trame temporelle k et pour chaque canal de stimulation i. La fonction poids $g(i,k)$ est estimée à partir de l'équation suivante :

$$g(i,n) = \exp\left(\frac{-\beta}{RSB(i,k)}\right) \tag{2.39}$$

où $\beta = 2$ et $0 < g(i,k) < 1$.
Le signal de parole est aussi estimé tel que $s(i,k) = g(i,k)x(i,k)$ où $x(i,k)$ est le signal de parole buité.
Cet algorithme de réduction du bruit a éte testé subjectivement avec 9 implantés portant l'implant cochléaire Clarion CII. Un premier bruit de type speech shaped pour deux valeurs du RSB=0 et 5 dB est considéré. Le deuxième est un bruit de type Babble considéré à deux valeurs du RSB=5 et 10 dB. Dans le cas du bruit de type Babble, les résultats ont montré un effet significatif de l'algorithme de réduction du bruit considéré pour les deux RSBs considérés comparé au cas des signaux non traités. Une amélioration plus importante du pourcentage de reconnaissance est notée pour la valeur du RSB la plus importante. Dans le cas du bruit de type speech shaped, le comportement est variable en fonction du RSB. En effet, comparé aux signaux non traité, l'algorithme de réduction du bruit présente un effet significatif pour un RSB= 0 dB, par contre l'effet n'est pas significatif pour un RSB= 5 dB.

2.7.4 Exploitation de la fonction de compression pour le débruitage du signal de parole

La fonction de compression est une fonction très importante dans l'implant cochléaire puisqu'elle permet de compresser (par bande de fréquences) la dynamique acoustique du signal de parole à la dynamique électrique de notre système auditif. L'effet de la forme de la fonction de compression sur les pourcentages de reconnaissance dans le bruit a été étudié par Fu et Shannon (Fu et Shannon, 1999), particulièrement la fonction de compression sous forme logarithmique. Les résultats ont montré que l'application d'une forte fonction de compression entre les amplitudes acoustiques et les amplitudes électriques produit de meilleures performances en milieux silencieux. Par contre, l'application d'une faible fonction de compression est parfois plus bénéfique en milieu bruité.

Kasturi et Loizou (Kasturi et Loizou, 2007) ont proposé une nouvelle fonction de compression sous forme sigmoïdale pour remplacer la fonction logarithmique classique. Cette fonction permet d'amplifier les faibles dynamiques du signal d'entrée jusqu'à un point seuil et compresser les dynamiques au-delà. Ce point seuil de la fonction de compression sigmoïdale est calculé dynamiquement suivant le plancher de bruit estimé. Les performances de cet algorithme de débruitage ont été étudiées avec neuf patients porteurs d'un implant cochléaire. Deux types de bruit sont considérés : Bruit de type Babble multi-locuteur à des RSB=5 et 10 dB et un bruit de type speech-shaped noise à un RSB de =5 dB. Les résultats ont indiqué que l'utilisation de la fonction de compression sigmoïdale n'a pas donné d'améliorations significatives pour tous les utilisateurs d'implants cochléaires. Par contre, des améliorations significatives de l'intelligibilité de la parole ont été observées lorsque la fonction de compression sigmoïdale est optimisée d'une manière individuelle pour chaque patient implanté.

3 Aperçu général sur les techniques de réduction du bruit bi-voie

Bien que les systèmes de débruitage mono-voie fournissent des résultats intéressants dans certaines applications audio et même si une augmentation des performances est souvent rapporté pour les appareils auditifs, ils ne fournissent pas de bénéfice en terme d'intelligibilité de la parole (Loizou et Kim, 2011). Cela indique que les méthodes de réduction du bruit mono-voie sont contraintes de l'hypothèse particulièrement pénalisant qui est l'hypothèse de stationnarité (à long-terme) du bruit. Cet hypothèse, qui n'est pas valide dans la plupart des cas, est en partie à l'origine du bruit musical généralement observé. Les informations fournies par d'autres voies d'observation permet d'envisager une meilleure discrimination des signaux de parole et de bruit en se basant sur des critères autres que celui de la stationnarité : Un capteur supplémentaire permet notamment d'introduire des considérations spatiales sur les signaux.
Nous supposons la présence de parole et de bruit sur les deux microphones, ce qui mène à deux signaux d'observation $x_i(n), i = 1, 2$, composés d'un mélange additif de bruit $b_i(n)$

et de parole $s_i(n)$ (issus d'un même signal $s(n)$) supposés indépendants :

$$x_i(n) = s_i(n) + b_i(n) \tag{2.40}$$

où n est l'indice temporel.
Nous désignons les Transformées de Fourier à Court-Terme (TFCT) des signaux par des lettres capitales $X_i(f,k) = FFT(x_i(n))$, où f est la fréquence et k l'indice de trame. Cet indice sera parfois omis pour la lisibilité des expressions. Les densités spectrales et interspectrales des signaux captés par les deux microphones sont notées

$$\begin{aligned} P_{xi}(f) = E\{X_i(f).X_i^*(f)\}, \quad i = \{1,2\} \\ P_{x1x2}(f) = E\{X_1(f).X_2^*(f)\} \end{aligned} \tag{2.41}$$

où $E(.)$ est l'opérateur d'espérance mathématique. Nous définissons de même les dsp des signaux $b_1(n)$, $b_2(n)$, $s_1(n)$ et $s_2(n)$ respectivement par $P_{b1b1}(f)$, $P_{b2b2}(f)$, $P_{s1s1}(f)$ et $P_{s2s2}(f)$. De manière identique, nous notons $P_{b1b2}(f)$ la dip entre bruits et $P_{s1s2}(f)$ la dip entre signaux de parole.
En toute rigueur, les transformées de Fourier dépendent également du temps car les signaux observés ne sont généralement stationnaires qu'à court-terme. La décomposition des signaux est réalisée par un fenêtrage (Hanning) et une transformée de Fourier. Étant donné que la plupart des filtres qui seront définis et exploités dans ce travail opèrent dans le domaine spectral, nous définissons la relation 2.42 dans l'espace de Fourier :

$$X_i(f,k) = S_i(f,k) + B_i(f,k) \tag{2.42}$$

Les signaux de parole des deux voies sont fortement corrélés du fait de la distance raisonnable entre les microphones et de la ponctualité de la source. Nous supposons généralement qu'il existe soit un canal acoustique liant les deux voies, soit un simple retard, supposant donc une corrélation parfaite. Quant au bruit, trois hypothèses se présentent suivant ses caractéristiques spatiales :

- Soit il est issu d'une source ponctuelle, ce qui assure une forte corrélation entre les microphones.
- Soit il est complètement décorrélé entre les deux voies. C'est l'hypothèse émise par le filtre de Wiener bi-voies, la fonction de cohérence, la soustraction spectrale...
- Soit il existe une certaine corrélation, mais non parfaite (bruit diffus). Pour ce dernier point, nous décrirons la méthode de la soustraction interspectrale.

3.1 Hypothèse des bruits corrélés

Sous l'hypothèse des bruits corrélés, les signaux de bruits sont corrélés, i.e. issus d'une même source $b(n)$. Nous considèrons ici une première approche qui prend comme hypothèse que nous disposons d'une première observation formée du signal utile $s_1(n)$ et une deuxième observation du bruit de référence sans la présence de parole $b_2(n)$. Le modèle est représenté sur la figure 2.6.

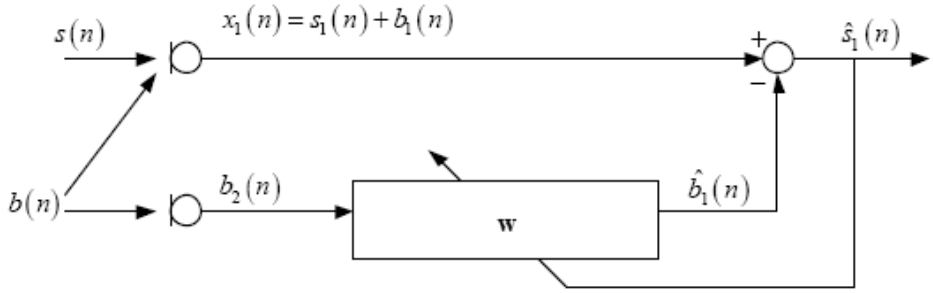


FIGURE 2.6 – Annulation de bruit avec référence de bruit seul

L'objectif est d'estimer une relation, généralement supposée linéaire convolutive, liant les deux bruits $b_1(n)$ et $b_2(n)$ de telle sorte qu'après identification de cette relation, l'estimée du bruit $\hat{b}_1(n)$ est soustraite à la voie une pour donner une estimation du signal de parole utile $\hat{s}_1(n)$. Sous cette hypothèse linéaire, le canal acoustique entre la première et la deuxième voies peut être modélisé comme une somme de couples retard-atténuation représentés par un filtre transversal de taille N, $W = [w_1, ..., w_n]^T$. Le bruit estimé est alors donné par la convolution discrète :

$$\hat{b}_1(n) = W^T.b_2(n) \tag{2.43}$$

Le calcul des coefficients du filtre W est réalisé en minimisant l'erreur quadratique moyenne (EQM) en sortie du soustracteur et fait appel directement à des notions de filtrage optimal (le filtre de Wiener).
Réellement, la voie de référence avec une source de bruit seul n'existe pas. La difficulté réside dans l'estimation de cette voie de bruit de référence. Lorsque les bruits sont supposés quasi-parfaitement corrélés entre les voies, l'objectif est de retrouver la situation précédente avec référence de bruit seul (Mezalek, 1990).

3.2 Hypothèse de bruits décorrélés

Dans ce paragraphe, nous présentons différentes technique de débruitage bi-voie qui peuvent être utilisées sous l'hypothèse de la décorrélation parfaite des signaux de bruits captés par les deux microphones. Nous présentons particulièrement les méthodes de débruitage basées sur la fonction de cohérence, la soustraction spectrale et le beamformer.

3.2.1 Cohérence

La cohérence entre deux signaux $x_1(n)$ et $x_2(n)$ supposés stationnaires s'exprime dans le domaine spectral comme la corrélation normalisée de leurs coefficients de Fourier suivant l'équation 2.44.

$$\rho(f) = \frac{E\{X_1(f).X_2^*(f)\}}{\sqrt{E\{|X_1(f)|^2\}.E\{|X_2(f)|^2\}}} = \frac{P_{x1x2}(f)}{\sqrt{P_{x1x1}(f).P_{x2x2}(f)}} \tag{2.44}$$

La cohérence fournit une information sur la corrélation entre deux signaux à partir de la corrélation de phase de leurs composantes fréquentielles : pour deux signaux statistiquement indépendants, $\rho(f)$ est nulle. La cohérence peut être également interprétée comme une mesure de corrélation spatiale, la phase étant intimement liée à la position de la source du signal. Nous introduisons généralement la grandeur réelle MSC (Magnitude Squared Coherence), qui est le module de la cohérence au carré.

$$MSC(f) = |\rho(f)|^2 \tag{2.45}$$

Nous pouvons grossièrement distinguer deux familles de signaux discriminés par la cohérence :

- les signaux cohérents : par définition, ces signaux présentent une cohérence proche de 1, indiquant une forte corrélation spatiale. C'est principalement le cas des signaux dont la source est ponctuelle, comme la parole par exemple, mais aussi de certains bruits de l'environnement réel.
- les signaux incohérents : ce sont des signaux qui sont généralement issus d'une multitude de sources considérées comme indépendantes entre elles. Les phases de ces sources se compensent mutuellement pour donner une valeur de cohérence proche de zéro pour des distances entre les capteurs de l'ordre de quelques longueurs d'onde.

Hypothèses simplificatrices :
Il est possible de montrer sous certaines hypothèses que la fonction de cohérence peut servir d'estimateur du filtre optimal de Wiener. Considérons les hypothèses suivantes :

Hypothèse 1 : $E\{S_i(f).B_j(f)\} = 0 \quad \forall(i,j) \in \{1,2\}^2$.
Justification : les signaux de parole et les bruits sont statistiquement indépendants.

Hypothèse 2 : $P_{s1s1}(f) = P_{s2s2}(f) = P_{ss}(f)$ *et* $P_{b1b1}(f) = P_{b2b2}(f) = P_{bb}(f)$.
Justification : du fait de la faible distance entre les microphones et le locuteur proche, nous pouvons considérer que le spectre de la parole locale est identique sur les deux voies. Dans le cas du bruit, si l'égalité exacte des dsp sur les deux microphones est moins valide, elle est souvent restreinte à une différence de gain, gardant la même distribution spectrale.

Hypothèse 3 : $P_{s1s2}(f) = P_{s1s1}(f) = P_{s2s2}(f) = P_{ss}(f)$
Justification : du fait de la ponctualité de la source de parole et du faible caractère réverbérant de l'environnement buité, cette hypothèse est pratiquement vérifiée.

Ces hypothèses permettent de simplifier l'expression de la cohérence. A partir de l'équation 2.44, nous obtenons la relation suivante :

$$\rho(f) = \frac{P_{ss}(f) - P_{b1b2}(f)}{P_{ss}(f) + P_{bb}(f)} \tag{2.46}$$

Dans le cas des bruits parfaitement décorrélés, le module de la cohérence s'exprime par :

$$\rho(f) = \frac{P_{ss}(f)}{P_{ss}(f) + P_{bb}(f)} \tag{2.47}$$

C'est le gain optimal à phase nulle qui correspond donc au filtre de Wiener.
La cohérence, sous forme de filtre, est largement utilisée en réduction de bruit. D'un point de vue plus général, le module de la cohérence, appliqué comme un gain sur une des voies, permet d'éliminer la composante de bruit non corrélée, sans autre hypothèse que l'indépendance de la parole et du bruit. Dans un contexte réel, les deux voies contiennent du bruit ambiant (la corrélation du bruit n'est pas nulle). L'information de cohérence seule ne permet donc pas de discriminer la parole du bruit surtout dans les zones de fréquence où le bruit est fortement énergétique. Il n'en reste pas moins que cet opérateur est largement utilisé du fait de son lien avec le filtre de Wiener dans le cas de bruits décorrélés.

3.2.2 Formation de voie

La technqiue de formation de voie ou beamforming est une technique basée sur la combinaison linéaire des signaux d'entrée captés par les différents microphones afin de rehausser ou supprimer les signaux provenant de certaines directions. La réalisation pratique de ce type de filtre spatial impose l'utilisation d'au moins deux microphones. Principalement, deux techniques de formation de voie existent : formation de voie fixes et formation de voie adaptative.

3.2.2.a Formation de voie fixe

Dans le cas de la technique de formation de voie (beamformer en terminologie anglaise) fixe, un filtrage des signaux captés par les différents microphones est considéré. Les filtres utilisés dans ce cas sont fixes et par la suite indépendants des informations à traiter. Les coefficients du filtre sont optimisés afin de focalier le beam vers la direction d'intérêt qui correspond, typiquement dans les cas des aides auditives et des implants cochléaires, au champ frontal de l'observateur. La formation de faisceau réalise alors un filtrage supprimant les signaux qui ne proviennent pas de cette direction d'intérêt. Différentes techniques de formation de voie fixe ont été développées telles que la technique retard et soustraction, la technique retard et sommation et la technique filtrage et sommation. La figure 2.7 montre le principe de la technique de formation de voie fixe de type retard et soustraction (Dillon, 2001a). Le signal de sortie est généré par soustraction du signal provenant du premier microphone frontal du signal provenant du microphone arrière décalé de τ (Ricketts et Henry, 2002). La fonction de transfert du Beamformer fixe est définie par l'équation suivante :

$$\bar{H}(f, \theta) = A_1 - A_2 e^{j2\pi f_0 \tau} \tag{2.48}$$

où θ est l'angle d'incidence et $A1$ et $A2$ représentent les amplitudes des signaux de parole au niveau des microphones avant et arrière respectivement. Le paramètre τ représente le

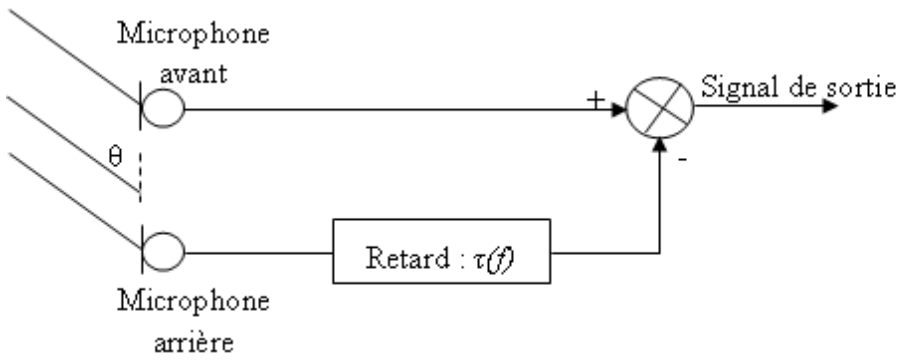


FIGURE 2.7 – Formation de voie fixe de type retard et soustraction

retard total engendré par le microphone arrière. Ce retard est la somme du retard fixe τ_0 et le retard dû à la distance inter-microphone d, soit $\tau = \tau_0 + \frac{d}{c}\cos(\theta)$ avec c est la célérité du son. Toutefois, ces algorithmes ne sont pas couramment utilisés vu qu'ils nécessitent un nombre de microphones assez important et que leurs performances ne deviennent efficaces que si la distance inter-microphone est au moins inférieure au quart de longueur d'onde de la fréquence d'intérêt (Dillon, 2001b).

3.2.2.b Formation de voie adaptative

L'algorithme de formation de voie fixe est un algorithme caractérisé par sa simplicité mais ses performances ne sont valables que dans des situations bien déterminées. Dans le cas où l'emplacement des sources interférentes devient inconnu et variable en fonction du temps, ce type d'algorithme devient inefficace et par la suite l'utilisation d'un algorithme de formation de voie adaptative devient primordiale. En effet, l'algorithme de formation de voie adaptatif est capable de mettre à jour les coefficients du filtre utilisé et par la suite de ses caractéristiques spatiales d'une manière instantannée en fonction de l'environnement acoustique considéré. En effet, les algorithmes de formation de voie adaptatifs ont montré de meilleures performances que les algorithmes de formation de voie fixe, particulièrement dans les cas où le nombre de sources interférentes est plus important que le nombre de microphones. Différents travaux de recherche ont traité le problème de réduction du bruit à base de la technique de formation de voie adaptative (Veen et Buckley, 1988; Compernolle et Gerven, 1995; Spriet, 2004).
Pour éviter toute distorsion du signal de parole, ces algorithmes de réduction du bruit adaptatifs basés sur la technique de formation de voie sont généralement limités à préserver les signaux arrivant juste dans le champ frontal. Ces algorithmes conduisent généralement à un problème d'optimisation avec contraintes. La figure 2.8 présente la structure d'un Beamformer adaptatif à deux microphones utilisée par (Luo *et al.*, 2002) pour la réduction du bruit dans le cas des aides auditives.
Cet algorithme, caractérisée par sa faible complexité, comporte principalement trois parties de traitement. Une première partie basée sur un algorithme de formation de voie fixe de type décalage et soustraction avec un paramètre de décalage noté τ qui permet de générer un signal de bruit de référence. Une deuxième partie qui utilise aussi un algorithme de formation de voie fixe de type décalage et soustraction avec un paramètre de décalage τ_f et qui permet de générer un signal de parole de référence. La troisième partie combine

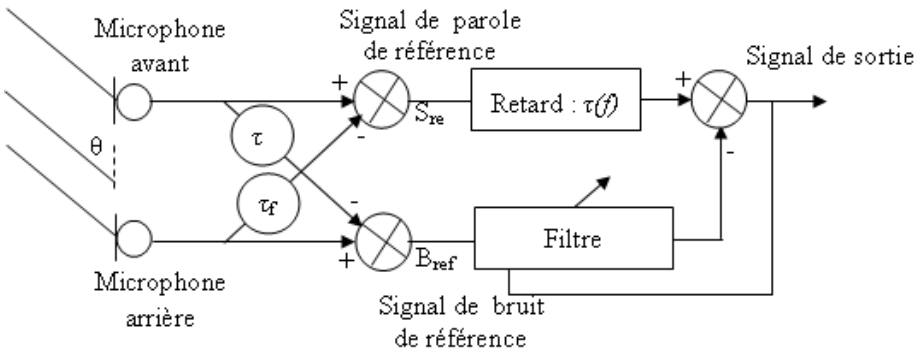


FIGURE 2.8 – Représentation du Beamformer adaptatif

les deux configurations spatiales obtenues suivant un traitement adaptatif pour fournir en sortie un signal de parole rehaussé.
Les performances de l'algorithme du Beamformer adaptatif bi-voie ont été étudiées dans le cas de l'implant cochléaire par différent travaux de recherche (Spriet *et al.*, 2007; Chung *et al.*, 2006). Dans leur travail, Spriet et al. (Spriet *et al.*, 2007) on étudié les performances de l'algorithme du Beamformer adaptatif bi-voie avec cinq patients porteurs de l'implant cochléaire Nucleus Freedom speech processor en présence de deux types de bruit : bruit blanc et bruit de type babble à deux niveaux du RSB. Les résultats ont montré que cet algorithme de débruitage permet une amélioration du niveau de perception de la parole entre 5 et 8 dB.
Les performances de l'algorithme du Beamformer adaptatif ont été aussi exploités dans le cas de l'implant cochléaire en présence de quatre microphones (deux microphones omnidirectionnels et deux autres directionnels)(Kokkinakis et Loizou, 2010). Les résultats obtenus avec des sujets implantés ont montré que cet algorithme de réduction du bruit permet d'améliorer l'intelligibilité de la parole et présente des performances supérieures à l'algorithme du Beamformer adaptatif bi-voie en présence d'une seule et de plusieurs sources de bruit.

3.2.2.c Algorithme de la soustraction spectrale bi-voie

Les algorithmes classiques de réduction du bruit basés sur un traitement mono-voie souffrent du problème du bruit musical. Un algorithme bi-voie pour la réduction du bruit a été proposé par Dörbecker et Ernst (Dörbecker et Ernst, 1996). Cet algorithme combine un estimateur bi-voie de la densité spectre de puissance du signal de bruit au niveau de la première voie $P_{b1b1}(f,k)$ et la deuxième voie $P_{b2b2}(f,k)$ avec l'algorithme de la soustraction spectrale proposé par Ephraim et Malah (Ephraim et Malah, 1985).
Puisque la cohérence entre les deux signaux de parole est supposée proche de 1, le canal acoustique entre le locuteur et les microphones peut être modélisé par deux fonctions de transfert $H_1(f)$ et $H_2(f)$. En tenant compte de la faible cohérence du bruit diffus aux basses fréquences, les signaux de bruit $b_1(n)$ et $b_2(n)$ reçus par les deux microphones peuvent être représentés par deux sources de bruit additives et décorrélées dont les spectres à court terme sont représentés respectivement par $B_1(f,k)$ et $B_2(f,k)$ et la densité interspectrale de puissance noté $P_{b1b2}(f,k)$ est nulle. Les TFCT des signaux bruités au niveau

des deux microphones, notées $X_1(f,k)$ et $X_2(f,k)$, sont alors données par les équations suivantes.

$$\begin{aligned} X_1(f,k) &= H_1(f,k)S_1(f,k) + B_1(f,k) \\ X_2(f,k) &= H_2(f,k)S_2(f,k) + B_2(f,k) \end{aligned} \tag{2.49}$$

Les deux hypothèses suivantes sont prises en considération dans le travail de Dörbecker et Ernst (Dörbecker et Ernst, 1996) :

Hypothèse 1 : $S_1(f,k) = S_2(f,k) = S(f,k)$ et $P_{b1b1}(f,k) = P_{b2b2}(f,k) = P_{bb}(f,k)$.
Justification : la distance entre les deux microphones ainsi que celle entre le locuteur et les microphone sont suffisamment faibles.

Hypothèse 2 : $H_1(f,k) = H_2(f,k) = H(f,k)$.
Justification : l'atténuation des signaux de parole est identique pour les deux microphones.

En tenant compte de ces hypothèses, la dip entre les deux signaux de parole bruités peut s'écrire sous la forme suivante.

$$\begin{aligned} |P_{x1x2}(f,k)| &= |E\{X_1(f,k)X_2(f,k)\}|^2 \\ &= E\{|S(f,k)|^2\}^2 \, |H_1(f,k)|^2 \, |H_2(f,k)|^2 \\ &= E\{|S(f,k)|^2\}^2 \, |H_1(f,k)|^4 \end{aligned} \tag{2.50}$$

Nous obtenons de la même manière le produit des dsp des signaux bruités :

$$\begin{aligned} P_{x1}(f,k)P_{x2}(f,k) &= E\{|X_1(f,k)|^2\}E\{|X_2(f,k)|^2\} \\ &= E\{|B_1(f,k)|^2\}E\{|B_2(f,k)|^2\} + E\{|S(f,k)|^2\}^2 \, |H_1(f,k)|^2 \, |H_2(f,k)|^2 \\ &+ E\{|S(f,k)|^2\}\{|H_1(f,k)|^2 \, E\{|B_2(f,k)|^2\} + |H_2(f,k)|^2 \, E\{|B_1(f,k)|^2\}\} \\ &= \left(E\{|B(f,k)|^2\}\{|H(f,k)|^2 \, E\{|S(f,k)|^2\}\right)^2 \end{aligned} \tag{2.51}$$

En combinant les équations 2.50 et 2.51, l'expression de la dsp du bruit est donnée par l'équation suivante :

$$E\{|B(f,k)|^2\} = P_{bb}(f,k) = \sqrt{P_{x1x1}(f,k)P_{x2x2}(f,k)} - |P_{x1x2}(f,k)| \tag{2.52}$$

Les spectres de puissance à cout-terme du signal de parole rehaussé au niveau de la première ($\hat{X}_1(f,k)$) et la deuxième ($\hat{X}_2(f,k)$) voie sont estimés en utilisant l'algorithme de la soustraction spectrale proposé par Ephraim et Malah (Ephraim et Malah, 1985). Afin de minimiser le bruit résiduel au niveau des deux voies, une deuxième étape de post-traitement basée sur un filtrage de Wiener est considérée. L'expression de ce post filtre est donnée par l'équation suivante :

$$W(f,k) = \frac{|P_{\hat{x}_1\hat{x}_2}(f,k)|^2}{\left(P_{\hat{x}_1\hat{x}_1}(f,k) + P_{\hat{x}_2\hat{x}_2}(f,k)\right)^2} \tag{2.53}$$

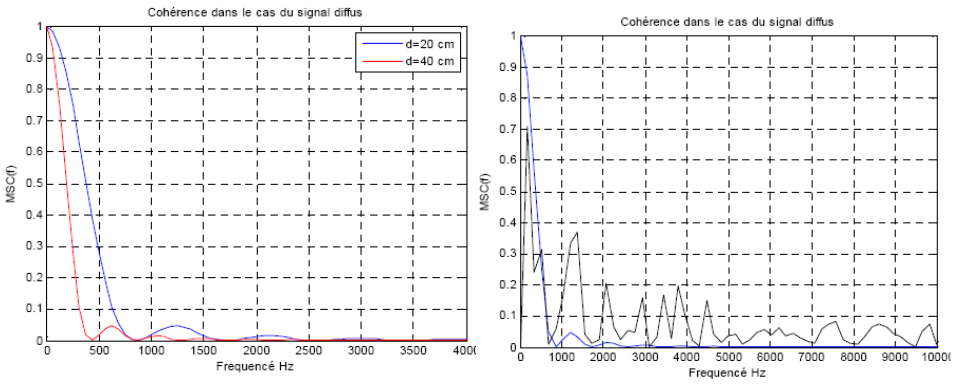
où $P_{\hat{x}_1\hat{x}_1}(f,k)$, $P_{\hat{x}_2\hat{x}_2}(f,k)$ et $P_{\hat{x}_1\hat{x}_2}(f,k)$ représentent respectivement les dsp des signaux rehaussés (issus de l'algorithme de rehaussement précédant) au niveau de la première et la deuxième voie ainsi que leur dip.
Ce filtrage de Wiener est appliqué simultanément sur les deux voies. Le résultats de cette opération de filtage correspond aux spectres de puissance des signaux rehaussés $\hat{S}_1(f,k)$ et $\hat{S}_2(f,k)$. Finalement, les signaux temporels sont obtenus par une Transformée de Fourier Rapide Inverse (TFRI) et une reconstruction par la méthode OverLap and Add (OLA).

3.3 Hypothèse de bruits diffus ou peu corrélés

Nous avons vu que les hypothèses visant à considérer les bruits parfaitement corrélés ou idéalement décorrélés ne se justifiaient plus dans les cas réels : de part leur configuration spatiale et la position des microphones, les perturbations présentent un profil se situant entre les deux extrêmes. Parmi les signaux corrélés, nous définissons un signal particulier dit signal diffus, dont les caractéristiques correspondent peu ou prou à celles de signaux réels rencontrés en environnement réel : un signal est diffus s'il est constitué d'une multitude de sources indépendantes, de même puissance et uniformément réparties sur un cercle autour du dipôle formé des deux microphones. A partir de cette définition, nous pouvons montrer que l'expression théorique de la cohérence d'un tel signal est donnée par l'équation 2.54 (Guérin, 2002) :

$$\rho(f) = \frac{\sin(2\pi f dc)}{2\pi f dc} \tag{2.54}$$

où f est la fréquence, d est la distance entre les microphones et c est la célérité du son dans l'air. La Figure 2.9 représente la MSC pour un signal théoriquement diffus ainsi que la MSC dans le cas d'un bruit réel correspondant à un bruit de type Babble. Différentes observations se dégagent : la MSC sur signaux réels est sensiblement plus élevée, ce qui est dû d'une part au biais de la méthode d'estimation des spectres (la méthode du périodogramme récursivement lissé), et d'autre part parce que le bruit considéré n'est pas réellement diffus.



(a) MSC théorique pour d = 20 et 40 cm (b) MSC dans le cas d'un bruit Babble

FIGURE 2.9 – MSC théorique et réelle dans le cas d'un bruit de type Babble

Sous l'hypothèse des bruits diffus, différent algorithmes bi-voie ont été développés en vue d'une estimation du signal débruité. Une des méthodes très souvent utilisée est basée sur l'algorithme de la Soustraction InterSpectrale (SIS). Cette méthode est basée sur la détermination des coefficient du filtre de Wiener utilisant la fonction de cohérence donnée par l'équation 2.55 en présence de bruits corrélés. Si nous supposons la dip des bruits $\hat{b}_1(n)$ connue, nous remarquons que le filtre de Wiener s'obtient simplement en soustrayant ce dernier du numérateur de la cohérence. La cohérence modifiée tenant compte de la corrélation du bruit s'écrit sous la forme suivante :

$$\hat{\rho}(f) = \frac{P_{s1s2}(f) + P_{b1b2}(f) - \hat{P}_{b1b2}(f)}{P_{ss}(f) + P_{bb}(f)} \tag{2.55}$$

Cependant, il est important de noter que les dip possèdent une phase non nulle : si la phase de $\hat{P}_{b1b2}(f)$ est opposée à celle de $P_{b1b2}(f)$, l'opération de soustraction de l'équation 2.55 peut devenir une addition de bruit. Afin d'éviter un rehaussement éventuel du bruit, nous préfèrons effectuer une soustraction des modules des dip, ce qui assure une réelle soustraction de la puissance des bruits corrélés. Le gain du filtre, noté $HSIS(f)$ est alors donné par l'équation 2.56 (Azirani, 1995) :

$$H_{SIS}(f) = \frac{|P_{x1x2}(f)| - \left|\hat{P}_{b1b2}(f)\right|}{\sqrt{P_{x1x1}(f)P_{x2x2}(f)}} \tag{2.56}$$

Nous pouvons donc assimiler cette méthode à une soustraction spectrale de puissance avec moyennage du spectre instantané étendu à deux voies : le spectre lissé de la soustraction spectrale $\left|X(\bar{f},k)\right|^2$ est équivalent à la dip $\hat{P}_{b1b2}(f)$. Il faut noter que ce filtre ne peut pas se justifier de manière théorique puisque, à l'image de la soustraction spectrale en amplitude, il subsiste une indécision sur les phases des signaux, indécision levée par l'emploi des modules. Comme dans la soustraction spectrale, si le gain prend des valeurs négatives, il est mis à zéro. Les signaux de sortie sont estimés en appliquant indifféremment le gain réel aux deux voies suivant l'équation 2.57.

$$\hat{S}_i(f,k) = H_{SIS}(f,k).X_i(f,k) \quad i = \{1,2\} \tag{2.57}$$

D'après l'équation 2.57, nous remarquons que l'expression du gain du filtre de la soustraction interspectrale H_{SIS} est basée sur l'estimation de la dsp des observations bruitées $P_{x1x1}(f,k)$ et $P_{x2x2}(f,k)$ ainsi que la dip entre les deux observations bruitées $P_{x1x2}(f,k)$ et la dip estimée entre les deux signaux de bruit $P_{b1b2}(f,k)$. Cette méthode de la soustraction interspectrale sera décrite plus en détail dans le dernier chapitre.

4 Conclusion

Nous avons présenté au cours de ce chapitre les techniques de base pour la réduction du bruit : en mono -voie et en bi-voie. Pour les techniques mono-voie, différentes approches pour l'estimation de la densité spectrale de puissance du bruit et des signaux rehaussés sont présentées. Nous nous sommes intéressés particulièrement à l'approche des

statistiques minimales pour l'estimation de la dsp du bruit, technique qui sera exploitée dans le dernier chapitre pour l'estimation de la dip des bruits, et l'approche de la soustraction spectrale, technique qui sera exploitée et étendue au cas bi-voie dans le quatrième chapitre.
Dans le cas des techniques bi-voie, différentes méthodes ont été présentées par classes différenciées suivant les hypothèses émises sur la distribution du bruit : corrélé, décorrélé ou peu corrélé. Cette classification nous permet d'effectuer un tri en nous appuyant sur les caractéristiques des bruits rencontrés en environnements réels. Il en ressort que les méthodes basées sur la fonction de cohérence, et plus précisément la soustraction interspectrale, sont des approches attractives car elles sont adaptées aux contraintes : adéquation avec l'environnement (faible corrélation du bruit) et complexité réduite.
Dans le cas de l'implant cochléaire bilatéral, les travaux visant l'amélioration de l'intelligibilité de la parole en milieux bruités sont minimes et la majorité est basée sur la technique du Beamformer adaptatif. Nous présentons dans le chapitre suivant une première approche pour l'amélioration de l'intelligibilité de la parole dans le cas de l'implant cochléaire bilatéral : la stimulation bilatérale décalée.

