N° 338-2012 Année 2012

thèse de l'université de Lyon

délivrée par

l'université claude bernard lyon 1

École doctorale E.E.A.

diplôme de doctorat
(Arrêté du 07 août 2006)
mention : Traitement du signal

Soutenue publiquement le
17 Décembre 2012
par

Mr. Arnaud JEANVOINE

Intérêt des algorithmes de réduction de bruit dans l'implant cochléaire : Application à la binauralité.

Composition du jury :

Directeur de thèse :	Pr. Christian BERGER-VACHON	Lyon 1
Co-directeur :	Pr. Hung THAI-VAN	Lyon 1
Rapporteur	Pr. Paul AVAN	Clermont-Ferrand
Rapporteur	Pr. Jaime LOPEZ-KRAHE	Paris 8
	Pr. Eric TRUY	Lyon 1
	Pr. Bruno FRACHET	Paris 13
	Dr. Dan GNANSIA	Neurelec

Citations

« C'est avec l'intuition qu'on trouve et les maths qu'on prouve ! »

Henri Poincaré,
Mathématicien Français
(1854-1912)

Phrases Cultes :

« C'est "identique" ... aux différences près ! »
« Ya qu'a ... faut qu'on ! »
« Courage !!! Fuyons !!! »
« Cela va sans dire ..... Mais encore mieux en le disant ! »
« Tout est simlpe ..... Quand on a compris ! »

Christian Berger-Vachon,
Professeur des Universités,
Directeur de thèse.

Remerciements

Le professeur Christian BERGER-VACHON, cela fait déjà quelques années que j'ai la chance de le connaitre. Notre première rencontre fut lors d'un cours de statistique de L3 GBM. Je lui avais alors demandé si je pouvais faire mes stages dans le laboratoire, il m'avait répondu « Oui » et depuis, l'histoire s'est poursuivie. Je le remercie pour la confiance qu'il m'a accordée durant toutes ces années.

Le professeur Hung THAI-VAN, je suis arrivé au laboratoire en même temps que lui. La première présentation que j'avais faite en réunion devant l'ensemble du laboratoire, était la présentation d'un beamformer...pour laquelle j'avais fait une animation qu'il avait particulièrement appréciée. Je le remercie de faire partie de mon jury

Le professeur Paul AVAN, je le remercie d'avoir accepté de juger mon travail, c'est un honneur de le compter dans ce jury.

Le professeur Jaime LOPEZ-KRAHE, je le remercie d'avoir accepté de juger mon travail, c'est un honneur de le compter dans ce jury et de saluer son action dans la lutte contre le handicap.

Le professeur Eric TRUY, sans qui aucune bourse portant sur les implants cochléaires et aucune validation clinique ne pourraient exister dans ce laboratoire. Je le remercie d'avoir eu confiance en moi ainsi que de son soutien.

Le professeur Bruno FRACHET, je le remercie d'avoir accepté de participer au jury de cette thèse.

Mr. Dan GNANSIA, qui m'a soutenu durant ces trois années de thèse, il m'a donné d'excellentes idées, de très bons conseils et m'a appris à

comprendre un article pour créer les algorithmes derrières. Je le remercie également pour la correction des articles, sa rapidité de réponse ainsi que pour son humour et sa bonne humeur.

Toute l'équipe de NEURELEC, pour m'avoir fourni une bourse CIRFRE et fait confiance pour mon travail de thèse. Je la remercie également pour la bonne humeur, l'ambiance de travail, les congrès, et plus particulièrement Maggie NATTA, David VERAN, Nicolas VEAU, Guillaume ROSANI et Cédric BRIAND. Je vous remercie tous.

Le professeur Emmanuel PERRIN, qui m'a fait confiance pour les TP d'acoustique, d'électronique et statistique que j'ai pu enseigner aux étudiants de Polytech Lyon, Pour les différents cours de programmation et les heures de discussion autour d'un café. Je le remercie.

Mes amis de GBM, Anoop qui m'a appris tout mon vocabulaire d'anglais, Thomas pour les nombreuses nuits passées à NHL en révisant les cours de statistiques, je vous remercie tous les deux.

Le 4^{ieme} étage de Puvis de Chavannes, Sophie, Virginie, Antoine, Cédric, Julien, Nicolas, qui ont toujours été présents lors des différents problèmes, des soirées crêpes ou tout simplement pour sortir un coup entre amis, je vous remercie.

Le train de 7h24 pour lyon, André, Florent, Guillaume, Hubert, Jacques, avec le défis de finir plus vite que Hubert les mots croisés du « 20 minutes » sans jamais réussir. Pour l'humour et la bonne ambiance je vous remercie.

Mes parents, frère et soeurs, Chantal et Denis, Elsa, Ophélie, et Simon, qui malgré la distance, ont toujours été là pour moi. Je vous remercie tous les cinq.

Ma femme et notre fils, qui sont tout pour moi, sans eux, je n'en serais pas là. Je vous remercie de croire en moi.

Résumé & Abstract

Résumé

Les implants cochléaires sont des appareilles destinés à la réhabilitation des surdités profondes et totales. Ils assurent la stimulation du nerf auditif en plaçant des électrodes dans la cochlée. Différentes études ont été établis afin d'améliorer l'intelligibilité de la parole dans le bruit chez le patient porteur de cet appareil. Les techniques bilatérales et binaurales permettent de reproduire une audition binaurale, car les deux oreilles sont simulées (comme pour les personnes normo-entendantes). Ainsi la localisation et la perception des sons environnants sont améliorées par rapport à une implantation monaurale. Toutefois, les capacités de reconnaissances des mots sont très vite limitées en présence de bruits. Nous avons développés des techniques de réduction de bruit afin d'augmenter les performances de reconnaissance. Des améliorations de 10% à 15% suivant les conditions ont été observées. Néanmoins, si la perception est améliorée par les algorithmes, ils focalisent sur une direction, et ainsi, la localisation est alors réduite à l'angle de localisation. Une seconde étude a alors été effectuée pour mesurer l'effet des algorithmes sur la localisation. Ainsi, le beamformer donne les meilleurs résultats de compréhension mais la moins bonne localisation. La ré-injection d'un pourcentage du signal d'entrée sur la sortie a permis de compenser la perte de la localisation sans détériorer l'intelligibilité.

Le résultat de ces deux expériences montre qu'il faut un compromis entre la perception et la localisation des sons pour obtenir les meilleures performances.

Mots clefs

Implant cochléaire binaural, Soustraction Spectrale, Filtre de Wiener, Beamformer, Ré-injection, localisation, intelligibilité

Abstract

Cochlear implants are to sail for the rehabilitation of deep and total deafness. They provide stimulation of the auditory nerve by placing electrodes in the cochlea. Various studies have been established to improve the speech intelligibility in noise in the patient of this device. Bilateral and binaural techniques allow reproducing a binaural hearing, since both ears are simulated (as for normal hearing people). Thus localization and the perception of the surrounding sounds are improved from a monaural implantation. However, the recognition of the words capabilities are limited very quickly in the presence of noise. We developed noise reduction techniques to increase the performance of recognition. Improvements of 10% to 15% depending on the conditions were observed. Nevertheless, if the perception is enhanced by algorithms, they focus on a direction, and thus the location is then reduced at the corner of localization. Then, a second study was made to measure the effect of localization algorithms. Thus, the beamformer gives the best results of understanding but the less good location. The re-injection of a percentage of the input to the output signal helped offset the loss of the location without damaging the intelligibility.

The result of these two experiments shows that it takes a compromise between perception and sound localization for best performance.

Key-words

Binaural cochlear implant, Spectral Subtraction, Winer filter, Beamforming, Re-injection, localization, speech intelligibility

Abréviations

Par ordre alphabétique :

CCE	:	Cellules ciliées externes
CCI	:	Cellules ciliées internes
dB	:	Décibel
DNI	:	Différence de Niveau Interaural
DSP	:	Densité Spectrale de Puissance
DTI	:	Différence de Temps Interaural
Fe	:	Fréquence d'échantillonnage
FFT	:	Fast Fourier Transform
iFFT	:	inverse Fast Fourier Transform
KEMAR	:	Knowles Electronic Manikin for Acoustic Research
MEQM	:	Minimum de l'Erreur Quadratique Moyenne
MMSE	:	Minimum Mean Square Method
MMSE-LSA	:	Minimum Mean Square Method - Log Spectral Amplitude
RSB	:	Rapport signal sur bruit
TF	:	Transformée de Fourier
TFD	:	Transformée de Fourier Discrète
TFR	:	Transformée de Fourier Rapide
TOP	:	Traitement Optimisé de la Parole

Table des matières

Citations i

Remerciements ii

Résumé & Abstract iv

Abréviations vi

Introduction 1
Composition de l'oreille . . . 1
Description de l'oreille externe . . . 2
Description de l'oreille moyenne . . . 2
Description de l'oreille interne . . . 3
Qu'est ce qu'un son ? . . . 5
Environnement du son . . . 5
Sensation de l'intensité . . . 7
Description de la parole . . . 8
Caractéristiques . . . 8
Fréquence fondamentale . . . 9
Formants . . . 9
Implant cochléaire . . . 10
Présentation du problème . . . 11
Plan du document . . . 12

1 Différents aspects des algorithmes de réduction de bruit. 13
Introduction . . . 13
1.1 Base de la Transformée de Fourier rapide . . . 14
1.2 Réduction du bruit pour les signaux mono-voie . . . 18
1.2.1 Estimation du bruit . . . 20
1.2.2 Soustraction spectrale . . . 22
1.2.3 Filtrage de Wiener . . . 27

1.2.4 Détection de la parole par le TOP 28
1.2.5 Réduction du bruit musical 29
1.3 Réduction du bruit pour les signaux multi-voies 30
Introduction 30
1.3.1 Choix de la direction de la localisation 32
1.3.2 Estimation du bruit stéréophonique 32
1.3.3 Technique du beamforming 33
1.3.4 Algorithme de Doerbecker 37
1.3.5 Récapitulatif des formules de gain de réduction du bruit dans le cas du binaural 43
1.3.6 Coefficient de ré-injection 43
1.4 Et sans les mathématiques ! 44
Conclusion 45

2 Implant Cochléaire, application à la binauralité et simulation. 46
Introduction 46
2.1 Aides auditives 47
2.2 Implant cochléaire 47
2.2.1 Introduction 47
2.2.2 Historique 48
2.2.3 Caractéristiques de l'implant cochléaire 48
2.2.4 Principe du fonctionnement de l'implant cochléaire . . 49
2.2.5 Descriptions des différentes stratégies de codage 53
2.3 Implant cochléaire bilatéral 56
2.4 Implant cochléaire binaural 56
2.4.1 Introduction à l'implant binaural 56
2.4.2 Traitement du signal effectué par l'implant bilatéral . . 58
2.5 Vocodeur de l'implant cochléaire 59
Conclusion 62