Chapitre 3

Effet de la distribution fréquentielle sur l'intelligibilité de la parole en milieux silencieux et bruité : Cas de l'implant cochléaire bilatéral

Sommaire

1 Introduction 53
2 Stimulation cochléaire bilatérale vs unilatérale 53
3 Simulateur d'implant 55
4 Comparaison des performances d'une stimulation bilatérale symétrique et décalée 61
5 Conclusion 70

1 Introduction

La prothèse cochléaire peut être implantée suivant un mode unilatéral où une seule oreille est stimulée ou suivant un mode bilatéral où les deux oreilles sont stimulées. Différent travaux de recherches ont comparé les performances d'une stimulation unilatérale par rapport à une stimulation bilatérale en environnement silencieux et bruité. Les résultats ont montré que la stimulation bilatérale présente des performances significativement meilleures que la stimulation unilatérale.
La stimulation bilatérale classique est une stimulation symétrique, c'est à dire les deux signaux de parole captés par les deux microphones sont traités en utilisant des bancs de filtres identiques et par la suite les deux oreilles sont stimulées avec les même signaux. Dans ce travail, nous considérons le cas d'une stimulation bilatérale décalée où les signaux de parole captés par les deux microphones sont traités moyennant des bancs de filtres fréquentiellement décalés. Une étude comparative des performances d'une stimulation bilatérale décalée, une stimulation bilatérale symétrique et une stimulation unilatérale est alors considérée. Cette étude comparative est réalisée en milieu silencieux ainsi qu'en milieu bruité à différents RSB. Deux vitesses d'analyse du signal de parole sont aussi considérées, à savoir 250Hz et 500 Hz.

2 Stimulation cochléaire bilatérale vs unilatérale

L'implant cochléaire unilatéral a permis aux personnes souffrantes d'une surdité profonde ou totale de bénéficier en milieu silencieux d'une qualité excellente de perception de la parole (Zeng et Fay, 2004) se rapprochant de celle obtenue avec des sujets normoentendants. D'autre part, différents autres travaux de recherche ont étudié les performances d'une stimulation unilatérale en milieu bruité et ont montré que les performances se dégradent d'une manière significative. Ceci pose un problème pour les patients unilatéralement implantés vu qu'ils souffrent de grands problèmes de perception de la parole en milieux bruité et réverbérant (Won *et al.*, 2007).

Les travaux de recherche actuels tendent à développer progressivement les performances de cet appareillage afin de fournir à ses utilisateurs une meilleure qualité de perception en améliorant l'intelligibilité de la parole surtout en environnement bruité. En effet, l'audition humaine est une audition binaurale grâce aux deux oreilles. Cette audition binaurale est assurée grâce à différents paramètres dont principalement les indices binauraux (ITD, ILD). Ces indices nous permettent de bénéficier principalement de deux avantages. D'une part, l'audition binaurale prend son effet lorsque les signaux de parole et du bruit proviennent de directions différentes. Dans ce cas, le système auditif humain exploite les deux signaux de parole provenant des deux oreilles pour favoriser ensuite l'oreille qui présente le meilleur rapport signal à bruit. D'autre part, l'audition binaurale permet de combiner les signaux de parole provenant des deux oreilles et par la suite assurer une meilleure perception de la parole surtout en milieu bruité (Loizou *et al.*, 2009). En plus, elle permet une meilleure localisation des sources de parole et de bruit dans l'espace.

Pour les patients souffrant d'une surdité profonde ou totale bilatérale, la seule manière qui permet de bénéficier d'une audition binaurale est l'implantation cochléaire bilatérale (Dillon, 2001b). Différents travaux de recherche ont étudié l'effet d'une stimulation bilatérale sur l'intelligibilité de la parole en milieux silencieux et bruité par rapport à la stimulation unilatérale. Les résultats ont montré que les performances sont variables d'un sujet à un autre, et les avantages de la stimulation bilatérale sont observés dans certaines conditions (Dunn *et al.*, 2008; VanHoesel, 2004). Dans le travail de Dunn et al. (Dunn *et al.*, 2010), une population de 60 sujets a été considérée. Trente sujets sont unilatéralement implantés et 30 autres sont bilatéralement implantés. Un ensemble de mots isolés noyés dans le bruit à différents niveaux du RSB sont considérés comme stimuli. Les résultats ont montré que les sujets bilatéralement implantés présentent des améliorations significatives des pourcentages de reconnaissance des mots par rapport aux sujets unilatéralement implantés pour tous les RSB considérés. Dans le travail présenté par Ricketts et al. (Ricketts *et al.*, 2006), les performances des deux modes de stimulation unilatérale et bilatérale ont été aussi comparées en milieu bruité. Seize adultes portant l'implant bilatéral MED-EL C40 ont participé à cette étude expérimentale. Les pourcentages de reconnaissance sont calculés dans le cas d'une stimulation unilatérale ensuite dans le cas d'une stimulation bilatérale. Dans une première expérience, les performances sont évaluées en milieu bruité en présence de cinq sources de bruit décorrelées et spatialement séparées. Le niveau du RSB est fixé à 10 dB en premier lieu, puis des niveaux adaptatifs du RSBs sont considérés. Les résultats de cette première expérience ont montré une amélioration significative des pourcentages de reconnaissance de 3.3 dB dans le cas du mode de stimulation bilatérale par rapport au mode de stimulation unilatérale lorsque le RSB est déterminé d'une manière adaptative. Lorsque le RSB est fixé à 10 dB, une amélioration des pourcentages de reconnaissance de l'ordre de 10% est observée.
Dans une deuxième expérience, les performances sont évaluées dans le même environnement bruité utilisé dans la première expérience mais à des niveaux du RSB égaux à 5, 10, 15, et 20 dB. Les résultats ont montré des performances largement supérieures pour le mode de stimulation bilatérale par rapport au mode de stimulation unilatérale pour les différents niveaux du RSB considérés. En plus, des performances nettement supérieures sont notées en milieu silencieux par rapport à celles obtenues en milieu bruité même à un niveau du RSB de 20 dB.

Une autre étude qui a pour but la comparaison des performances des deux modes de stimulation unilatérale et bilatérale en environnements silencieux et bruité à été présentée par Müller et al. (Müller *et al.*, 2002). Neufs adultes bilatéralement implantés avec l'implant cochléaire MED-EL multicanaux ont participé à cette expérience. Les pourcentages de reconnaissance sont calculés pour tous les sujets participants dans le cas d'une stimulation unilatérale (l'implant droit puis l'implant gauche sont testés séparément) ensuite dans le cas d'une stimulation bilatérale. Le matériel de test utilisé est formé d'un ensemble de mots monosyllabiques enregistrés en milieux silencieux et bruité à un niveau du RSB de 10 dB. Le signal de parole est présenté par un haut-parleur placé dans le plan frontal (en face) et le signal de bruit est présenté par deux haut-parleurs placés à des directions de 90° et de 270°. Les résultats ont montré que tous les sujets participants ont présenté des pourcentages de reconnaissance meilleurs pour le mode de stimulation bilatérale par rapport au mode de stimulation unilatérale. Dans ce cas, une amélioration moyenne de

18.7% du pourcentage de reconnaissance des mots monosyllabiques est notée. En plus, une amélioration du pourcentage de reconnaissance de 31.1% est notée avec une stimulation cochléaire bilatérale comparée à une stimulation unilatérale du côté latéral du signal de bruit. Les performances d'une stimulation cochléaire bilatérale sont aussi comparées à une stimulation unilatérale du côté controlatéral du signal de bruit et les résultats ont montré seulement une amélioration de 10.7%.

A partir des résultats présentés dans ces différents travaux de recherche, nous constatons que la stimulation bilatérale présente des améliorations au niveau des pourcentages de reconnaissance, souvent faibles mais significatives, par rapport à la stimulation unilatérale. Ces résultats sont variables suivant la valeur du RSB considérée et aussi suivant le nombre et la direction des sources de bruit. Nous remarquons aussi que les performances de la stimulation bilatérale et unilatérale sont comparées en milieux faiblement bruités (des valeurs positives du RSB) ; le cas des milieux fortement bruités n'est pas considéré. Dans ce chapitre, nous présentons une étude expérimentale où nous comparons les performances des modes de stimulation unilatérale et bilatérale (symètrique et décalée) en environnements silencieux et bruités à différents niveaux du RSB.

3 Simulateur d'implant

Dans ce travail de recherche, nous nous sommes intéressés au système d'implant Digisonic SP de la société Française Neurelec à 20 électrodes dans le cas d'une stimulation unilatérale et au système Digisonic SP Binaural à 24 électrodes en totalité dans le cas d'une stimulation bilatérale (deux implants, chacun à 12 électrodes). Trois modes de stimulations sont considérés dans cette étude expérimentale : Stimulation Unilatérale (SU), Stimulation Bilatérale Symétrique (SBS) et Stimulation Bilatérale Décalée (SBD). Les performances de ces trois modes de stimulation sont comparées en simulation. Un simulateur d'implant cochléaire est alors associé pour chaque mode de stimulation.
Le signal de parole capté par le microphone est traité suivant une stratégie de codage ACE (Advanced Combinaison Encoder) avec une sélection de huit maximums. Toutes les expérimentations présentées dans ce travail ont été validées en simulation avec des sujets normoentendants. Dans ce cas, un simulateur d'implant à bruit bandes étroites est considéré. On rappelle que pour ce type de simulateur d'implant, les signaux de parole sont synthétisés en utilisant des signaux de bruit filtrés comme signaux d'excitation. La zone cochléaire stimulée est limitée entre les fréquences 300 Hz et 6055 Hz.
Le signal de parole $s(n)$ capté par le microphone (n est l'indice temporel discret) est échantillonné à une fréquence $F_e = 16000$ Hz et segmenté en trames moyennant une fenêtre de Hanning de largeur $L = 128$ échantillons (8 ms de parole) moyennant l'équation 3.1. Un chevauchement entre les trames d'analyse successives est considéré.

$$s_w(l) = s(l).w(l) \quad l = 1...L \tag{3.1}$$

où $w(l)$ est l'expression de la fenêtre de Hanning donnée par l'équation suivante :

$$w(l) = 0.5\left(1 - \cos(2\pi\frac{l}{L})\right) \tag{3.2}$$

La Transformée de Fourier Rapide (TFR) du signal de parole fenêtré ($s_w(l)$ est donnée par l'équation suivante

$$\begin{aligned} S(k,\lambda) &= TFR\,(s_w(l)) \\ &= \sum_{l=0}^{L-1} s(l+\lambda U)w(l)e^{-j2\pi\frac{l}{L}k} \\ &= Re\{S(k,\lambda)\} + jIm\{S(k,\lambda)\}, \quad k = 0...L-1 \\ &= S_r(k,\lambda) + jS_i(k,\lambda) \end{aligned} \tag{3.3}$$

où $Re\{S(k,\lambda)\}$ et $Im\{S(k,\lambda)\}$ représentent respectivement les parties réelle et imaginaire de $S(k,\lambda)$, k est l'indice de la fréquence discrete, U est le taux de chevauchement entre trames et λ est l'indice de la trame.
L'application de cette TFR pour chaque trame donne naissance à 64 raies spectrales. Le spectre de puissance d'une raie spectrale k noté ($e(k)$) est calculé suivant l'équation suivante :

$$e(k) = S_r(k)^2 + S_i(k)^2 \tag{3.4}$$

Le spectre de puissance à court terme du signal de parole est ensuite divisé en Nb bandes d'analyse fréquentielles ($Nb = 20$ dans le cas d'une stimulation unilatérale et $Nb = 12$ dans le cas d'une stimulation bilatérale). Ces bandes de fréquences sont réparties suivant une échelle lin-log (échelle en barks). Le modèle approximatif permettant le passage du domaine de représentation fréquentiel f (en Hz) vers le domaine de représentation en bandes critiques B (en barks) est donnée par l'équation 3.5 (Tranmüller, 1990).

$$B(f) = 6.7Argsh(\frac{f-20}{600}) \tag{3.5}$$

La fréquence minimale F_{min} est fixée à 300 Hz alors que la fréquence maximale F_{max} est fixée à 6055 Hz. La figure 3.1 donne une représentation du banc de filtres considéré.

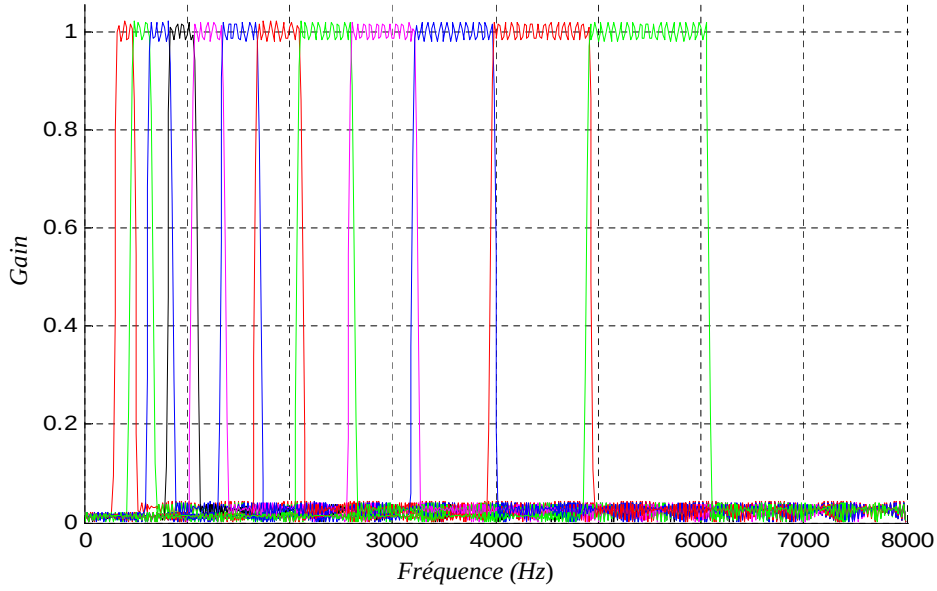


FIGURE 3.1 – Exemple de représentation fréquentielle du banc de filtres d'analyse (Nb=12)

L'algorithme 1 décrit le principe de calcul du nombre de raies spectrales (N_z), la raie spectrale de début ($ndébut_z$), la raie spectrale de fin ($nfin_z$) et la fréquence centrale (f_{centre}) pour chaque bande fréquentielle z ($z = 1...Nb$). Les limites inférieures et supérieures en bark de chaque bande d'analyse z sont notées respectivement B_z et $B_z + 1$. Les limites fréquentielles pour chaque bande d'analyse sont notées en F_z.

Algorithme 1 Répartition fréquentielle des différentes bandes d'analyse

Initialisation

Étape 1 : Calcul de la limite inférieure en bark de la première bande d'analyse (B_1) en se basant sur l'équation 3.5 : $B_1 = 6.7\left(\frac{F_{min}-20}{600}\right)$.

Étape 2 : Calcul de la limite supérieure en bark de la dernière bande d'analyse (B_{Nb+1}) en se basant sur l'équation 3.5 :$B_{Nb+1} = 6.7(\frac{F_{max}-20}{600})$.

Étape 3 : Calcul de la largeur des différentes bandes d'analyse en bark : $\delta_B = \frac{B_{Nb+1}-B_1}{N_b}$.

Limites inférieures et supérieures en bark pour chaque bande d'analyse

Étape 4 : Pour z variant de 1 jusqu'à N_b faire

$B_{z+1} = B_z + \delta_B$,

Transformation du plan des barks vers le plan des fréquences en Hz

Étape 5 : Pour z variant de 1 jusqu'à $N_b + 1$ faire

$F_z = 600sh\left(\frac{B_z}{6.7}\right) + 20$

Indices des raies spectrales de début et de fin pour chaque bande d'analyse

Étape 6 :

- $ndébut_1 = ceil\left(\frac{F_1.L}{F_e}\right)$,
- $nfin_{Nb} = floor\left(\frac{F_{N_b+1}.L}{F_e}\right)$,
- Pour z variant de 2 jusqu'à N_b faire
- $ndébut_z = ceil\left(\frac{F_z.L}{F_e}\right)$,
- $nfin_{z-1} = floor\left(\frac{F_z.L}{F_e}\right)$,

Nombre de raies spectrales pour chaque bande d'analyse

Étape 7 : Pour z variant de 1 jusqu'à N_b faire

N_z=$nfin_z - ndébut_z + 1$,

Fréquences centrales pour chaque bande d'analyse

Étape 8 : Pour z variant de 1 jusqu'à N_b faire

$f_{centre} = \left(\frac{ndébut_z+nfin_z}{2}\right)\frac{F_e}{L}$,

Le tableau 3.1 représente les différentes fréquences de coupure inférieures ($Fmin_z$) et supérieures ($Fmax_z$), les valeurs de N_z, $ndébut_z$ ainsi que la fréquente centrale (f_{centre}) pour chaque bande fréquentielle z ($z = 1...N_b$) dans le cas d'une stimulation unilatérale. Dans le cas de la stimulation bilatérale (modes SBS et SBD), le même algorithme de traitement décrit précédemment (algorithme 1) est considéré sauf que le nombre de bandes d'analyse est égale à $Nb = 12$ et la répartition fréquentielle des différents filtres d'analyse est différente. En effet, dans le cas du mode SBS, la répartition fréquentielle de ces différents filtres au niveau de la voie droite et la voie gauche est identique. Les fréquences de coupure minimales ($Fmin_z$) et maximales ($Fmax_z$), le nombre de raies spectrales (N_z),

TABLE 3.1 – *Configuration des différentes bandes d'analyse : mode SU*

Bande d'analyse z	$Fmin_z - Fmax_z(Hz)$	N_z	$ndébut_z$	$f_{centre}(Hz)$
1	300-387	1	3	375
2	387-480	1	4	438
3	480-581	1	4	500
4	581-690	1	5	625
5	690-810	1	6	750
6	810-944	1	7	875
7	944-10931	1	8	1000
8	1093-1259	2	9	1187
9	1259-1445	1	11	1375
10	1445-1655	2	12	1562
11	1655-1891	2	14	1812
12	1891-2158	2	16	2062
13	2158-2459	2	18	2312
14	2459-2801	3	20	2625
15	2801-3188	3	23	3000
16	3188-3626	4	26	3437
17	3626-4123	3	30	3875
18	4123-4688	5	33	4375
19	4688-5328	5	38	5000
20	5328-6055	11	43	5687

l'indice du raie spectrale de début ($ndébut_z$) et la fréquente centrale ($fcentre_z$) pour chaque bande fréquentielle d'analyse z sont donnés dans le tableau 3.2.

TABLE 3.2 – *Configuration des différentes bandes d'analyse : mode SBS*

Bande z	1	2	3	4	5	6	7	8	9	10	11	12
Voie droite ≡ Voie gauche												
$Fmin_z$ (Hz)	300	448	616	810	1041	1318	1654	2065	2568	3187	3950	4892
$Fmax_z$ (Hz)	448	616	810	1041	1318	1654	2065	2568	3187	3950	4892	6055
$ndébut_z$	3	4	5	7	9	11	14	17	21	26	32	40
N_z	1	1	2	2	2	3	3	4	5	6	8	9
$fcentre_z$ (Hz)	375	500	687	937	1187	1500	1875	2312	2875	3562	4437	5500

Dans le cas du mode de stimulation bilatérale SBD, la répartition fréquentielle des différentes filtres d'analyse au niveau de la première et la seconde voie est différente. En effet, les filtres d'analyse utilisés au niveau de la voie droite sont fréquentiellement décalés par rapport à ceux utilisés au niveau de la voie gauche. La figure 3.2 illustre les limites

fréquencielles en Hz et en barks pour chacune des douze bandes d'analyse considérées dans le cas du mode SBD.

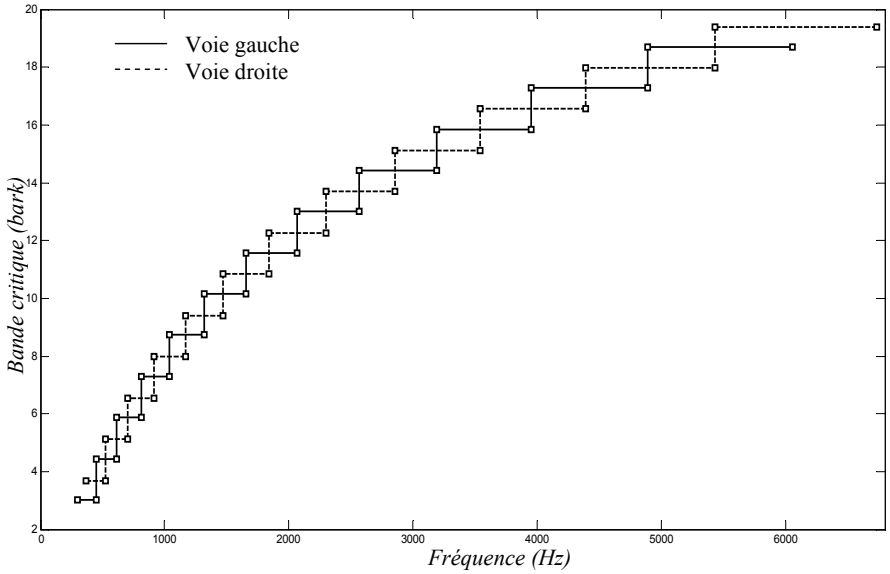


FIGURE 3.2 – Répartition en Hz et en Barks des différentes bandes d'analyse dans le cas du mode SBD

Nous notons ici que le décalage est variable d'une bande à une autre suivant la largeur de la bande d'analyse considérée. Ce décalage est toujours égal à la moitié de la largeur de la bande fréquencielle considérée. Le tableau 3.3 indique les fréquences de coupure minimales ($Fmin_z$) et maximales ($Fmax_z$), le nombre de raies spectrales (N_z), l'indice du raie spectrale de début ($ndébut_z$) et la fréquente centrale ($fcentre_z$) pour chaque bande fréquentielle d'analyse z pour chacune des voies droite et gauche.

TABLE 3.3 – *Configuration des différentes bandes d'analyse : mode SBD*

Bande z	1	2	3	4	5	6	7	8	9	10	11	12
Voie droite												
$Fmin_z$ (Hz)	300	448	616	810	1041	1318	1654	2065	2568	3187	3950	4892
$Fmax_z$ (Hz)	448	616	810	1041	1318	1654	2065	2568	3187	3950	4892	6055
$ndébut_z$	3	4	5	7	9	11	14	17	21	26	32	40
N_z	1	1	2	2	2	3	3	4	5	6	8	9
$fcentre_z$ (Hz)	375	500	687	937	1187	1500	1875	2312	2875	3562	4437	5500
Voie gauche												
$Fmin_z$ (Hz)	365	523	704	916	1168	1473	1843	2297	2854	3540	4387	5432
$Fmax_z$ (Hz)	523	704	916	1168	1473	1843	2297	2854	3540	4387	5432	6735
$ndébut_z$	3	5	6	8	10	12	15	19	23	29	36	44
N_z	2	1	2	2	2	3	4	4	6	7	8	10
$fcentre_z$ (Hz)	437	625	812	1062	1312	1625	2062	2562	3187	4000	4937	6062

L'énergie au niveau de chaque bande d'analyse fréquentielle z est ensuite calculée moyennant le théorème de Parseval suivant l'équation suivante :

$$E(z) = \frac{1}{N_z} \sum_{k=ndébut_z}^{ndébut_z+N_z-1} e(k) \tag{3.6}$$

A l'issue de cette étape, un vecteur à Nb valeurs d'énergie est obtenu. Nous rappelons ici que l'analyse du signal de parole dans le cas de l'implant cochléaire considéré est basé sur une stratégie de codage ACE (Cochlear Corporation, 2002). Dans ce cas, une étape de recherche et de sélection de M valeurs d'énergie les plus importantes parmi les Nb valeurs calculées est nécessaire. Seulement les bandes d'analyse fréquentielles présentant ces valeurs d'énergie maximales sont utilisées pour la reconstruction du signal de parole vocodé. Les énergies des "$Nb - M$" bandes d'analyse restantes sont mises à zéro. L'enveloppe au niveau de chaque bande d'analyse z est alors estimée en modulant une fenêtre de Hanning $w(l)$ avec l'énergie $E(z)$ relative à cette bande. Cette enveloppe $env(z, l)$ peut être calculée en utilisant tout simplement l'équation suivante :

$$env(z, l) = E(z)w(l) \quad z = 1...Nb, \; l = 1...L \tag{3.7}$$

Le signal enveloppe au niveau de chaque bande est ensuite lissé moyennant un filtrage passe-bas à une fréquence de coupure égale à 150 Hz. L'étape suivante consiste à synthétiser le signal de parole modifié (vocodé) au niveau de chaque bande. Ceci est assuré en modulant, au niveau de chaque bande d'analyse, le signal d'enveloppe ainsi estimé avec un bruit blanc filtré (à bande étroite). Ce bruit blanc à bande étroite est synthétisé pour chaque bande de fréquences et il est obtenu en filtrant un bruit blanc par un filtre passe-bande à réponse impulsionnelle infini de type Butterworth d'ordre 3. Dans le cas d'une stimulation unilatérale par exemple, les fréquences de coupure des ces différents filtres sont données par le tableau 3.1. Le signal de parole vocodé est obtenu par la sommation des différents signaux de parole vocodés issue des différentes bandes d'analyse fréquentielles. Enfin, la puissance du signal de parole vocodé est ajustée afin qu'elle soit égale à celle du signal de parole d'entrée. La figure 3.3 présente un schéma bloc regroupant les différentes étapes de traitement pour le simulateur d'implant considéré.

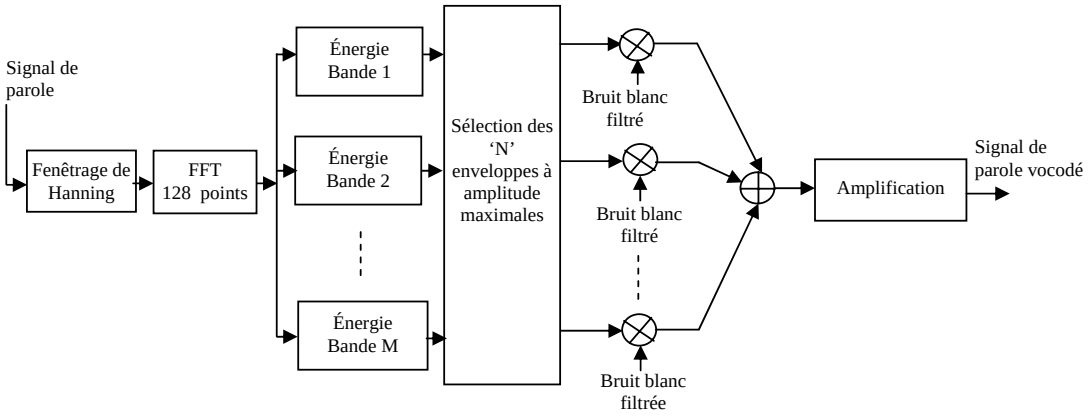


FIGURE 3.3 – Représentation schématique des différents blocs du simulateur d'implant cochléaire considéré

Le deuxième paramètre considéré dans cette étude expérimentale est la vitesse d'analyse. Afin d'étudier l'effet de ce paramètre sur l'intelligibilité de la parole chez les sujets implantés, un modèle décrivant ce paramètre est proposé. Ce modèle est basé sur la variation du degré de chevauchement entre les trames successives d'analyse temporelle du signal de parole comme indiqué par la figure 3.4.

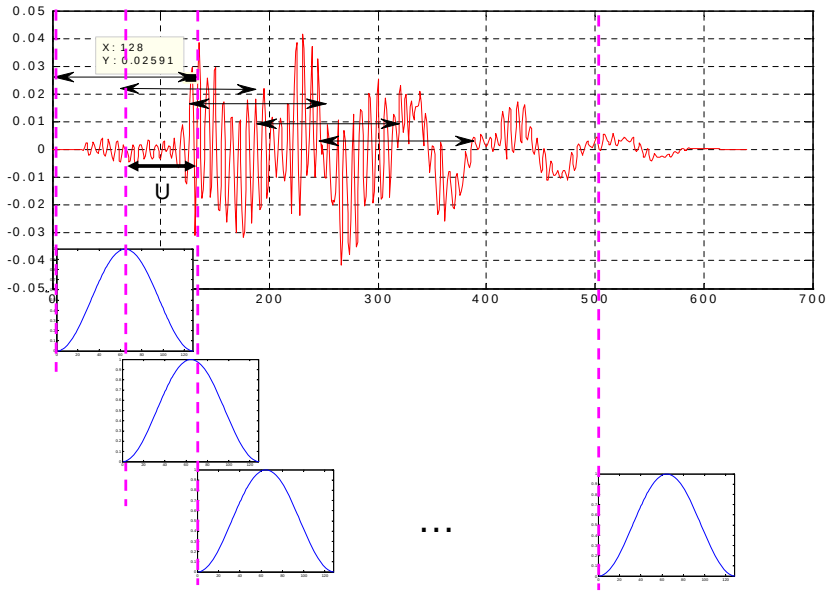


FIGURE 3.4 – Principe de simulation de la vitesse d'analyse

Dans cette figure, le paramètre U indique le taux de chevauchement entre les trames d'analyse successives, et il est définie pour chacune des vitesses d'analyse considérées.
La largeur de la fenêtre d'analyse est fixée à 128 échantillons (Durée=8ms). Deux vitesses d'analyse sont considérées : une faible vitesse d'analyse fixée à 250 Hz et une moyenne vitesse d'analyse fixée à 500 Hz. Dans le cas d'une vitesse d'analyse ègale à 250 Hz, une mise à jours de la trame de parole est assurée toute les 4 millisecondes, ce qui correspond à la moitié de la durée de la fenêtre d'analyse. Un chevauchement de 50% entre les trames d'analyse consécutives est donc considéré. Dans le cas où la vitesse d'analyse est fixée à 500 Hz, une mise à jours de la trame d'analyse du signal de parole est effectuée toute les 2 ms, ce qui correspond au quart de la durée de la fenêtre d'analyse, et donc un chevauchement de 25% entre les trames consécutives est nécessaire.

4 Comparaison des performances d'une stimulation bilatérale symétrique et décalée

Dans ce paragraphe, nous nous intéressons à la comparaison, en simulation, des performances d'une stimulation bilatérale par rapport à une stimulation unilatérale. Dans le cas de la stimulation bilatérale, deux modes ont été considérés : Stimulation bilatérale symétrique et stimulation bilatérale décalée. Cette étude comparative a été faite dans le

cas de deux vitesses d'analyse et à différents niveaux du RSB. Dans le cas d'une stimulation unilatérale, une seule oreille a été stimulée. Dans le cas d'une stimulation bilatérale symétrique, deux oreilles ont été stimulées d'une manière identique (les mêmes filtres fréquentiels ont été utilisés pour analyser les signaux de parole captés par les microphones droit et gauche, et d'autre part, les mêmes signaux vocodés ont été présentés aux deux oreilles). Dans le cas d'une stimulation bilatérale décalée, les deux oreilles ont été stimulées par des stimuli différents. En effet, les vocodeurs utilisés pour analyser les signaux de parole d'origine n'étaient pas identiques. Plus précisément, un décalage fréquentiel entre les filtres d'analyse utilisés au niveau du vocodeur droit par rapport au vocodeur gauche a été considéré.

4.1 Matériel et méthode

4.1.1 Sujets

Une population de cinquante sujets présentant une audition normale a été indemnisée pour participer à cette expérience. Ces sujets ont été âgés entre 18 et 34 ans. La langue maternelle de tous les participants était le français. L'ensemble des tests d'écoute ont été déroulés dans le centre d'implantation cochléaire de l'hôpital Edouard Herriot de Lyon.

4.1.2 Matériel

Le matériel de test utilisé dans cette expérience était l'ensemble des vingt listes phonétiques de Lafon (Lafon, 1964). Chaque liste comporte 17 mots triphonémiques qui sont prononcés par un seul locuteur. Le niveau seuil du signal de parole a été fixé à 70 dB. Toutes les listes ont été enregistrées dans une chambre sourde du centre d'implantation cochléaire de l'hôpital Edouard Herriot de Lyon. Les listes ont été tout d'abord enregistrées sans bruit ajouté (milieu silencieux), ensuite dans un environnement bruité à différents niveaux de RSB. Le signal de bruit qui a été utilisé dans cette expérience était de type Babble prononcé par 2 locuteurs. Les signaux ont été enregistrés à l'aide d'une tête artificielle (KEMAR : Knowles Electronic Manikin for Acoustical Research) équipée de deux microphones, un au niveau de chaque oreille. La distance entre les deux microphones était de 20 cm. La source de parole a été mise dans le plan transversal en face du KEMAR (0 °), alors que les deux sources de bruit ont été placées de part et d'autre du KEMAR à 60° et −60°, toujours au niveau du plan transversal. La figure 3.5 récapitule le protocole expérimental adopté.

4.1.3 Protocole expérimental

Les tests d'écoute ont été déroulés dans une chambre sourde en utilisant un ordinateur connecté à un lecteur CD (PHILIPS-CD723). Un audiomètre numérique (MADSEN-Orbiter 922) a été utilisé pour le calibrage et l'ajustement des intensités des différents signaux sonores. Dans les cas des modes SBS et SBD, les stimuli ont été présentés d'une manière bilatérale à chaque sujet. Dans le cas du mode SU, les stimuli ont été présentés d'une manière unilatérale (uniquement à l'oreille droite). Ces stimuli ont été délivrés

moyennant un casque 'Sennheiser HD250' à un niveau sonore confortable calibré à 70 dB SPL.

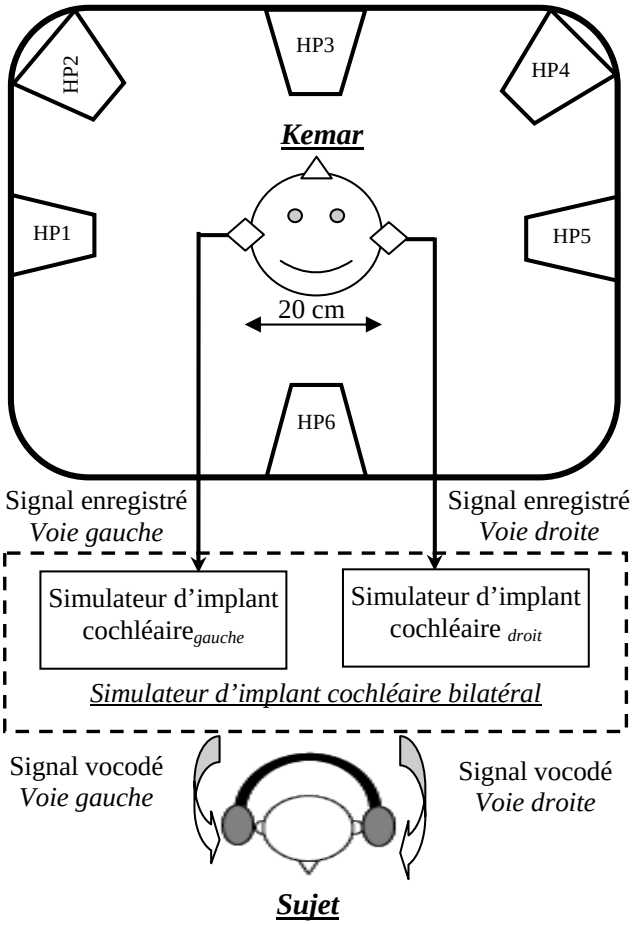


FIGURE 3.5 – Protocole expérimental

Avant le début des tests d'écoute, une session d'apprentissage a été tout d'abord considérée. Pendant cette session, le sujet écoutait un ensemble de 10 mots choisis aléatoirement de l'ensemble des listes afin de se familiariser avec les différents types de stimuli. Aucun score n'a été calculé pendant cette session d'apprentissage. A la suite de cette session préliminaire, chacun des cinquante sujets a été testé dans les différentes conditions expérimentales. Durant cette session d'écoute, chaque sujet avait comme consigne de répéter les mots entendus. En tout, l'expérience comportait 24 conditions expérimentales différentes (3 modes de stimulation x 2 vitesses d'analyse x 4 RSBs).
Afin d'aider les sujets de se familiariser et de s'adapter progressivement aux différentes conditions expérimentales, les différents tests d'écoute ont été déroulés en commençant par les listes traitées en absence de bruit, suivies par les listes traitées en milieu bruité. A la fin de chaque test, les réponses du sujet ont été enregistrées et le pourcentage de reconnaissance de phonèmes correctement identifiés a été enfin calculé.

4.2 Résultats

4.2.1 Représentation des signaux

Le mot 'souche' extrait de de la première liste de Lafon a été traité suivant le mode de stimulation SBS à une vitesse d'analyse égale à 250 Hz. Une représentation des différents signaux relatifs aux différentes bandes d'analyse est donnée par la figure 3.6.

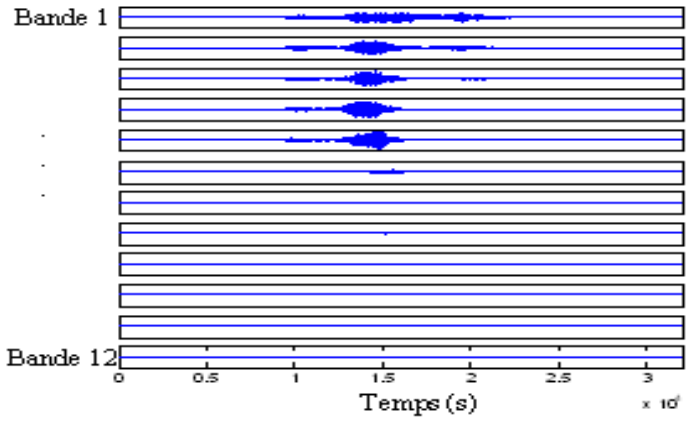


FIGURE 3.6 – Représentation temporelle des signaux de parole synthétisés pour chaque canal de stimulation

La figure 3.7 montre la variation temporelle du signal de parole vocodé (signal représenté en couleur bleue), synthétisé à partir du simulateur d'implant précédemment décrit, dans le cas du mode de stimulation SBS. Le signal d'enveloppe correspondant est aussi représenté (signal représenté en couleur rouge)

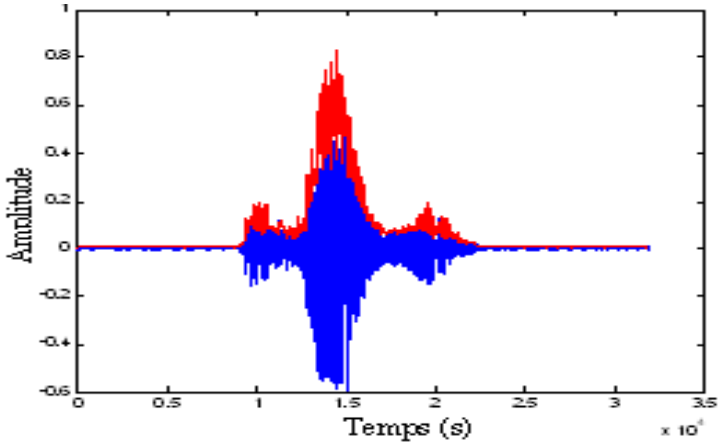


FIGURE 3.7 – Représentation temporelle du signal enveloppe estimé

Une représentation temps-fréquence du mot 'agis' traité suivant les trois modes de stimulation est enfin donnée par la figure 3.8.

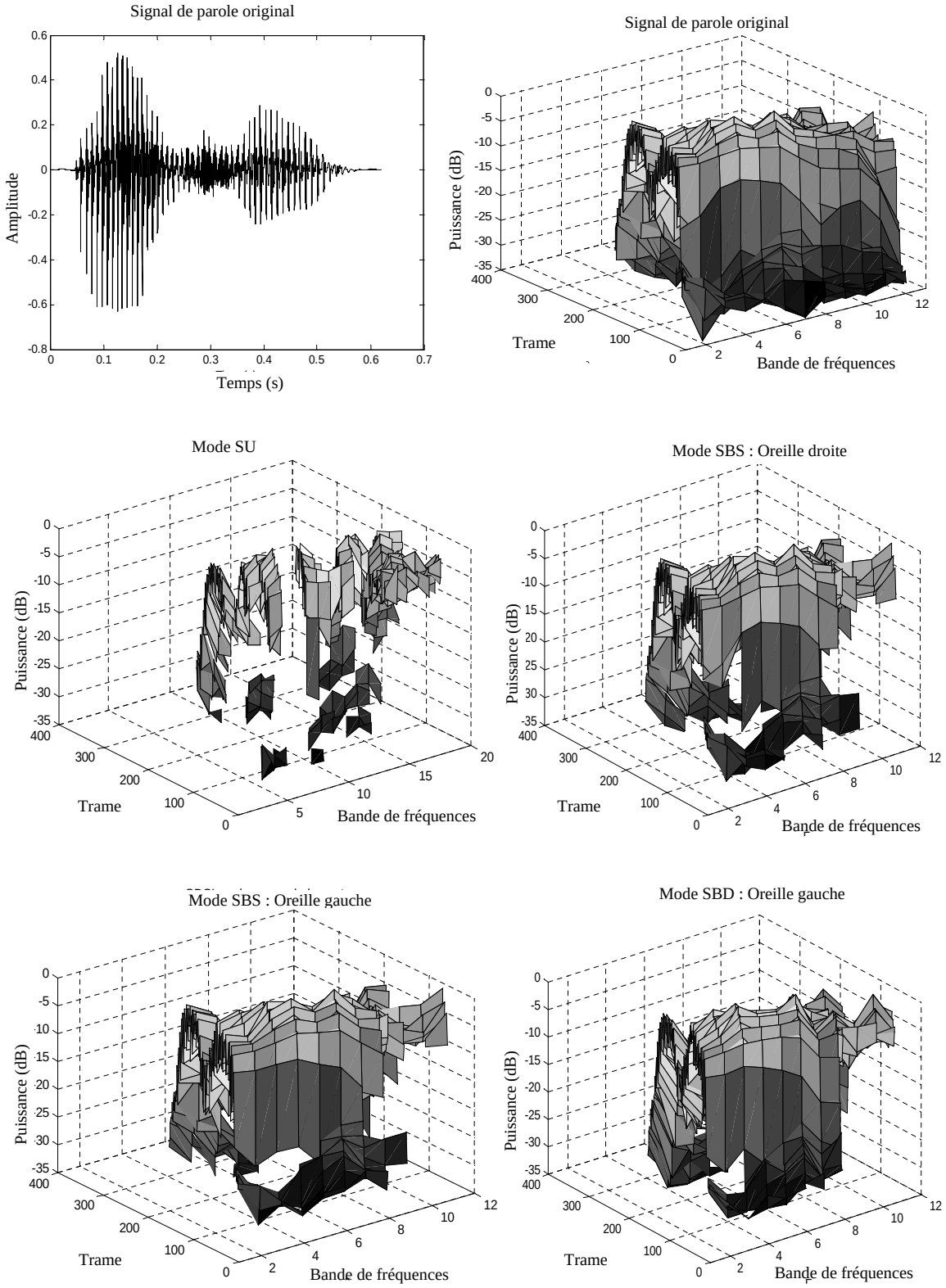


FIGURE 3.8 – Représentation temps-fréquence du mot 'agis' pour les différents modes de stimulation

4.2.2 Evaluation subjective

Les résultats sont donnés sous forme d'un pourcentage de reconnaissance des phonèmes pour chaque condition expérimentale. Une moyenne des scores obtenus par l'ensemble des 50 sujets a été calculée dans chaque condition. Les résultats moyens sont représentés sous forme d'histogrammes avec des barres indiquant l'erreur standard. La figure 3.9 illustre les résultats obtenus dans le cas de la première vitesse d'analyse fixée à 250 Hz.

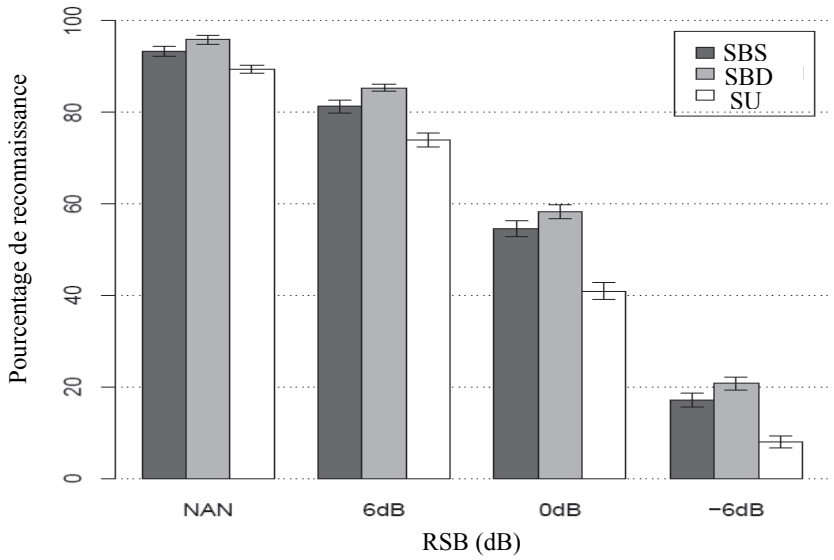


FIGURE 3.9 – Performances obtenues dans le cas des trois modes de stimulation pour différents RSB à 250 Hz

Dans le cas de la deuxième vitesse d'analyse, les résultats obtenus sont illustrés par la figure 3.10 .

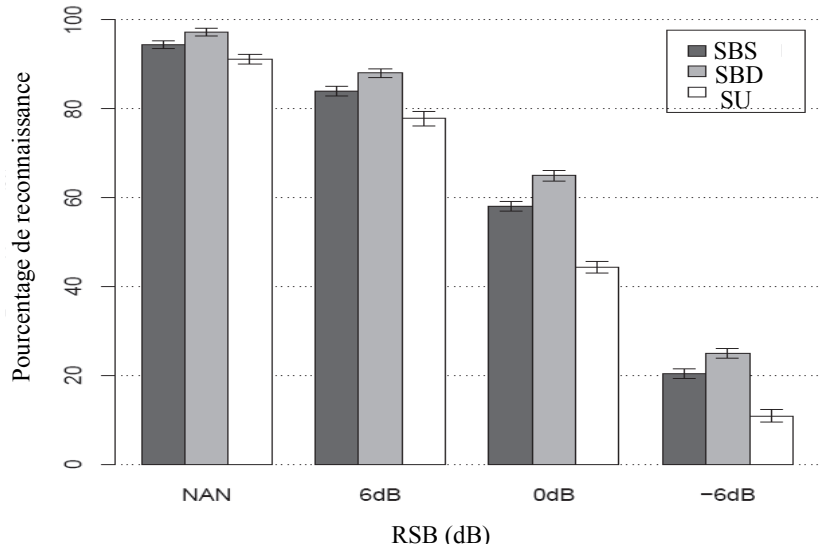


FIGURE 3.10 – Performances obtenues dans le cas des trois modes de stimulation pour différents RSB à 500 Hz

A partir de ces deux figures, nous remarquons que les modes de stimulation bilatérale présentaient des performances nettement supérieures au mode de stimulation unilatérale. De meilleures performances ont été aussi notées dans le cas d'une stimulation bilatérale décalée (mode SBD) par rapport au mode SBS. Ces résultats ont été observés aussi bien en milieux silencieux qu'en milieu bruité aux différents RSB. En plus, nous remarquons que les performances étaient variables en fonction de la vitesse d'analyse. En effet, lorsque la vitesse d'analyse 500 Hz a été considérée, les performances étaient meilleures par rapport à celles obtenues à une vitesse d'analyse égale à 250 Hz.
Le tableau 3.4 illustre les pourcentages de reconnaissance moyens obtenus en milieux silencieux et bruité à différents RSB pour chacun des modes de stimulation unilatérale et bilatérale (SBS et SBD) considérés aux deux vitesses d'analyse 250 Hz et 500 Hz.

TABLE 3.4 – *Pourcentages de reconnaissance moyens en milieux silencieux et bruités aux différents niveaux du RSB*

Vitesse d'analyse	**250 Hz**				**500 Hz**			
$RSB(dB)$	Silence	6	0	-6	Silence	6	0	-6
$SU(\%)$	89.5	74	41	8	91	78	44	11
$SBS(\%)$	93.3	81.3	54.6	17	94.5	84	58	20.4
$SBD(\%)$	95.9	85.4	58.3	20.7	97.3	88	65	25

Afin d'étudier l'effet de ces différents facteurs (mode de stimulation, vitesse d'analyse et RSB) sur l'intelligibilité de la parole, une analyse statistique a été considérée. Cette analyse a été basée sur des tests chi2 avec les paramètres suivants :

- Mesures répétées : Chacun des 50 sujets participait à l'ensemble des 24 conditions considérées.

- Variable dépendante : le pourcentage de reconnaissance de phonèmes.

- Trois facteurs :
  - Mode de stimulation (SBS, SBD et SU)
  - Vitesse d'analyse (250 Hz et 500 Hz)
  - RSB (Silence, 6 dB, 0 dB et -6 dB)

Les deux premiers facteurs sont à effets fixes, alors que le troisième est à effet aléatoire. Des modèles regroupant simultanément des effets fixes et aléatoires sont connus sous le nom de modèles à effets mixtes. Ce type de modèle est très souvent utilisé pour l'analyse des données (Baayen, 2008). Dans notre modèle, nous avons utilisé le programme lmer du package lme4 (Bates, 2007) implémenté sous le logiciel R qui est un logiciel de programmation dédié pour les analyses statistiques (R, 2007).
Les résultats ont montré un effet significatif du mode de stimulation (Chi2[2]=1139, $p<0.0001$), de la vitesse d'analyse (Chi2[1]=101, $p<0.0001$) et du RSB (Chi2[2]=25524, $p<0.0001$). En plus, un effet d'interaction significatif entre le mode de stimulation et le RSB a été observé (Chi2[6]=41, $p<0.0001$). Par contre, aucun effet significatif n'a été noté

entre la vitesse d'analyse et le mode de stimulation (Chi2[2]=3, p=0.25) ainsi qu'entre la vitesse d'analyse et le RSB (Chi2[3]=1.5, p=0.7). Les résultats donnés par ces tests de chi2 précisaient les résultats donnés par les figures 3.9 et 3.10.
Les résultats de cette première analyse statistique ont montré qu'il existe une interaction significative entre les facteurs mode de stimulation et RSB. Ceci peut être expliqué par le fait que chaque mode de stimulation avait un comportement différent suivant la valeur du RSB. Afin de mieux étudier ce comportement et chercher les paires (RSB, mode de stimulation) pour lesquelles les différences ont été significatives, des tests de comparaison multiples appelés aussi des tests post-hoc ont été établis. Ce type de test peut être réalisé en utilisant le test de Tukey par exemple, appelé aussi le test de la différence franchement significative (HSD : honestly significative difference). Le logiciel R dispose de la fonction glht (general linear hypothesis test) du package multcomp pour réaliser ce type de test. Les résultats obtenus à la suite de l'application de cette fonction d'analyse statistique dans le cas de notre modèle sont donnés comme suit.

4.2.2.a Environnement silencieux

En absence du bruit, la reconnaissance des phonèmes a été relativement sensible au mode de stimulation. En effet, les tests de comparaison post-hoc effectués pour étudier l'effet du mode de stimulation en environnement silencieux ont montré que les pourcentages de reconnaissance ont été améliorés lorsque les modes de stimulation bilatérale (SBS et SBD) ont été comparés aux pourcentages obtenus dans le cas d'une stimulation unilatérale ($p<0.005$). Les résultats ont montré aussi de meilleures performances pour le mode SBD par rapport au mode SBS. En effet, comparée au pourcentage de reconnaissance calculé dans le cas du mode SU, une amélioration significative de l'ordre de 6.3% dans le cas du mode SBD a été notée. Cette amélioration était plus faible dans le cas du mode SBS et elle ne dépassait pas 3.6%.

4.2.2.b RSB=6 dB

En environnement bruité, et à un RSB=6 dB, les résultats ont montré un effet significatif du mode de stimulation bilatérale (SBS et SBD) par rapport au mode de stimulation unilatérale ($p<0.001$). En plus, une amélioration significative des pourcentages des reconnaissance ont été obtenues dans le cas du mode SBD par rapport au mode SBS ($p<0.001$). En effet, dans le cas d'une stimulation bilatérale, de meilleurs pourcentages de reconnaissance ont été notés lorsque le mode de stimulation SBD a été considéré. Comparée au mode SU, l'amélioration était de 11% dans le cas du mode de stimulation SBS et de 15% dans le cas du mode de stimulation SBD. Nous notons ici que l'amélioration était plus importante que celle obtenue en environnement silencieux.

4.2.2.c RSB=0 dB

Dans le cas où le RSB est fixé à 0 dB, les résultats ont montré que les modes de stimulation bilatérale SBS et SBD permettaient une amélioration significative des performances par rapport au mode SU ($p<0.001$) ; mais aucune amélioration significative entre

les modes de stimulation SBS et SBD n'a été observée (p=0.24). En effet, comparées au mode de stimulation SU, les améliorations étaient de l'ordre de 13% dans le cas du mode SBS et 22% dans le cas du mode SBD.

4.2.2.d RSB=-6 dB

Un pattern similaire à celui obtenu à un RSB=0 dB a été observé dans le cas du RSB=-6 dB. En effet, une amélioration significative des performances a été notée dans le cas des modes de stimulation SBS et SBD par rapport au mode SU ($p<0.001$), mais aucune interaction significative n'a été observée entre les modes de stimulation SBS et SBD (p=0.16). Dans ce cas, comparée au mode SU, les améliorations ont été limitées à 13% dans le cas du mode SBD et 6% dans le cas du mode SBS.

4.2.3 Discussion

Dans ce chapitre, une étude expérimentale visant l'étude des performances de deux modes de stimulation bilatérale (SBS et SBD) par rapport à une stimulation unilatérale (SU) dans le cas de deux vitesses d'analyse (250 Hz et 500 Hz) a été présentée. Ces performances ont été évaluées en comparant les pourcentages de reconnaissance des phonèmes en milieux silencieux et bruité à différents niveaux du RSB.

4.2.3.a Stimulation bilatérale vs Unilatérale

A partir des pourcentages de reconnaissance moyens calculés dans les différentes conditions, nous avons remarqué une amélioration significative des performances lorsque les deux modes de stimulation bilatérale ont été considérés par rapport au mode de stimulation unilatérale. Ces résultats trouvés sont consistants avec différents travaux de recherche publiés (Veekmans *et al.*, 2009; Dunn *et al.*, 2008). En effet, dans le cas d'une stimulation bilatérale, deux versions identiques du signal de parole ont été combinées et additionnées (effet de sommation binaurale) puis présentées au système auditif humain. Cette redondance d'informations engendrait une meilleure représentation de l'information, et donc une meilleure intelligibilité de la parole pour le patient implanté. Dans le cas du mode de stimulation SBD, les deux oreilles ont été stimulées mais la version du signal de parole présentée au niveau de l'oreille latérale est différente de celle présentée au niveau de l'oreille controlatérale. Cette stimulation pourrait être plus riche en information, ce qui pourrait permettre au système auditif humain de mieux exploiter les informations et expliquer l'avantage du mode SBD par rapport au mode SBS.

4.2.3.b Vitesse d'analye

La vitesse d'analyse était le deuxième facteur considéré dans cette étude expérimentale. Deux vitesses d'analyse ont été considérées : 250 Hz et 500 Hz. Comparé au pourcentage de reconnaissance moyen obtenu à 250 Hz, une amélioration moyenne de 3.5% a été observée lorsque la deuxième vitesse d'analyse égale à 500 Hz a été considérée. En

environnement silencieux, la variation de la vitesse d'analyse de 250 Hz à 500 Hz a entrainé une amélioration très faible du pourcentage de reconnaissance des phonèmes (la différence est de 1.5%). Ces améliorations ont été plus importantes en milieu bruité. En effet, comparés aux pourcentages de reconnaissance obtenus à la vitesse d'analyse 250 Hz, des améliorations égales à 3% à 6 dB, 4.5% à 0 dB et 3.5% à −6 dB ont été notées lorsque la vitesse d'analyse 500 Hz a été considérée.
Différents travaux de recherche ont étudié l'effet de la vitesse de stimulation dans le cas où la stratégie de codage utilisée est la stratégie ACE (Vandali *et al.*, 2000; Plant *et al.*, 2007; Weber *et al.*, 2007). Les résultats trouvés ont été approximativement similaires à ceux obtenus dans cette présente étude expérimentale en milieu silencieux. En milieu bruité, quelques différences ont été observées. Dans un autre travail de recherche proposé par Balkany et al. (Balkany *et al.*, 2007), les auteurs ont étudié l'effet de quatre vitesses de stimulation (275, 350, 500 et 900 pps) en milieux silencieux et bruité. Les résultats ont montré que pour la plupart des sujets implantés, la vitesse de stimulation n'avait pas un effet significatif en absence du bruit. Par contre, les résultats ont montré que les vitesses 500 et 900 pps permettaient d'améliorer significativement les pourcentages de reconnaissance en présence du bruit par rapport aux deux vitesses 275 et 350 pps. Ces derniers résultats ont été en concordance avec les résultats obtenus dans notre étude expérimentale.

5 Conclusion

Classiquement, la stimulation bilatérale est symétrique puisque les stimuli présentés au niveau des deux oreilles sont identiques. Dans ce travail, nous avons considéré le cas de la stimulation bilatérale décalée où les stimuli sont différents. En effet, dans ce dernier cas, les signaux de parole ont été traités avec des bancs de filtres fréquentiellement décalés : le banc de filtres utilisé pour l'analyse des signaux de parole au niveau de la voie droite est fréquentiellement décalé par rapport à celui utilisé au niveau de la voie gauche. Dans ce chapitre, nous avons présenté une étude comparative des performances d'une stimulation bilatérale sous ses deux modes symétrique et décalée par rapport à la stimulation unilatérale. La comparaison a été réalisée pour deux vitesses d'analyse fixées à 250 Hz et 500 Hz, en milieux silencieux et bruité à différents niveaux du RSB. Les résultats ont montré que la stimulation bilatérale présente des performances nettement supérieures à la stimulation unilatérale. Ces résultats ont été conformes avec les résultats publiés dans la littérature. Il a été noté aussi que la stimulation bilatérale décalée présente des performances supérieures à la stimulation bilatérale symétrique. Les améliorations étaient variables d'un RSB à un autre. L'effet de la vitesse d'analyse à été aussi considéré et les résultats ont montré que l'augmentation de la vitesse d'analyse permette d'améliorer significativement l'intelligibilité de la parole.
La stimulation cochléaire bilatérale décalée permet une légère amélioration significative de l'intelligibilité de la parole, mais ces amélioration restent toujours très faible en milieu bruité. Plusieurs algorithmes de rehaussement des signaux de parole bruité ont été exploités dans le cas de la stimulation cochléaire unilatérale ; par contre ils sont peu dans le cas de la stimulation bilatérale. Dans le chapitre suivant, nous présentons les résultats d'une étude expérimentale où les performances de l'algorithme de la soustraction spectrale bi-voie ont été evaluées dans le cas de la stimulation cochléaire bilatérale.

Chapitre 4

Exploitation de la technique de soustraction spectrale bi-voie pour la réduction du bruit : Cas de l'implant cochléaire bilatéral

Sommaire

1	**Introduction**	**72**
2	**Algorithmes de la soustraction spectrale bi-voie pour la réduction du bruit**	**72**
3	**Algorithmes d'estimation des signaux de parole rehaussés**	**77**
4	**Etude comparative et évaluation des performances**	**80**
5	**Conclusion**	**91**

1 Introduction

Comme mentionné dans le chapitre précédant, la stimulation bilatérale décalée engendre une amélioration de l'intelligibilité de la parole mais les performances restent toujours inférieures à celles obtenues avec des sujets normoentendants, spécifiquement en milieu bruité. L'étude bibliographique présentée dans le deuxième chapitre nous a permis d'identifier différents algorithmes de réduction du bruit pour l'amélioration de l'intelligibilité de la parole dans le cas de l'implant cochléaire. Les algorithmes mono-voie ont montré une modeste amélioration significative de l'intelligibilité. De meilleures performances ont été observées dans le cas des techniques de réduction du bruit bi-voie. Généralement, ces algorithmes sont basés sur l'algorithme du Beamformer adaptatif dans le cas de l'implant cochléaire.
Dans ce chapitre, nous présentons une nouvelle approche de réduction du bruit bi-voie dédiée pour l'implant cochléaire bilatéral. Cet algorithme est basé sur l'extension de deux techniques de soustraction spectrale, particulièrement l'algorithme de la soustraction spectrale non linéaire et l'algorithme de la soustraction spectrale multi-bande au cas bi-voie. Les performances de ces deux algorithmes sont validées d'une manière objective puis subjective dans le cas de l'implant cochléaire, en simulation avec des sujets normoentendants, puis avec des patients implantés.

2 Algorithmes de la soustraction spectrale bi-voie pour la réduction du bruit

Dans ce paragraphe, nous présentons le principe de l'algorithme de réduction du bruit bi-voie proposé dans le cas de l'implant cochléaire bilatéral. Le principe de cet algorithme est illustré dans la figure 4.1 (Kallel *et al.*, 2012a). Deux microphones omnidirectionnels bien calibrés sont utilisés (un microphone au niveau de chaque oreille) pour l'acquisition des signaux bruités provenant du milieu extérieur. Nous désignons par $i = 1$ l'indice du microphone installé au niveau de l'oreille droite et $i = 2$ l'indice du microphone installé au niveau de l'oreille gauche. Nous supposons que les signaux de bruit acquis par les deux microphones sont additifs et parfaitement décorrélés. Dans ce cas, les signaux bruités $x_i(n)$ consistent en la somme du signal original propre $s_i(n)$ et des composantes du bruit $b_i(n)$ tel que : $x_i(n) = s_i(n) + b_i(n),\ i = \{1, 2\}$ où i désigne l'indice du microphone et n est l'indice du temps.
Le traitement sera effectué trame par trame. L'analyse des trames entrelacées des signaux bruités est implémentée en utilisant une Transformée de Fourier Rapide (TFR) précédée par une fenêtrage de Hamming. Le spectre de puissance des signaux bruités peuvent être exprimés comme suit :

$$\begin{aligned} |X_i(f,k)|^2 &= |S_i(f,k)|^2 + |B_i(f,k)|^2 \\ P_{xixi}(f,k) &= P_{sisi}(f,k) + P_{bibi}(f,k) \end{aligned} \tag{4.1}$$

où $X_i(f,k)$, $S_i(f,k)$ et $B_i(f,k)$ dénotent respectivement les TFR des signaux $x_i(n)$, $s_i(n)$ et $b_i(n)$ au niveau de la trame k. f désigne la composante fréquentielle discrète et i l'indice

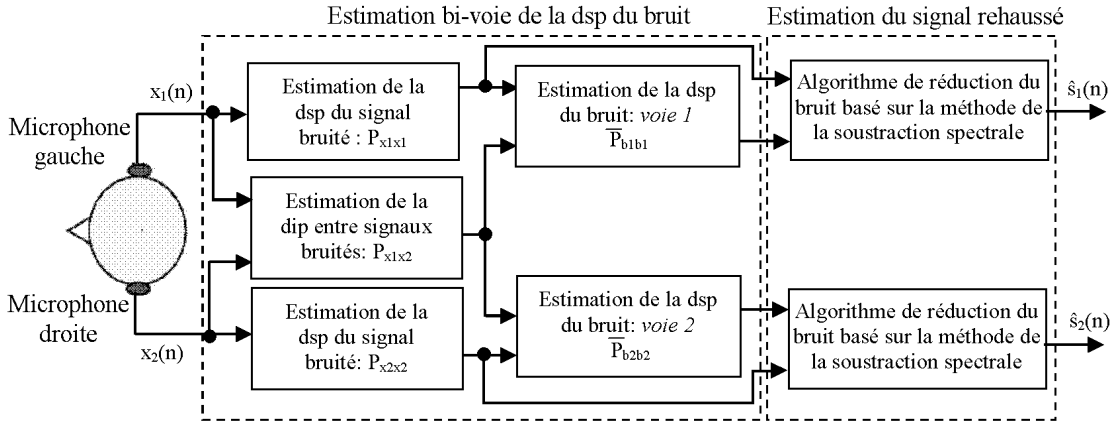


FIGURE 4.1 – Schéma global du système de réduction de bruit considéré

du microphone. D'autre part, $P_{xixi}(f,k)$, $P_{sisi}(f,k)$ et $P_{bibi}(f,k)$ représentent respectivement les densités spectrales de puissances (dsp) des signaux $x_i(n)$, $s_i(n)$ et $b_i(n)$. Comme les spectres des bruits ne peuvent pas être directement déterminés, alors une estimation de ces spectres doit être considérée.

2.1 Estimation des dsp : un point crucial

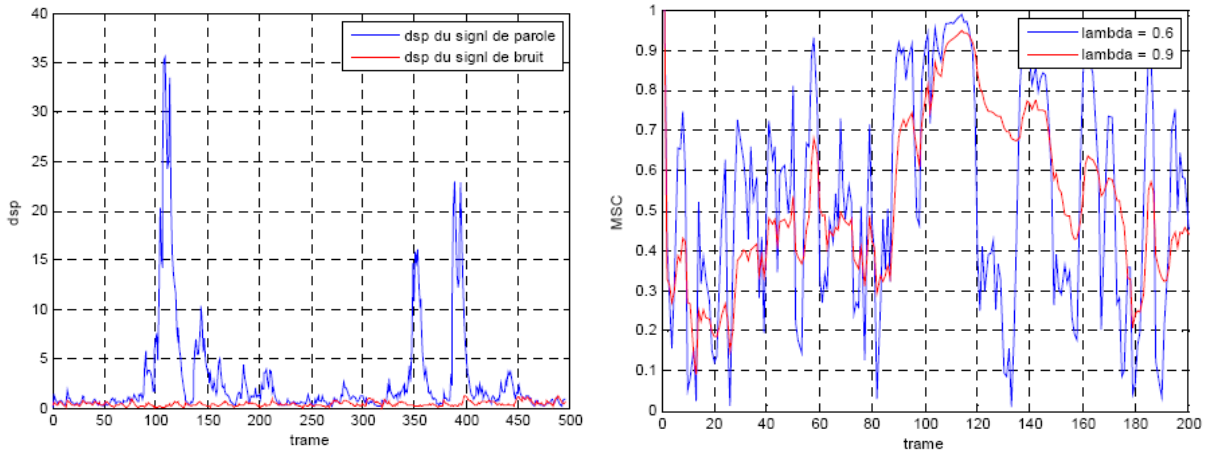
L'estimation des dsp et de la densité interspectrale de puissance (dip) des signaux d'observation est nécessaire au calcul des dsp du bruit au niveau des voies droite et gauche. Ce point se révèle être une des pierres angulaires du système de réduction de bruit considéré. Cela suppose implicitement que les signaux sont stationnaires, au moins sur la fenêtre temporelle d'analyse.
Puisque nous disposons des spectres instantanés, les densités spectrales et interspectrale sont estimées de manière récursive par les relations suivantes :

$$P_{x1x2}(f,k) = \lambda_x P_{x1x2}(f,k-1) + (1-\lambda_x) X_1(f,k) X_2^*(f,k) \tag{4.2}$$

$$P_{xixi}(f,k) = \lambda_x P_{xixi}(f,k-1) + (1-\lambda_x) X_i(f,k) X_i^*(f,k), \quad i = 1,2 \tag{4.3}$$

où λ_x est un coefficient d'oubli proche de 1. Ce coefficient d'oubli joue un rôle prépondérant dans le sens donné à ces estimations. Pour un coefficient d'oubli fort ($\lambda_x = 0.9$), l'estimation de la dsp tend à privilégier les statistiques à long-terme des signaux, en lissant les composantes à court-terme. Ce sont les ordres typiques de valeurs utilisées pour l'estimation de la dsp du bruit en absence de parole. Dans le cas de la parole, λ_x ne doit au contraire pas prendre des valeurs trop proches de 1 afin que les dsp calculées par l'équation 4.3 rendent compte de son caractère dynamique et faiblement stationnaire. Pour une fréquence d'échantillonnage égale à 16000 Hz, des fenêtres d'analyse à 128 échantillons et un chevauchant de 50% entre les trames d'analyse successives, Guérin et al. (Guérin *et al.*, 2003) ont montré qu'un coefficient d'oubli entre 0.6 et 0.7 constitue une valeur maximale. Pour ces valeurs, la variance des dsp et de la dip calculées d'une manière récursive est très importante. Si l'influence sur les dsp se limite à une augmentation de leur variance, cela se traduit en outre par une augmentation de la valeur moyenne du module de la dip.

La figure 4.2 représente les composantes spectrales des signaux de parole et de bruit pour une fréquence de 300 Hz ainsi que les courbes de la MSC associées pour deux coefficients d'oubli, $\lambda_x = 0.9$ et $\lambda_x = 0.6$.



(a) dsp des signaux de parole et de bruit

(b) MSC pour $\lambda_x = 0.9$ et $\lambda_x = 0.6$

FIGURE 4.2 – Effet du coefficient d'oubli sur l'estimation de la MSC, pour $f = 300Hz$

Pour les fortes valeurs du coefficient de lissage $\lambda_x = 0.9$, une réverbération est introduite : l'information de corrélation, due principalement à la parole, se propage alors que celle-ci n'est plus présente. Ceci rejoint l'observation précédente sur la limite haute du coefficient d'oubli pour la parole. Pour cette même valeur, la cohérence du bruit est très faible à cette fréquence. Alors que pour des valeurs plus faible $\lambda_x = 0,6$, il n'y a plus d'effet de mémoire, mais la cohérence du bruit précédant la parole fluctue très largement pour atteindre des valeurs proches du maximum (0,8 ici), autour d'une moyenne qui est clairement plus élevée que dans le cas $\lambda_x = 0.9$.
A la suite de cette estimation récursive des dsps et de la dip des signaux bruités, la densité spectrale de puissance du bruit est déterminée en se basant sur un estimateur bi-voie. Les signaux de parole rehaussés au niveau de chaque canal sont ensuite estimés en se basant sur la technique de la soustraction spectrale.