3 Reconnaissance de phonèmes après un débruitage binaural par l'intermédiaire d'un vocodeur. 63
Introduction 63
3.1 Matériel & Méthodes 66
3.1.1 Matériel acoustique 66
3.1.2 Enregistrement du signal 67
3.1.3 Algorithmes de traitement du signal 68
3.1.4 Vocodeur 70
3.1.5 Session 71
3.1.6 Analyses statistiques 72

3.2 Résultats 73
3.3 Discussion 77
Conclusion 78

4 Performance de localisation de source sonore après un débruitage binaural par l'intermédiaire d'un vocodeur. 80
Introduction 80
4.1 Matériel & Méthodes 81
4.1.1 Soustraction spectrale 82
4.1.2 Filtrage de Wiener 82
4.1.3 Ré-injection 82
4.1.4 Vocodeur 83
4.1.5 Sujets 83
4.1.6 Protocole 83
4.2 Résultats 86
4.2.1 Courbes de localisation 86
4.2.2 Statistiques 92
4.3 Discussion 99
Conclusion 99

Conclusion 101

Pour continuer 104

Annexes 106

Bibliographie 137

Table des figures

1 Physiologie de l'audition 1
2 Coupe de la cochlée 3
3 Tonotopie de l'oreille 4
4 Bruit VS Signal ! 6
5 Triangle formantique de Delattre 10

1.1 Différence entre les différents types de fenêtres 17
1.2 Modèle de débruitage mono-voie 19
1.3 Modèle de débruitage binaural 31
1.4 Synotptique du beamformer fixe 34
1.5 Synoptique de Doerbecker combiné avec Ephraim & Malah . . 39
1.6 Synoptique de Doerbecker combiné avec Scalart 41
1.7 Synoptique de Doerbecker, filtre de Wiener commun a Ephraim & Malah et Scalart 42

2.1 Représentation des parties externe et interne de l'implant cochléaire 49
2.2 Principe de fonctionnement de l'implant cochléaire 50
2.3 Implant binaural 57
2.4 Traitement du son dans l'implant binaural 59
2.5 Vocodeur simulant l'implant cochléaire 60
2.6 Répartition fréquentielle suivant l'échelle de bark 61

3.1 Illustration des informations binaurales. 65
3.2 Plan d'enregistrement des différentes pistes 67
3.3 Mr KEMAR 68
3.4 Fonction de transfert $H(f)$ de la pré-emphase 69
3.5 Récapitulatif de l'expérience 71
3.6 RSB $= -6dB$ 75
3.7 RSB $= 0dB$ 75
3.8 RSB $= +6dB$ 75
3.9 Pourcentage obtenus par rapport à la condition VOC 76

4.1 Récapitulatif de l'expérience 84
4.2 Logiciel de localisation 85
4.3 Courbe de localisation pour la condition « Normale » 86
4.4 Courbe de localisation pour la condition « Diago » de référence 87
4.5 Courbe de localisation pour la condition « Voc » 88
4.6 Courbe de localisation pour la condition « Beam » 89
4.7 Courbe de localisation pour la condition « Doem » 90
4.8 Courbe de localisation pour la condition « Dos » 91
4.9 Paramètre c_0 93
4.10 Paramètre c_1 94
4.11 Paramètre c_2 94
4.12 Paramètre c_3 95
4.13 Erreur de localisation 97
4.14 Erreur pour les différentes conditions en fonction de l'angle . . 98

Liste des tableaux

1 Correspondance entre niveau de pression acoustique et pression 8
2 Exemple de valeurs des formants pour différentes voyelles . . . 9

1.1 Récapitulatif des différentes méthodes de réduction de bruit . 43

2.1 Implants cochléaires commercialisés 55
2.2 Comparaison bilatéral VS binaural 58
2.3 Tableau des bancs de filtres pour la séparation des douze bandes de fréquences 61

3.1 Récapitulatif des différentes conditions 73
3.2 Résultats en fonction du RSB par condition 74

4.1 Récapitulatif des différentes conditions 85
4.2 Erreur par rapport à la diagonale 96

Introduction

L'ouie est un sens, tout comme les quatre autres, elle permet de percevoir et de comprendre ce qui nous entoure afin de mieux discerner notre environnement. Il conviendra de la préserver au maximum.

Composition de l'oreille.

L'Homme est composé de deux oreilles séparées d'une distance moyenne égale à $d = 20cm$. Elle peut se décomposer en trois parties que sont l'oreille externe, l'oreille moyenne et l'oreille interne [64], ces trois parties sont représentées dans la figure 1. Chaque partie à un rôle bien défini pour la compréhension des sons qui arrivent aux oreilles.

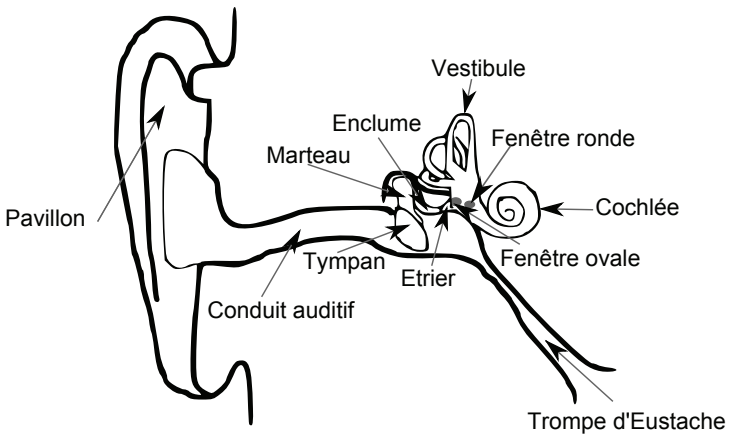


FIGURE 1 – Physiologie de l'audition

Description de l'oreille externe.

L'oreille externe permet de localiser les sons dans l'espace par l'intermédiaire de nos deux oreilles et du pavillon [62, 64]. Grâce à la distance qui les sépare, la perception des sons est en trois dimensions (elle joue le rôle d'antenne acoustique). Une fois le son capté, l'oreille externe joue alors un rôle de sélection (faible mais dépendant des fréquences) du signal en fonction de l'angle d'incidence. L'amplification des sons est faible pour les basses fréquences ($G < 10dB$) quel que soit l'angle de provenance du son. Pour les fréquences situées entre $2kHz$ et $5kHz$, l'angle de provenance joue un rôle déterminant, car l'amplification peut aller jusqu'à $20dB$ pour un angle azimutal de $\pm 45^o$ et retombe très vite pour arriver à une amplification de $10dB$ pour un angle de $\pm 90^o$.

Description de l'oreille moyenne.

Elle est située entre l'oreille externe et l'oreille interne. Elle part du tympan et se termine par la fenêtre ovale. Elle comprend toute la chaîne des osselets (marteau, enclume et étrier). La mastoïde et la trompe d'Eustache sont considérées comme des annexes mais en font parties. La trompe d'Eustache permet l'adaptation de l'impédance de l'air entre la partie externe du tympan et sa partie interne [62, 64]. Le rôle de l'oreille moyenne est surtout une amplification du signal transmis par l'oreille externe. L'amplification se fait de la même façon que pour un levier en mécanique. La seconde fonction est la protection pour l'oreille interne (réflexe stapédien pour un son $> 80dBSPL$). La troisième fonction est un rôle d'adaptation de l'impédance du milieu acoustique en un milieu liquidien. La dernière fonction est un filtrage passe haut pour compenser le filtrage passe bas de la parole en sortie de la bouche.

Le rapport des surfaces permet une amplification du signal acoustique (air) dans le milieu liquidien (la cochlée). Cette adaptation se fait par un passage mécanique, et grâce aux mouvements des osselets assimilés à un levier (une amplification d'environ $30dB$ est alors obtenue) [64].

Description de l'oreille interne.

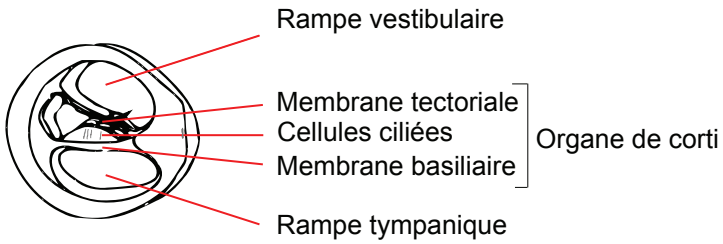


FIGURE 2 – Coupe de la cochlée

L'oreille interne regroupe deux organes distincts. Le vestibule (dont on ne parlera plus par la suite) qui est l'organe de l'équilibre est composé des canaux semi-circulaires, et la cochlée qui est l'organe « propre » de l'audition. En effet, sans cochlée, il est impossible au cerveau d'interpréter un son. Elle fait une rotation de deux tours et demie. C'est l'organe le plus important de l'oreille. La cochlée est reliée sur l'oreille moyenne par la fenêtre ovale. Elle se termine par la fenêtre ronde.

En prenant une coupe de la cochlée, celle-ci peut se décomposer en trois parties (figure 2). La partie supérieure est la rampe vestibulaire, la partie inférieure est la rampe tympanique qui sont toutes deux remplies par un liquide appelé la périlymphe. Ensuite, situé entre ces deux parties, l'organe de corti, qui est l'élément maître de la cochlée est rempli par l'endolymphe. Il contient les cellules ciliées internes et externes qui transforment la vibration du liquide en potentiel d'action sur le nerf auditif. Les rôles des cellules ciliées internes (CCI) et externes (CCE) ne sont pas les mêmes. En effet, les cellules ciliées externes (au nombre de trois rangées cce) ont un rôle d'amplification « des fréquences locales » alors que les cellules ciliées internes (cci) transforment le signal en potentiel d'action.

L'interprétation du son par l'oreille est extrêmement complexe. En effet, celle-ci doit décomposer le signal qui arrive à la fenêtre ovale en une représentation fréquentielle. Pour cela, deux modèles ont été mis en œuvre.

Le premier proposé par Helmotz, [16, 35] propose une théorie sur des phénomènes de résonance selon la fréquence à la manière des cordes d'un piano. Ensuite est apparu le second modèle de Georges von Békésy [66] qui montre que l'onde se propage dans la cochlée et produit une amplitude maximale en un point donné.