2.2 Estimation bi-voie de la dsp du bruit

Différent estimateurs mono-voie ont été proposés pour l'estimation de la dsp du bruit. Ces estimateurs présentent des performances acceptables à faibles rapports signal-sur-bruit (RSB) et se dégradent en environnements fortement bruités. Des estimateurs bi-voie ont été proposés et ont montré de meilleures performances que les estimateurs mono-voie. En effet, les estimateurs bi-voie présentent l'avantage de l'utilisation de deux observations et par la suite une meilleure exploitation des caractéristiques spatiales des signaux. Dörbeckeret et Ernst (Dörbecker et Ernst, 1996) ont proposé un estimation bi-voie de la dsp du bruit dont le développement mathématique a été détaillé dans le deuxième chapitre.

Cet algorithme est valable sous certaines hypothèses :

Hypothèse 1 : $E\{S_i(f).B_j(f) = 0\} \quad \forall(i,j) \in \{1,2\}^2$.
Justification : les signaux de parole et les bruits sont statistiquement indépendants.

Hypothèse 2 : $P_{s1s1}(f) = P_{s2s2}(f) = P_{ss}(f)$ *et* $P_{b1b1}(f) = P_{b2b2}(f) = P_{bb}(f)$.
Justification : du fait de la faible distance entre les microphones et le locuteur proche, nous pouvons considérer que le spectre de la parole locale est identique sur les deux voies. Dans le cas du bruit, si l'égalité exacte des dsp sur les deux microphones est moins valide, elle est souvent restreinte à une différence de gain, gardant la même distribution spectrale.

Hypothèse 3 : $P_{b1b2}(f) = 0$.
Justification : La distance entre les deux microphones est suffisamment importante. En plus, les valeurs importantes de la cohérence du bruit diffus vers les basses fréquences sont négligées.

En tenant compte de ces différents hypothèses, l'estimateur bi-voie de la dsp du bruit proposé par Dörbeckeret et Ernst (Dörbecker et Ernst, 1996) peut être donné par l'équation suivante :

$$P_{bb}(f,k) = \sqrt{P_{x1x1}(f,k)P_{x2x2}(f,k)} - |P_{x1x2}(f,k)| \tag{4.4}$$

où $P_{x1x1}(f)$, $P_{x2x2}(f)$ et $P_{x1x2}(f)$ représentent respectivement les dsp des signaux de parole bruités captés par le premier et le deuxième microphone ainsi que leur dip.
La figure 4.3 montre la dsp réelle du bruit au niveau des voies droite et gauche. En effet, la dsp réelle du bruit est calculée d'une manière récursive, directement à partir des composantes du bruit enregistrés en absence de parole.

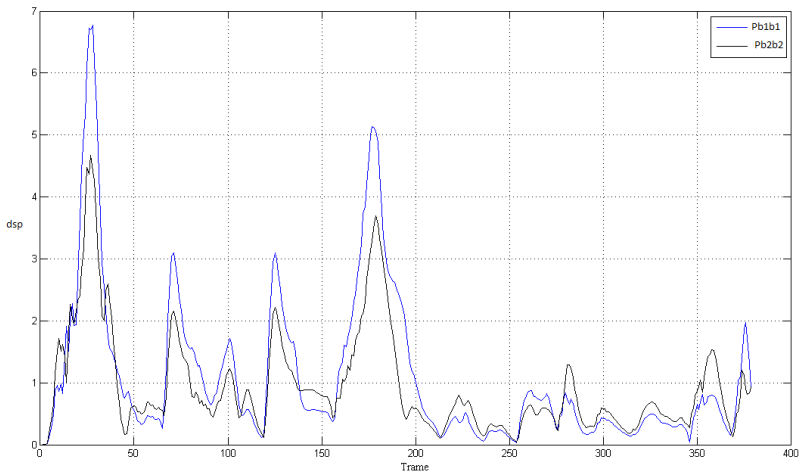


FIGURE 4.3 – Variation temporelle de la dsp réelle du bruit au niveau des voies droite et gauche

Nous remarquons à partir de cette figure qu'il existe une différence entre la dsp réelle du bruit au niveau de la voie droite et celle de la voie gauche. Par conséquent, quelque modifications ont été apportées à l'estimateur proposé par Dörbeckeret et Ernst (Dörbecker et Ernst, 1996). En effet, nous présentons ici une généralisation de cet estimateur en estimant séparément la dsp du bruit au niveau de la voie droite, notée $P_{b1b1}(f,k)$, et la dsp du bruit au niveau de la voie gauche, notée $P_{b2b2}(f,k)$. Cette estimation nécessite tout d'abord le développement de l'expression du carré du module de la dip des signaux bruités comme indiqué par l'équation 4.5 :

$$\begin{aligned}
|P_{x1x2}(f)|^2 &= |E\{X_1^*(f)X_2(f)\}|^2 \\
&= |E\{(S(f)+B_1(f))^*(S(f)+B_2(f))\}|^2 \\
&= |E\{S(f)S^*(f)+S(f)B_1^*(f)+S^*(f)B_2(f)+B_1^*(f)B_2(f)\}|^2 \\
&= \left|E\{|S(f)|^2\}\right|^2
\end{aligned} \tag{4.5}$$

L'expression du carré du module de la dsp du signal bruité capté par le premier microphone est ensuite développée et est donnée par l'équation suivante :

$$\begin{aligned}
|P_{x1x1}(f)|^2 &= \left|E\{|X_1(f)|^2\}\right|^2 \\
&= \left|E\{|(S(f)+B_1(f))|^2\}\right|^2 \\
&= \left|E\{|S(f)|^2+2\,|S(f)|\,|B_1(f)|+|B_1|^2\}\right|^2 \\
&= \left|E\{|S(f)|^2+|B_1|^2\}\right|^2 \\
&= ||P_{x1x2}(f)|+P_{b1b1}(f)|^2
\end{aligned} \tag{4.6}$$

La dsp du bruit au niveau du premier microphone peut être alors estimée à partir de l'équation 4.7.

$$P_{b1b1}(f,k) = |P_{x1x1}(f,k)| - |P_{x1x2}(f,k)| \tag{4.7}$$

De la même manière nous déterminons l'expression du carré du module de la dsp du signal bruité capté par le deuxième microphone :

$$\begin{aligned}
|P_{x2x2}(f)|^2 &= \left|E\{|X_2(f)|^2\}\right|^2 \\
&= \left|E\{|(S(f)+B_2(f))|^2\}\right|^2 \\
&= \left|E\{|S(f)|^2+2\,|S(f)|\,|B_2(f)|+|B_2|^2\}\right|^2 \\
&= \left|E\{|S(f)|^2+|B_2|^2\}\right|^2 \\
&= ||P_{x1x2}(f)|+P_{b2b2}(f)|^2
\end{aligned} \tag{4.8}$$

La dsp du bruit au niveau du deuxième microphone peut être alors estimée à partir de l'équation 4.9

$$P_{b2b2}(f,k) = |P_{x2x2}(f,k)| - |P_{x1x2}(f,k)| \tag{4.9}$$

Afin d'empêcher les variations brusques de la dsp du bruit au niveau des voies droite et gauche, une estimation récursive de la dsp du bruit suivant l'équation 4.10 est enfin considérée .

$$\begin{aligned}\bar{P}_{b1b1}(f,k) &= \lambda_b \bar{P}_{b1b1}(f,k-1) + (1-\lambda_b)\, P_{b1b1}(f,k)\\ \bar{P}_{b2b2}(f,k) &= \lambda_b \bar{P}_{b2b2}(f,k-1) + (1-\lambda_b)\, P_{b2b2}(f,k)\end{aligned} \tag{4.10}$$

où λ_b est un coefficient de lissage proche de 1 ($\lambda_b = 0.9$ dans notre cas). La figure suivante montre la dsp réelle du bruit ainsi que celle estimée suivant l'équations 4.10 dans le cas d'un bruit de type Babble à un RSB=0 dB :

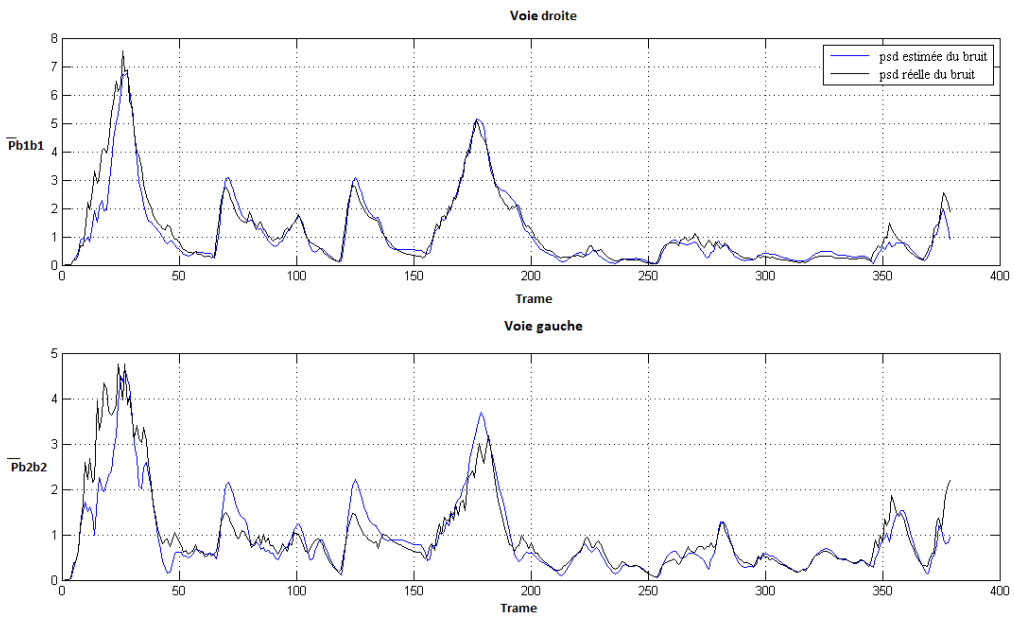


FIGURE 4.4 – Représentation temporelle des dsp réelle et estimée du bruit au niveau des voies droite et gauche

A partir de la figure 4.4, nous remarquons que la dsp estimée du bruit suit bien la dsp réelle, ce qui montre les bonnes performances de l'estimateur proposé.

3 Algorithmes d'estimation des signaux de parole rehaussés

L'étude bibliographique présentée dans le deuxième chapitre nous a permit l'identification de différents algorithmes de réduction du bruit à base de la technique de la soustraction spectrale, particulièrement l'algorithme de la soustraction spectrale multi-bande. Cet algorithme est caractérisé aussi par sa simplicité et son faible temps de traitement. Dans ce paragraphe, nous présentons une extension de deux techniques de soustraction spectrale au cas bi-voie pour l'estimation des signaux débruités. Le premier est l'algorithme

de la soustraction spectrale non linéaire proposé par Boll (Boll, 1979) et le deuxième est l'algorithme de la soustraction spectrale multi-bande proposé par Kamath et Loizou (Kamath et Loizou, 2002) puis adopté par différents autres travaux de recherche (Udrea *et al.*, 2005; Li *et al.*, 2010).

3.1 Algorithme de la soustarction spectrale non linéaire bi-voie

L'algorithme de la soustraction spectrale non linéaire, initialement proposé par Boll (Boll, 1979) dans le cas mono-voie, est tout d'abord considéré. Un facteur de sur-estimation α permet de soustraire du spectre de puissance du signal bruité une sur-estimation de la dsp du bruit. Cette opération aide à minimiser la présence du bruit résiduel en réduisant l'excursion spectrale du spectre rehaussé. Ce facteur de sur-estimation α peut être vu comme une fonction dépendante du temps qui fournit un degré de contrôle sur le processus de réduction de bruit.
Dans ce paragraphe, nous présentons une extension de cet algorithme au cas bi-voie pour l'estimation des signaux de parole rehaussés au niveau des voies droite et gauche. Le spectre de puissance du signal de parole rehaussé au niveau de la voie i ($i = \{1, 2\}$) est donné par l'équation 4.11.

$$|S_i(f,k)|^2 = \begin{cases} |X_i(f,k)|^2 - \alpha_i \bar{P}_{bibi}(f,k) & \text{si } |S_i(f,k)|^2 > \beta \bar{P}_{bibi}(f,k) \\ \beta \bar{P}_{bibi}(f,k) & \text{sinon} \end{cases} \tag{4.11}$$

où α_i ($\alpha_i > 1$) est le facteur de sur-estimation et β ($0 < \beta < 1$) est le plancher fréquentiel du bruit.
Les valeurs de α_i sont déterminées en utilisant l'équation 4.12 pour chaque trame temporelle k en fonction du rapport signal sur bruit segmenté $RSBS$ calculé au niveau de chaque voie :

$$\alpha_i = \begin{cases} 4.75 & RSBS_i < -5 \text{ dB} \\ 4 - \frac{3}{20} RSBS_i & -5 \text{ dB} < RSBS_i < 20 \text{ dB}, \quad i = \{1, 2\} \\ 1 & RSBS_i > 20 \text{ dB} \end{cases} \tag{4.12}$$

3.2 Algorithme de la soustraction spectrale multi-bande bi-voie

Dans l'algorithme de la soustraction spectrale tel que proposé par Boll (Boll, 1979), il est supposé que le bruit affecte le signal de parole d'une manière fréquentiellement uniforme. Dans ce cas, le facteur de surestimation α est indépendant de la fréquence et affecte uniformément la dsp du siganl de bruit. Cependant, dans le cas du bruit de type Babble par exemple, l'énergie du bruit est en grande proportion concentrée vers les basses fréquences. La figure 4.5 montre les variations du RSBS dans le cas d'un découpage en 4 bandes fréquentielles linéairement réparties entre 100 Hz et 4000 Hz. La première bande (bande1) est limitée entre 100 Hz et 1000 Hz, alors que la quatrième bande est limitée entre 3000 Hz et 4000 Hz. A partir de cette figure, nous remarquons que les valeurs du RSBS sont variables d'une bande fréquentielle à une autre. Les valeurs les plus importantes du RSBS sont observées dans le cas de la première bande fréquentielle (basses

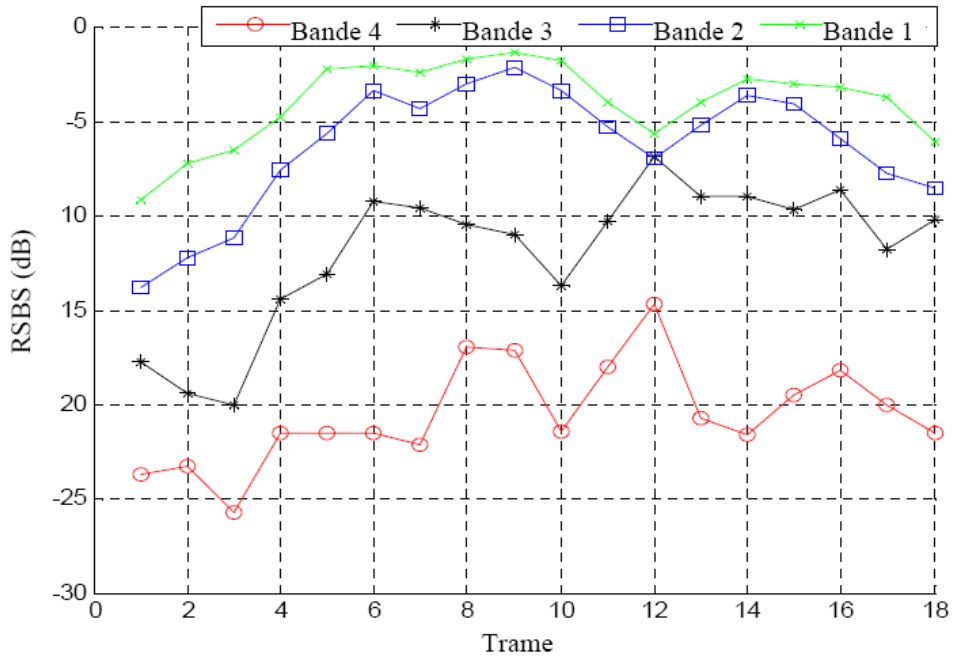


FIGURE 4.5 – RSBS pour différentes bandes fréquentielles en fonction des trames

fréquences) et atteignent des valeurs négatives au niveau de la quatrième bande fréquentielle. Ceci montre que le bruit de type Babble affecte le signal de parole d'une manière non uniforme. Par conséquent, une estimation indépendante de la fréquence du facteur de surestimation α peut engendrer une suppression des composantes du signal de parole, ce qui favorise l'apparition du bruit musical. A partir de cette constatation, et afin de minimiser les artéfacts et proposer une meilleure estimation du signal rehaussé, il est impératif d'estimer les valeurs de α par bandes de fréquences au niveau de chaque voie i. Ceci a donné naissance à l'algorithme de la soustraction spectrale multi-bande (SSMB) initialement proposé par Kamath et Loizou (Kamath et Loizou, 2002) et dont le principe est détaillé dans le deuxième chapitre.
Dans ce paragraphe, nous proposons une extension de cet algorithme au cas bi-voie donnant ainsi naissance à l'algorithme de la soustraction spectrale multi-bande bi-voie (SSMB-BV). Dans ce cas, une estimation multi-bandes du spectre de puissance du signal rehaussé au niveau des voies droite et gauche est considérée. L'approche de la SSMB est basée sur le découpage du spectre du signal de parole bruité ainsi que celui du bruit en L bandes de fréquences non chevauchées. L'algorithme de la soustraction spectrale est ensuite appliqué indépendamment au niveau de chaque bande fréquentielle. Le spectre de puissance du signal de parole rehaussé au niveau de chaque voie i ($i = \{1, 2\}$) est estimé à la $l^{ème}$ bande fréquentielle $\{l = 1, ..., L\}$ suivant l'équation 4.13.

$$|S_{i,l}(f,k)|^2 = |X_{i,l}(f,k)|^2 - \alpha_{i,l}\delta_l \bar{P}_{bibi,l}(f,k), \quad b_l < f < e_l \tag{4.13}$$

où b_l et e_l représentent respectivement les limites fréquentielles inférieures et supérieures de chaque bande de fréquences l, $\alpha_{i,l}$ est un facteur de sur-estimation déterminé pour chaque bande fréquentielle l et au niveau de chacune des voies droite et gauche. Le paramètre δ_l est un facteur de tweaking (lissage) déterminé empiriquement pour chaque bande fréquentielle

permettant une suppression plus flexible des composantes du bruit.
Les valeurs du facteur de sur-estimation du bruit $\alpha_{i,l}$ sont calculées en fonction du $RSBS_{i,l}$ suivant l'équation 4.14.

$$\alpha_{i,l} = \begin{cases} 4.75 & RSBS_{i,l} < -5 \text{ dB} \\ 4 - \frac{3}{20} RSBS_{i,l} & -5 \text{ dB} < RSBS_{i,l} < 20 \text{ dB} \\ 1 & RSBS_{i,l} > 20 \text{ dB} \end{cases} \tag{4.14}$$

Les valeurs du $RSBS_{i,l}(f)$ sont estimées pour chaque bande de fréquences l et au niveau de chaque voie i suivant l'équation suivante :

$$RSBS_{i,l} = 10 \log 10 \left(\frac{\sum_{f=b_l}^{e_l} |X_{i,l}(f)|^2}{\sum_{f=b_l}^{e_l} \left|\bar{P}_{bibi,l}(f)\right|^2} \right) \tag{4.15}$$

Les valeurs de δ_l, déterminées empiriquement, sont données par l'équation suivante (Kamath et Loizou, 2002) :

$$\delta_l = \begin{cases} 1 & 60 \text{ Hz} < f_l < 300 \text{ Hz} \\ 1.3 & 300 \text{ Hz} < f_l < 1000 \text{ Hz} \\ 1.6 & 1000 \text{ Hz} < f_l < 2000 \text{ Hz} \\ 1.8 & 2000 \text{ Hz} < f_l < 3000 \text{ Hz} \\ 1.3 & 3000 \text{ Hz} < f_l < 8000 \text{ Hz} \end{cases} \tag{4.16}$$

Le spectre de puissance du signal de parole rehaussé au niveau de la voie i ($i = \{1, 2\}$) pour la $l^{ème}$ bande fréquentielle est enfin estimé à partir de l'équation 4.17.

$$|S_{i,l}(f,k)|^2 \begin{cases} |S_{i,l}(f,k)|^2 & \text{si } |S_{i,l}(f,k)|^2 > \beta \bar{P}_{bibi,l}(f,k) \\ \beta \bar{P}_{bibi,l}(f,k) & \text{sinon} \end{cases} \tag{4.17}$$

où le paramètre du plancher du bruit est fixé à $\beta = 0.01$. Les spectres de puissance modifiés au niveau des différentes bandes fréquentielles, $|S_{i,l}(f)|^2$, sont recombinés afin d'estimer le spectre de puissance du signal de parole rehaussé $|S_i(f)|^2$ au niveau de chaque voie i. Chacun des spectres modifiés ainsi obtenu est combiné avec la phase du signal de parole bruité associé ($\theta_{Xi}(f) = \angle X_i(f)$) et le résultat est exploité afin de reconstruire le signal de parole rehaussé dans le domaine temporel moyennant une transformée de Fourier rapide inverse (TFRI) conjointement avec la méthode recouvrement et addition (OLA).

$$\hat{s}_i(n) = \left(|S_i(f)| \, e^{j\theta_{X_i}(f)} \right) \quad i = 1, 2 \tag{4.18}$$

4 Etude comparative et évaluation des performances

Dans ce paragraphe, les deux algorithmes de réduction du bruit proposés (SSNL-BV et SSMB-BV) sont évalués en utilisant deux métriques d'évaluation objectives et des tests d'écoute subjectifs.

4.1 Matériel phonétique

L'ensemble des vingt listes phonétiques de Lafon (Lafon, 1964) ont été utilisées dans cette étude expérimntale. Le signal de bruit considéré était un bruit additif de type Babble. Le niveau du signal sonore a été calibré à 70 dB SPL. Les signaux de parole bruités ont été enregistrés à différents niveaux du RSB variant de −3 dB à 6 dB avec un pas de 3 dB. Le protocole expérimental est donné par la figure 4.6.

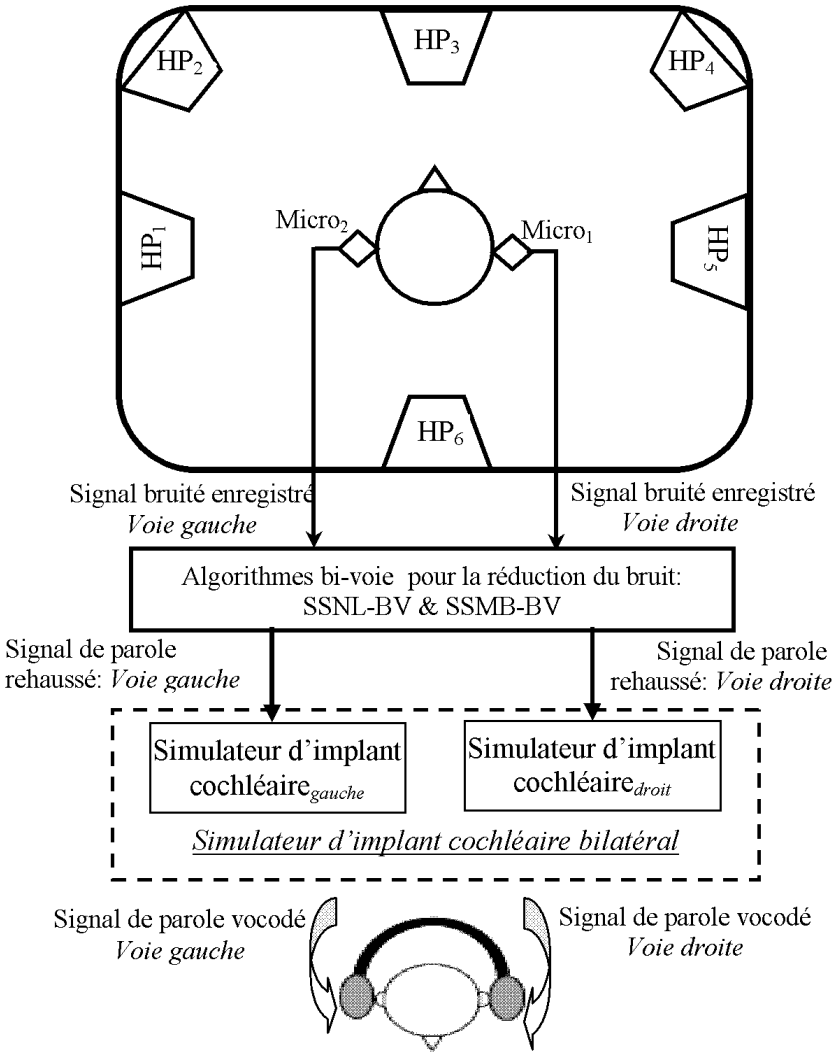


FIGURE 4.6 – Protocole expérimental

Les différentes manipulations d'enregistrement ont été effectuées dans une chambre sourde du service ORL de l'hôpital Edouard-Herriot de Lyon-France. Cette chambre est équipé de six haut-parleurs (HP) répartis sur toute la surface. Un lecteur CD (PHILIPS-CD723) a été connecté à un audiomètre (MADSEN-Orbiter 922) pour le calibrage et l'ajustement des niveaux des signaux de parole et du bruit. Ces deux appareillages ne sont pas schématisés dans la figure 4.6. Un KEMAR équipée de deux microphones omni-

directionnels a été utilisé pour l'enregistrement des différentes pistes sonores. Ce KEMAR a été placé au milieu de la salle audiométrique. La source de parole était placée dans le plan transversal en face du KEMAR (0°) et correspondait au haut parleur HP_3 . Deux configurations de sources de bruit ont été considérées dans cette expérience. Dans la première configuration, une seule source de bruit placée à droite du KEMAR à un angle de 90° a été considérée. Dans ce cas, le haut-parleur HP_5 a été utilisé. Dans la deuxième configuration, trois sources de bruit placées asymétriquement à droite et à gauche du KEMAR ont été considérées. Dans ce cas, deux sources de bruit ont été placées à droite du KEMAR à des angles de 90° et 60° correspondant respectivement aux deux haut-parleurs HP_4 et HP_4 ; alors que la troisième source a été placée à gauche du KEMAR à un angle de −60° et correspondait au haut-parleur HP_2.

4.2 Evaluation objective

Nous étudions ici les performances de l'algorithme de la SSMB-BV pour différents nombres de bandes d'analyse fréquentielle L variant de 1 jusqu'à 8 ($L = 1...8$). Le spectre du signal de parole bruité ainsi que le spectre du bruit estimé ont été découpés en différentes bandes de fréquences réparties suivant une échelle linéaire vers les basses fréquences et logarithmique vers les hautes fréquences. Le principe de ce découpage a été décrit dans le premier chapitre (algorithme 1). Pour chaque valeur de L, le nombre de raies spectrales N_l ainsi que l'indice de la raie spectrale de début $ndébut_l$ pour chaque bande de fréquences l tel que $l = 1...L$ sont données dans le tableau 4.1.

TABLE 4.1 – *Nombre de raies spectrales pour chaque bande de fréquences et pour différents nombres de bandes d'analyse*

Nombre de bandes L		**Bande**							
		1	2	3	4	5	6	7	8
2	N_l	11	53						
	$ndébut_l$	1	12						
3	N_l	6	14	43					
	$ndébut_l$	1	7	22					
4	N_l	4	7	16	37				
	$ndébut_l$	1	5	12	28				
5	N_l	3	5	8	17	31			
	$ndébut_l$	1	4	9	17	34			
6	N_l	2	4	5	10	15	28		
	$ndébut_l$	1	3	7	12	22	37		
7	N_l	2	3	4	6	9	15	25	
	$ndébut_l$	1	3	6	10	16	25	40	
8	N_l	2	2	3	4	7	9	15	22
	$ndébut_l$	1	3	5	8	12	19	28	43

Les signaux de parole bruités ont été enregistrés dans le cas d'un bruit additif de type Babble à une valeur du RSB égale à 0 dB en présence d'une seule source de bruit. Les résultats indiquant les valeurs moyennes des scores PESQ (Perceptual Evaluation of Speech Quality) (voir annex) pour différents nombres de bandes au niveau de la voie droite et la voie gauche sont donnés dans la figure 4.7.

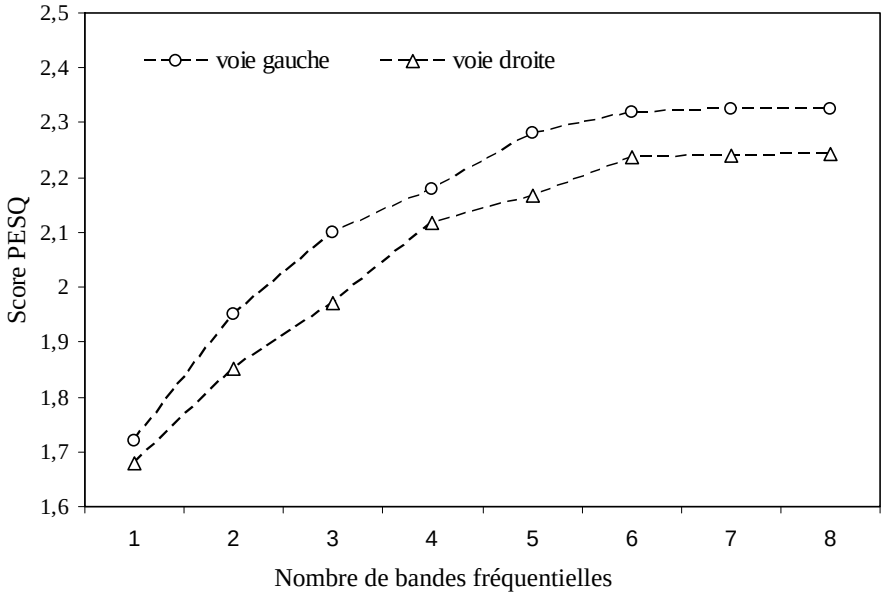


FIGURE 4.7 – SSMB-BV : Scores PESQ pour différent nombre de bandes fréquentielles

Les résultats indiquant les valeurs moyennes des distances IS (Itakura Saito) (voir annex) pour différents nombres de bandes au niveau de la voie droite et la voie gauche sont donnés dans la figure 4.8.

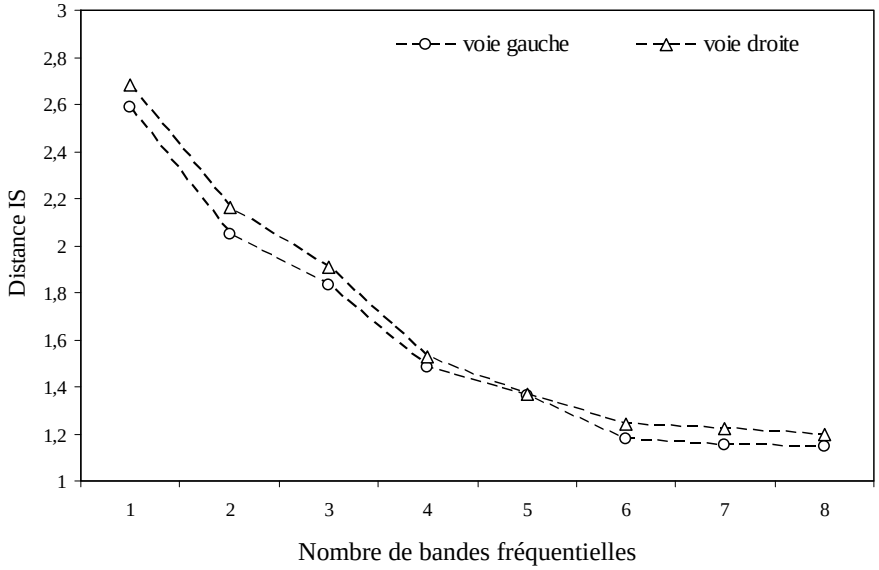


FIGURE 4.8 – SSMB-BV : Distances IS pour différent nombre de bandes fréquentielles

Nous notons que lorsque le nombre de bandes est égal à 1, l'algorithme de la SSMB-BV est équivalent à l'algorithme de la SSNL-BV. A partir de la figure 4.7 nous remarquons que les scores PESQ augmentent proportionnellement avec le nombre de bandes fréquentielles jusqu'à un nombre de bandes $L = 6$. Au delà, une augmentation du nombre de bandes n'est pas bénéfique vu que les valeurs du score PESQ restent quasiment constantes. Nous remarquons aussi que les valeurs du score PESQ au niveau de la voie gauche sont supérieures aux scores PESQ au niveau de la voie droite. Ceci peut être expliqué par la position de la source du bruit qui est plus proche du coté droit du KEMAR. La figure 4.8 montre que la distance IS décroit avec le nombre de bandes fréquentielles. Ceci étant observé pour chacune des voies droite et gauche. A partir de ces mesures objectives nous pouvons conclure que l'algorithme de la SSMB-BV présente des performances de plus en plus meilleures lorsque le nombre de bandes fréquentielles augmente jusqu'à six. Au delà, les performances sont quasiment constantes. En effet, un traitement par bande assure une meilleure estimation des facteurs de sur-estimation de la dsp du bruit et par la suite une meilleure estimation du signal rehaussé.
Nous donnons enfin une représentation temps-fréquence du signal de parole brut, bruité et rehaussés par les algorithmes SSNL-BV et SSMB-BV. Le signal de parole brut est extrait de la première liste de Lafon. Le signal bruité correspond au signal brut affecté par un bruit Babble additif à un RSB=0 dB en présence d'une seule source de bruit.

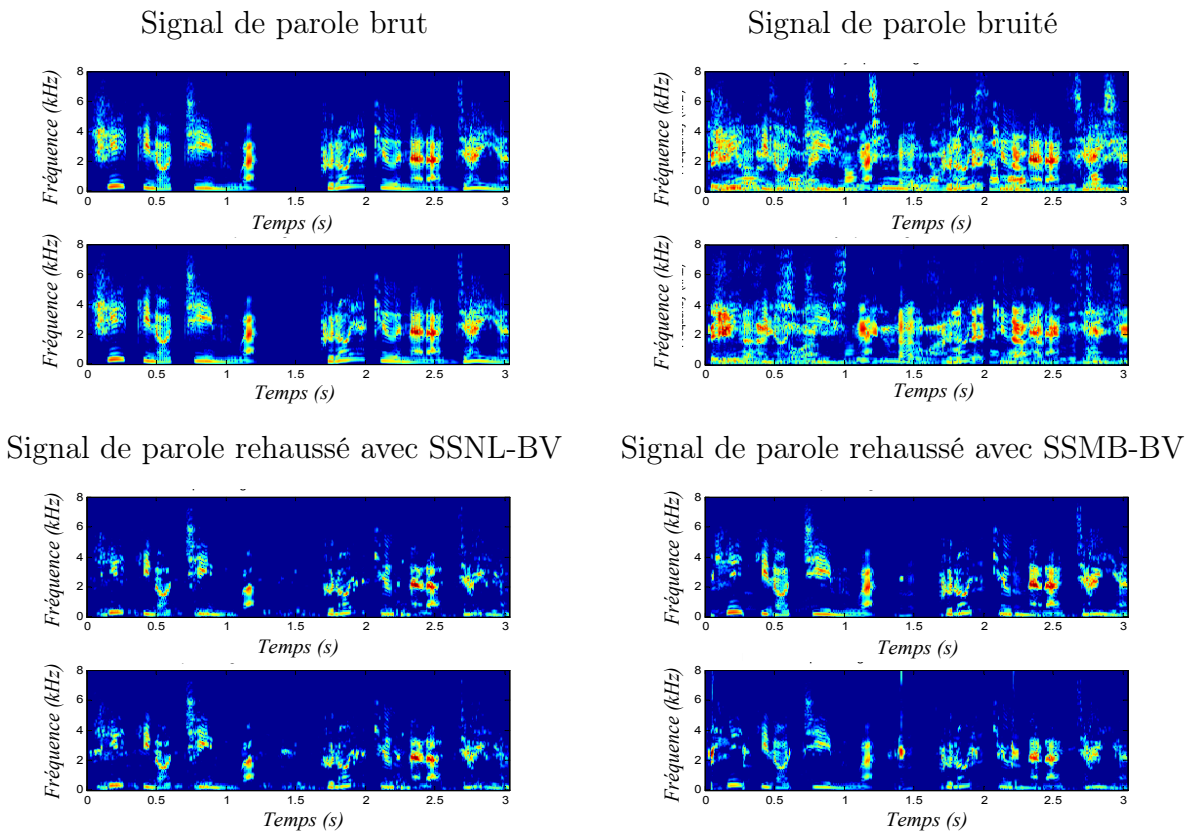


FIGURE 4.9 – Représentation temps-fréquence des différents signaux de parole

A partir de cette représentation temps-fréquence des différents signaux, nous remarquons que les deux algorithmes de débruitage proposés permettent de réduire le niveau de bruit introduit au niveau des deux voies. Les artéfacts liés à la distorsion du signal de parole et le bruit musical sont moins importants lorsque l'algorithme de débruitage SSMB-BV est considéré. Ceci peut être justifié par le fait que cet algorithme de débruitage traite le signal de parole par bande et les coefficients de sur-estimation sont estimés d'une manière fréquentiellement dépendante.

4.3 Evaluation subjective

Dans ce paragraphe, nous étudions les avantages potentiels du traitement du signal de parole bruité avec les deux algorithmes de débruitage proposés (SSNL-BV et SSMB-BV). Dans le cas de l'algorithme SSMB-BV, un découpage du spectre d'amplitude du signal de parole bruité ainsi que le spectre d'amplitude du bruit en six bandes fréquentielles d'analyse sont considérés. L'évaluation subjective de ces deux algorithmes de débruitage a été assurée grâce à des tests d'écoute permettant la détermination des pourcentages de reconnaissance des phonèmes. Cette étude expérimentale a été réalisé tout d'abord en simulation avec des sujets normoentendants, puis avec des patients bilatéralement implantés.

4.3.1 Sujets

Les performances des algorithmes de débruitage bi-voie proposés ont été évaluées avec une population de 50 sujets normoentendants et 3 patients bilatéralement implantés. L'âge des sujets normoentendants était comprise entre 18 et 32 ans. Les patients implantés portait l'implant cochléaire binaural Digisonic SP de la société Française Neurelec et leurs données biographiques sont illustrées dans le tableau 4.2. Tous les sujets avait la langue française comme langue maternelle.

TABLE 4.2 – *Données biographiques des patients implantés*

	Sujet 1	**Sujet 2**	**Sujet 3**
Age (ans)	56	36	48
Sexe	F	M	F
Nombre d'années d'implantation (ans)	4	4	4

4.3.2 Procédure

Les tests d'écoute ont eu lieu dans une chambre sourde du centre d'implantation cochléaire de l'hôpital Edouard Herriot de Lyon en utilisant un ordinateur connecté à un lecteur CD (PHILIPS-CD723). Un audiomètre numérique (MADSEN-Orbiter 922) a été utilisé pour la calibrage et l'ajustement des intensités des différents signaux sonores. Les stimuli ont été présentés d'une manière bilatérale pour chaque sujet moyennant un casque

'Sennheiser HD250' à un niveau sonore confortable calibré à 70 dB SPL.

Avant le début des tests d'écoute proprement dit, une session d'apprentissage a été tout d'abord considérée. Pendant cette session, chaque sujet entendait à un ensemble de 10 mots afin de se familiariser avec les différents types de stimuli. Aucun score n'a été calculé pendant cette session d'apprentissage. Après cette session préliminaire, chacun des cinquante sujets a été testé dans les différentes conditions expérimentales. Durant cette session d'écoute, il a été demandé de chaque sujet de répéter à haute et intelligible voix les mots qu'il a entendu. Un ensemble de 24 conditions expérimentales ont été considérées en totalité (3 méthodes de traitement x 4 RSBs x 2 configurations de sources de bruit). Une liste de 17 mots a été utilisée pour chacune des conditions expérimentales. Ces listes ont été choisies aléatoirement. Pour les 20 premières conditions, l'ensemble des 20 listes de Lafon ont été utilisées. Pour les 4 conditions expérimentales restantes, 4 listes de Lafon choisies aléatoirement ont été reconsidérées. Afin d'aider les sujets de se familiariser et de s'adapter progressivement aux différentes conditions de bruit, les différents tests d'écoute se déroulaient en commençant par les listes bruités au niveau du RSB le plus élevé (6 dB) et arrivant jusqu'au aux listes les plus fortement bruités (RSB$=-3$ dB). A la fin de chaque test expérimental, les réponses de chaque sujet ont été rassemblées et le pourcentage de reconnaissance des phonèmes correctement identifiés a été calculé.

4.3.3 Résultats

Les performances des deux algorithmes de réduction du bruit considérés (SSNL-BV et SSMB-BV) ont été évaluées et comparées d'une manière subjective avec des sujets normoentendants en simulation puis avec des sujets bilatéralement implantés.

4.3.3.a Résultats en simulation avec des sujets normoentendants

Dans ce paragraphe, nous présentons les résultats sous forme de pourcentage de reconnaissance des phonèmes dans chaque condition expérimentale. Une moyenne des scores obtenus par l'ensemble des 50 sujets a été calculée dans chaque condition. Les résultats moyens sont présentés sous forme d'histogrammes avec des barres indiquant l'erreur standard. La figure 5.8 illustre les résultats obtenus en présence d'une seule source de bruit. En présence de trois source de bruit (deuxième configuration), les résultats obtenus sont donnés par la figure 5.9.

A partir de ces deux figures, nous remarquons que l'algorithme de réduction du bruit basé sur la soustraction spectrale bi-voie permet d'améliorer les pourcentages de reconnaissance des phonèmes pour différents RSB et pour les deux configurations des sources de bruit. Les performances sont moins importantes en présence de plusieurs sources de bruit, en plus elles se dégradent pour les faibles valeurs du RSB. On remarque aussi que les performances sont variables suivant l'algorithme de la soustraction spectrale considéré. En effet, de meilleures pourcentages de reconnaissance sont notées lorsque l'algorithme de la soustraction spectrale multi-bande bi-voie est considéré.

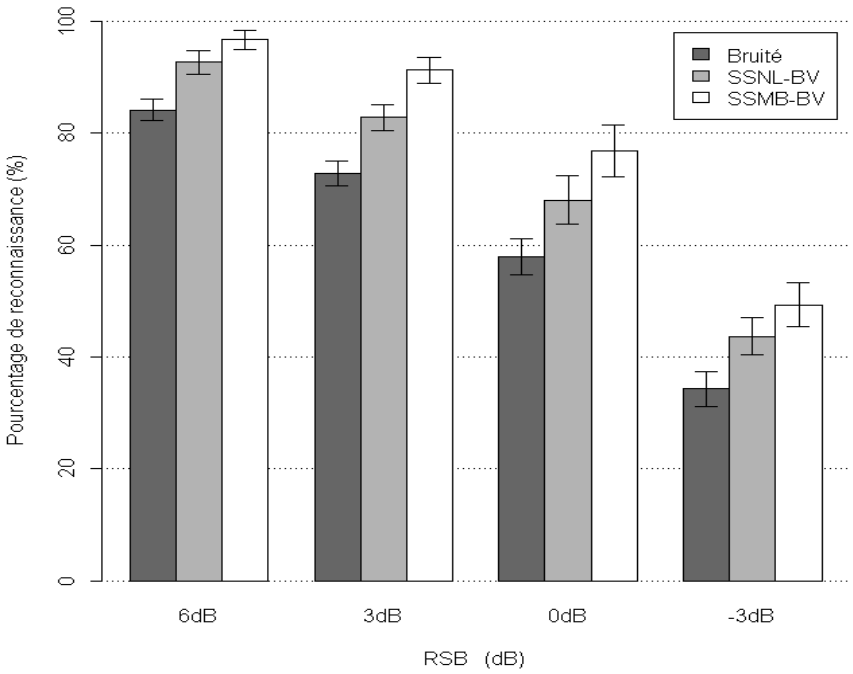


FIGURE 4.10 – Performances des algorithmes SSNL-BV et SSMB-BV en présence d'une seule source de bruit à différents niveaux du RSB

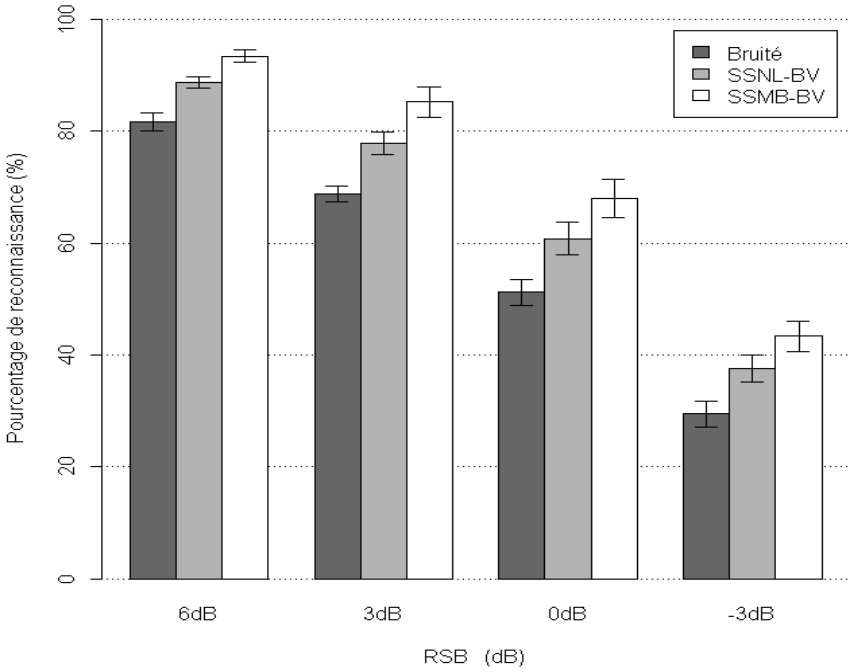


FIGURE 4.11 – Performances des algorithmes SSNL-BV et SSMB-BV en présence de trois sources de bruit à différents niveaux du RSB

Afin d'étudier l'effet des différents facteurs (algorithme, configuration des sources de bruit, RSB) ainsi que les interactions entre eux sur l'intelligibilité des phonèmes chez les patients bilatéralement implantés (en simulation), une analyse statistique des résultats obtenus a été considérée. En effet, la significativité de nos résultats a été mesurée à l'aide d'une analyse statistique basée sur des tests chi2 avec les paramètres suivants :

- Mesures répétées : Chacun des 50 sujets participent à l'ensemble des 24 conditions considérées.
- Variable dépendante : le pourcentage de reconnaissance des phonèmes.
- Trois facteurs :
  - Algorithme (Bruité, SSNL-BV et SSMB-BV)
  - Configuration des sources de bruit ('90°' et '−60°60°90°')
  - RSB (6 dB, 3 dB, 0 dB et -3 dB)

Ce type de modèle à effets mixtes est assez utilisé pour l'analyse des données (Baayen, 2008). Dans notre modèle, nous avons utilisé le programme lmer du package lme4 (Bates, 2007) implémenté sous le logiciel R (R, 2007).

Les analyses statistiques ont montré un effet significatif des facteurs algorithme (Chi2[2] =526, p<0.001), RSB (Chi2[3]=4178, p<0.001) et configuration des sources de bruit (Chi2[1]=97, p<0.001). En plus, un effet d'interaction significatif est observé entre les facteurs algorithme et RSB (Chi2[6]=45, p<0.001). Par contre, aucun effet d'intéraction significatif n'est noté entre les facteurs algorithme et configuration des sources de bruit (Chi2[2]=3, p=0.122) et entre les facteurs RSB et configuration des sources de bruit (Chi2[3]=5, p=0.4).

Les résultats ont montré qu'il existe une interaction significative entre les facteurs algorithme et RSB. Afin de mieux étudier ce comportement et chercher les paires (algorithme, RSB) pour lesquelles les différences étaient significatives, des tests de comparaison multiples ont été établis. Ce type de test peut être réalisé en utilisant le test HSD de Tukey par exemple. Le logiciel R dispose de la fonction glht du package multcomp pour réaliser ce type de test. Comparés aux pourcentages de reconnaissance obtenus avec les signaux bruités, les résultats ont montré que les pourcentages de reconnaissance sont significativement améliorés lorsque l'algorithme de réduction du bruit SSNL-BV est considéré pour tous les niveaux du RSB (p<0.001). En outre, comparés à l'algorithme SSNL-BV, les pourcentages de reconnaissance des phonèmes ont été significativement améliorés lorsque l'algorithme de réduction du bruit SSMB-BV est considéré à des valeurs du RSB égales à 3 dB, 0 dB et −3 dB (p<0.001) ; par contre, aucune amélioration significative n'a été notée à 6 dB (p=0.142).

4.3.3.b Résultats avec des sujets bilatéralement implantés

Afin d'étudier davantage les performances des algorithmes de réduction du bruit proposés dans ce chapitre et valider les résultats obtenus en simulation avec des sujets nor-

moentendanrs, les signaux rehaussés suivant les algorithmes SSNL-BV et SSMB-BV ont été présentés à des patients bilatéralement implantés. Les résultats sont aussi présentés sous forme de pourcentages de reconnaissance des phonèmes pour chacun des trois sujets bilatéralement implantés pour différents niveaux du RSB et dans le cas des deux configurations des sources de bruit précédemment considérées. La figure 4.12 montre les résultats individuels en présence d'une seule source de bruit.

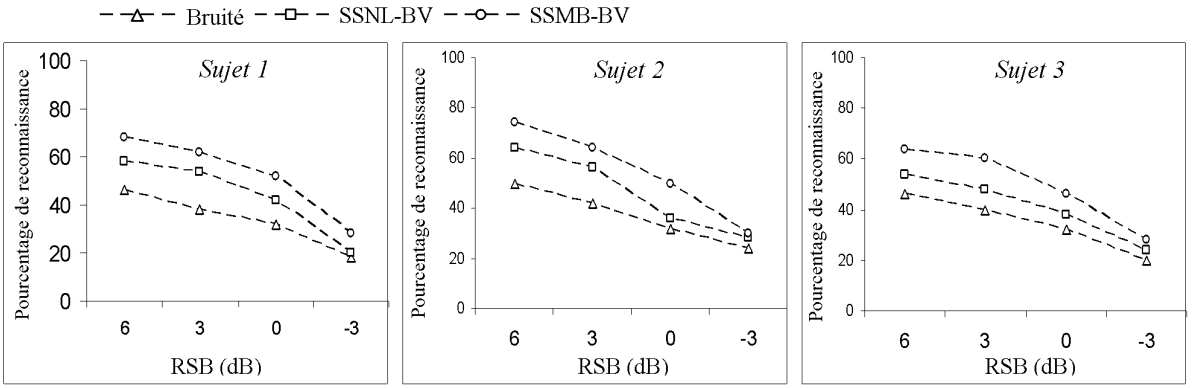


FIGURE 4.12 – Pourcentage de reconnaissance des phonèmes par sujet pour les algorithmes SSNL-BV et SSMB-BV à différents RSB : Cas d'une seule source de bruit

En présence de trois sources de bruit, les résultats sont donnés par la figure 4.13

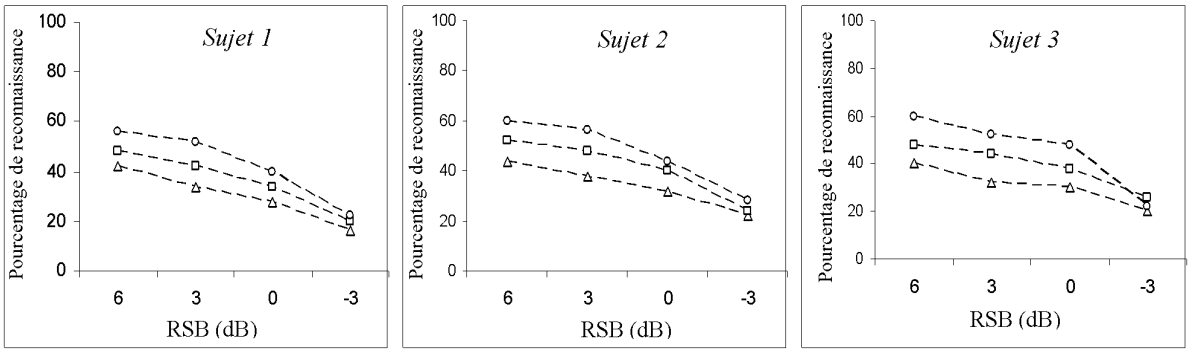


FIGURE 4.13 – Pourcentage de reconnaissance des phonèmes par sujet pour les algorithmes SSNL-BV et SSMB-BV à différents RSB : Cas de trois sources de bruit

La même analyse statistique décrite précédemment est reconsidérée pour analyser les résultats obtenues avec les sujets bilatéralement implantés. L'analyses statistique a montré un effet significatif des facteurs algorithme (Chi2[2]=70, $p<0.001$), RSB (Chi2[3]=206, $p<0.001$) et configuration des sources de bruit (Chi2[1]=10, $p<0.001$). Par contre, aucune interaction significative n'est observée entre les facteurs algorithme et RSB (Chi2[6]=3, $p=0.7$) ainsi qu'entre les facteurs algorithme et configuration des sources de bruit (Chi2[2] =1, $p=0.8$).

4.3.4 Discussion

Dans ce chapitre, un algorithme bi-voie pour la réduction du bruit a été considéré dans le cas de l'implant cochléaire bilatéral. Dans cet algorithme, l'estimation de la dsp du bruit a été basé sur une généralisation de l'algorithme bi-voie proposé par Dörbecker et Ernst (Dörbecker et Ernst, 1996). Les algorithmes de la soustraction spectrale non linéaire bi-voie (SSNL-BV) et la soustraction spectrale multi-bande bi-voie (SSMB-BV) ont été exploités afin d'estimer les spectres d'amplitude des signaux rehaussés. Une étude comparative des performances de ces deux algorithmes de débruitage a été réalisé en premier lieu en se basant sur deux mesures objectives. Les résultats préliminaires de cette étude ont montré que les performances de l'algorithme SSMB-BV s'améliorent lorsque le nombre de bandes varient de 1 à 6 ; au delà, aucun bénéfice n'a été observé. Les performances des deux algorithmes ont été par la suite évaluées et comparées d'une manière subjective avec des sujets normoentendants (en simulation) puis avec des sujets bilatéralement implantés.

4.3.4.a Algorithme de réduction du bruit vs RSB

Les résultats ont montré que les pourcentages de reconnaissance des phonèmes en présence d'une seule et de trois sources de bruit étaient significativement sensibles au RSB (figures 4.12 et 4.13). Comparés aux pourcentages de reconnaissance obtenus dans le cas des signaux bruités, des améliorations moyennes de 13% et 7% ont été respectivement observées avec les algorithmes SSNL-BV et SSMB-BV. Nous avons également noté que les améliorations étaient approximativement constantes pour des valeurs élevées du RSB (6 dB et 3 dB), mais décroissaient pour les faibles valeurs (0 dB et -3 dB). Ce comportement était confirmé avec les sujets bilatéralement implantés comme illustré par les figures 4.12 et 4.13. En effet, avec les sujets bilatéralement implantés, nous avons constaté une amélioration moyenne des pourcentages de reconnaissance de 9% pour le RSB=6 dB et de 4% pour le RSB=-3 dB lorsque l'algorithme SSNL-BV était considéré. Nous avons également observé que l'algorithme de la SSMB-BV présentait des performances supérieures à l'algorithme de la SSNL-BV. En effet, des améliorations de 8% (pour les valeurs positives du RSB) et 6% (pour le RSB=-3 dB) ont été notées par rapport à celles obtenues avec l'algorithme de la SSNL-BV.

4.3.4.b Algorithme de réduction du bruit vs configuration des sources de bruit

Deux configurations des sources de bruit ont été considérées dans cette étude. Dans la première configuration, une seule source de bruit a été utilisée. Dans la deuxième configuration, trois sources de bruit ont été considérées. L'effet du nombre de sources de bruit sur l'intelligibilité des phonèmes dans le cas des sujets bilatéralement implantés à été étudié tout d'abord avec des sujets normoentendants en utilisant un simulateur d'implant puis avec des sujets bilatéralement implantés. En présence d'une seule source de bruit (première configuration), comparés aux scores obtenus avec les signaux bruités, les améliorations moyennes étaient de l'ordre de 8% et 16% dans le cas des algorithmes SSNL-BV et SSMB-BV respectivement. Ces améliorations ont été limitées seulement à 6% et 13% en présence de trois sources de bruit. En second lieu, les perfromances ont

été évaluées avec trois sujets bilatéralement implantés. En présence d'une seule source de bruit, les résultats ont montré que l'algorithme SSMB-BV présentait une amélioration moyenne des pourcentages de reconnaissance de l'ordre 18% par rapport à l'algorithme SSNL-BV. Ces améliorations ont été plus faibles en présence de trois sources de bruit et ne dépassaient pas les 14%. Il est clair à partir des résultats obtenus que les performances des algorithmes de réduction du bruit se dégradent en présence de plusieurs sources de bruit. En effet, ceci peut être expliqué par le fait que dans le cas de l'audition binaurale, en présence de deux implants, les patients bilatéralement implantés bénéficie du phénomène d'ombre de la tête qui peut se produire principalement lorsque les signaux de parole et de bruit sont spatialement séparés.

5 Conclusion

Dans ce chapitre, une extension des algorithmes de la soustraction spectrale non linéaire (SSNL-BV) et soustraction spectrale multi-bande (SSMB-BV) au cas bi-voie a été présentée. Dans une première étape, une évaluation objective des performances de l'algorithme de la SSMB-BV pour différent nombre de bandes a montré que les performances sont maximales pour un nombre de bandes égal à six. Dans une deuxième étape, une évaluation subjective des performances des algorithmes SSNL-BV et SSMB-BV dans le cadre de l'implant cochléaire bilatéral en simulation a montré que ces algorithmes permettent d'améliorer significativement l'intelligibilité de la parole. Une supériorité des performances est notée lorsque l'algorithme SSMB-BV est considéré. Cette évaluation subjective est ensuite validée avec trois sujets portant l'implant cochléaire binaural Digisonic SP de Neurelec. Ces derniers résultats confirment ceux obtenus avec les sujets normoentendants en simulation.
L'algorithme de la soustraction spectrale bi-voie, développé sous l'hypothèse des bruits parfaitement décorrélés, a permis une amélioration de l'intelligibilité de la parole dans le cas de l'implant cochléaire. Ces améliorations se dégradent en environnements fortement bruités et en présence de plusieurs sources de bruits. Dans le chapitre suivant, l'algorithme de la soustraction interspectrale, développé sous l'hypothèse des bruits peu corrélé, sera considéré en vue d'une meilleure amélioration de l'intelligibilité de la parole dans le cas de la stimulation cochléaire bilatérale.

Chapitre 5

Exploitation de la technique de soustraction interspectrale pour la réduction du bruit : Cas de l'implant cochléaire bilatéral

Sommaire

1	**Introduction**	**93**
2	**Présentation**	**93**
3	**Algorithme de rehaussement**	**94**
4	**Estimation du bruit**	**97**
5	**Evaluation et résultats**	**106**
6	**Conclusion**	**116**

1 Introduction

Sous l'hypothèse des bruits décorrélés, l'algorithme de la soustraction spectrale bi-voie a été proposé et validé dans le cas de l'implant cochléaire bilatéral dans le chapitre précédant. Malgré que les résultats soient encourageants, les performances restent toujours insuffisantes surtout en présences de plusieurs sources de bruit et à faibles rapport signal sur bruit (RSB). L'état de l'art nous a permis d'identifier un algorithme, la soustraction interspectrale (SIS), qui est adapté à la fois aux conditions de bruits rencontrées en environnements réels, et aux contraintes temps-réel. L'algorithme de la soustraction interspectrale a été proposé sous l'hypothèse du bruit diffus. L'expression du filtre de SIS repose sur deux composantes clés qui sont :

- Estimation des densités spectrales de puissance (dsp) et de la densité interspectrale de puissance (dip) des signaux d'observation. Cette estimation doit faire un compromis entre d'une part respecter les caractéristiques non stationnaires de la parole, et d'autre part contrôler la fluctuation de la dsp du bruit présent dans les observations et qui peut être source de bruit musical.
- Estimation de la dip des bruits. Classiquement, le bruit est appris sur des périodes de silence, puis gelé pendant les phases de parole. Une nouvelle méthode pour l'estimation de la dip du bruit sera présentée dans ce chapitre. Cette méthode est une extension de l'algorithme proposé par Farsi (Farsi, 2010) au cas bi-voie. Les performances de cet estimateur sont comparées à deux autres méthodes basées sur la détection de l'activité vocale (Guérin, 2002) et les statistiques minimales (Rahmani *et al.*, 2009b).

Dans ce chapitre, nous proposons une étude comparative des performances de l'algorithme de la SIS dans le cas des trois estimateurs de la dip des bruits mentionnés dans le cas de l'implant cochléaire bilatéral en simulation. Des mètriques de mesure objectives ainsi que des tests d'écoute subjectifs ont été considérés dans cette étude comparative.

2 Présentation

Comme annoncé auparavant, l'objectif est de développer un algorithme bi-voie, dont les performances doivent être supérieures à celles de l'algorithme de la soustraction spectrale bi-voie qui ont montré leurs limites en environnement fortement bruité. En effet, l'algorithme de la soustraction spectrale bi-voie est développé en se basant sur l'hypothèse de la décorrélation parfaite des signaux de bruit, hypothèse qui n'est pas adapté principalement aux conditions de bruits rencontrées quotidiennement. L'algorithme de la soustraction interspectrale peut être une solution vu qu'une estimation robuste et permanente de la dip entre les signaux de bruit est primordiale. Nous disposons de deux observations ($x_1(n)$ et $x_2(n)$), captés par deux microphones omnidirectionnels, sommes des composantes de parole locale $s_1(n)$ et $s_2(n)$ et des bruits ambiants $b_1(n)$ et $b_2(n)$:

$$x_i(n) = s_i(n) + b_i(n), \quad i = 1, 2 \tag{5.1}$$

où n est l'indice temporel.

Nous désignons par $X_i(f,k)$, $S_i(f,k)$ et $B_i(f,k)$ les Transformées de Fourier Rapide (TFR) à Court-Terme respectives des signaux $x_i(n)$, $s_i(n)$ et $b_i(n)$, où f désigne la fréquence discrète ($f = 1...N$) et k indique l'indice de trame. La décomposition des signaux est réalisée par un fenêtrage de hamming. Les densités spectrales de puissance (dsp) ainsi que la densité interspectrale de puissance (dip) des signaux microphoniques sont données comme suit :

$$P_{xixi}(f,k) = E\{X_i(f,k)X_i^*(f,k)\} \quad i = 1,2 \tag{5.2}$$

$$P_{x1x2}(f,k) = E\{X_1(f,k)X_2^*(f,k)\} \tag{5.3}$$

Nous notons de manière identique $P_{s1s1}(f,k)$, $P_{s2s2}(f,k)$, $P_{b1b1}(f,k)$ et $P_{b2b2}(f,k)$ les dsp des signaux de parole $s_1(n)$ et $s_2(n)$ et de bruit $b_1(n)$ et $b_2(n)$ respectivement. les densités interspectrales entre bruits et entre signaux de parole sont notées respectivement par $P_{b1b2}(f,k)$ et $P_{s1s2}(f,k)$.
En toute rigueur, les transformées de Fourier dépendent également du temps car les signaux bruités ne sont généralement stationnaires qu'à court-terme. A partir des observations dans l'espace de Fourier, l'objectif est de dériver le filtre optimal de la soustraction interspectrale (H_{SIS}) pour un certain critère, à appliquer à chacune des voies droite et gauche dans le but d'extraire les spectres des signaux utiles $\hat{S}_1(n)$ et $\hat{S}_2(n)$. Finalement, les signaux temporels, $\hat{s}_1(n)$ et $\hat{s}_2(n)$, sont obtenue par une Transformée de Fourier Rapide Inverse (TFRI) et une reconstruction par la méthode OverLap and Add (OLA). Un schéma global du système de réduction de bruit bi-voie basé sur l'algorithme de la soustraction interspectrale est donnée par la figure 5.1.

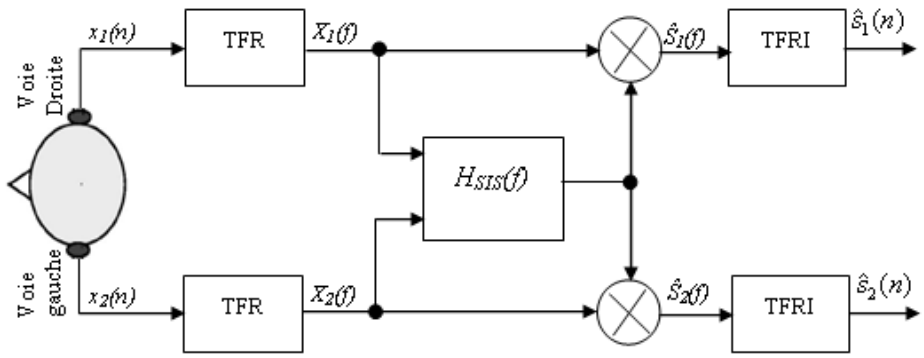


FIGURE 5.1 – Représentation schématique du système de réduction de bruit

3 Algorithme de rehaussement

L'algorithme de la soustraction interspectrale considéré s'appuie sur la fonction de cohérence en posant les hypothèses suivantes :

- Le signal de parole, issu d'une source ponctuelle, présente une corrélation très forte entre les capteurs,

- Le bruit, supposé diffus, est relativement décorrélé. La cohérence permet théoriquement d'effectuer une discrimination des signaux sur la base de la corrélation (ou la non-corrélation) spatiale. Nous avons vu que le problème des basses fréquences nous menait à considérer la dip des bruits suivant la technique de soustraction interspectrale.

La dérivation du filtre de la soustraction interspectrale repose principalement sur les deux critères suivants :

- Estimation des dsp des signaux bruités et de la dip entre ces signaux en se basant sur une méthode de filtrage récursive avec coefficient d'oubli constant,

- Estimation de la dip entre les bruits, paramètre fondamental du système de réduction de bruit. Cette estimation est traitée avec attention : nous proposons deux approches, l'une basée sur une détection d'activité vocale, l'autre sur la technique des statistiques minimales.

3.1 Soustraction interspectrale

La dérivation de l'algorithme de la soustraction interspectrale est détaillée dans le deuxième chapitre. Nous rappelons ici l'expression du filtre de soustraction interspectrale noté H_{SIS} :

$$H_{SIS} = \frac{|P_{x1x2}(f,k)| - |P_{b1b2}(f,k)|}{\sqrt{P_{x1x1}(f,k)P_{x2x2}(f,k)}} \tag{5.4}$$

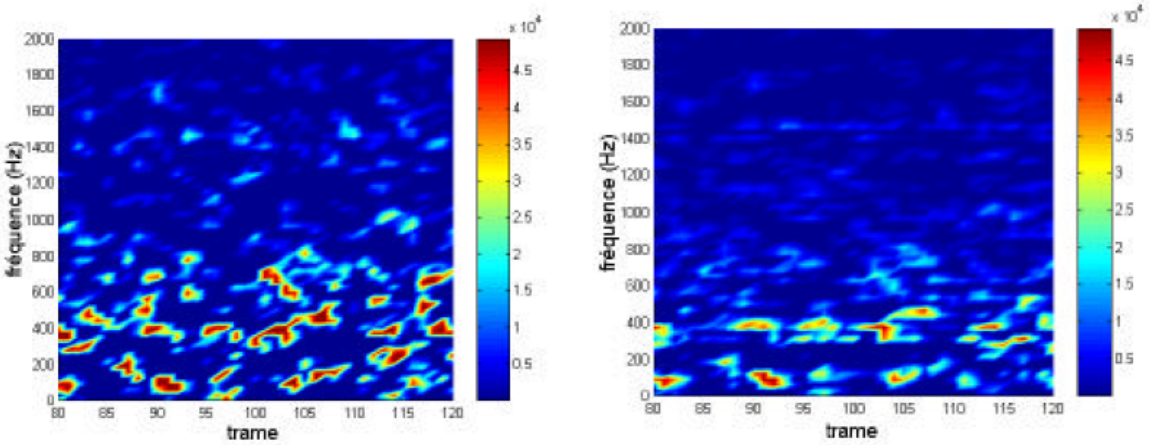
Comme dans la soustraction spectrale, si le gain $H_{SIS}(f,k)$ prend des valeurs négatives, il est mis à zéro. La composante spectrale du signal de sortie est alors estimée en appliquant le gain réel $H_{SIS}(f,k)$ à chacune des deux voies pour donner les spectres d'amplitude des signaux rehaussés :

$$\hat{S}_i(f,k) = H_{SIS}(f,k)X_i(f,k), \quad i = 1,2 \tag{5.5}$$

3.2 Soustraction interspectrale et bruit musical

La densité interspectrale des signaux de bruit est classiquement estimée pendant les périodes de bruit seul avec l'appui d'une Détection d'Activité Vocale (DAV) et gelée pendant les phases de parole. Il est notoire que, dans le cas des algorithmes de réduction du bruit mono-voie, la soustraction spectrale est source de bruit musical, conséquence d'une sous-estimation ponctuelle de la dsp du bruit due à sa non-stationnarité. Dans le cas des algorithmes de rehaussement bi-voie, la situation est moins critique car il ne s'agit pas d'une simple soustraction spectrale au vrai sens du terme à cause de l'utilisation des densités spectrales en lieu et place du spectre instantané comme indiqué dans l'équation 5.4 : les variations du gain sont finalement moins importantes.
Afin de comparer les performances de l'algorithme de la soustraction spectrale et interspectrale en terme de minimisation du bruit musical, la simulation suivante a été considérée :

à partir d'un enregistrement bi-voies d'un signal de parole en présence d'un bruit de type Babble à un RSB=5 dB, nous déterminons la dsp du bruit au niveau de la première voie $P_{b1b1}(f,k)$ ainsi que la dip entre les bruits $P_{b1b2}(f,k)$ en se basant sur la méthode de moyennage du périodogramme proposé par Welch (Welch, 1967). Le gain correspondant à la soustraction interspectrale est obtenu par l'équation 5.4, les dsp et dip étant calculées d'une manière récursive suivant les équations 5.7. Le gain du filtre de soustraction spectrale de puissance H_{ss} est donné quant à lui par l'équation 2.16 du deuxième chapitre. Des représentations temps-fréquence des bruits résiduels sont données par la figure 5.2a pour la soustraction spectrale et la figure 5.2b pour la soustraction interspectrale.