La vibration du liquide entraîne une vibration de la cellule ciliée à la fréquence correspondante. Les cellules ciliées vont alors transmettre ce signal acoustique au nerf auditif. Cette décomposition de l'oreille suivant les fréquences s'appelle la tonotopie. Les hautes fréquences sont codées à la « base » de la cochlée et les basses fréquences à « l'apex ». On peut en effet décomposer la tonotopie de l'oreille comme un piano déroulé où chaque touche correspond à une fréquence (figure 3).

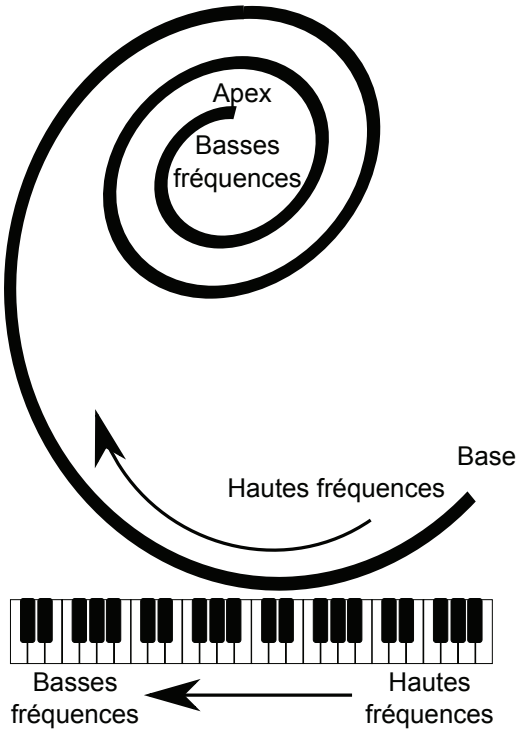


FIGURE 3 – Tonotopie de l'oreille

Le signal sonore est transmis de façon fréquentielle le long du nerf auditif. On peut assimiler ce traitement à une transformée de Fourier en temps réel.

Maintenant que nous avons rappelé les bases de la physiologie de l'oreille, intéressons-nous au comportement d'un signal sonore (un son) qui arrive à l'oreille.

Qu'est-ce qu'un son ?

Une onde de pression située entre 20 Hz et 20 kHz devient un son car il est situé dans le domaine audible (c'est le spectre de l'audition). Il peut alors être interprété comme un signal sonore par le cerveau.

Le son a plusieurs origines. Il est dépendant du lieu dans lequel il est produit. Il peut être de la parole, de la musique, du bruit et la plupart du temps, un mixage de ces trois informations. Notre oreille capte le son, qui est ensuite transféré au cerveau pour le comprendre et l'interpréter.

Environnement du son

Un son sans un milieu matériel ne pourrait pas être entendu, s'il n'y avait pas de vibration de pression. Il est situé dans notre milieu. Il provient de toutes les directions à la fois et pourtant, notre organisme, notre oreille et le cerveau savent faire le nécessaire pour extraire ce que l'on souhaite entendre. De nombreux organisme sont mis en œuvre pour exécuter correctement le débruitage et ne garder que le signal utile.

Notion de bruit.

Le bruit peut être caractérisé par de nombreuses choses. Il est très différent en fonction du lieu où la personne se trouve. L'importance d'avoir un système de débruitage (cerveau ou algorithme) pour améliorer la perception de ce que l'on voudrait écouter prend alors toute sa place (figure 4).



FIGURE 4 – Bruit VS Signal !

Les personnes normo-entendantes peuvent facilement débruiter et écouter ce qu'elles ont envie d'entendre, en sélectionnant par exemple, le meilleur rapport signal sur bruit qui l'entoure [62]. Si la personne veut écouter de la musique et que des personnes parlent à coté, il faut alors pouvoir atténuer le signal de la parole. A contrario, si l'on veut avoir une discussion au restaurant avec la personne en face de nous, le signal correspondant au bruit est alors inversé par rapport à la situation précédente, c'est l'autre cas.

Une petite règle générale peut s'appliquer pour la plupart des débruitages : ce que nous regardons, c'est ce que nous écoutons. Nous avons tendance à privilégier et regarder devant nous : c'est la direction de face. Les algorithmes de débruitage binaural que nous verrons, rehaussent le signal provenant de face (ils seront expliqués dans le chapitre 1).

Les algorithmes binauraux de réduction de bruit présentés dans ce manuscrit privilégient automatiquement la direction de face (« front

target »). Le but des études n'étant pas de trouver la direction de provenance des sources mais de déterminer les meilleurs paramètres pour appliquer un débruitage.

Les personnes ayant des difficultés de compréhension de la parole dans le bruit, ne reconnaissent pas ou très difficilement les « bonnes » informations. Un outil de réduction de bruit est alors nécessaire pour essayer de rétablir ce manque.

Réducteur de bruit.

Il existe plusieurs façons de débruiter un signal de parole. Soit le signal est monophonique (une seul piste à traiter), soit le signal est stéréophonique (deux pistes à traiter). Dans tous les cas un estimateur du bruit devra être effectué. Il estimera la proportion de bruit dans le signal. La plupart des algorithmes de débruitage fonctionnent dans le domaine spectral (fréquentiel) comme la soustraction spectrale ou bien le filtrage de Wiener, mais d'autres comme le beamformer peuvent directement être appliqués dans le domaine temporel.

Sensation de l'intensité.

Le système auditif humain est sensible à cette pression puisqu'elle représente le son. Donc, toute variation atmosphérique est interprétée comme un son. La définition en acoustique de cette variation est le décibel (dB) en niveau de pression acoustique (dB SPL, Sound Pressure Level). On peut le calculer de deux façons :
– En intensité, I_0
– En pression, P_0

La pression P_0 de référence est choisie comme étant la plus petite variation audible et I_0 son intensité correspondante.

Le tableau suivant montre quelques ordres de grandeur de niveau dB SPL.

dB SPL	Pression (μPa)	Exemples
0	20	Référence
5	36	Chambre anéchoïque
10	63	Minimum extérieur
20	200	Studio d'enregistrement
30	630	Résidence tranquille
40	2000	Conversation normale
50	6300	Musique douce
60	20.10^3	Conversation vive
80	20.10^4	Rue bruyante
110	63.10^5	Marteau piqueur $\approx 2m$
120	20.10^6	Réacteur d'avion $\approx 10m$
140	20.10^7	Seuil d'intolérance

TABLE 1 – Correspondance entre niveau de pression acoustique et pression

Description de la parole.

Par définition, la parole c'est : « *La faculté d'exprimer la pensée par le langage articulé.* »

Caractéristiques.

Comme tous les autres sons arrivant aux oreilles, c'est une onde de pression. Elle varie suivant les personnes (hommes, femmes ou enfants), suivant l'intensité (énervement, chuchotement)... La parole est décrite dans plusieurs domaines :

- Le domaine temporel :
  - L'enveloppe, qui contient l'énergie du signal
  - La structure fine, qui contient les différentes variations du signal
- Le domaine fréquentiel :
  - La fréquence fondamentale, F_0, et ses harmoniques
  - Les formants, F_1, F_2, ...

Des méthodes existent pour trouver ces différentes informations. L'enveloppe temporelle peut être obtenue en calculant le module de la transformée de Hilbert. La F_0 se calcul par une auto-corrélation du signal,

[8]. Et les formants par la méthode de la LPC (Linear Predicting Coding, [8]).

Fréquence fondamentale.

La fréquence fondamentale dans le signal de parole « représente » la vibration des cordes vocales (le signal glottique). Ce signal est composé de la fondamentale et de ses harmoniques. Elle caractérise la personne qui parle :
– Pour un homme : $\approx 100Hz$
– Pour une femme : $\approx 200Hz$
– Pour un enfant : 300 à $400Hz$

Formants.

On peut les caractériser comme étant le filtre analogique du signal glottique. En effet les formants représentent les résonances entre le signal glottique et la sortie de la bouche. Ils caractérisent les voyelles. Ce sont les maximums d'énergie dans le signal. On peut représenter les deux premiers formants (F_1 et F_2) dans le triangle de Delattre (table et figure).

Voyelle (API)	$F_1(Hz)$	$F_2(Hz)$(Hz)
$[u]$	320	800
$[o]$	500	1000
$[\alpha]$	700	1150
$[a]$	1000	1400
$[\emptyset]$	500	1500
$[y]$	320	1650
$[\epsilon]$	700	1800
$[e]$	500	2300
$[i]$	320	3200

TABLE 2 – Exemple de valeurs des formants pour différentes voyelles

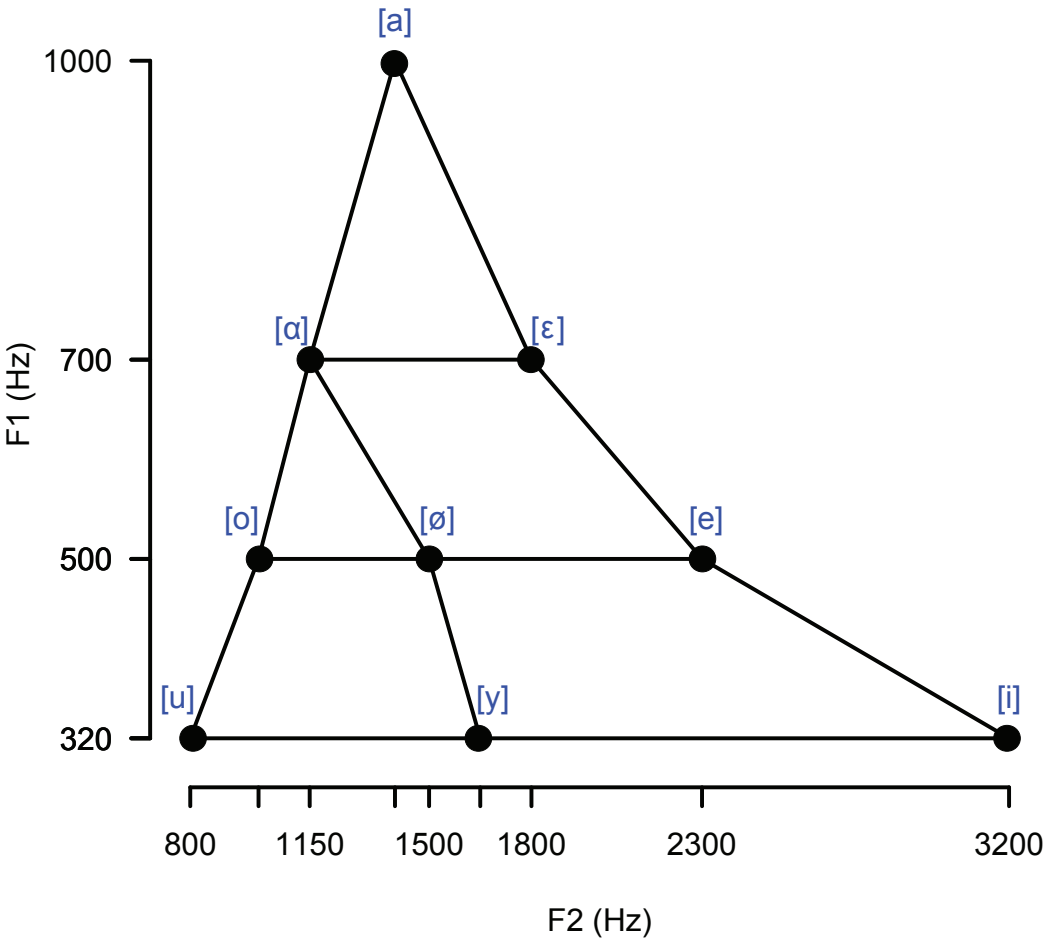


FIGURE 5 – Triangle formantique de Delattre

Implant cochléaire.