(a) Soustraction spectrale (b) Soustraction interspectrale

FIGURE 5.2 – Représentation temps-fréquence des bruit résiduels dans le cas de l'algorithme de la (a)soustraction spectrale (b) soustraction interspectrale

Comme nous pouvons s'y attendre, le signal résiduel issu de la soustraction spectrale présente des variations ponctuelles du spectre en temps et en fréquence, symptomatiques du bruit musical. Ce comportement n'est pas aussi marqué dans le cas de la soustraction interspectrale. Il faut néanmoins remarquer qu'il subsiste un certain bruit musical, confirmé par des écoutes. Cet artefact doit être absolument éliminé car il est inacceptable dans un système de réhabilitation à la surdité.

3.3 Estimation des dsp et de la dip des signaux bruités

Outre la dip des bruits, l'estimation des dsp et dip des signaux d'observation est nécessaire au calcul de la fonction de cohérence. Puisque nous disposons des spectres instantanés, les densités spectrales et interspectrale sont estimées de manière récursive par les relations suivantes :

$$P_{x1x2}(f,k) = \lambda_x P_{x1x2}(f,k-1) + (1-\lambda_x) X_1(f,k) X_2^*(f,k) \tag{5.6}$$

$$P_{xixi}(f,k) = \lambda_x P_{xixi}(f,k-1) + (1-\lambda_x) X_i(f,k) X_i^*(f,k), \quad i = 1,2 \tag{5.7}$$

où λ_x est un coefficient d'oubli proche de 1.

Les fortes valeurs de ce coefficient d'oubli représentent des ordres typiques pour l'estimation de la dsp du bruit en absence de parole. Dans le cas de la parole, λ_x ne doit au contraire pas prendre des valeurs trop proches de 1 afin que les dsp calculées par l'équation 5.7 rendent compte de son caractère dynamique et faiblement stationnaire. Bref, ce coefficient d'oubli doit prendre en compte le signal présent : en présence de parole, il doit être faible, tandis que pendant les périodes de bruit seul, il doit prendre des valeurs élevées. Une estimation récursive adaptative des dsp et dip des signaux d'observation peut être intéressante. La méthode 'decision directed' a été proposé par Beaugeant (Beaugeant, 1999) dans le cadre d'annulation d'écho pour l'adaptation du coefficient de lissage. Ce type d'adaptation du coefficient de lissage a également été proposé par Guérin et al. (Guérin *et al.*, 2003) et il est donné par l'expression suivante : $\lambda_x(f,k) = 0.98 - 0.3H_{SIS}(f,k-1)$. Dans Guérin et al. (Guérin *et al.*, 2003), les performances des deux méthodes mentionnées pour l'estimation adaptative du coefficient d'oubli ont été comparées. Pour cela, des signaux de parole bruités ont été considérés avec différentes valeurs du RSB. Pour ces signaux, deux critères d'évaluation basés sur le calcul de la distance cepstrale pour la mesure de distorsion, et des écoutes (informelles) pour le bruit musical ont été définit. Les mesures de distance cepstrale ont montré que les performances moyennes des deux estimateurs sont à peu près identiques et les différences ne sont pas significatives pour favoriser une méthode ou l'autre. En revanche, les écoutes mettent nettement en évidence que le bruit musical persiste avec la méthode 'decision directed', alors qu'il est totalement éliminé dans le cas du deuxième estimateur.

4 Estimation du bruit

Les filtres de la soustraction interspectrale, tel que décrit dans le paragraphe précédant, prennent comme hypothèse que la densité interspectrale des bruits $P_{b1b2}(f,k)$ est connue, et ce n'est qu'avec l'estimation de cette composante que les lois d'atténuation peuvent être calculées. Différentes techniques d'estimation de la dip des bruits ont été proposées. Dans ce qui suit, nous présentons le principe des méthodes les récentes pour l'estimation de la dip des bruits.

4.1 Estimation à base d'une détection d'activité vocale

La première approche pour l'estimation de la dip des bruits consiste à estimer le bruit pendant les phases de silence, puis de le geler pendant les périodes d'activité vocale. C'est la méthode classique et relativement répandue, mais qui suppose d'avoir une Détection d'Activité Vocale (DAV) fiable. L'estimation de la dip des bruits $P_{b1b2}(f,k)$ est déterminée récursivement en se basant sur l'équation 5.8 en absence de la parole :

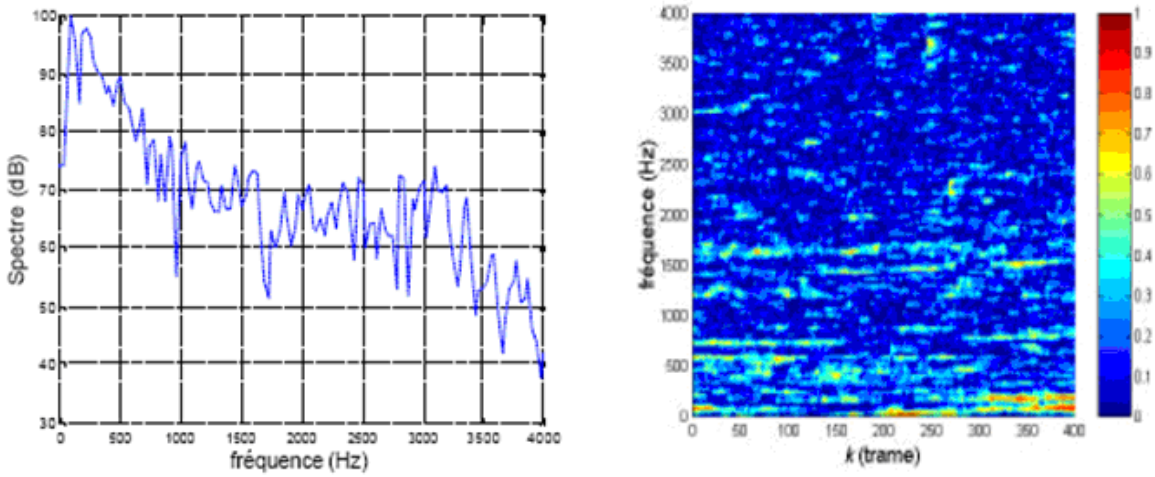
$$P_{b1b2}(f,k) = \lambda_b P_{b1b2}(f,k-1) + (1-\lambda_b)X_1(f,k)X_2^*(f,k) \tag{5.8}$$

où λ_b est un coefficient de lissage adapté aux caractéristiques du bruit.
Généralement, il est fixé à des valeurs très proches de 1. De plus, l'estimation de la dip par filtrage récursif nécessite des coefficients d'oubli relativement forts, de manière à

minimiser le biais dû aux comportements dynamiques énergétiques, mais éventuellement non corrélés. De plus, nous faisons implicitement l'hypothèse que le bruit est stationnaire sur toute la durée de la période de parole : l'estimation ne doit pas être modifiée par un événement ponctuel non cohérent précédant l'apparition d'activité vocale. Pour ces raisons, les valeurs considérées sont de l'ordre de 0.9, voire plus.
L'approche binaire de la DAV (parole/silence) induit des contraintes particulièrement fortes : elle doit être en mesure d'assurer un taux de non détections nul. Dans le cas contraire, le bruit est réactualisé avec des composantes de parole, ce qui mène à une élimination ou au moins à une forte distorsion du signal de parole.

4.1.1 Critère de détection

La DAV proposé par Guérin (Guérin, 2002) utilise les propriétés de la fonction de cohérence appliquée aux bruits en environnement de conversation réel (bruit de type Babble par exemple). La figure 5.3a montre le spectre instantané du bruit alors qu'une représentation temps-fréquence de la fonction de cohérence calculée suivant une estimation récursive avec un coefficient d'oubli $\lambda = 0.9$ est donnée par la figure 5.3b. Il est clair à partir de cette figure que le bruit n'est corrélé que pour des fréquences très faibles, en dessous de $250Hz$ environ, d'ou la validation de l'hypothèse du bruit diffus. Au-delà de $250Hz$, les mesures de corrélation sont très faibles, voire nulles pour la partie haute du spectre.



(a) Spectre instantané

(b) MSC(f,k)

FIGURE 5.3 – Caractéristiques du bruit diffus

A la lueur de ces observations, Guérin (Guérin, 2002) a proposé de prendre comme paramètre de décision la MSC (Magnitude Squared Coherence) moyenne sur la partie haute du spectre en utilisant l'équation 5.9 :

$$\bar{MSC}(k) = \frac{2}{N} \sum_{f=N/4}^{N/2+1} MSC(f,k) \tag{5.9}$$

L'avantage de ne considérer que la partie haute du spectre découle directement des observations menées dans le paragraphe précédent. Si nous pouvons présenter des mesures de cohérence très proches de 1 au niveau des basses fréquences, nous avons vu que ce n'était plus le cas pour la partie haute du spectre. Par contre, dans le cas de la parole, cette cohérence est quasiment égale à 1 sur tout le spectre, du fait du caractère ponctuel de la source.

4.1.2 Seuil de détection

Bouquin-Jannes (Bouquin-Jeannés, 1991) a proposé d'utiliser un seuil constant pour la détection d'activité vocale, sachant que la valeur de $\bar{MSC}(k)$ est toujours comprise entre 0 et 1. Cependant, comme vu précédemment, la MSC moyenne peut varier sensiblement d'une condition de bruit à l'autre. De manière très simple, Guérin (Guérin, 2002) a proposé d'utiliser comme seuil de détection la moyenne temporelle de la $\bar{MSC}(k)$ du bruit, notée $\mu_{\bar{MSC}}(k)$, initialement apprise sur une période de bruit seul. Un intervalle de confiance autour de cette moyenne qui dépend de l'écart-type de $\bar{MSC}(k)$, notée $\sigma_{\bar{MSC}}(k)$ est ensuite considéré. Si $\bar{MSC}(k)$ est considérée comme une variable aléatoire gaussienne, la probabilité que $\bar{MSC}(k)$ reste dans l'intervalle $\mu_{\bar{MSC}}(k) \mp 3\sigma_{\bar{MSC}}(k)$ est de 0,997. Bien que l'hypothèse gaussienne pour $\bar{MSC}(k)$ ne soit en aucun cas garantie, Guérin a choisi le seuil de détection donné par l'équation 5.10 :

$$\begin{cases} \bar{MSC}(k) > \mu_{\bar{MSC}}(k) + 3\sigma_{\bar{MSC}}(k) \rightarrow \text{parole locale} \\ \bar{MSC}(k) < \mu_{\bar{MSC}}(k) + 3\sigma_{\bar{MSC}}(k) \rightarrow \text{bruit seul} \end{cases} \tag{5.10}$$

La moyenne et l'écart-type du bruit, notés respectivement par $\mu_{\bar{MSC}}(k)$ et $\sigma_{\bar{MSC}}(k)$, sont estimés en se basant sur un filtrage récursif du premier ordre. Dans ce cas, le DAV est décrit en suivant les étapes suivantes :

- Si $\bar{MSC}(k) < \mu_{\bar{MSC}}(k-1) + 3\sigma_{\bar{MSC}}(k-1)$: DAV=0
$$\begin{cases} \mu_{\bar{MSC}(k)} = \lambda_\mu \mu_{\bar{MSC}}(k-1) + (1-\lambda_\mu)\bar{MSC}(k) \\ \sigma_{\bar{MSC}(k)} = \lambda_\sigma \sigma_{\bar{MSC}}(k-1) + (1-\lambda_\sigma)\left|\bar{MSC}(k) - \mu_{\bar{MSC}}(k-1)\right| \end{cases}$$
- sinon : DAV=1
$$\begin{cases} \mu_{\bar{MSC}(k)} = \mu_{\bar{MSC}(k-1)} \\ \sigma_{\bar{MSC}(k)} = \sigma_{\bar{MSC}(k-1)} \end{cases}$$

où λ_μ et λ_σ sont des coefficients d'oubli proches de 1.

Pour le choix de ces constantes de temps, il faut respecter la relation d'ordre $\lambda_\mu < \lambda_\sigma$ afin que l'écart-type ne soit pas trop sensible à des changements de moyenne, notamment au début des périodes d'activité vocale. Dans le cas contraire, les estimateurs sont surestimés, les valeurs de seuil et de temps de montée également, ce qui peut aboutir à une estimation biaisée de P_{b1b2}. La figure suivante présente un exemple d'estimation de la dip du bruit à partir du signal bruité à un RSB=0 dB.

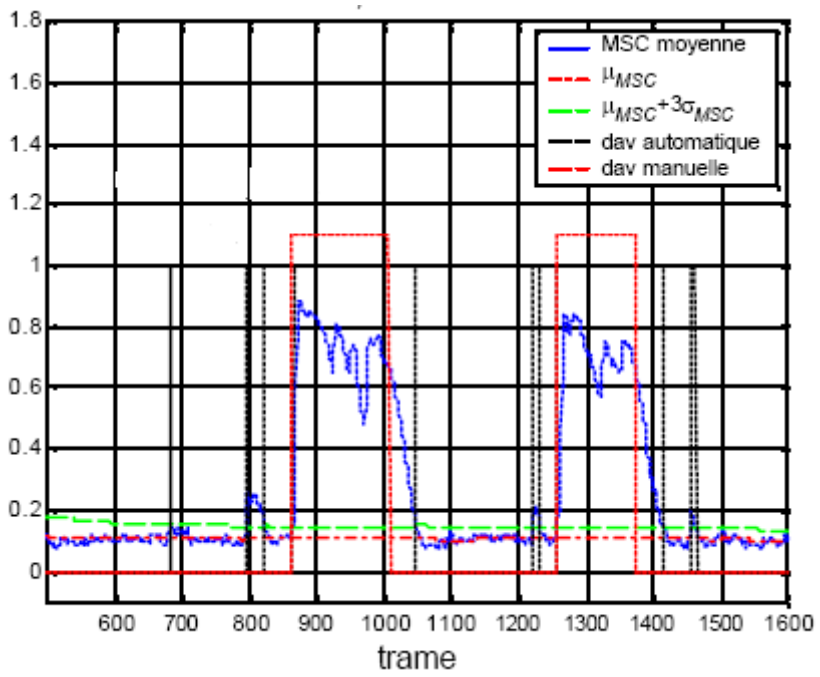


FIGURE 5.4 – Evaluation des performances du DAV considérée dans le cas d'un bruit Babble à un RSB=0dB

Nous remarquons que les temps de montée sont en moyenne faibles ($\approx$ 36 ms). Nous notons ici que l'estimation du bruit pendant ces périodes peut amener à une distorsion du signal utile puisque l'observation contient des composantes de parole. Par contre, les temps de descente sont très importants ($\approx$ 0.3 s), principalement dans les conditions moyennement ou faiblement bruitées. Ces retards peuvent s'avérer critiques pour le bon fonctionnement du débruiteur, notamment pour les faibles et moyennes valeurs du RSB. En effet, dans ces conditions, le bruit n'ayant pas été réactualisé depuis le début de l'activité vocale. Son estimation peut être fortement biaisée surtout en fin d'activité vocale, et le comportement du filtre de la soustraction interspectrale n'est alors plus optimal. Ceci nous mène tout naturellement à nous pencher sur la réactualisation du bruit en période d'activité vocale et par suite une estimation permanente de la dip des bruits.

4.2 Estimation permanente de la dip des bruits

En se basant sur l'information de présence de parole déterminée à partir du DAV décrit précédemment, le filtre de la soustraction interspectrale présente une estimation fiable du bruit pendant les périodes de silence (présence du bruit seul), ce qui garantit une bonne réduction de bruit, tout en conservant les caractéristiques spectrales de ce dernier. Cependant, Guérin (Guérin, 2002) a indiqué que les tests d'écoute informelles ont montré que les signaux en sortie des filtres présentent certains artefacts pendant les phases de parole. La soustraction interspectrale présente un fort bruit musical, qui se manifeste également lors des périodes de mauvaise estimation de la dip des bruits. Nous en concluons que le gel du bruit pendant les phases de parole est une solution valide uniquement à court-terme pour des périodes ne dépassant pas quelques dizaines de millisecondes dans des contextes fortement stationnaires. La DAV proposée par Guérin (Guérin, 2002) ne peut apporter

acune solution à ce problème car elle rend une décision globale :

- d'une part sur le plan temporel dans le sens où, du fait des forts coefficients de lissage, elle ne peut discriminer les phases inter-syllabiques qui pourraient être mises à profit pour la réactualisation de la dip des bruits,

- d'autre part sur le plan fréquentiel : la DAV est estimée au niveau de chaque trame temporelle mais non par fréquence. Or, du fait de sa structure formantique, la parole possède un spectre présentant des pics et des vallées, caractéristiques qui pourraient être également utilisées.

Ce problème a été abordé largement dans la littérature. A partir du système de détection d'activité locale décrit dans le paragraphe précédant, Akbari Azirani (Azirani, 1995) a proposé d'augmenter la résolution temporelle en choisissant un coefficient de lissage plus faible ($\lambda = 0.7$) afin de prendre en compte plus aisément les phases inter-syllabiques. En considérant cette approche, l'estimation de la dip des bruits est fréquenciellement indépendante. Différentes autres approches ont été proposées pour une meilleure estimation de la dip des bruit, en se basant sur une estimation fréquenciellement dépendante.

4.2.1 Estimation récursive

Un algorithme d'estimation récursive de la dip du bruit a été proposé par Rahmani et al. (Rahmani *et al.*, 2009a). Cet algorithme est une extension de l'algorithme mono-voie proposé par Sovka et al. (Sovka *et al.*, 1996) pour l'estimation de la psd du bruit. Une expression modifiée du filtre de la soustraction interspectrale est considérée :

$$H_{SIS} = \frac{P_{x1x2}(f,n)}{\sqrt{P_{x1x1}(f,k)P_{x2x2}(f,k)}} \frac{R_{prio}(f,k)}{1 + R_{prio}(f,k)} \tag{5.11}$$

où $R_{prio}(f,k)$ est l'estimation a priori du rapport de la dip entre les signaux propres $P_{s1s2}(f,k)$ et celle des bruits $P_{b1b2}(f,k)$.
L'estimation de la dip du bruit $P_{b1b2}(f,k)$ est faite en deux étapes :

- Première étape : Estimation récursive de la dip des bruit :
  $P_{b1b2}(f,k) = \lambda_b P_{b1b2}(f,k-1) + (1-\lambda_b)\,|X_2(f,k)X_1^*(f,k)|\,(1 - H_{sis})$
- Deuxième étape. : Estimation la dip des signaux propres :
  $P_{s1s2}(f,k) = P[|X_2(f,k)X_1^*(f,k)| - |P_{b1b2}(f,k)|]$

où λ_b est un coefficient de lissage entre 0 et 1 et $P[.]$ est la fonction de redressement simple alternance. Cet algorithme a été proposé dans le cadre des algorithmes de réduction du bruit dédiés pour les communications en mains libres.

4.2.2 Estimation basée sur les informations de phase

Une autre approche pour l'estimation de la dip des bruits été proposée par Rahmani et al. (Rahmani *et al.*, 2009c). Cette estimation est valable dans le cas des bruits diffus

et elle est basée sur l'exploitation des informations données par la phase des signaux. En partant de l'hypothèse que $P_{x1x2} = P_{s1s2} + P_{b1b2}$, les auteurs ont démontré que l'estimation de la dip des bruit peut être établie en utilisant l'équation suivante :

$$P_{b1b2}(f,k) = |P_{x1x2}(f,k)| \left| \frac{\sin\left(\angle P_{s1s2} - \angle P_{x1x2}\right)}{\sin\left(\angle P_{s1s2} - \angle P_{b1b2}\right)} \right| \tag{5.12}$$

où $\angle P_{x1x2}$, $\angle P_{s1s2}$ et $\angle P_{s1s2}$ dénotent respectivement la phase des dip entre signaux bruités, signaux propres et bruits. Cet estimateur a été proposé dans le but de l'amélioration de l'intelligibilité de la parole dans le cadre des communications en mains libres.

4.2.3 Estimation en ligne de la dip des bruits

Il est notoire que la soustraction spectrale est source de bruit musical, et bien que la soustraction interspectrale soit moins sensible, elle n'y échappe pas. Une des solutions permettant de limiter cet artefact est de surestimer le bruit. Cette technique consiste à multiplier le spectre du bruit estimé par un coefficient constant α. Des études ont montré qu'un facteur de surestimation de l'ordre de $\alpha = 6$ peut éliminer complètement le bruit musical, mais en contrepartie, de fortes distorsions du signal utile apparaissent (Berouti *et al.*, 1979). Dans (Guérin *et al.*, 2003), les auteurs ont proposé de faire une surestimation du bruit en approchant la dip court-terme des bruits par la moyenne géomètrique des densités spectrales long-terme des bruits (appelé spectre bi-voies), soit $|P_{b1b2}(f,k)| = \sqrt{P_{b1}(f)P_{b2}(f)}$. Un estimateur heuristique du spectre bi-voie basé sur des a priori énergétiques a été proposé par Guérin et al. (Guérin *et al.*, 2003) et il est donné par l'équation suivante :

$$\sqrt{P_{b1}(f,k)P_{b2}(f,k)} = \alpha\left(\widetilde{RSB}_{2,post}(f,k)\right)\sqrt{P_{b1}(f,k-1)P_{b2}(f,k-1)} \tag{5.13}$$

où la fonction $\alpha\left(\widetilde{R}SB_{2,post}(f,k)\right)$ est définie par :

$$\alpha\left(\widetilde{RSB}_{2,post}(f,k)\right) = L + (1-L)\frac{1}{1+\frac{1}{g.\widetilde{RSB}_{2,post}}}\left(1+\frac{1}{1+g.b.\widetilde{RSB}_{2,post}}\right) \tag{5.14}$$

Le rapport signal à bruit modifié, $\widetilde{RSB}_{2,post}$, est donné par le rapport entre la composante spectrale du signal bruité et du spectre bi-voies des bruits estimé de la trame précédente :

$$\widetilde{RSB}_{2,post}(f,k) = \frac{|X_1(f,k)X_2(f,k)|}{\sqrt{P_{b1}(f,k-1)P_{b2}(f,k-1)}} \tag{5.15}$$

où b, g et L sont des constantes qui déterminent la forme de la fonction $\alpha\left(\widetilde{RSB}_{2,post}(f,k)\right)$. Le coefficient $L \in [0,1]$ règle la décroissance exponentielle du spectre bi-voies, la constante b fixe la valeur maximale atteinte par le coefficient α, alors que la constante g permet d'ajuster ces valeurs maximales pour une valeure du $\widetilde{RSB}_{2,post}(f,k)$ donnée.

4.2.4 Estimation basée sur la technique des statistiques minimales

Dans (Martin, 1994), Martin a proposé une méthode d'estimation de la dsp du bruit dans le cas monovoie avec une réactualisation des composantes du bruit pour chaque bande de fréquences. Cette méthode est basée sur la technique des statistiques minimales du signal observé (Minimum Statistics). Cette méthode repose sur l'hypothèse que, pour chaque composante fréquentielle, la puissance minimale sur une période d'observation donnée correspond à une composante de bruit. Un mécanisme de compensation du biais d'estimation a été introduit par l'auteur grâce à un estimateur des puissances basé sur un lissage autorégressif. En effet, ce facteur de compensation dépend des caractéristiques statistiques de bruit et doit être appris hors-ligne, ce qui impose d'introduire différentes compensations pour les différentes situations de bruit. L'intérêt de cette méthode repose sur la possibilité d'extraire des informations de bruit en présence de parole.
Cet estimateur de la dsp du bruit proposé par Martin (Martin, 1994) dans le cas monovoie a été étendu au cas bi-voie par Rahmani et al. (Rahmani *et al.*, 2009b) pour l'estimation de la dip des bruits. Une estimation récursive avec un coefficient de lissage adaptatif de la dip des signaux bruités a été tout d'abord considérée :

$$R_{x1x2}(f,k) = \lambda_x(f,k)R_{x1x2}(f,k-1) + (1-\lambda_x(f,k))X_1(f,k)X_2^*(f,k) \tag{5.16}$$

où les valeurs de $\lambda_x(f,k)$ sont déterminées empiriquement en utilisant l'équation 5.17 :

$$\lambda_x(f,k) = \begin{cases} \sqrt{1-t_1}, & \text{si } \rho(f,k) \leq t_1 \\ \sqrt{1-\rho(f,k)} & \text{si } t_1 < \rho(f,k) < t_2 \\ \sqrt{1-t_2} & \text{si } si\rho(f,k) \geq t_2 \end{cases} \tag{5.17}$$

où t_1 et t_2 sont deux seuils fixés empiriquement à 0.2 et 0.8 respectivement, $\rho(f,k)$ est la fonction de cohérence tel que $\rho(f,k) = \frac{P_{x1x2}(f)}{\sqrt{P_{x1x1}(f).P_{x2x2}(f)}}$.
La dip à court-terme des bruits est déterminée en calculant les minima locaux de la dip des signaux bruités, notés $R_{x1x2min}(f,k)$, estimés à partir de l'équation 5.16. Ceci est assuré en appliquant un algorithme de suivi des minima (SM) locaux. En effet, ces minima sont identifiés en appliquant un algorithme de recherche appliqué sur un ensemble de L trames en se basant sur l'équation suivante :

$$R_{x1x2min}(f,k) = min\{|R_{x1x2}(f,m)|, k-L+1 < m < k\} \tag{5.18}$$

Généralement, les valeurs des minimums locaux estimées à partir de l'équation 5.18 sont inférieures à la moyenne des dip des signaux bruités, et par la suite $R_{x1x2min}(f,k)$ représente une estimation biaisée de la dip du bruit. Afin de fournir une estimation non biaisais de la dip des bruit, les valeurs de $R_{x1x2min}(f,k)$ sont multipliés par un facteur B_{min} appelé facteur de compensation du biais. La dip à court-terme des bruits est estimée en utilisant l'équation 5.19.

$$\hat{P}_{b1b2}(f,k) = B_{min}R_{x1x2min}(f,k) \tag{5.19}$$

En se basant sur le travail de Le Bouquin-Jeannes et al. (Bouquin-Jeannés *et al.*, 1997) proposé pour l'estimation de la dsp du bruit dans le cas mono-voie, Rahmani et al. (Rahmani *et al.*, 2009b) ont choisie B_{min} un facteur de compensation du biais d'estimation constant égal à 2. Martin (Martin, 2006) a montré que ce facteur est fonction de la largeur de la fenêtre de recherche L.

4.2.5 Algorithme proposé pour l'estimation de la dip des bruits

Lorsque la technique proposée par Martin (Martin, 1994) basée sur la technique des statistiques minimales est considérée, la mise à jour des composantes spectrales de la dip des bruits sont directement dépendantes de la largeur de la fenêtre de recherche utilisée pour le suivi des minima locaux. L'inconvénient majeur de cette méthode se présente lors des variations brusques des niveaux du bruit. Afin de surmonter ce problème, Farsi (Farsi, 2010) a proposé une nouvelle méthode basée aussi sur la technique des statistiques minimales et le suivi des minima locaux est fait d'une manière non linéaire en considérant une moyenne des composantes spectrales antérieures. Dans cet algorithme, la réactualisation du minimum est faite d'une manière temporellement continue et non dépendante de la largeur fixe d'une fenêtre de recherche. Cet algorithme a été proposé en partie par Doblinger (Doblinger, 1995). Nous présentons ici une extension de cet algorithme au cas bi-voie où le suivi des minima locaux et par la suite l'estimation de la dip des bruits est déterminé en se basant sur l'approche des statistiques minimales améliorées (SMA) (Kallel *et al.*, 2012b).
Une estimation récursive avec un coefficient de lissage λ_b constant de la dip des signaux bruités, donnée par l'équation 5.20 a été tout d'abord considérée.

$$R_{x1x2}(f,k) = \lambda_b R_{x1x2}(f,k-1) + (1-\lambda_b) X_1(f,k) X_2^*(f,k) \tag{5.20}$$

où λ_b est un coefficient de lissage fixé à la valeur $\lambda_b = 0.9$.
Une estimation non linéaire et récursive du minimum local $R_{x1x2min}(f,k)$ est déterminé au niveau de chaque trame et pour chaque composante fréquentielle en se basant sur l'équation suivante :

$$\begin{array}{l} si \;\; R_{x1x2min}(f,k-1) < R_{x1x2}(f,k) \\ R_{x1x2min}(f,k) = \gamma R_{x1x2min}(f,k-1) + \frac{1-\gamma}{1-\beta}(R_{x1x2}(f,k) - \beta R_{x1x2}(f,k-1)) \\ sinon \\ R_{x1x2min}(f,k) = \gamma R_{x1x2min}(f,k-1) - \mu(R_{x1x2min}(f,k-1) - \zeta R_{x1x2}(f,k-1)) \end{array} \tag{5.21}$$

où α, β, μ et ζ sont des constantes déterminées empiriquement par Farsi (Farsi, 2010) tel que $\alpha = 0.997$, $\beta = 0.8$, $\mu = 0.01$ et $\zeta = 0.9$.
Afin de compenser le biais d'estimation, la dip des bruits est enfin estimée en multipliant le minimum local $R_{x1x2min}(f,k)$ par un facteur de compensation du biais d'estimation, choisi constant ($B_{min} = 2$), tel que donné par l'équation 5.19.
La figure 5.5 présente un schéma bloc de l'algorithme d'estimation de la dip des bruits proposé et qui est basé sur la technique des statisiques minimales améliorée.
Dans cet algorithme, les psd ainsi que la dip des signaux bruités sont tout d'abord calculées en utilisant l'équation 5.7. Dans l'étape suivante, une estimation lissée de la dip des signaux bruités est déterminée en se basant sur l'équation 5.20. La dip des bruits est enfin estimée en se basant sur les équations 5.21 et 5.19. Cet estimé est exploité pour calculer le filtre de la soustraction interspectrale H_{SIS}. Ce filtre est appliqué sur les composantes spectrales des signaux bruités de chacune des deux voies droite et gauche. Les spectres modifiés $\hat{S}_1(f,k)$ et $\hat{S}_2(f,k)$ sont utilisés pour reconstruire les signaux rehaussés dans le domaine temporel en utilisant une TFRI conjointement avec la méthode OLA.
L'algorithme 2 présente les différentes étapes nécessaires pour l'estimation des spectres d'amplitude des signaux rehaussés en se basant sur l'approche proposée SMA pour le

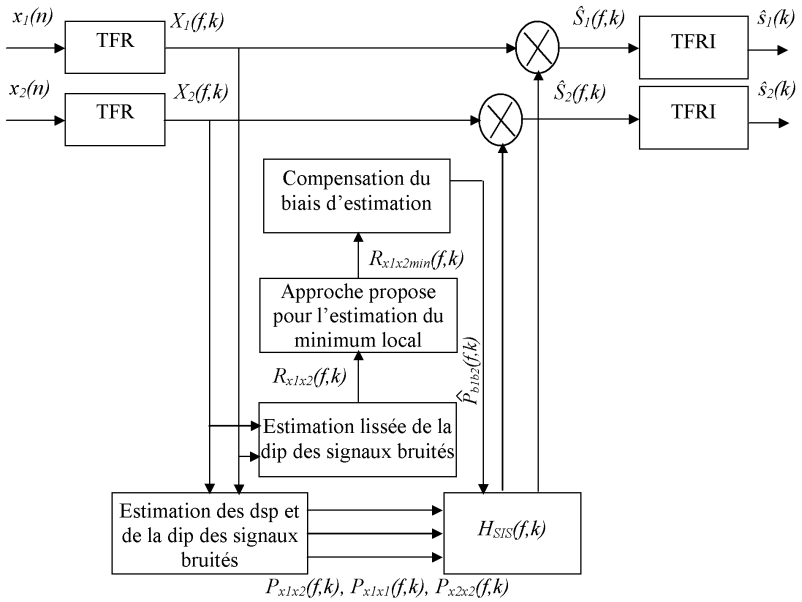


FIGURE 5.5 – Schéma bloc de l'estimateur de la dip des bruits proposé

suivi des minima locaux et l'estimation de la dip des bruits. Nous notons ici qu'une phase d'initialisation est considérée au niveau de la première trame. En effet, cette trame est considérée comme une trame ne contenant que du bruit (absence de la parole) et par conséquent, la dip des signaux bruités correspond à la dip des bruits et le filtre de la soustraction interspectrale est mis à zéro afin d'atténuer toutes les composantes du bruit.

Algorithme 2 Différentes étapes nécessaires pour l'estimation des signaux rehaussés

Paramètres : $\lambda_x = 0.7$, $\lambda_b = 0.9$, $\alpha = 0.997$, $\beta = 0.8$, $\mu = 0.01$, $\zeta = 0.9$, $B_{min} = 2$

Première trame (k=1) : Initialisation

Étape 1 : $P_{x1x1}(f,1) = X_1(f,1).X_1*(f,1)$.

Étape 2 : $P_{x2x2}(f,1) = X_2(f,1).X_2*(f,1)$.

Étape 3 : $P_{x1x2}(f,1) = X_1(f,1).X_2*(f,1)$.

Étape 4 : $R_{x1x2}(f,1) = X_1(f,1).X_2*(f,1)$.

Étape 5 : $R_{x1x2min}(f,1) = R_{x1x2}(f,1)$.

Étape 6 : $H_{SIS} = 0$.

Les autres trames (k>1)

Étape 7 : $P_{x1x1}(f,k) = \lambda_x P_{x1x1}(f,k-1) + (1-\lambda_x)X_1(f,k).X_1*(f,k)$.

Étape 8 : $P_{x2x2}(f,k) = \lambda_x P_{x2x2}(f,k-1) + (1-\lambda_x)X_2(f,k).X_2*(f,k)$.