L'implant cochléaire (décrit dans le chapitre 2), permet à l'aide d'électrodes insérées dans la cochlée, de « compenser », réhabiliter, la perte de l'audition. Les électrodes déchargeant des impulsions électriques sur le nerf auditif, va permettre de « jouer » le rôle des cellules ciliées.

Plusieurs façon existent pour la réhabilitation par l'implant cochléaire. Le choix des méthodes se pose sur un seul implant cochléaire(cas du monaural), deux implants cochléaires (cas du bilatéral), ou un seul implant pour les deux oreilles (cas du binaural).

La première méthode consiste à positionner un implant cochléaire monaural dans chaque oreille. Cette méthode est dite bilatérale car il n'y a aucun lien, aucun traitement croisé, et aucune stimulation entre les deux implants [1, 45, 50]. La seconde méthode est de positionner un seul implant cochléaire binaurale [20, 86]. Ces deux méthodes donnent actuellement les mêmes résultats [101] si les deux voies sont indépendantes.

Lorsqu'il y a un seul processeur pour la stimulation des deux oreilles. On peut d'obtenir un parfait contrôle entre les deux oreilles et une parfaite synchronisation. L'avantage de ce dernier est que l'on peut alors appliquer des méthodes de « traitement du signal » croisées qui permettent d'améliorer, comme dans l'aide auditive actuelle, la parole dans les milieux bruités.

De nombreux travaux ont porté sur l'étude de l'implant cochléaire et l'intelligibilité de la parole. Il en ressort que dans le silence, le score de reconnaissance de mots est supérieur à 80% [76]. De plus, dans le silence, quatre canaux suffisent pour reconnaître la parole. Les performances de reconnaissance, diminuent ensuite très rapidement lorsque l'on ajoute du bruit dans le signal de parole. Le nombre de canaux dans ce cas, pour reconnaître un minimum la parole, est de huit, [76]. L'importance pour retrouver une meilleure intelligibilité de la parole est de mettre un algorithme de débruitage (autrement dit un algorithme de réduction de bruit) qui rehaussera la parole par rapport aux autres signaux.

Présentation du problème

Les problèmes de débruitage de signaux pour la parole ou bien pour les autres domaines ne sont pas récents. L'objectif étant de restaurer un signal utile à partir d'un signal bruité qui le plus souvent est considéré comme additif.

Les méthodes classiques comme la soustraction spectrale ou le filtrage de Wiener améliorent grandement la qualité du signal mais introduisent en

contre partie, un bruit résiduel que l'on appel le bruit musical. Ce dernier étant très gênant pour la perception humaine. Il faut donc arriver à diminuer ce bruit tout en gardant l'intelligibilité de la parole.

Les progrès pour la perception humaine en codage de la parole sont obtenus, en exploitant les différentes propriétés de la perception à travers différents phénomènes de masquage. En effet, si le son existe, il n'est pas forcement perçu. Ceci est du à la possibilité qu'il soit masqué par la présence d'un autre bruit plus puissant. On peut alors récupérer ce petit signal.

La localisation après une réhabilitation de la parole est aussi utile. En effet, il est bon de savoir d'où viennent les nombreuses informations arrivant aux microphones. L'apport de la binauralité dans ce cas est un atout majeur.

Plan du document

Le manuscript sera présenté de la façon suivante. Le premier chapitre parlera de la littérature et les explications des différents algorithmes mono-voie et binaural qui permettent d'isoler le bruit d'une trace utile. Puis dans le second chapitre on rappellera, l'histoire de l'implant cochléaire avec la simulation du vocodeur de l'implant cochléaire. Les deux chapitres suivant seront consacrés à nos travaux. Le chapitre trois présente les résultats obtenus après débruitage du signal en reconnaissance de la parole et dans le dernier chapitre on étudiera, la localisation d'une source sonore après un débruitage binaural. Enfin une conclusion avec les perspectives viendra terminer le manuscript de thèse.

Chapitre 1

Différents aspects des algorithmes de réduction de bruit.

Introduction

L'importance de la compréhension de la parole en milieu bruité pour les personnes normo-entendantes, malentendantes ainsi que sourdes profondes est primordiale. En effet, comme dit dans l'introduction, la compréhension de la parole permet aux humains, de pouvoir communiquer. Si la compréhension entre une où plusieurs personnes n'est alors pas possible, il y a un risque pour que celle-ci ne communique plus et amène à un isolement. Il existe des systèmes qui permettent de « compenser » ce manque. Ce sont les algorithmes de rehaussement de la parole ou dit autrement les algorithmes de réduction de bruit (appelé aussi débruitage) permettent de rehausser la parole ou bien, d'atténuer le bruit environnent.

L'aspect principal de la réduction de bruit est justement d'améliorer le signal mais par la même occasion de garder l'aspect dit : temps réel. En effet, si le débruiteur enlève parfaitement le bruit dans le signal capté par les microphones, mais qu'il lui faut environ une seconde de traitement pour

l'effectuer alors il ne sert à rien car il est impossible de tenir une conversation avec des latences aussi grandes. De plus si l'atténuateur permet d'augmenter le rapport signal sur bruit de la trace capté par les microphones, au final, le signal est toujours détérioré. Parfois il n'est pas nécessaire d'avoir un atténuateur (dans le silence par exemple).

La méthode pour débruiter un signal peut être séparé en deux étapes ; La première consiste à estimer le bruit dans le signal et la seconde à rehausser le signal utile (la parole) ou atténuer le bruit. Généralement ces méthodes sont effectuées dans le domaine spectral par l'intermédiaire de la transformée de Fourier rapide (FFT, Fast Fourier Transform).

Il existe plusieurs façon de traiter le signal ou les signaux arrivants aux microphones. On peut distinguer deux catégories qui seront décrites dans les prochaines sections. La première méthode concerne un seul microphone. Dans ce cas, c'est la méthode de réduction du bruit mono-voie. La seconde méthode porte sur plusieurs microphones (dans notre cas, le signal sera limité à deux microphones décalés d'une distance égale à la distance entre les deux oreilles prise généralement égale à d = 20 cm). Cette méthode est dite réduction du bruit stéréophonique ou bi-voie.

1.1 Base de la Transformée de Fourier rapide

La Transformée de Fourier (TF) permet de passer d'un espace temporel à une dimension fréquentielle. Cette fonction est basée (Fourier) sur la décomposition d'un signal en somme de sinusoïdes (ou cosinus) qui sont des périodes multiples de la fréquence fondamentale (fréquence de base). Les coefficients de chaque fréquence représentent le niveau d'énergie.

La formule de la TF est :

$$G(f) = \int_{-\infty}^{+\infty} g(t).e^{-j2\pi ft}dt \tag{1.1}$$

Où :
– t : représente le temps
– f : représente les fréquences
– $g(t)$ représente la fonction temporelle

Cette TF peut se décomposer en somme de sinusoïdes et consinusoïdes par définition :

$$\begin{aligned} g(t) \quad = A_0 \quad & +A_1.sin(2\pi ft) + A_2.sin(4\pi ft) + ... + A_n.sin(2n\pi ft) \\ & +B_1.cos(2\pi ft) + B_2.cos(4\pi ft) + ... + B_n.cos(2n\pi ft) \end{aligned} \tag{1.2}$$

Où :
– A_0 : représente le coefficient de Fourier à la fréquence $f_i = \frac{i}{T}$
– B_0 : est pris comme étant égale à 0

Une fois le signal échantillonné, l'intégrale devient une somme sur un environnement finis, l'équation devient alors :

$$G_N(k) = \frac{1}{N} \sum_{n=0}^{N-1} g(n).e^{-j\frac{2\pi kn}{N}} \tag{1.3}$$

Où :
– k : se réfère à f_k
– n : est l'indice de l'échantillon
– N : est le nombre total d'échantillons de la trame

Attention cependant au repliement après l'échantillonnage pour respecter le théorème de Shanon, $f_{max} \leq \frac{f_e}{2}$, qui indique que la fréquence maximale du signal d'entrée ne doit pas dépasser la fréquence d'échantillonnage divisée par deux.

En pratique, la TFD n'est appliquée que sur une partie du signal et non sur le total. Cette partie est appelée « trame ». Une fenêtre de pondération est ensuite appliquée pour éviter les discontinuités au début et à la fin de la trame. Introduisant dans le spectre du signal des informations non existantes

et donc l'introduction d'erreur. Il existe plusieurs types de fenêtre (figure 1.1) :

– Rectangulaire : 0, lorsque ce n'est pas la trame du signal et 1, pendant la durée de la trame. Si le signal commence et finit très proche de 0 (type signaux impulsionnels) alors il n'y a pas beaucoup de déformation au niveau du signal fréquentiel car une telle fenêtre engendre un sinus cardinal au niveau du spectre.

– Hanning : C'est de loin la fenêtre la plus utilisée. Elle est particulièrement utile pour les signaux pseudo stationnaires. Partant de 0 et finissant à 0, elle rend le signal périodique sur la durée de la trame. C'est elle qui introduit le meilleur compromis entre résolution fréquentielle et temporelle en amplitude.

– Hamming : Elle est similaire à Hanning, c'est en fait la dérivée de la fenêtre de Hanning, mais conduit une meilleure réponse en fréquence.

– Kaiser-Bessel : C'est l'inverse de la fenêtre de Hamming, elle donne une meilleure réponse sur les amplitudes.

– Flat Top : c'est la seule fenêtre qui donne l'amplitude exacte pour les raies spectrales. Elle est utilisée pour l'étalonnage des signaux.

– Blackman : Elle permet de réduire fortement les lobes secondaires. C'est la fenêtre la plus étroite.

La multiplication d'une fenêtre avec le signal de la trame entraîne automatiquement un sinus cardinal (figure 1.1). Celui-ci est plus ou moins atténué selon les caractéristiques de la fenêtre.

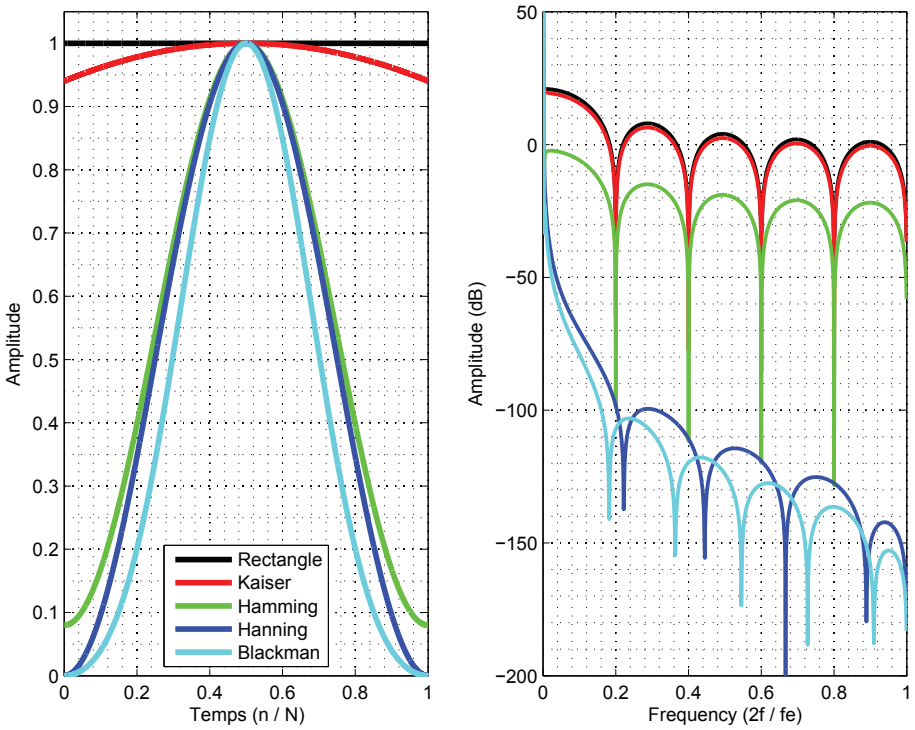


FIGURE 1.1 – Différence entre les différents types de fenêtres

Algorithme de calcul de la FFT

Le but de l'algorithme de la transformée de Fourier rapide est d'optimiser le temps de calcul de la TFD. En effet, la TFD classique demande énormément de calcul. Il faut N multiplications complexes et N-1 additions et il y a N composantes à calculer. Ceci conduit à N^2 multiplications complexes et à N(N-1) additions complexes. Il a fallu rechercher un algorithme qui permette de diminuer fortement ce nombre d'opération pour optimiser le temps de calcul dans les processeurs.

Le principe de la FFT repose sur des séparations des échantillons deux à deux du signal (échantillons pairs et impairs). On a au départ une trame de taille N, il y a alors $k = 2r$ éléments de rang pair et $k = 2r + 1$ de rang

impair. On peut alors décomposer en deux TFD le signal :

$$G_n(k) = \sum_{r=0}^{\frac{N}{2}-1} g_{2r}(n).e^{-j\frac{2\pi n(r)}{\frac{N}{2}}} + \sum_{r=0}^{\frac{N}{2}-1} g_{2r+1}(n).e^{-j\frac{2\pi n(2r+1)}{\frac{N}{2}}} \tag{1.4}$$

Où :

$$G_n(k) = \sum_{r=0}^{\frac{N}{2}-1} g_{2r}(n).e^{-j\frac{2\pi nr}{\frac{N}{2}}} + e^{-j\frac{2\pi n}{N}} \sum_{r=0}^{\frac{N}{2}-1} g_{2r}(n).e^{-j\frac{2\pi nr}{\frac{N}{2}}} \tag{1.5}$$

Où le premier terme représente les rangs pairs et le second les rangs impairs. Ceci représente la décomposition de la TFD d'ordre 2. On peut généraliser en prenant :

$$G_n = G_n^e + e^{-j\frac{2\pi n}{N}} G_n^0 \tag{1.6}$$

Le calcul se fait ensuite sur le résultat de l'ordre deux pour obtenir le rang 4 et ainsi de suite. Les temps de calculs sont alors de $N = 2p$ avec $p = log_2(N)$. Ils sont fortement diminués. Le coût de traitement de l'algorithme est de $N.log_2(N)$ additions et $\frac{N}{2}.(log_2(N) - 1)$ multiplications.

Ainsi la FFT conduit à un gain de temps de calcul non négligeable. Il conviendra ensuite d'interpréter le résultat. En effet, le résultat étant sous forme complexe, on peut alors facilement calculer le module et la phase des fréquences correspondantes à la trame.

1.2 Réduction du bruit pour les signaux mono-voie

Dans ce paragraphe, la base de la technique de réduction de bruit dans un système mono-voie va être présentée. La plupart des méthodes de réduction de bruit sont issues de deux principales techniques. La première est la « Soustraction Spectrale » et la seconde est le « filtrage de Wiener ». La réduction de bruit mono-voie est effectuée à partir d'un seul microphone, il n'y aura alors qu'un seul signal à traiter. Les démonstrations qui vont

suivre seront décrites de la même façon pour les deux méthodes. Dans tous les cas, une amélioration de la compréhension de la parole est nécessaire, ces algorithmes sont faits pour améliorer la parole par rapport au reste du signal.

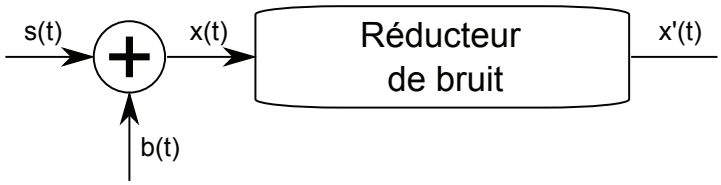


FIGURE 1.2 – Modèle de débruitage mono-voie

Au départ, on peut considérer que le signal qui arrive au microphone est composé d'un signal utile qui est la parole et d'un bruit qui est ce que l'on doit atténuer. Le but d'un tel algorithme dans l'aide auditive ou bien dans l'implant cochléaire est d'améliorer le rapport signal sur bruit en sachant que le signal est considéré comme étant de la parole et que le reste est un bruit. Ces algorithmes sont implantés dans l'aide auditive et améliorent la parole.

On considère le signal arrivant au microphone est un mixage entre un signal utile et un signal bruité :

$$x(t) = s(t) + b(t) \tag{1.7}$$

Où :
– t : représente le temps
– x : représente le signal capté par le microphone
– s : représente le signal utile
– b : représente le bruit
Cette équation (eq 1.7) doit être modifiée pour traiter le signal. Il doit être échantillonné. De ce fait, l'équation devient :

$$x(n) = s(n) + b(n) \tag{1.8}$$

Où :
– n : représente le numéro de l'échantillon
– x : représente le signal capté par le microphone
– s : représente le signal utile
– b : représente le bruit

Une « simple » soustraction du bruit $b(n)$ au signal $x(n)$ pourrait alors être faite pour enlever le bruit dans $x(n)$. Cependant, le bruit n'est jamais connu par avance. Il faut poser des hypothèses. Le bruit $b(n)$ doit être estimé à partir de l'échantillon en cours et des échantillons précédents. Un estimateur de bruit est alors appliqué pour estimer le bruit. Le signal que l'on connaît est $x(n)$, mais il peut aussi avoir des hypothèses sur la parole, par exemple la connaissance du spectre à long terme de la parole, n'est pas le même que celui du bruit. Il est actuellement très difficile de travailler dans le domaine temporel par échantillon car le filtrage doit être un produit de convolution.

1.2.1 Estimation du bruit

La première étape consiste à transformer le signal temporel par l'intermédiaire de la FFT en nombre complexe et plus particulièrement le module. Il est très rare d'utiliser la phase du signal car l'amplitude suffit habituellement. L'équation 1.8 devient alors :

$$X(f) = S(f) + B(f) \tag{1.9}$$

Et en module :

$$|X(f)| = |S(f)| + |B(f)| \tag{1.10}$$

Où :
– f : représente les fréquences
– X : représente le spectre du signal capté par le microphone
– S : représente le spectre du signal utile
– B : représente le bruit

Le but de cette sous-section est d'estimer de bruit $B(f)$ afin d'appliquer en sortie, un algorithme de réduction de bruit (figure 1.2). Pour trouver le spectre du bruit, la première hypothèse consiste à dire qu'il est stationnaire (hypothèse qui n'est pas toujours valable) et indépendant sur la trame d'analyse. La seconde hypothèse est que la parole varie beaucoup plus rapidement que le bruit. On peut dire par contre que sur des trames d'analyse courte (durée de la trame $<$ 25 ms) la parole est stationnaire. C'est pourquoi une trame d'analyse longue renseigne sur le bruit alors qu'une trame d'analyse courte donne des informations sur la parole.

Le calcul du bruit se fait généralement sur la densité spectrale de puissance (DSP) qui est obtenue en calculant l'énergie par fréquence du signal (théorème de Parseval).

$$P_X(f) = |X(f)|^2 \tag{1.11}$$

Où :
– $P_X(f)$: représente la DSP du signal

L'équation devient alors :

$$P_X(f) = P_S(f) + P_B(f) \tag{1.12}$$

Il existe plusieurs façons pour estimer cette DSB au niveau du bruit ($P_B(f)$). Le plus simple est de prendre une trame d'acquisition longue de 200 ms pour connaître le spectre moyen assimilé au bruit et des trames courtes pour le signal de parole d'une durée de 8 ms. La comparaison ensuite de ces deux aspects permet d'estimer le bruit des trames à l'entrée du signal.

Une fois le module du bruit considéré, il faut essayer de l'atténuer dans la trame correspondante, il existe plusieurs méthodes (classiquement utilisées) pour exécuter ce débruitage. Les trois principaux filtre en mono-voie seront présentés dans les sous sections suivantes.