Étape 9 : $P_{x1x2}(f,k) = \lambda_x P_{x1x2}(f,k-1) + (1-\lambda_x)X_1(f,k).X_2*(f,k)$.

Étape 10 : $R_{x1x2}(f,k) = \lambda_b R_{x1x2}(f,k-1) + (1-\lambda_b)X_1(f,k).X_2*(f,k)$.

Étape 11 : Estimation de $R_{x1x2min}(f,k)$ en utilisant l'équation 5.21.

Étape 12 : $\hat{P}_{b1b2}(f,k) = B_{min} R_{x1x2min}(f,k)$.

Étape 13 : $H_{SIS} = \frac{|P_{x1x2}(f,k)| - \left|\hat{P}_{b1b2}(f,k)\right|}{\sqrt{P_{x1x1}(f,k)P_{x2x2}(f,k)}}$.

Étape 14 : $S_i(f,k) = H_{SIS}(f,k)X_i(f,k), \quad i = 1,2$.

5 Evaluation et résultats

Les performances du filtre de la soustraction interspectrale sont directement liées au performances de l'estimateur de la dip des bruits $P_{b1b2}(f,k)$. Nons proposons ici comparer les performances de trois approches d'estimation de cette dip des bruits. La première approche est basée sur l'utilisation d'une détection d'activité vocale dont l'algorithme a été proposé par Guérin (Guérin, 2002) et décrit dans le paragraphe 4.1. La deuxième approche considérée est basée sur la technique des statistiques minimales où la dip des bruit est déterminée par suivi des minima locaux de la dip des signaux bruités suivant l'approche SM. Cette approche a été proposée par Rahmani (Rahmani *et al.*, 2009b) et décrite dans le paragraphe 4.2.4. La troisième approche est aussi basée sur la technique des statistiques minimales et les minima locaux ont été estimés en se basant sur l'approche SMA proposée. Le principe de cette approche a été décrit dans le paragraphe 4.2.5.
Une étude comparative préliminaire des performances de ces trois approches d'estimation de la dip des bruits a été tout d'abord considérée. Les performances du filtre de la SIS H_{SIS} ont été ensuite évaluées d'une manière objective puis subjective en considérant à chaque fois une des méthodes d'estimation de la dip des bruits déjà mentionnées.

5.1 Etude comparative préliminaire

La figure 5.6 montre les variations fréquentielles de la dip moyenne (sur toutes les trames) des bruits. Cette dip est estimée dans le cas d'un signal de parole (le mot rogne) échantillonné à une fréquence égale à 16 kHz et en présence de trois sources de bruit interférentes de type Babble. Le RSB est fixé à 0 dB. La dip réelle des bruits est illustrée en ligne bleue. La dip des bruits estimée avec la technique de détection d'activité vocale (DAV) est représentée en ligne noire. Les dip des bruits estimées suivant la technique des statistiques minimales selon les approches SM et SMA pour le suivi des minima sont représentées respectivement en lignes verte et rouge. Dans le cas de l'approche SM, la largeur de la fenêtre de recherche est fixée à $L = 30$ trames (0.48 s).

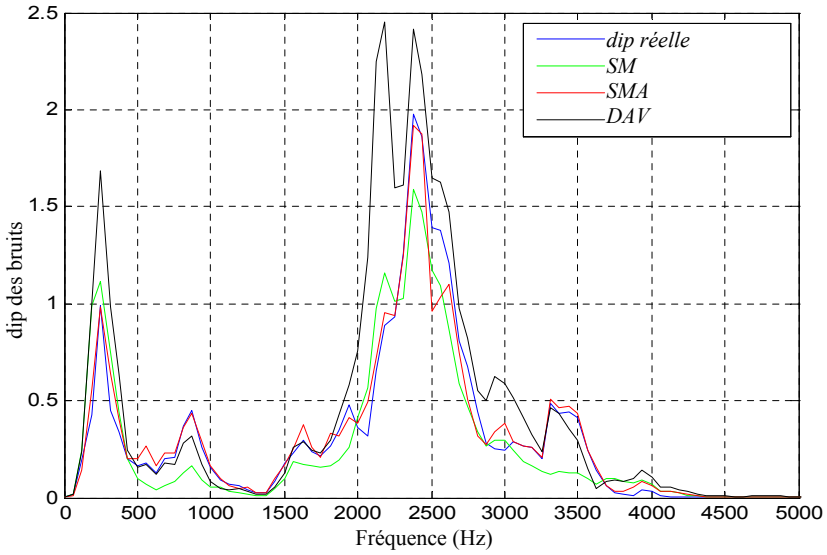


FIGURE 5.6 – dip réelle et estimées des bruits en utilisant les approches DAV, SM et SMA

A partir de cette figure, nous remarquons que l'approche proposée (SMA) pour le suivi des minima et l'estimation de la dip des bruits présente de meilleures performances par rapport aux deux autres approches DAV et SM. Ces observations sont particulièrement plus importantes au niveau de la bande fréquentielle [1500 Hz-3500 Hz]. En effet, comparée aux dip estimées avec les méthodes SM et DAV, la dip des bruits estimée avec la méthode SMA est plus proche de la dip réelle.
Les trois approches d'estimation de la dip des bruits ont été implémentées sur une machine équipée d'un processeur de type AMD à une fréquence d'horloge de 2.4Ghz et une mémoire vive RAM de 1 Go. L'implémentation a été faite en utilisant le logiciel Matab. Les performances de ces trois approches ont été comparées en termes de temps de traitement. Le tableau 5.1 illustre les résultats pour chacune des trois méthodes considérées. Dans le cas de la méthode SM, des fenêtres de recherche à différentes largeurs sont considérées : (L=10, 20, 30, 40, 50 et 60 trames).

TABLE 5.1 – *Temps de traitement pour l'estimation de la dip des bruits avec les méthodes DAV, SM et SMA*

Méthode	Largeur de la fenêtre de recherche (trames)	Temps de traitement (s)
	10	1.038
	20	1.35
SM	30	1.55
	40	1.82
	50	2.06
	60	2.32
DAV	***	1.22
SMA	***	1.13

On remarque à partir de ce tableau que le temps nécessaire pour l'estimation de la dip des bruits avec la méthode SM est proportionnel à la largeur de la fenêtre de recherche choisie. En plus, un temps de traitement minimum est obtenu lorsque la méthode SMA est considérée.

5.2 Evaluation de l'intelligibilité

5.2.1 Matériel phonétique

L'ensemble des vingt listes phonétiques de Lafon (Lafon, 1964) ont été utilisées dans cette étude expérimntale. Le niveau du signal sonore a été calibré à 70 dB SPL. Le signal de bruit considéré était un bruit additif de type Babble. Les signaux de parole bruités ont été enregistrés à différents RSB variant de -3 dB à 6 dB avec un pas de 3 dB. Les différentes manipulations d'enregistrement ont été effectuées dans une chambre sourde du service ORL de l'hôpital Edouard-Herriot. Le protocole expérimental est donnée par la figure 5.7.

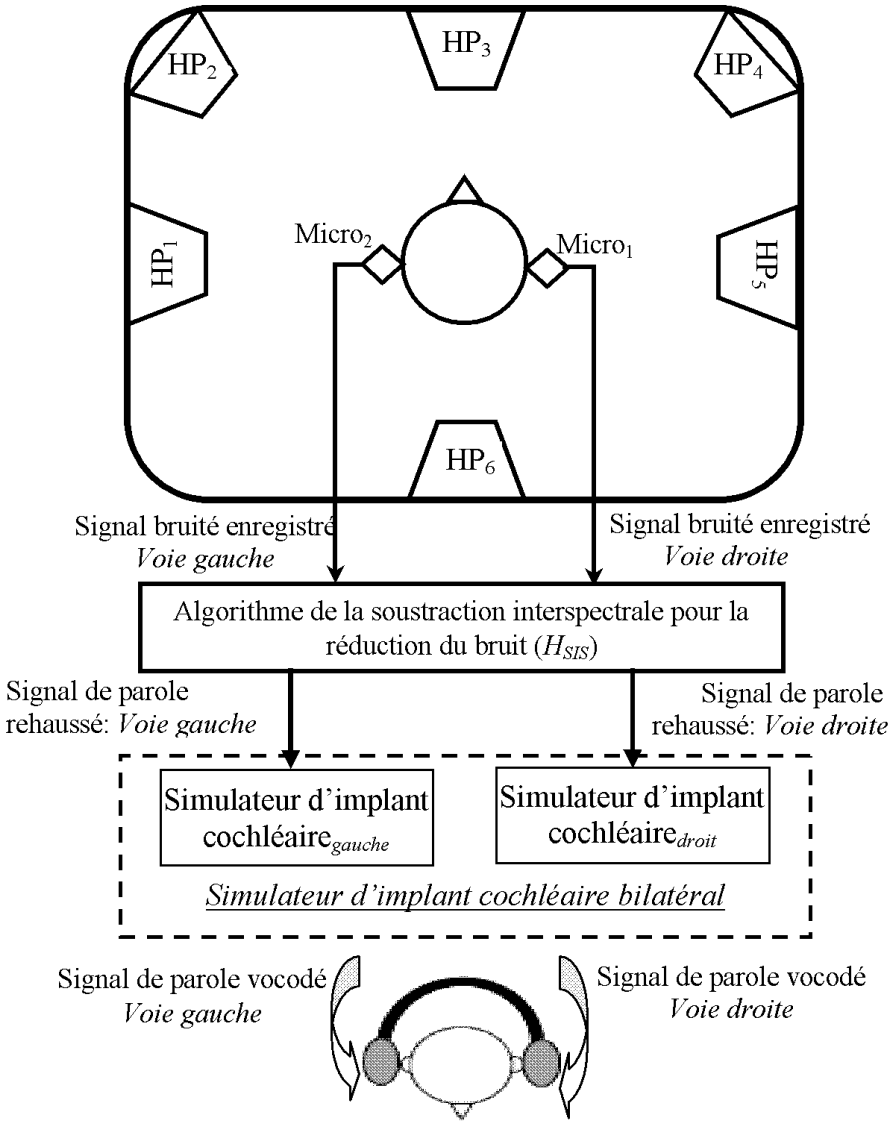


FIGURE 5.7 – Protocole expérimental

Cette chambre est équipé de six haut-parleurs (HP) répartis sur toute la surface. Un lecteur CD (PHILIPS-CD723) a été connecté à un audiomètre (MADSEN-Orbiter 922) pour le calibrage et l'ajustement des niveaux des signaux de parole et du bruit. Ces deux appareillages ne sont pas schématisés dans la figure 5.7. Un KEMAR équipée de deux microphones omnidirectionnels a été utilisée pour l'enregistrement des différentes pistes sonores. Ce KEMAR a été placé au milieu de la salle audiométrique. Le signal de parole brut a été émis par un haut-parleur placé devant le KEMAR (0°) et correspondait au haut parleur HP_3. Deux configurations de sources de bruit ont été considérées dans cette expérience. Dans la première configuration, une seule source de bruit placée à droite du KEMAR à un angle de 90° a été considérée. Dans ce cas, le haut-parleur HP_5 a été utilisé. Dans la deuxième configuration, trois sources de bruit placées asymétriquement à droite et à gauche du KEMAR ont été considérées. Dans ce cas, deux sources de bruit ont été placées à droite du KEMAR à des angles de 90° et 60° correspondant respectivement aux

deux haut-parleurs HP_4 et HP_4 ; alors que la troisième source a été placée à gauche du KEMAR à un angle de $-60°$ et correspondait au haut-parleur HP_2.

5.2.2 Evaluation objective

Les performances de l'algorithme de la soustraction interspectrale ont été comparées d'une manière objective dans le cas de chacun des trois estimateurs de la dip des bruits considérés (DAV, SM et SMA) par rapport au signal de parole bruité. Pour chacun des signaux de parole bruité et rehaussés, un score PESQ a été calculé suivant le standard ITU-T Recommendation P. 862 (P.862., 2001) (voir annexe 1). Les résultats sont donnés dans le cas de trois niveaux du RSB (-3 dB, 0 dB, 3 dB et 6 dB) en présence d'une seule source de bruit et de trois sources de bruit. Le tableau 5.2 illustre les différents résultats obtenus.

TABLE 5.2 – *Scores PESQ des signaux bruité et rehaussés*

	PESQ							
Configuration	**Une seule source de bruit**				**Trois sources de bruit**			
RSB	-3 dB	0 dB	3 dB	6 dB	-3 dB	0 dB	3 dB	6 dB
Signal bruité	1.23	1.62	1.98	2.08	1.13	1.36	1.52	1.94
DAV	1.32	1.84	2.26	2.71	1.06	1.70	2.09	2.31
SM	1.86	2.13	2.44	2.62	1.52	1.96	2.28	2.48
SMA	2.04	2.31	2.58	2.79	1.77	2.04	2.40	2.62

A partir des résultats donnés par le tableau 5.2, nous remarquons que les scores PESQ obtenus dans le cas des estimateurs de la dip des bruits basés sur la technique des statistiques minimales (SM et SMA) sont nettement supérieurs à ceux obtenus avec la technique de détection d'activité vocale. Ceci peut être expliqué par le fait que l'estimation de la dip des bruits suivant la méthode DAV est faite seulement pendant les périodes de silence (absence de parole) ; alors que dans le cas de la méthode des statistiques minimales, cette dip est estimée en présence et en absence de la parole. Il est à noter aussi que les scores PESQ obtenus dans le cas de la méthode SMA sont supérieurs à ceux obtenus avec la méthode SM. En outre, les performances se dégradent en environnements fortement bruité et en présence de plusieurs sources de bruits.

5.2.3 Evaluation subjective

Dans ce paragraphe, nous proposons une étudie des performances de l'algorithme de la soustraction interspectrale dans le cas de l'implant cochléaire bilatéral. Des tests subjectifs basés sur le calcul des pourcentages de reconnaissance des phonèmes ont été menés dans le cas d'une stimulation bilatérale en simulation avec des sujets normoentendants (voir figure 5.7).

5.2.3.a Sujets

Les performances de l'algorithme de débruitage bi-voie basé sur la technique de la soustraction interspectrale ont été évaluées avec une population de 50 sujets normoentendants dont leur âge variait entre 20 et 32 ans. Tous les participants ont été testés individuellement. Tous les sujets avait la langue française comme langue maternelle. Tous les participants étaient volontaires et ont été dédommagés pour leur participation. La session durait environ 50 minutes au total.

5.2.3.b Procédure

Les tests d'écoute ont eu lieu dans une chambre sourde du centre d'implantation cochléaire de l'hôpital Edouard Herriot de Lyon en utilisant un ordinateur connecté à un lecteur CD (PHILIPS-CD723). Un audiomètre numérique (MADSEN-Orbiter 922) a été utilisé pour le calibrage et l'ajustement des intensités des différents signaux sonores. Les stimuli ont été présentés d'une manière bilatérale à chaque sujet moyennant un casque 'Sennheiser HD250' à un niveau sonore confortable calibré à 70 dB SPL.
Avant la phase de test, l'expérience débutait par une session d'entraînement qui n'a pas été prise en compte parmi les résultats. Après cette session préliminaire, chacun des cinquante sujets a été testé dans les différentes conditions expérimentales. Durant cette session d'écoute, il a été demandé de chaque sujet de répéter les mots qu'il a entendu. Un ensemble de 32 conditions expérimentales ont été considérées en totalité (4 méthodes de traitement x 4 RSBs x 2 configurations des sources de bruit). Une liste de 17 mots a été utilisée pour chacune des conditions expérimentales. Ces listes ont été choisies aléatoirement. Pour les 20 premières conditions, l'ensemble des 20 listes de Lafon ont été utilisées. Pour les 16 conditions expérimentales restantes, 16 listes de Lafon choisies aléatoirement ont été reconsidérées. Afin d'aider les sujets de se familiariser et de s'adapter progressivement aux différentes conditions de bruit, les différents tests d'écoute se déroulaient en commençant par les listes bruités au niveau du RSB le plus élevé (6 dB) et arrivant jusqu'au aux listes les plus fortement bruités (RSB$=-3$ dB). A la fin de chaque test expérimental, les réponses de chaque sujet ont été rassemblées et le pourcentage de reconnaissance des phonèmes correctement identifiés a été calculé.

5.2.3.c Résultats

Les performances de l'algorithme de réduction de bruit basé sur la technique de la soustraction interspectrale ont été évaluées dans le cas de l'implant cochléaire bilatéral en simulation. Trois estimateurs de la dip des bruits ont été considérés : détection d'activité vocale, statistique minimale et statistique minimale amélioré. Les performances de ces trois estimateurs ont été comparées d'une manière subjective en simulation avec des sujets normoentendants. Les résultats sont donnés sous forme de pourcentages de reconnaissance des phonèmes dans chaque condition expérimentale. Une moyenne des scores obtenus par les 50 sujets a été calculée dans chaque condition. Les résultats moyens sont représentés sous forme d'histogrammes avec des barres indiquant l'erreur standard pour différents niveaux du RSB et dans le cas de chacune des configurations des sources de bruit.

La figure 5.8 illustre les résultats obtenus en présence d'une seule source de bruit.

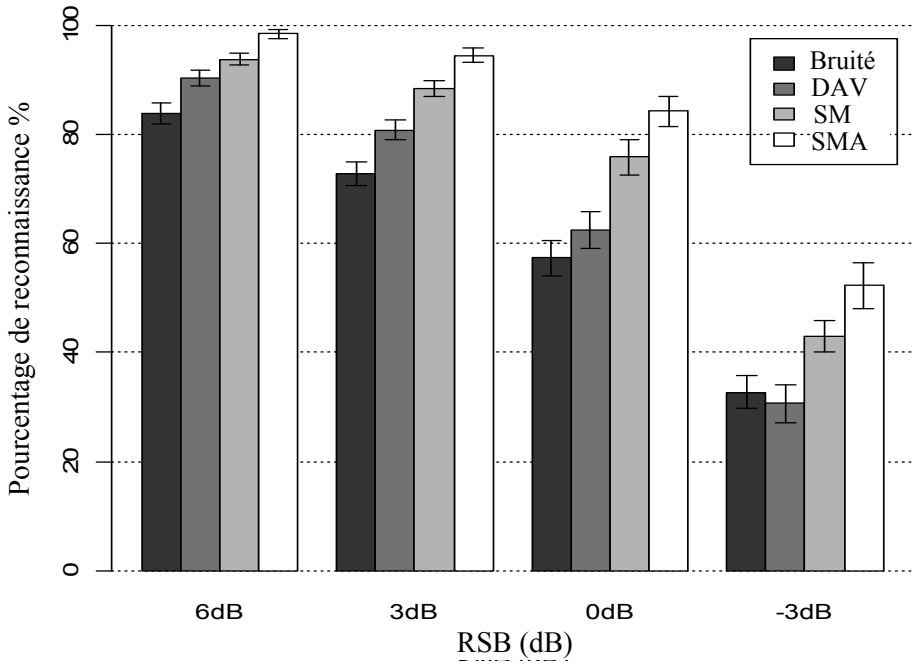


FIGURE 5.8 – Performances de l'algorithme de la SIS dans le cas de chacun des estimateurs de la dip des bruits considérés en présence d'une seule source de bruit

En présence de trois source de bruit (deuxième configuration), les résultats obtenus sont donnés par la figure 5.9.

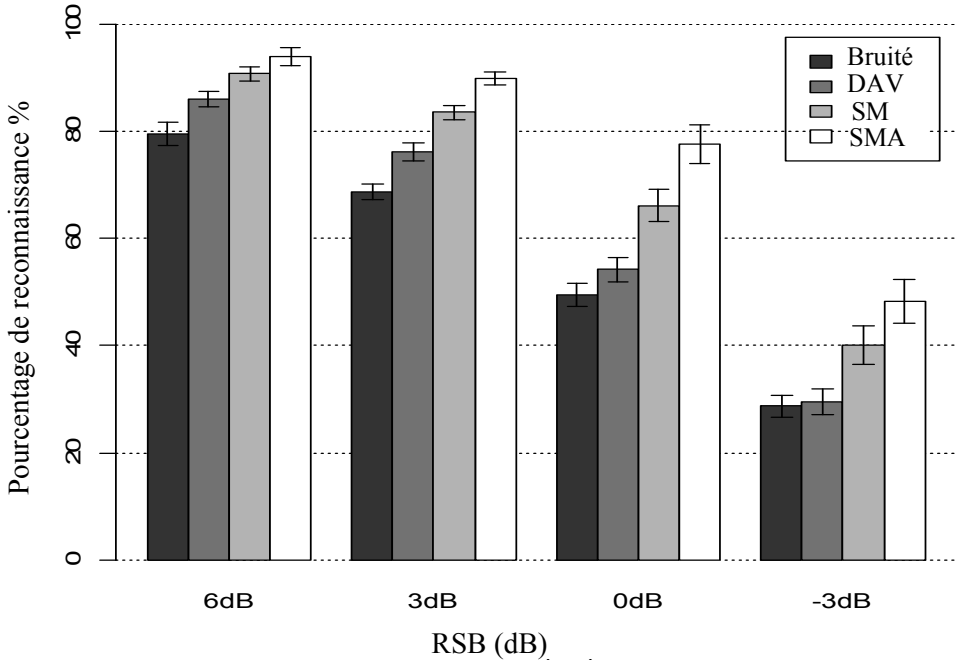


FIGURE 5.9 – Performances de l'algorithme de la SIS dans le cas de chacun des estimateurs de la dip des bruits considérés en présence de trois sources de bruit

Les résultats de cette étude expérimentale ont montré que l'algorithme de la SIS permet d'améliorer l'intelligibilité de la parole dans le cas de l'implant cochléaire bilatéral. Ces améliorations ont été variables suivant l'estimateur de la dip des bruits considérés (DAV, SM et SMA). En effet, par rapport aux résultats obtenus avec la technique de détection d'activité vocale, de meilleures performances ont été observées lors d'une estimation basée sur la technique des statistiques minimales (SM et SMA) avec une supériorité pour l'approche SMA. Il a été noté aussi que les pourcentages de reconnaissance diminuaient en fonctions du niveau du RSB et du nombre de sources de bruit. Une meilleure robustesse aux conditions expérimentales les plus difficiles a été observée lorsque l'approche SMA a été considérée pour l'estimation de la dip des bruits.
Afin d'étudier l'effet des différents facteurs (Algorithme, RSB et Configuration des sources de bruit) ainsi que les interactions entre eux sur l'intelligibilité des phonèmes dans la cas de l'implant cochléaire bilatéral, une analyse statistique a été considérée. Cette analyse a été basée sur des tests chi2 avec les paramètres suivants :

- Mesures répétées : Chacun des 50 sujets participaient à l'ensemble des 32 conditions considérées.

- Variable dépendante : le pourcentage de reconnaissance de phonèmes.

- Trois facteurs :
  - Algorithme (Bruité, DAV, SM, SMA)
  - Configuration des sources de bruit ('90°' et '−60°60°90°')
  - RSBs (6 dB, 3 dB 0 dB et -3 dB)

Ce type de modèle à effets mixtes est assez utilisé pour l'analyse des données (Baayen, 2008). Dans notre modèle, nous avons utilisé le programme lmer du package lme4 (Bates, 2007) implémenté sous le logiciel R (R, 2007).

Les analyses statistiques ont montré un effet significatif des facteurs algorithme (Chi2[3] =1244, p<0.001), RSB (Chi2[3]=6461, p<0.001) et configuration des sources de bruit (Chi2[1]=121, p<0.001). En plus, un effet d'interaction significatif entre les facteurs algorithme et RSB a été observé (Chi2[9]=77, p<0.001). Par contre, l'interaction n'a pas été significative entre les facteurs algorithme et configuration des sources de bruit (Chi2[3]=8, p=0.03) et entre RSB et configuration des sources de bruit (Chi2[3]=8, p=0.2).
Afin de mieux étudier la significativité de nos résultats et chercher les paires (Algorithme, RSB) pour lesquelles les différences étaient significatives, des tests de comparaison multiples ont été établis. Le logiciel R dispose de la fonction glht du package multcomp pour réaliser le test HSD de Tukey. Comparés aux pourcentages de reconnaissances obtenus avec les signaux bruités, les résultats ont montré que les performances ont été significativement améliorées lorsque la dip des bruits a été estimée en se basant sur l'algorithme DAV pour les niveaux du RSB 6 dB, 3 dB et 0 dB (p<0.001) ; par contre les améliorations étaient non significatives à un RSB=−3 dB (p=0.08). Dans le cas des estimateurs basés sur la technique des statistiques minimales, les deux algorithmes SM et SMA présentaient des améliorations significatives des pourcentages de reconnaissances par rapport à celles obtenues avec la méthode de détection d'activité vocale pour tous les niveaux du RSB

($p<0.001$). En plus, les résultats ont montré que les pourcentages de reconnaissance obtenus avec l'algorithme SMA étaient significativement meilleurs que ceux obtenus avec la méthode SM pour les niveaux du RSB égaux à 3 dB, 0 dB et −3 dB ($p<0.001$). Ces améliorations étaient non significatives pour le RSB=6 dB ($p=0.142$).

5.3 Discussion

Dans ce chapitre, l'algorithme de réduction du bruit basé sur la technique de la soustraction interspectrale a été considéré dans le cas de l'implant cochléaire bilatéral. L'estimation de la densité interspectrale des bruits est une composante prépondérante au calcul du filtre de la soustraction interspectrale. Différents estimateurs ont été proposés dans la littérature. Trois estimateurs, basés sur la technique de détection d'activité vocale proposé par Guérin (Guérin, 2002), la technique des statistiques minimales (approche SM) proposée par Rahmani et al. (Rahmani *et al.*, 2009b) et l'approche des statistiques minimales améliorées proposée par Kallel et al. (Kallel *et al.*, 2012b), ont été considérés dans ce travail. Une évaluation subjective des performances de ces algorithmes a été réalisée dans le cas de l'implant cochléaire bilatéral.

5.3.1 Signaux rehaussés vs signaux bruités

Il est clair à partir des figures 5.8 et 5.9 que l'algorithme de la soustraction interspectrale permette d'améliorer l'intelligibilité de la parole dans le cas de l'implant cochléaire bilatéral. En effet, comparés aux pourcentages de reconnaissance moyens des phonèmes obtenus avec les signaux bruités, des améliorations de 5.3%, 12.1% et 23% ont été observées avec les signaux rehaussés lorsque la dip des bruits est estimée respectivement avec les techniques DAV, SM et SMA. Ces résultats subjectifs confirment les mesures objectives obtenues et montrent la supériorité de l'algorithme proposé pour l'estimation de la dip des bruits.

5.3.2 Algorithme de réduction du bruit vs RSB

L'effet de l'algorithme de la soustraction interspectrale était variable suivant la technique d'estimation de la dip des bruits considérée pour les différentes valeurs du RSB considérées. En effet, comparées aux pourcentages de reconnaissance moyens calculés en environnement bruité aux différents niveaux du RSB, des améliorations moyennes de 5.2%, 13.6% et 20.1% ont été obtenues respectivement lorsque les algorithmes DAV, SM et SMA ont été considérés pour l'estimation de la dip des bruits. Ceci montre la supériorité des performances de l'algorithme SMA proposé. En plus, nous avons constaté que, dans le cas d'une estimation de la dip des bruits basée sur l'approche de détection d'activité vocale, les améliorations étaient pratiquement constantes pour les niveaux du RSB positifs et diminuaient pour la valeur du RSB négative. Dans le cas des algorithmes SM et SMA, les améliorations étaient maximales à 3 dB et 0 dB, mais légèrement moins importantes à 6 dB et −3 dB.

5.3.3 Algorithme de réduction du bruit vs configuration des sources de bruit

Les résultats ont montré que la configuration des sources de bruits avaient un effet sur l'intelligibilité de la parole dans le cas de l'implant cochléaire bilatéral : Les pourcentages de reconnaissance des phonèmes en présence d'une seule source de bruit (première configuration) étaient significativement dégradés en présence de trois sources de bruit (deuxième configuration). En effet, en présence d'une seule source de bruit, des améliorations moyennes des pourcentages de reconnaissance de 6%, 13.5% et 24.6% ont été observées respectivement aves les algorithmes DAV, SM et SMA. En présence de trois sources de bruit, les améliorations étaient seulement de l'ordre de 4.5%, 10.7% et 21.2% respectivement avec les algorithmes DAV, SM et SMA.

5.3.4 Consistance des résultats

Différents travaux de recherche (Monica et Ruth, 2004; Laszig *et al.*, 2004) ont montré que la stimulation bilatérale permette d'améliorer l'intelligibilité de la parole en milieux silencieux et bruité par rapport à la stimulation unilatérale. Dans le troisième chapitre, nous avons montré que la stimulation bilatérale décalée permette d'améliorer l'intelligibilité de la parole par rapport à la stimulation bilatérale symétrique en milieux silencieux et bruité. Afin d'améliorer davantage l'intelligibilité de la parole chez les patients implantés en milieu bruité, différents algorithmes de réduction de bruit utilisant un, deux ou plusieurs microphones ont été proposés dans le cas de l'implant cochléaire (Loizou *et al.*, 2005; Hu *et al.*, 2007; Spriet *et al.*, 2007). Dans le cas de l'implant cochléaire bilatéral, la plupart des algorithmes proposés ont été basés sur la technique du beamformer adaptatif à deux signaux d'entrée (deux microphones) (Chung *et al.*, 2006) ou à quatre signaux d'entrée (Kokkinakis et Loizou, 2010). Les résultats ont montré des améliorations des pourcentages de reconnaissance de la parole variant de 10% à 40%.
Dans notre travail de recherche, sous l'hypothèse des bruits diffus, les performances de l'algorithme de la soustraction interspectrale ont été évaluées dans le cas de l'implant cochléaire bilatéral. Les performances de cet algorithme dépendent énormément de l'approche considérée pour l'estimation de la dip des bruits. Un nouveau algorithme basé sur la technique des statistiques minimales a été proposé pour le suivi des minima locaux de la dip des signaux bruités et par la suite l'estimation de la dip des bruits. Les performances de cette technique proposée ont été comparées à une autre approche basée sur la technique des statistiques minimales proposée par Rahmani et al. (Rahmani *et al.*, 2009b) et une autre approche basée sur la détection d'activité vocale. Les résultats ont montré de meilleures performances lorsque l'algorithme proposé a été considéré.
Le tableau 5.3 présente une étude comparative des performances de l'algorithme de la soustraction interspectrale avec l'estimateur de la dip des bruits SMA et l'algorithme de la soustraction spectrale multi-bande bi-voie (SSMB-BV) présenté dans le chapitre précédant. Les résultats sont donnés pour différents niveaux du RSB et dans le cas de deux configurations des sources de bruit. Comparé aux résultats obtenus avec l'algorithme SSMB-BV, nous avons remarqué que l'algorithme de la soustraction interspectrale présente de meilleures performances. Ce résultat est évident vu que l'algorithme de la soustraction spectrale bi-voie est développé sous l'hypothèse des bruits parfaitement décorrélés (dip des bruits nulle).

TABLE 5.3 – *Comparaison des performances des algorithmes de la soustraction interspectrale et la soustraction spectrale*

Configuiration	**Une seule source de bruit**				**Trois sources de bruit**			
$RSB(dB)$	6	3	0	-3	6	3	0	-3
H_{SIS} (%)	15.6	22.15	27.2	21.6	13.7	18.5	26.55	15.05
$SSMB-BV$ (%)	13	18	18	12	11	14	15	12

Nous présentons enfin une étude comparative des performances des trois approches proposées dans ce travail et qui visent l'amélioration de l'intelligibilité de la parole dans le cas de l'implant cochléaire bilatéral. On rappelle ici que la première approche est basée sur la technique de la stimulation bilatérale décalée (SBD), la seconde est basée sur l'algorithme de la soustraction spectrale bi-voie (particulièrement l'algorithme de la soustraction spectrale multi-bande bi-voie (SSMB-BV)) et l'algorithme de la soustraction interspectrale (SIS) avec l'estimateur SMA pour l'estimation de la dip des bruits. Les améliorations sont toujours déterminées par rapport aux scores calculés dans le cas de la stimulation bilatérale symètrique à l'environnement bruité considéré. Les résultats sont donnés pour différents niveaux du RSB tel que illustré par la figure 5.10.

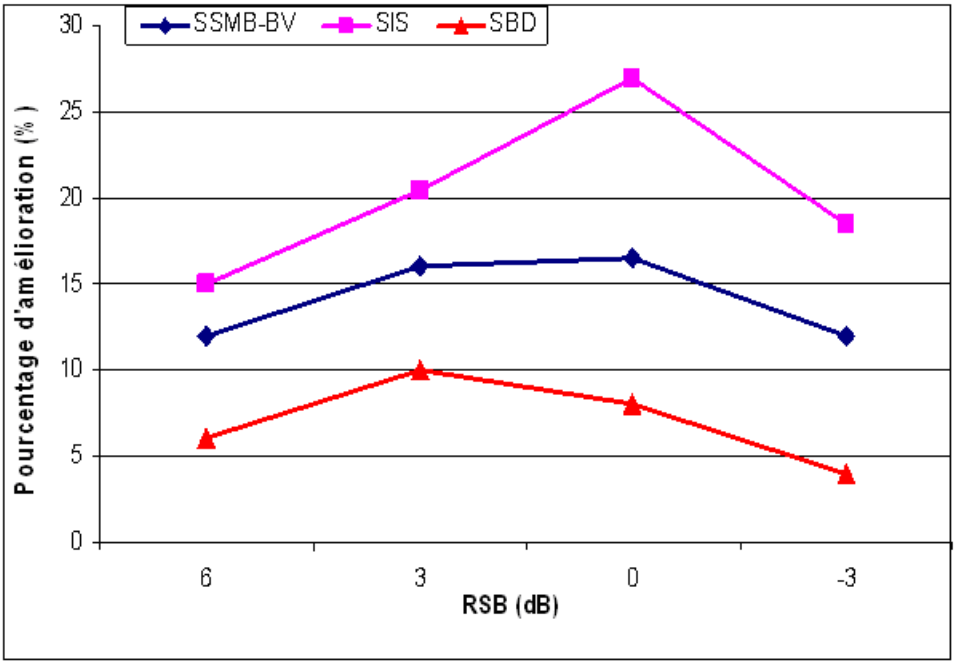


FIGURE 5.10 – Comparaison des performances des approches SBD, SSMB-BV et SIS

Nous remarquons que les améliorations sont variables en fonction du RSB. L'approche SBD permettait des améliorations entre 4% et 10%. L'approche SSMB-BV présentait des améliorations variables entre 10% et 17% approximativement. De meilleures performances ont été obtenues dans le cas de l'algorithme SIS où les améliorations sont entre 15% et 27%.