1.2.2 Soustraction spectrale

La soustraction spectrale est le débruiteur de plus ancien. Elle a été introduite par Boll [9]. Comme son nom l'indique, elle effectue son travail dans le domaine spectral et a pour principe de soustraire le bruit estimé au signal. L'estimation du bruit se fait sur plusieurs trames d'acquisition ($\simeq 300ms$).

$$P_{X'}(f) = P_S(f) + P_B(f) - EP_B(f) \tag{1.13}$$

Où :
- $EP_B(f)$: représente l'estimation du bruit
- $P_{X'}(f)$: représente la DSP après la soustraction spectrale

Principe

Il existe deux versions de base pour la soustraction spectrale se différenciant par l'amplitude ou la puissance.

$$|X'(f)| = |X(f)| - |EB(f)| \tag{1.14}$$

Dans ce cas, il s'agit de la « soustraction spectrale d'amplitude ». Mais le plus souvent comme indiqué dans le paragraphe précédent, la soustraction se fait au niveau de la puissance.

$$|X'(f)|^2 = |X(f)|^2 - |EB(f)|^2 \tag{1.15}$$

Le problème de ces deux équations 1.14 et 1.15, est que le second terme peut être négatif. On peut le rendre positif en changeant de signe ou bien en l'annulant comme dans l'équation 1.16. C'est la première amélioration que l'on peut proposer.

$$|X'(f)|^2 = \begin{cases} |X(f)|^2 - |EB(f)|^2 & \text{si } |X(f)|^2 > |EB(f)|^2 \\ 0 & \text{sinon} \end{cases} \tag{1.16}$$

Le passage dans le domaine temporel se fait par une iFFT. La phase du signal d'entrée est gardée, une estimation du bruit de la phase s'avère être

une tache très compliquée.

Soustraction spectrale par filtrage

En se basant sur un filtre, et en gardant l'estimation du bruit on peut assimiler la soustraction spectrale à un filtrage.

$$|X'(f)| = G(f).|X(f)| \qquad 0 \leq G(f) \leq 1 \tag{1.17}$$

Et pour la soustraction spectrale de puissance, on obtient :

$$G(f) = \begin{cases} \sqrt{1 - \frac{|EB(f)|^2}{|X(f)|^2}} & \text{si } |X(f)|^2 > |EB(f)|^2 \\ 0 & \text{sinon} \end{cases} \tag{1.18}$$

Dans la littérature, la soustraction spectrale est très utilisée car elle est très simple à mettre en oeuvre. Néanmoins, elle génère des artéfacts après la réduction du bruit ainsi qu'une distorsion du signal et le bruit musical.

Amélioration de la soustraction spectrale par Berouti

Berouti [6], a trouvé qu'après une soustraction spectrale, le bruit résiduel contient deux types de pics spectraux. Ce sont des pics larges perçus comme étant un bruit large bande et des pics étroit comme étant des tonales. Il qualifie ce dernier de bruit musical. Il propose alors dans la soustraction spectrale de rajouter à la surestimation du bruit une quantité $\beta |EB(f)|$ au lieu de 0 afin d'éviter que le seuil de tolérance du bruit dépasse la puissance du bruit. On obtient alors l'équation suivante :

$$|X'(f)|^2 = \begin{cases} |X(f)|^2 - \alpha |EB(f)|^2 & \text{si } \left(|X(f)|^2 - \alpha |EB(f)|^2\right) > \beta |EB(f)|^2 \\ \beta |EB(f)|^2 & \text{sinon} \end{cases} \tag{1.19}$$

Ces travaux ont montré que le paramètre α dépend du RSB de la trame correspondante.

$$\alpha = \alpha_0 - \frac{RSB}{s} \tag{1.20}$$

Où :
- $s = \frac{20}{3}$
- $\alpha_0 = 4$

Ces valeurs sont données pour un RSB compris entre $\pm 6dB$, α_0 est compris entre 3 et 6. La variable β est très sensible au bruit. Pour une fort niveau de bruit, $RSB = -6dB$, β doit être compris entre $0.02 \leq \beta \leq 0.06$. Et pour un bruit faible, $0.005 \leq \beta \leq 0.02$.

Amélioration de la soustraction spectrale par Boll

Boll [9] a décomposé la soustraction spectrale en quatre parties. La moyenne d'amplitude, la correction de l'estimateur, la réduction du bruit résiduel et l'atténuation du signal pendant les périodes de silences.

1. La moyenne d'amplitude

$$X'(f) = [|X(f)| - \mu(f)]\, e^{i\theta X(f)} \tag{1.21}$$

Où : $\mu(f) = EB(f)$ est la moyenne du bruit calculé pendant les silences. Le filtre est alors de la forme suivante :

$$H(f) = 1 - \frac{\mu(f)}{|X(f)|} \tag{1.22}$$

On en déduit l'erreur de l'estimation :

$$e(f) = X(f) - \hat{X}(f) = B(f) - \mu(f) e^{i\theta X(f)} \tag{1.23}$$

Où $e(f)$ dépend à la fois de $B(f)$ et de la moyenne de $\mu(f)$. Il faut réduire au minimum $e(f)$ pour avoir la meilleure estimation du bruit. Pour cela, il faut que : $B(f) \simeq \mu(f)$. L'introduction de la moyenne de l'amplitude dans le signal bruité $\overline{|X(f)|} = \frac{1}{N} \sum_{f=0}^{N-1} X(f)$ permet

d'obtenir une nouvelle équation 1.24.

$$X'(f) = \left[\overline{|X(f)|} - \mu(f)\right] e^{i\theta X(f)} \tag{1.24}$$

L'erreur devient alors :

$$e(f) = X(f) - \hat{X}(f) \simeq \mu(f) - \overline{|B(f)|} \tag{1.25}$$

Où, $\overline{|B(f)|} = \frac{1}{N}\sum_{f=0}^{N-1} B(f)$. De cette façon, si l'on moyenne sur une grande période, l'erreur se réduit. Mais dans ce cas, le bruit doit être considéré comme stationnaire. La parole n'étant pas stationnaire, il y a une limite à cette méthode.

2. La réduction du bruit résiduel

La méthode consiste à remplacer les valeurs négative dans $X'(f)$ par des zéros. C'est une méthode de rectification demi-onde. La nouvelle expression de $H(f)$ devient :

$$H^R(f) = \frac{H(f) + |H(f)|}{2} \tag{1.26}$$

Et l'équation devient $X'(f) = H^R(f)X(f)$

3. La réduction du bruit résiduel

On peut résumer cette partie par l'équation suivante :

$$|X'_k(f)| = \begin{cases} |X'_k(f)| & si X'(f) \geq max\left|B^R(f)\right| \\ min\left|X'_j(f)\right|, j = k-1, k, k+1 & si X'(f) < max\left|B^R(f)\right| \end{cases} \tag{1.27}$$

Où :

– k représente le numéro de la trame

– $max\left|B^R(f)\right|$ représente le maximum du bruit résiduel mesuré pendant les silences

La réduction du bruit résiduel s'effectue ainsi en sélectionnant le minimum de l'amplitude estimée durant trois trames j.

4. Atténuation du signal pendant les périodes de silences

Boll propose de manière empirique un seuil de détection d'activité vocale.

$$\tau = 20.log_{10}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{X'(f)}{\mu(f)}\right|df\right] \tag{1.28}$$

Si $\tau < -12dB$, Boll considère qu'il n'y a pas d'activité vocale. Dans ce cas, au lieu d'enlever le signal totalement, il propose de l'atténuer.

Amélioration de la soustraction spectrale par Virag

Virag [102] a mixé les deux précédentes méthodes : Boll et Berouti avec la généralisation de Lim [53] pour obtenir une généralisation de la soustraction spectrale. L'intérêt revient à trouver un compromis entre la réduction du bruit et la distorsion du signal. Le gain $G_k(f)$ est donné par :

$$G_k(f) = \begin{cases} \left(1-\alpha\left(\left|\frac{B(f)}{X(f)}\right|\right)^{\eta_1}\right)^{\eta_2} & si\ \left(\left|\frac{B(f)}{X(f)}\right|\right)^{\eta_1} < \frac{1}{\alpha+\beta} \\ \left(\beta\left(\left|\frac{B(f)}{X(f)}\right|\right)^{\eta_1}\right)^{\eta_2} & sinon \end{cases} \tag{1.29}$$

Où :
– α est le facteur de sur-soustraction ($\alpha > 1$)
– β avec $0 \simeq \beta << 1$, est un facteur qui permet d'introduire un léger bruit de fond

On peut remarquer que le choix de α et β est beaucoup plus critique que $\eta_{1,2}$. Si $\eta_1 = \eta_2 = 1$, il s'agit d'une soustraction spectrale d'amplitude. Si $\eta_1 = \eta_2 = 0.5$, il s'agit d'une soustraction spectrale de puissance et si $\eta_1 = 2$, $\eta_2 = 1$, il s'agit d'un filtrage de Wiener.

Amélioration psychoacoustique

La soustraction spectrale est très simple à mettre en oeuvre. Cependant, elle reste une méthode limitée par l'apparition du bruit musical. Des améliorations ont été étudiées sur les aspects psychoacoustiques [7, 90, 102]. Le but est de rendre le bruit musical moins gênant en exploitant un phénomène de masquage. Ce dernier est pour rappel obtenu par le calcul d'un seuil et que tout bruit au-dessous de ce seuil est imperceptible.

1.2.3 Filtrage de Wiener

Le filtrage de Wiener a été introduit à la fin des années 60 [103] pour essayer d'améliorer la qualité de la trace recueillie dans les potentiels évoqués. Le problème du type de filtrage proposé est qu'il n'est pas applicable sur une moyenne d'acquisition mais pour chaque trace. Doyle [25], propose une modification pour pouvoir l'adapter à la moyenne (le calcul dans ce cas est beaucoup plus rapide), il n'y a pas besoin de filtrer chaque trace. Cependant, il faut considérer que le bruit est stationnaire dans ce cas [25].

Tout comme la soustraction spectrale, les calculs sont effectués dans le domaine fréquentiel. La DSP est calculée comme précédemment dans l'équation 1.11. Le système peut être isolé entre le bruit et le signal utile. Ce qui revient à dire que le rapport $\frac{P_S(f)}{P_B(f)}$ doit être maximisé pour obtenir le signal utile. Le filtre de Wiener est défini de la façon suivante :

Principe

La DSP du bruit est prise dans les périodes de silence. La DSP du signal utile est quant à elle calculée sur chaque trame d'acquisition.

$$W(f) = \frac{P_S(f)}{P_S(f) + P_B(f)} = \frac{1}{1 + \frac{P_B(f)}{P_S(f)}} \tag{1.30}$$

En utilisant les définitions du RSB_{prio}, le filtre de Wiener peut s'exprimer

sous cette forme :

$$W(f) = \frac{RSB_{prio}(f)}{1 + RSB_{prio}(f)} \tag{1.31}$$

La remarque de ce filtre, est que si le bruit est très bien estimé alors le signal transmis sera directement le signal utile. Le filtrage de Wiener est le filtrage optimal au sens du minimum de l'erreur quadratique moyenne (MEQM). Il adapte le rapport signal sur bruit pour chaque trame traitée. Cependant, le régime du bruit doit être stationnaire et non transitoire car l'estimation ne sera pas bonne dans ce second cas. Ce type de filtrage est utilisé en général derrière une soustraction spectrale ou autre réducteur de bruit pour améliorer le rapport signal sur bruit de la trace.

Amélioration psychoacoustique

Lin [54], a proposé de décomposer le signal acoustique suivant les bandes de fréquences de l'audition avec un banc de filtres pour simuler le comportement non linéaire de la cochlée. L'erreur quadratique moyenne de chaque bande est ensuite calculée ce qui revient à faire un filtrage de Wiener. Un autre travail proposé par Beaugeant [5], propose lui un concept différent. C'est de débruiter les canaux contenant du bruit et de ne pas toucher aux autres. Le filtrage sera alors appliqué uniquement dans les bandes spectrales ou le bruit est audible.

$$W(f) \begin{cases} \frac{P_S(f)}{P_S(f)+P_B(f)} & \text{ailleurs} \\ 1 & \text{si} P_S(f) > P_B(f) \end{cases} \tag{1.32}$$

1.2.4 Détection de la parole par le TOP

Principe

Le but de la détection d'activité de parole par le traitement optimisé de la parole (TOP) est que pendant les périodes de silence, il n'y a aucune activité et pendant les périodes de parole, le signal est transmis. Sohn [80], propose une méthode statistique. La DSP du bruit est considéré constante pendant tout le traitement jusqu'à une nouvelle période de silence. Cette hypothèse ne fonctionne qu'avec des trames à court terme. Le TOP peut être amélioré

par la méthode de Martin [58], qui n'est autre qu'une soustraction spectrale avec un suivi du minimum statistique de la DSP du bruit. Il est basé sur chaque composante spectrale supposé comme du bruit.

1.2.5 Réduction du bruit musical

Définition

Le bruit musical est par définition un bruit résiduel perceptuellement gênant suite à un débruitage d'un signal par des algorithmes. Il est généralement induit par ce sont des réducteurs de bruit à court terme tel que les deux algorithmes cités précédemment, la soustraction spectrale et le filtrage de Wiener. Le spectre de ce bruit est tonal d'où le caractère musical. Sa valeur moyenne est plus faible que le bruit du signal d'entrée mais la dispersion fréquentielle est beaucoup plus grande. Le bruit est étalé sur les différentes bandes de fréquences.

D'un point de vu perceptif, le bruit musical est beaucoup plus gênant que le bruit de base, [38]. Les principales raisons de l'apparition de ce type de bruit sont :
– Le traitement non linéaire des composantes négatives du signal bruité,
– L'évaluation non précise de la densité spectrale de bruit,
– L'estimation basée sur les spectrogrammes,
– La variabilité de la fonction de gain appliquée au signal bruité,
– La variance des estimateurs locaux de la DSP du signal...

Estimation et réduction du bruit

Si l'on prend un problème de débruitage classique linéaire où le but est de trouver $H(f)$ qui est l'estimateur, l'erreur dûe à ce filtrage est :

$$\begin{aligned} e_k(f) &= S_k(f) - \tilde{S}_k(f) \\ &= (H_k(f) - 1).S_k(f) + H_k(f).B_k(f) \end{aligned} \tag{1.33}$$

Où :
– $(H_k(f) - 1).S_k(f)$: représente la distorsion du signal
– $H_k(f).B_k(f)$: représente le bruit résiduel qui contient le bruit musical
– k : représente le numéro de la trame
– $\tilde{S}_k(f)$: représente la fonction estimée

Dans le cas ou $0 \leq H(f) \leq 1$, le bruit musical est très difficile à réduire sans augmenter et apporter de la distorsion au signal débruité. Comme dans beaucoup de cas, un compromis entre la distorsion et le bruit musical doit être trouvé pour que l'intelligibilité de la parole soit bonne et que le bruit musical soit moins gênant.

1.3 Réduction du bruit pour les signaux multi-voies

Introduction

Les premières réductions de bruit basées sur plusieurs microphones ont été réalisé pour des aides auditives, [40, 44, 52]. Les microphones sont placés dans le même contour d'oreille. Il n'y a pas de stéréophonie dans ce cas. Plusieurs microphones permettent de sélectionner (comme il sera démontré par la suite) une direction ou d'atténuer le bruit spatial, [91]. Mais ces microphones ne permettent pas de mieux localiser les sources. Ils sont au nombre de deux et généralement positionnés en position avant et arrière. De cette façon, une technique simple de beamforming pourra être implémentée pour débruiter le signal, [47].

Par la suite, ce deuxième microphone a été placé de l'autre coté de la tête, sur l'autre oreille, pour ainsi récolter les signaux stéréophoniques. Cette fonction stéréophonique permet notamment de localiser les sources sonores. De plus la réduction du bruit dans ce cas est binaurale (deux canaux).

Le but de la réduction de bruit stéréophonique (deux canaux placé chacun respectivement sur chaque oreille) contrairement a une réduction du bruit mono-voie (un seul canal), est de localiser une direction et d'atténuer les autres sur un plan donné. Les techniques classiques comme la soustraction spectrale, le filtrage de Wiener ou bien le beamformer peuvent être utilisées. L'algorithme prendra en compte les deux signaux d'entrée comme montré dans la figure 1.3. Ensuite un traitement croisé des voies permet de localiser et d'améliorer le signal utile.

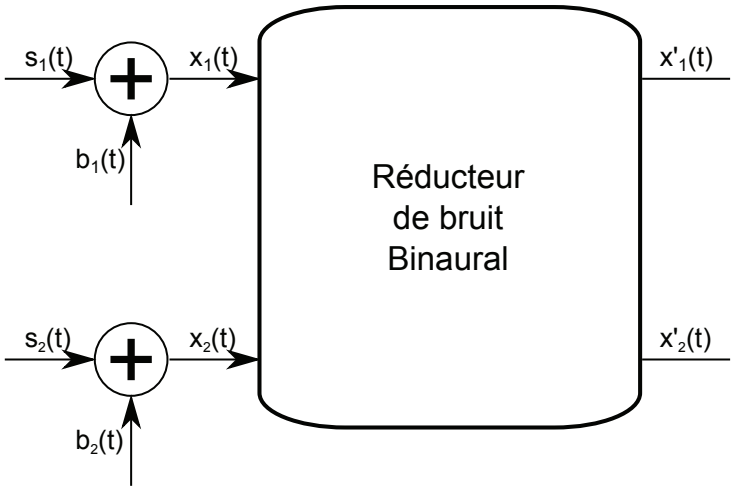


FIGURE 1.3 – Modèle de débruitage binaural

Au départ, on peut considérer que les signaux qui arrivent aux microphones sont composés d'un signal utile qui est la direction à garder et d'un signal bruité qui est celui que l'on doit atténuer. Le but d'un tel algorithme dans l'aide auditive ou bien dans l'implant cochléaire est d'améliorer la localisation d'un son par rapport aux autres.

Le but des algorithmes stéréophoniques, est d'améliorer suivant les préférences, la localisation, la parole, le rapport signal sur bruit. Ils fournissent en termes d'intelligibilité, le double de l'information mono-voie, [55], contrairement au bruit mono-voie, qui implique que le bruit soit

stationnaire à long terme. Le fait d'introduire un microphone de plus permet d'introduire l'espace environnant.

1.3.1 Choix de la direction de la localisation

Le choix de la direction pour tous les algorithmes présentés ci-dessous est situé sur le plan horizontal et centré sur la direction de face (cible de face). En effet, cette direction semble la plus efficace pour la compréhension de la parole, car « *ce que nous écoutons, c'est ce que nous regardons, et généralement, c'est ce qui provient de face* ». Les débruiteurs garderont ce qui est commun aux deux voies.

1.3.2 Estimation du bruit stéréophonique

Un traitement croisé des voies permet d'obtenir une estimation du bruit. Par exemple nous pouvons prendre le cas de la soustraction spectrale binaurale pour connaitre la DSP du bruit. La direction à garder est celle qui provient de face. Le bruit peut être ainsi estimé.

Cas du bruit corrélé

Supposons que les différents bruits soient corrélés au signal utile. Il faut au préalable disposer d'observations contenant le signal utile et une seconde observation contenant le bruit de référence. Le but est de soustraire les deux bruits (de chaque voies) $b1(t)$ et $b2(t)$ de telle sorte que l'estimateur donne le signal utile.

Cas du bruit décorrélé

Dans ce cas, la fonction de cohérence de phase permet d'estimer plus facilement le bruit. Elle s'exprime dans le domaine fréquentielle par l'équation suivante (ce qui revient à calculer la corrélation dans le domaine temporel) :

$$\rho(f) = \frac{E\left\{X_1(f).X_2(f)\right\}}{\sqrt{E\left\{|X_1(f)|^2\right\}.E\left\{|X_2(f)|^2\right\}}} = \frac{P_{x_1x_2}(f)}{\sqrt{P_{x_1x_1}(f).P_{x_2x_2}(f)}} \tag{1.34}$$

Cette équation 1.34, donne des informations sur les corrélations entre les deux signaux. Si $x_1(t)$ est totalement différent de $x_2(t)$ alors la cohérence sera nulle, $\rho = 0$. Cette cohérence est liée à la position des sources et elle est indiquée par la phase. On peut en déduire le module de la fonction de cohérence :

$$MSC(f) = |\rho(f)|^2 \tag{1.35}$$

De ce dernier résultat, à partir de l'équation 1.35, on peut distinguer deux catégories :

– Les signaux cohérents, par définition, présentent un module proche de 1. Ceci indique une source ponctuelle. Donc une forte corrélation spatiale, et peut représenter la parole.
– Les signaux incohérents, sont issus généralement de plusieurs sources indépendantes. Les phases se compensent entre elles pour donner une cohérence de phase proche de 0.

On peut souligner un cas particulier. Dans le cas de bruits parfaitement décorrélés, on peut exprimer la cohérence de phase par le filtrage de Wiener :

$$\rho(f) = \frac{Pss(f)}{Pss(f) + Pbb(f)} \tag{1.36}$$

Le gain optimal avec un déphasage nul correspond aux meilleurs coefficients de Wiener.