6 Conclusion

Dans ce chapitre, nous avons considéré le problème de réduction de bruit dans le cas de l'implant cochléaire bilatéral. Les études préliminaires ont montré que dans ce contexte, les signaux de bruits et de parole sont caractérisés par une distribution spatiale sensiblement différente. En effet, les signaux de parole issues d'une source ponctuelle sont corrélés, alors que les bruits sont globalement décorrélés. Ceci nous a amené à considérer la fonction de cohérence comme base de notre système de débruitage. Cette fonction permet de discriminer les signaux corrélés des signaux décorrélés, ce qui correspond approximativement à notre situation. D'autre part, les bruits rencontrés peuvent être considérés comme issus d'une multitude de sources réparties dans l'espace, et leur distribution considérée comme diffuse. Cette observation justifie la considération de l'algorithme de réduction de bruit basé sur l'algorithme de la soustraction interspectrale.
Au cours de ces développements, les analyses ont montré que l'estimation des densités spectrale et interspectrale de puissance, nécessaire au calcul du filtre de soustraction interspectrale, est prépondérante. Particulièrement, une estimation fiable de la dip des bruits permet d'améliorer de manière importante les qualités du filtrage. Une étude comparative des performances de trois estimateurs de la dip des bruits a été présentée. Le premier algorithme est basé sur la technique de détection d'activité vocale (DAV), le deuxième est basé sur la technique des statistiques minimales (SM), alors que le troisième est basé sur une version améliorée de la technique des statistiques minimales (SMA).
Les signaux de parole bruités à différents niveaux du RSB et dans le cas des deux configurations des sources de bruit ont été traités par le filtre de la soustraction interspectrale calculé pour chacun des estimateurs de la dip des bruits. Une évaluation objective basée sur la comparaison des scores PESQ calculés pour les différents signaux rehaussés a montré de meilleures performances pour les estimateurs de la dip des bruits basés sur la technqiue des statistique minimales par rapport à la technqiue du DAV. Une supériorité pour l'approche SMA par rapport à l'approche SM a été noté.
En termes de résultats aussi, une campagne rigoureuse de tests subjectifs est venue compléter les tests objectifs menés sur des signaux de parole bruités à différent niveaux du RSB et dans le cas de deux configuration des sources de bruit. Ils ont montré clairement que la soustraction interspectrale présentait des résultats particulièrement intéressants, notamment en termes de contrôle à la fois des bruits résiduels, mais aussi de la distorsion du signal utile. En outre, les performances de cet algorithme sont nettement supérieures à celles obtenues avec l'algorithme de la soustraction multi-bande bi-voie présentée dans le chapitre précédant, surtout en milieu fortement bruité.

Conclusion Générale

La prothèse cochléaire est un appareillage biomédical, permettant la stimulation directe des neurones cochléaires. La pose d'électrodes dans des zones bien définies dans la cochlée permet de stimuler sélectivement les cellules sensorielles suivant différentes fréquences et différentes intensités électriques, qui sont générées suite à un traitement spécifique du signal sonore. Suite à la rééducation du patient implanté, la stimulation électrique apportée par les électrodes permet un niveau de compréhension très intéressant tout en notant une adaptation progressive du patient à son appareillage. La stimulation peut être unilatérale dans le cas où une seule oreille est stimulée ou bilatérale dans le cas où deux oreilles sont simultanément stimulées
Les travaux présentés dans cette thèse ont traité le problème de l'intelligibilité de la parole chez les patients bilatéralement implantés en milieux silencieux et bruité. En effet, différents travaux de recherche ont montré que la stimulation bilatérale présente des performances supérieures à la stimulation unilatérale, particulièrement en milieux faiblement bruités. Ces performances se dégradent en milieux moyennement et fortement bruités. Nous avons de ce fait étudié les performances de trois approches de traitement du signal en vue d'une amélioration de l'intelligibilité de la parole dans le cas de l'implant cochléaire bilatéral : la stimulation bilatérale décalée, l'algorithme de la soustraction spectrale bi-voie et l'algorithme de la soustraction interspectrale.
Après avoir rapidement présenté le mécanisme de l'audition humaine et quelques outils de réhabilitation à la surdité, nous avons introduit la notion de débruitage et de réduction de bruit. En effet, les algorithmes de débruitage peuvent être classés en deux catégories : les algorithmes mono-voie et les algorithmes bi-voie. Pour les algorithmes mono-voie, une seule observation bruitée issue d'un seul microphone est exploité pour la réduction du bruit. Alors que dans le cas des algorithmes bi-voie, deux observations bruitées issues de deux microphones sont utilisés pour le rehaussement des signaux bruités. Dans ce travail, nous nous sommes particulièrement intéressés aux algorithmes bi-voie dans le cas de l'implant cochléaire bilatéral.

Dans la suite, nous résumons les travaux réalisés ainsi que les résultats obtenus.

Dans une première partie, nous avons comparé (en simulation) les performances de trois modes de stimulation cochléaire dans les cas de deux vitesses d'analyse et à différents niveaux du RSB. Le premier mode a été basé sur la stimulation cochléaire unilatérale. Le deuxième mode a été basé sur la stimulation bilatérale symétrique où les deux oreilles ont été stimulées d'une manière identique avec les mêmes signaux. Le troisième mode a été basé sur la stimulation bilatérale décalée où les deux oreilles ont été stimulées par des signaux différents (un décalage fréquentiel au niveau des filtres d'analyse du signal de

parole a été considéré). Les résultats ont montré que la stimulation bilatérale permette d'améliorer significativement l'intelligibilité de la parole par rapport à la stimulation unilatérale. En plus, de meilleurs résultats ont été observés dans le cas de la stimulation bilatérale décalée. Les améliorations étaient entre 4% et 10%.

Dans la deuxième partie, les performances de deux nouveaux algorithmes de réduction du bruit bi-voie dédiés pour implant cochléaire bilatéral ont été présentés. En effet, un algorithme bi-voie pour l'estimation de la densité spectrale de puissance (dsp) du bruit a été considéré. Cet algorithme a été basé sur le calcul des dsp et de la densité interspectrale (dip) des signaux bruités et il a été développé sous l'hypothèse de la décorrélation des bruits. Les algorithmes basés sur les techniques de la soustraction spectrale non linéaire et la soustraction spectrale multi-bande étendus au cas bi-voie ont été utilisés pour l'estimation des signaux rehaussés au niveau de chaque voie. Les performances de ces algorithmes ont été évaluées et comparées dans le cas de l'implant bilatéral, en simulation dans un premier temps avec des sujets normoentendants, puis avec des sujets portant l'implant cochléaire Digisonic SP binaural de Neurelec. Cette étude comparative a été établie à différents niveaux du RSB et en présence d'une seule source de bruit puis de trois sources de bruit de type Babble. Les résultats ont montré que les algorithmes de débruitage développés permette d'améliorer l'intelligibilité de la parole chez les sujets bilatéralement implantés avec une supériorité pour l'algorithme basée sur la technique de la soustraction spectrale multi-bande. Néanmoins, ces performances se dégradaient en présence de plusieurs sources de bruit et pour les faibles niveaux du RSB.

Dans la dernière partie, un deuxième algorithme de réduction de bruit dédié pour implant cochléaire bilatéral a été considéré. Cet algorithme a été basé sur la méthode de la soustraction interspectrale et il a été développé sous l'hypothèse des bruits diffus. L'exploitation de l'algorithme de la soustraction interspectrale reposait principalement sur l'estimation de la dip des bruits. Deux approches pour l'estimation de la dip des bruits basées sur la technique de détection d'activité vocale et la technique des statistiques minimales ont été considérées. Dans ce travail, une nouvelle approche basée aussi sur la technique des statistiques minimales a été proposée pour l'estimation de la dip des bruits. Les performances de l'algorithme de la soustraction interspectrale ont été comparées dans le cadre de l'implant cochléaire bilatéral en simulation dans le cas de chacun des estimateurs de la dip des bruits considérés. Cette étude comparative a été réalisée à différents niveaux du RSB et en présence d'une seule source de bruit puis de trois sources de bruit de type Babble. Les résultats ont montré la supériorité de l'approche proposée pour l'estimation de la dip des bruits. En plus, de meilleures performances ont été notées pour l'algorithme de la soustraction interspectrale par rapport à l'algorithme de la soustraction spectrale bi-voie précédemment présenté.

En conclusion, les deux nouveaux algorithmes bi-voie considérés pour la réduction de bruit dans le cadre de l'implant cochléaire bilatéral ont été validés et peuvent être implémentés au niveau des processeurs de traitement du signal dédiés pour l'implant cochléaire. Nous notons ici que ces algorithmes nécessitent une communication entre les deux processeurs et un échange des données captées par les deux microphones installés au niveau de chaque oreille.

Nous pouvons donc conclure ce travail de thèse en affirmant que de nombreuses voies restent à explorer. Pour poursuivre ces travaux, les pistes sont nombreuses :

- Tester les algorithmes de débruitage validés en simulation avec des patients bilatéralement implantés.

- Etudier les performances des algorithmes proposés avec d'autres types de bruits et dans le cas d'autres configurations des sources de bruit.

- Exploiter d'autres techniques de débruitage et de compensation du bruit bi-voie en vue d'une meilleure amélioration de l'intelligibilité de la parole dans le cas de l'implant cochléaire bilatéral.

- Etudier les performances des algorithmes de débruitage proposés en termes de localisation spatiale des sources de parole et de bruit (comparaison des différences de temps interaurale entre les signaux de parole bruité et rehaussé). Ceci a pour objectif d'exploiter ces algorithmes dans le cas de l'implant cochléaire binaural.

Annexe

Critères objectifs de qualité

Dans le contexte de la réduction de bruit, l'évaluation des performances représente une réelle problématique : en effet, c'est l'oreille humaine qui est le juge final de la qualité du système. De nombreuses mesures existent qui tentent de représenter le plus finement possible l'évaluation subjective. Les critères objectifs les plus communément utilisées peuvent être classés principalement en trois catégories :

- Mesures dans le domaine temporel : RSB, RSB segmental,
- Mesures dans le domaine fréquentiel : Mesure d'Itakura Saito (IS), Distance cepstrale,
- Mesures dans le domaine perceptuel : La mesure BSD (Bark Spectral Distortion), Le MBSD (Modified Bark Spectral Distortion), Le PSQM (Perceptual Speech Quality Measure), Le PESQ (Perceptual Evaluation of Speech Quality).

Nous retiendrons dans cette étude uniquement une mesure opérant dans le domaine fréquenciel qui est la mesure d'Itakura Saito (IS) et une deuxième mesure opérant dans le domaine perceptuel qui est le score PESQ. Nous présentons plus en détails ces deux mesures ci-après.

1. *Mesure d'Itakura Saito*

La mesure d'Itakura Saito repose sur l'analyse LPC (Quackenbush *et al.*, 1988). Son expression fait intervenir le modèle tout pôle du signal de référence (s) et celui du signal testé (y). Soient P(f) et $\hat{P}(f)$ les densités spectrales de puissance du modèle AR du signal de référence et du signal de test respectivement. La distance d'Itakura Saito est donnée par :

$$d_{IS}(P(f), \hat{P}(f)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{P(f)}{\hat{P}(f)} - \log \frac{P(f)}{\hat{P}(f)} - 1 \right) \mathrm{d}f \tag{5.22}$$

2. *Perceptual Evaluation of Speech Quality*

Le PESQ (Perceptual Evaluation of Speech Quality) est l'évaluation de la qualité vocale perçue désignée dans la norme *P*.862 (P.862., 2001) comme moyen adapté aux codecs vocaux et aux mesures de bout en bout. De ce fait, d'autres facteurs supplémentaires sont pris en considération pour mieux simuler les conditions réelles, à savoir le temps de propagation, les distorsions dûes aux erreurs de transmission... D'une manière schématique, cette mesure est basée sur un calcul de distance perceptuelle (différence audible entre la représentation perceptuelle du signal de référence et celle du signal de test) suivie d'un modèle cognitif qui permet de prendre en compte le fait qu'une dégradation n'a pas le même impact selon qu'elle est additive ou soustractive, ou selon son contexte (segment de parole ou non) et sa distribution (localisée ou non). La note d'évaluation PESQ finale est une combinaison linéaire de la valeur de perturbation moyenne et de la valeur de perturbation asymétrique moyenne. Les scores PESQ sont compris entre 0.5 pour les pires des cas et 4.5 pour le meilleur des cas.

Bibliographie

Azirani, A. A., 1995. *Rehaussement de la Parole en ambiance Bruit ée. Application aux Télécommunications Mains-Libres.* Thèse de doctorat, Université de Rennes1, Renne, France.

Baayen, R. H., 2008. *Analyzing Linguistic Data : A Practical Introduction to Statistics using R.* Cambridge University Press, 1 édition. ISBN 0521709180. URL `http ://www.worldcat.org/isbn/0521709180`.

Balkany, T., A. Hodges, C. Menapace, L. Hazard et Driscoll, 2007. Nucleus freedom north american clinical trial. *Otolaryngol Head Neck Surg*, 136 :757–762.

Bates, D., 2007. Linear mixed model implementation in lme4. http ://cran.r-project.org/web/packages/lme4/vignettes/Implementation.pdf.

Beaugeant, C., 1999. *Réduction de Bruit et Contrôle d'Echo pour les Applications Radio-mobiles.* Thèse de doctorat, Université de Rennes I, Renne, France.

Berouti, M., R. Schwartz et J. Makhoul, 1979. Enhancement of speech corrupted by acoustic noise. Dans *IEEE Int. Conf. Acoustics, Speech, and Signal Process.*

Boll, S. F., 1979. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27 :113–120.

Boothroyd, A., L. Hanin et T. Hnath, 1985. A sentence test of speech perception : Reliability, set equivalence and short-term learning. Rapport technique, New York : City University of New York.

Bouquin-Jeannés, R. L., 1991. *Traitements pour la Réduction du Bruit sur la Parole, Applications aux Communications Radio-Mobiles.* Thèse de doctorat, Université de Rennes 1, Rennes, France.

Bouquin-Jeannés, R. L., A. A. Azirani et G. Faucon, 1997. Enhancement of speech degraded by coherent and incoherent noise using a cross-spectral estimator. *IEEE Trans.Speech Audio Processing*, 5 :484–487.

Cappé, O., 1994. Elimination of the musical noise phenomenon with the ephraim and malah noise suppressor. *IEEE Trans. on speech and audio processing*, 2 :345–349.

Chung, K., F. G. Zeng et K. N. Acker, 2006. Effects of directional microphone and adaptive multichannel noise reduction algorithm on cochlear implant performance. *J. Acoust. Soc. Am.*, 120 :2216–2227.

Cochlear Corporation, 2002. Ace and cis dsp strategies. Software Requirements Specification N95287F Issue 1. Lane Cove, New South Wales, Australia.

Compernolle, D. V. et S. V. Gerven, 1995. *Applications of difital signal processing to Telecommunications*, chapitre Beamforming with microphone arrays, pages 107–131. COST 229.

Dillon, H., 2001a. *Advanced signal processing schemes in hearing aids*, chapitre 7, pages 187–208. Thieme.

Dillon, H., 2001b. Compression systems in hearing aids. *Hearing Aids*, 6 :159–186.

Doblinger, G., 1995. Computationally efficient speech enhancement by spectral minima tracking in subbands. Dans *Proc. Eurospeech*, pages 1513–1516.

Dorman, M., P. C.Loizou et R. Rainey, 1997. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *J Acoust Soc Am*, 102 :2403–2411.

Dörbecker, M. et S. Ernst, 1996. Combination of two-channel spectral subtraction and adaptive wiener post-filtering for noise reduction and dereverbcration. Dans *European Signal Processing Conference*.

Dunn, C. C., R. Tyler, S. Oakley, B. Gantz et W. Noble, 2008. Comparison of speech recognition and localization performance in bilateral and unilateral cochlear implant users matched on duration of deafness and age at implantation. *Ear & Hearing*, 29 :352–359.

Dunn, C. C., R. S. Tyler, S. Oakley, B. J. Gantz et W. Noble, 2010. Comparison of speech recognition and localization performance in bilateral and unilateral cochlear implant users matched on duration of deafness and age at implantation. *Ear & Hearing*, 29 :352–359.

Eddington, D. K., W. R. Rabinowitz, J. Tierney, V. Noel et M. Whearty, 1997. Speech processors for auditory prostheses. Rapport technique, National Institutes of Health, Bethesda.

Ephraim, Y. et D. Malah, 1984. Speech enhancement using a minimum mean square error short-time spectral amplitude estimator. *IEEE Trans. on ASSP*, 32 :1109–1121.

Ephraim, Y. et D. Malah, 1985. Speech enhancement using a minimum mean square log-spectral amplitude estimator. *IEEE Trans. on ASSP*, 33 :443–445.

Ephraim, Y. et H. L. V. Trees, 1995. A signal subspace approach for speech enhancement. *IEEE Trans. Speech Audio Process.*, 3 :251–266.

Fant, G., 1973. *Speech Sounds and Features.* Cambridge, MIT Press.

Farsi, H., 2010. Improvement of minimum tracking in minimum statistics noise estimation method. *Signal Processing : An International Journal*, 4 :17–22.

Fetterman, B. et E. Domico, 2002. Speech recognition in background noise of cochlear implant patients. *Otolaryngol. Head Neck Surg.*, 126 :257–263.

Friesen, M. L., V. R. Shannon, D. Baskent et X. Wang, 2001. Speech recognition in noise as a function of the number of spectral channels : Comparison of acoustic hearing and cochlear implants. *J. Acoust. Soc. Am*, 110(2) :1150–1163.

Fu, Q. J. et R. V. Shannon, 1999. Phoneme recognition by cochlear implant users as a function of signal-to-noise ratio and nonlinear amplitude mapping. *J. Acoust. Soc. Am*, 106 :L18–L23.

Fu, Q. J., R. V. Shannon et X. Wang, 1998. Effects of noise and spectral resolution on vowel and consonant recognition : Acoustic and electric hearing. *J. Acoust. Soc. Am*, 104 :3586–3596.

Galvin, J. J. et Q. J. Fu, 2009. Influence of stimulation rate and loudness growth on modulation detection and intensity discrimination in cochlear implant users. *Hearing Research*, 250 :46–54.

Ghorbel, M., 2007. *Etude et conception d'un stimulateur cochléaire programmable multi-électrodes*. Thèse de doctorat, Université de Sfax (ENIS), Sfax, Tunisie.

Guérin, A., 2002. *Rehaussement de la parole pour les communications mains-libres. Réduction de bruit et annulation d'écho non linéaire*. Thèse de doctorat, Université de Rennes 1, Rennes, France.

Guérin, A., R. L. Bouquin-Jeannés et G. Faucon, 2003. A two-sensor noise reduction system : Applications for hands-free car kit. *EURASIP Journal on Applied Signal Processing*, 11 :1125–1134.

Hartmann, W. M., 1997. Signals, sound and sensation. *American institute of Physics*, 10 :251–252.

Hillenbrand, J., L. Getty, M. Clark et K. Wheeler, 1995. Acoustic characteristics of american english vowels. *J. Acoust. Soc. Am*, 97 :3099–3111.

Hu, Y. et P. Loizou, 2002. A subspace approach for enhancing speech corrupted with colored noise. *IEEE Signal Process. Lett.*, 9 :204–206.

Hu, Y., P. C. Loizou, N. Li et K. Kasturi, 2007. Use of a sigmoidal-shaped function for noise attenuation in cochlear implants. *J Acoust Soc Am.*, 122 :EL128–134.

Kallel, F., M. Frikha, M. Ghorbel, A. B. Hamida et C. Berger-Vachon, 2012a. Dual-channel spectral subtraction algorithms based speech enhancement dedicated to a bilateral cochlear implant. *Applied Acoustics*, 73 :12–20.

Kallel, F., M. Ghorbel, M. Frikha, C. Berger-Vachon et A. B. Hamida, 2012b. A noise cross psd estimator based on improved minimum statistics method for two-microphone speech enhancement dedicated to a bilateral cochlear implant. *Applied Acoustics*, 73 :256–264.

Kamath, S. et P. C. Loizou, 2002. A multi-band spectral subtraction method for enhancing speech corrupted by colored noise. *IEEE International Conference ICASSP*, pages 4160–4164.

Kasturi, K. et P. C. Loizou, 2007. Use of s-shaped input-output functions for noise suppression in cochlear implants. *Ear & Hearing*, 28 :402–411.

Kokkinakis, K. et P. C. Loizou, 2010. Multi-microphone adaptive noise reduction strategies for coordinated stimulation in bilateral cochlear implant devices. *J. Acoust. Soc. Am*, 127 :3136–3144.

Komal, A., D. Pam, R. Richard et V. Andrew, 2009. Electrical stimulation rate effects on speech perception in cochlear implants. *International Journal of Audiology*, 48 :561–567.

Kuhn, G. F., 1977. Model for the interaural time differences in the azimuthal plane. *J. Acoust. Soc. Am*, 62(1) :157–167.

Lafon, J. C., 1964. *Le test phonétique et la mesure de l'audition.* Eindhoven.

Laszig, R., A. Aschendorff et M. Stecker, 2004. Benefits of bilateral electrical stimulation with the nucleus cochlear implant in adults : 6-month postoperative results. *Otology and Neuro-otology*, 25 :958–968.

Li, S., J. Q. Wang et X. J. Jing, 2010. The application of nonlinear spectral subtraction method on millimeter wave conducted speech enhancement. *Mathematical Problems in Engineering*, 2010 :1–12.

Lobo, A. P., F. Toledos, P. Loizou et M. F. M. F. Dorman, 2002. The effect of envelope cutoff frequency on consonant and melody recognition by ci listeners. *J. Acoust. Soc. Am*, 112 :2245.

Loizou, P. et G. Kim, 2011. Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions. *IEEE Trans. Audio, Speech, Language Processing*, 19 :47–56.

Loizou, P. C., Y. Hu, R. Litovsky, G. Yu, R. Peters, J. Lake et P. Roland, 2009. Speech recognition by bilateral cochlear implant users in a cocktail-party setting. *J. Acoust. Soc. Am.*, 125 :372–383.

Loizou, P. C., A. Lobo et Y. Hu, 2005. Subspace algorithms for noise reduction in cochlear implants. *J. Acoust. Soc. Am.*, 118 :2791–2793.

Loizou, P. C., O. Poroy et M. F. Dorman, 2000. The effect of parametric variations of cochlear implant processors on speech understanding. *J. Acoust. Soc. Am*, 108 :790–802.

Luo, F. L., J. Yang, C. Pavlovic et A. Nehorai, 2002. Adaptive null-forming scheme in digital hearing aids. *IEEE Transactions on Signal Processing*, 50 :1583–1590.

Macaulay, R. J. et M. L. Malpass, 1980. Speech enhancement using a softdecision noise suppression filter. *IEEE Trans. Acoust., Speech, Signal Process.*, 28 :137–145.

Martin, R., 1994. Spectral subtraction based on minimum statistics. Dans *Proc EUSIPCO*, pages 1182–1185. Edinburgh, Scotland.

Martin, R., 2001. Noise power spectral density estimation based on optimal smoothing and minimum statistics. *IEEE Transactions on Speech and Audio Processing*, 9 :504–512.

Martin, R., 2006. Bias compensation methods for minimum statistics noise power spectral density estimation. *Signal Processing*, 86 :1215–1229.

McAulay, R. J. et M. L. Malpass, 1980. Speech enhancement using a soft-decision noise suppression filter . *IEEE Trans. on ASSP*, 28 :137–145.

Mezalek, S. T., 1990. *Algorithmes de Débruitages de la Parole pour les Radio-Mobiles*. Thèse de doctorat, Université de Rennes 1, Rennes, France.

Müller, J., F. Schon et J. Helms, 2002. Speech understanding in quiet and noise in bilateral users of the med-el combi 40/40 cochlear implant system. *Ear & Hearing*, 23 :198–206.

Monica, L. H. et Y. L. Ruth, 2004. The benefit of binaural hearing in a cocktail party : Effect of location and type of interferer. *J. Acoust. Soc. Am.*, 115 :833–843.

Nie, K., A. Barco et F. G. Zeng, 2006. Spectral and temporal cues in cochlear implant speech perception. *Ear & Hearing*, 27 :208–17.

Nilsson, M., S. Soli et J. Sullivan, 1994. Development of hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *J. Acoust. Soc. Am.*, 95 :1085–1099.

P.862., I. R., 2001. Perceptual evaluation of speech quality (PESQ) : an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs.

Plant, K., L. Holden, M. Skinner, J. Arcaroli et L. Whitford, 2007. Clinical evaluation of higher stimulation rates in the nucleus research platform 8 system. *Ear & Hearing*, 28 :381–393.

Quackenbush, S., T. Barnwell et M. Clements, 1988. Objective measures for speech quality testing. Dans *Prentice-Hall*.

R, 2007. Development Core Team, R : A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. online. URL `http ://www.R-project.org`.

Rahmani, M., A. Akbari et B. Ayad, 2009a. An iterative noise cross-psd estimation for two-microphone speech enhancement. *Applied Acoustics*, 70 :514–521.

Rahmani, M., A. Akbari, B. Ayad et N. Derakhshan, 2009b. A noise cross psd estimator for dual-microphone speech enhancement based on minimum statistics. *Journal of Zhejiang University SCIENCE A*, 10 :805–809.

Rahmani, M., A. Akbari, B. Ayada et B.Lithgowb, 2009c. Noise cross psd estimation using phase information in diffuse noise field. *Signal Processing*, 89 :703–709.

Rangachari, S. et P. C. Loizou, 2006. A noise estimation algorithm for highly non-stationary environments. *Speech Communication*, 48 :220–231.

Rayleigh, L., 1907. On our perception of sound direction. *Philosophical Magazine*, 13 :214–232.

Ricketts, T. et P. Henry, 2002. Evaluation of an adaptive, directionalmicrophone hearing aid. *International Journal of Audiology*, 41 :100–112.

Ricketts, T. A., D. W. Grantham, H. D. Ashmead, D. H. Haynes et R. F. Labadie, 2006. Speech recognition for unilateral and bilateral cochlear implant modes in the presence of uncorrelated noise sources. *Ear & Hearing*, 27 :763–773.

Rosen, S., 1992. Temporal information in speech : Acoustic, auditory and linguistic aspects. *Phil Trans R Soc*, 336 :367–373.

Shannon, R., F. G. Zeng, V. Kamath, J. Wygonski et M. Ekelid, 1995. Speech recognition with primarily temporal cues. *Science*, 270 :303–304.

Skinner, M., L. Holden et T. Holden, 1995. Effect of frequency boundary assignment on speech recognition with the speak speech-coding strategy. *Annals of Otology, Rhinology and Laryngology*, 104(Suppl. 166) :307–311.

Sohn, J., N. S. Kim et W. Sung, 1999. A statistical model-based voice activity detection. *IEEE Signal Process. Lett*, 6 :1–3.

Sovka, P., P. Pollak et J. Kybic, 1996. Extended spectral subtraction. Dans *European signal processing conference.*

Spriet, A., 2004. *Adaptive filtering techniques for noise reduction and acoustic feedback cancellation in hearing aids.* Thèse de doctorat, K.U.Leuven, Belgique.

Spriet, A., L. V. Deun, K. Eftaxiadis, J. Laneau, M. Moonen, B. V. Dijk, A. V. Wieringen et J. Wouters, 2007. Speech understanding in background noise with the two-microphone adaptive beamformer beam in the nucleus freedom cochlear implant system. *Ear & Hearing*, 28 :62–72.

Tranmüller, H., 1990. Analytical expression for the tonotopic sensory scale. *J. Acoust. Soc. Am*, 85 :97–100.

Udrea, R. M., S. D. Ciochina et N. Vizireanu, 2005. Multi-band bark scale spectral over-subtraction for colored noise reduction. Dans *Proceedings of ISSCS.*

Vandali, A. E., L. A. Whitford, K. L. Plant et G. M. Clark, 2000. Speech perception as a function of electrical stimulation rate : using the nucleus 24 cochlear implant system. *Ear & Hearing*, 21 :608–624.

VanHoesel, R. J., 2004. Exploring the benefits of bilateral cochlear implants. *Audiol Neurootol*, 9 :234–246.

Vary, P., 1985. Noise suppression by spectral magnitude estimation-mechanism and theoretical limits. *Signal Processing*, 8 :387–400.

Veekmans, K., L. Ressel, J. Mueller, M. Vischer et S. J. Brockmeier, 2009. Comparison of music perception in bilateral and unilateral cochlear implant users and normal-hearing subjects. *Audiol Neurotol*, 14 :315–326.

Veen, B. D. V. et K. M. Buckley, 1988. Beamforming : A versatile approach to spatial filtering. *IEEE ASSP Magazine*, 5 :4–24.

Verschuur, C., 2005. Effect of stimulation rate on speech perception in adult users of the med-el cis speech processing strategy. *Int J Audiol*, 44 :58–63.

Virag, N., 1999. Single channel speech enhancement based on masking properties of the human auditory system. *IEEE Trans. Speech and Audio Processing*, 7 :126–137.

Wang, D. L. et J. S. Lim, 1982. The unimportance of phase in speech enhancement. *IEEE Trans. on ASSP*, 30 :679–681.

Weber, B. P., W. K. Lai, N. Dillier, E. L. V. Wallenberg et M. J. Killian, 2007. Performance and preference for ace stimulation rates obtained with nucleus rp 8 and freedom system. *Ear & Hearing*, 28 :465–485.

Welch, P. D., 1967. The use of fast fourier transform for the estimation of power spectra : A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio Electroacoustics*, 15 :70–73.

Whitmal, N. A., S. F. Poissant, R. L. Freyman et K. S. Helfer, 2007. Speech intelligibility in cochlear implant simulations : Effects of carrier type, interfering noise, and subject experience. *J. Acoust. Soc. Am.*, 122 :2376–2388.

Wightman, F. L. et D. J. Kistler, 1992. The dominant role of lowfrequency interaural time differences in sound localization. *Journal of the Acoustical Society of America*, 91(3) :1648–1661.

Won, J. H., W. R. Drennan et J. Rubinstein, 2007. Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. *J Assoc Res Otolaryngol.*, 8 :384–392.

Yang, L. P. et Q. J. Fu, 2005. Spectral subtraction-based speech enhancement for cochlear implant patients in background noise. *J. Acoust. Soc. Am.*, 117 :1001–1004.

Zeng, F. G. et R. R. Fay, 2004. Cochlear implants : auditory prostheses and electric hearing. *Springer*, 8 :334–376.

Zwicker, E. et H. Fastl, 1999. *Psychoacoustics : Facts and Models (Springer Series in Information Sciences) (v. 22)*. Springer, 2nd updated ed. édition.

Liste des Publications

REVUES INTERNATIONALES

- **Kallel, F.**, M. Frikha, M. Ghorbel, A. B. Hamida et C. Berger-Vachon, 2012. Dual-channel spectral subtraction algorithms based speech enhancement dedicated to a bilateral cochlear implant. *Applied Acoustics*, 73 :12-20.

- **Kallel, F.**, M. Ghorbel, M. Frikha, C. Berger-Vachon et A. B. Hamida, 2012. A noise cross psd estimator based on improved minimum statistics method for two-microphone speech enhancement dedicated to a bilateral cochlear implant. *Applied Acoustics*, 73 :256-264.

- A. Derbel, **Kallel, F.**, M. Samet et A. B. Hamida, 2008. Bionic wavelet transform based on speech processing dedicated to a fully programmable stimulation strategy for cochlear prostheses. *Asian Journal of Scientific Research*, 1 :293-309.

CONFERENCES INTERNATIONALES

- **Kallel, F.**, A. Jeanvoine, A. B. Hamida et C. Berger-Vachon, 2012. Etude de l'effet du mode de stimulation sur l'intelligibilité de la parole en milieu silencieux et en milieu bruité. *Handicap 2010*, Porte des Versailles, Paris-France, Juin 2010.

- **Kallel, F.**, D. Daoud, M. Ghorbel, et A. B. Hamida, 2009. Comparaison des différents algorithmes de débruitage du signal de parole pour les aides auditives binaurales. 5th *International Conference : Sciences of Electronic, Technologies of Information and Telecommunications*, Hammamet-Tunisie, March 2009.

- D. Daoud, **Kallel, F.**, M. Ghorbel, et A. B. Hamida, 2009. Spatial filtering based speech enhancement for binaural hearing aid. 6th *International Multi-Conference on Systems, Signals and Devices*, Djerba-Tunisie, March 2009.

Algorithmes de Réduction du Bruit en Vue d'une Amélioration de l'Intelligibilité de la Parole: Cas de la Prothèse Cochléaire

Fathi KALLEL

Résumé : La prothèse cochléaire est un appareillage destiné à la réhabilitation des surdités profondes et totales dont un appareillage conventionnel est inefficace. Elle assure la stimulation directe des neurones cochléaires à travers un faisceau d'électrodes. Différents travaux de recherches ont été établis afin d'évaluer l'intelligibilité de la parole chez les sujets bilatéralement implantés en environnements silencieux et bruité. Les résultats ont montré une bonne intelligibilité de la parole en milieu silencieux. Toutefois, les capacités de perception de la parole chez les patients implantés se dégradent en environnement bruité. Nous avons de ce fait proposé trois approches de traitement du signal en vue d'une amélioration de l'intelligibilité de la parole dans le cas de l'implant cochléaire bilatéral: la stimulation bilatérale décalée, l'algorithme de la soustraction spectrale bi-voie et l'algorithme de la soustraction interspectrale. Des améliorations de l'intelligibilité de la parole entre 4% et 10% ont été notées dans le cas de la stimulation bilatérale décalée par rapport à la stimulation bilatérale symétrique. L'approche basée sur l'algorithme de la soustraction spectrale bi-voie présentait des améliorations variables entre 10% et 17%. De meilleures performances ont été obtenues lorsque l'approche basée sur l'algorithme de la soustraction interspectrale est considérée où les améliorations étaient entre 15% et 27%.

Mots clés: Prothèse cochléaire, Intelligibilité de la parole, Décalage fréquentiel, Soustraction spectral bi-voie, Soustraction interspectrale.

Abstract: Cochlear prostheses are intended for persons suffering from deep or total deafness where conventional prostheses proved ineffective. In quiet listening conditions, most bilateral cochlear implant (BCI) users can now achieve even more than 80% word recognition scores regardless the used device. However, under more challenging listening conditions, BCI recipients perform poorly, compared to normal-hearing listeners. In this work, we proposed three speech processing approaches for speech intelligibility improvement. The first is based on shifted bilateral cochlear implant stimulation; the second is based on dual-channel spectral subtraction algorithm and finally the cross power spectral subtraction algorithm was considered. Experimental results showed a speech intelligibility improvement between 4% and 10% when the shifted bilateral cochlear implant stimulation is considered. Performance amelioration was observed when the dual-channel spectral subtraction based speech enhancement algorithm was considered and the improvement was between 10% and 17%. The better performance was obtained when noisy speech signals were processed using cross power spectral subtraction algorithm and the improvement was between 15% and 27%.

Key-words: Cochlear prostheses, Speech intelligibility, Frequency shift, Dual-channel spectral subtraction, Cross power spectral subtraction.