1.3.3 Technique du beamforming

Cette technique est basée sur les combinaisons linéaires des différents signaux d'entrée captés par les deux microphones. Cette technique sélectionne une direction parmi le plan horizontal. Il faut au moins deux microphones pour appliquer cette technique. C'est pourquoi elle n'a pas été décrite contrairement aux autres dans la section précédente. Il existe deux techniques distinctes, un « beamformer fixe » et un « beamformer adaptatif ».

Beamformer fixe

Dans ce cas, la direction est fixe. Elle ne détecte pas la source et ne change pas suivant les cas se présentant. L'avantage et le temps de calcul qui est beaucoup plus rapide car il n'y a pas d'adaptation des coefficients. Le principe repose sur un retard et une soustraction (ou addition suivant la direction voulue), [21]. Le signal de sortie généré par les retards provenant des retards ont été démontré par Ricketts, [72]. Cette technique est ensuite très décrite et améliorée dans la littérature, [10, 48].

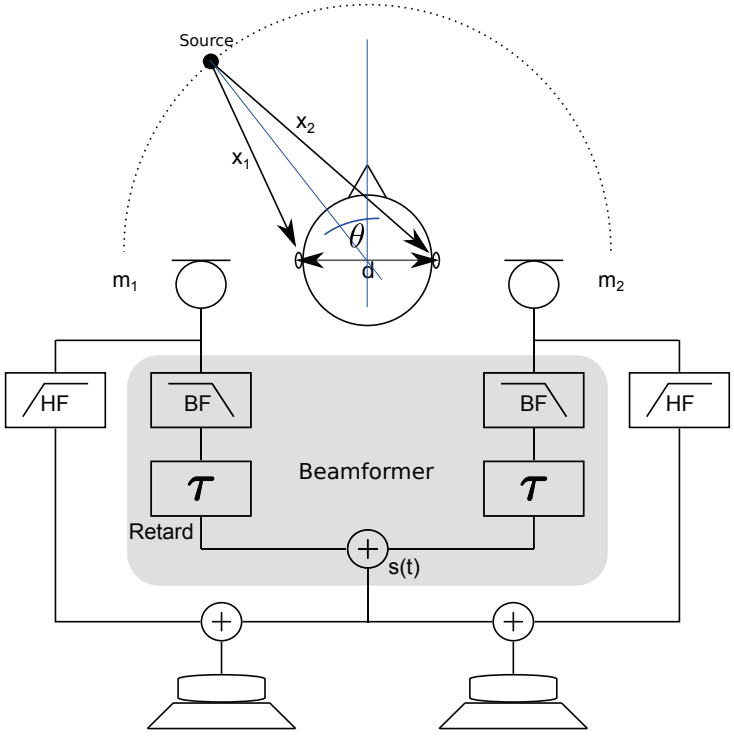


FIGURE 1.4 – Synotptique du beamformer fixe

Le beamformer travaille dans les basses fréquences (BF) car il ajoute un retard constant. Ce retard ne dois pas être plus grand que le double du temps de parcours de la distance entre les deux oreilles car le retard introduirait plus d'une période de la sinusoide et donc la phase ne pourrait pas être calculée. Dans la figure 1.4, la partie en gris représente le beamformer. On peut remarquer qu'il y a deux microphones m_1 et m_2 qui captent le signal

de la source. La distance « d » entre les deux capteurs est de $d = 20cm$. Le retard τ permet de retarder le signal d'entré pour obtenir la fréquence de coupure du filtre passe-bas et qui sera la même que le filtre passe-haut. La différence de phase entre les deux signaux d'entrée ne doit pas dépasser π, [22] :

$$|\Delta\phi| = \pi \tag{1.37}$$

On peut en déduire la phase entre les deux signaux :

$$|\Delta\phi| = \frac{2\pi}{\lambda}(x_1 - x_2) \simeq \frac{2\pi d}{\lambda} sin(\theta) \tag{1.38}$$

Où :
– x est la distance entre la source est les microphones
– θ est l'angle par rapport au plan horizontal (sagittal)
– λ est la longueur d'onde du signal

L'équation 1.38 découle des relations suivantes :
Les signaux arrivant aux microphones $e_1(t)$ et $e_2(t)$ perçuent sont :

$$\begin{aligned} e(t) &= A.e^{j\omega t-\tau} \\ &= A.e^{j\omega t-\frac{x}{c}} \\ &= A.e^{j\omega t-\frac{2\pi x}{cT}} \\ &= A.e^{j\omega t-\frac{2\pi x}{\lambda}} \\ &= A.e^{j\omega t-\phi} \end{aligned} \tag{1.39}$$

Où :
– A est l'amplitude du signal
– c est la célérité du son dans l'air
– T est la période du signal
– m est la distance moyenne entre la source jusqu'au centre de la tête
– $x_1 \simeq m - \frac{d}{2} sin(\theta)$

Il faut maintenant trouver les valeurs de τ pour garder la direction provenant de face. On peut remarquer que si $\tau_1 = \tau_2$, alors la direction

sera la face. Le beamformer va créer une cardioide centrée sur la direction de face. Comme le beamformer travaille dans les basses fréquences, on peut trouver la fréquence maximum de travail qui devra aussi vérifier la relation de Shannon :

$$f_{max} \leq \frac{Fe}{2} \tag{1.40}$$

Où :
– Fe est la fréquence d'échantillonage

En revenant à la relation 1.38 et en prenant $\theta = \pi/2$ et $|\Delta\phi| = \pi$, on écrit $\pi = \frac{2\pi d}{\lambda}$, donc la longueur d'onde est égale à $\lambda = 2 * d$

$$\lambda = 2 * d = cT = \frac{c}{f_{max}} \tag{1.41}$$

D'où :

$$f_{max} = \frac{c}{2 * d} \tag{1.42}$$

En prenant comme valeur $c = 340m.s^{-1}$ et $d = 20cm$. On en déduit $f_{max} = 860Hz$. La différence de retard est alors donnée par, $\tau = \tau_1 - \tau_2 = \frac{d}{c} \simeq 0.6ms$. Dans les programmes, la valeur τ correspond à 10 échantillons de retard pour une fréquence d'échantillonnage , $Fe = 16kHz$. Avant de ré-injecter le signal haute-fréquence, on additionne les deux voies. On peut remarquer que si la source ne provient pas de face alors lorsque les signaux seront retardés, ils ne seront plus en phase et lors de l'addition, ils seront atténués. Au contraire, si les signaux sont en phase et provenant de face, alors ils seront additionnés et amplifierons le gain de deux. C'est pourquoi un gain en sortie est appliqué pour garder les mêmes rapports d'énergie entre les basses fréquences et les hautes fréquences.

Lors d'un traitement de type beamformer, le seul moyen de localiser la source dans ce cas reste les hautes fréquences car les basses fréquences sont communes aux deux voies.

Beamformer adaptatif

Contrairement au beamformer fixe, qui localise sur une seule direction, le beamformer adaptatif localise la source et sélectionne ensuite la direction d'où le nom d'adaptatif. Les travaux de Veen, [81, 100] montrent que cette technique donne une meilleure reconnaissance au niveau de la parole que pour le beamformer fixe. Néanmoins ces algorithmes sont limités généralement au champ frontal pour limiter les distorsions. Dans nos travaux, nous n'avons pas utilisé cet algorithme car cette technique fonctionne correctement lorsque les microphones sont situés sur la même oreille en position avant-arrière, et non un sur chaque oreille.

Le beamformer adaptatif a aussi été décrit dans l'implant cochléaire en présence de quatre microphones, [47, 82].

1.3.4 Algorithme de Doerbecker

L'algorithme souvent appelé de Doerbecker, [23], est une méthode de réduction de bruit classiquement utilisée dans l'aide auditive. Elle permet par l'intermédiaire de chaque voie d'atténuer le bruit musical et augmenter la parole. Cet algorithme est issu d'une soustraction spectrale et d'un filtrage de Wiener.

Il propose plusieurs solutions pour atténuer le bruit musical, la permière est d'utiliser le filtre d'Ephraim & Malah, [30], figure 1.5 et la seconde d'utiliser le filtre de Scalart, [74], figure 1.6.

La principale hypothèse est que la cohérence de phase doit être proche de 1. Il considère donc que les bruits environnants sont décorrélés. Il faut aussi que la distance entre les deux microphones soit très inférieures à celle du locuteur. La dernière hypothèse repose que sur le fait que les signaux de parole arrivant aux microphones sont identiques sur chaque voie (la source sonore est située à une distance supérieure à $1m$).

La DSP sur chaque voie est alors calculée comme pour la soustraction spectrale mono-voie. Elle est notée $\phi_{X_L X_L}$ pour la voie de gauche et $\phi_{X_R X_R}$ pour la voie de droite. La DSP « croisée » est obtenue en faisant le produit des deux FFT de chaque voie : $\phi_{X_L XR} = X_L.X_R$. Les notations entre parenthèse sont omises pour une meilleure compréhension.

L'estimation du bruit ϕ_{NN} est obtenue en faisant le rapport entre les deux voies et la voie croisée. Le bruit dans ce cas est ce qui n'est pas commun aux deux voies. Une fois que le bruit est connu, une correction va être appliquée, soit par la méthode d'Ephraim & Malah, [29, 30], soit par la méthode de Scalart, [74]. Ces méthodes sont décrites dans l'article de Cappé, [13] et récapitulé dans Scalart, [74].

Correction du bruit par la méthode d'Ephraim & Malah

Le principe de la réduction de bruit introduit par Ephraim & Malah repose sur la technique de MMSE,[29] et MMSE-LSA, [30], respectivement « Minimum Mean Square Method » et « Minimum Mean Square Method - Log Spectral Amplitude ». Elle est fondée sur des périodes d'estimations à court terme. C'est une méthode très populaire qui donne des résultats très satisfaisants pour la réduction du bruit musical.

La fonction de gain pour la trame k est la suivante :

$$G_k(f) = \frac{\sqrt{\pi}}{2}\sqrt{\frac{1}{1+R_{post}(f)}x\frac{R_{prio}(f)}{1+R_prio}(f)}M\left[(1+R_{post}(f))(\frac{R_{prio}(f)}{1+R_prio}(f))\right] \tag{1.43}$$

Où :

– $R_{prio}(f)$ est le RSB du signal a *priori* dans la trame donné par l'équation 1.44

$$R_{prio(f)} = (1-\alpha)h(R_{post}(f)) + \frac{\alpha\left|G_{k-1}(f)Y_{k-1}(f)\right|^2}{\gamma_k(f)} \tag{1.44}$$

– α est le lissage temporel souvent égale à 0.99

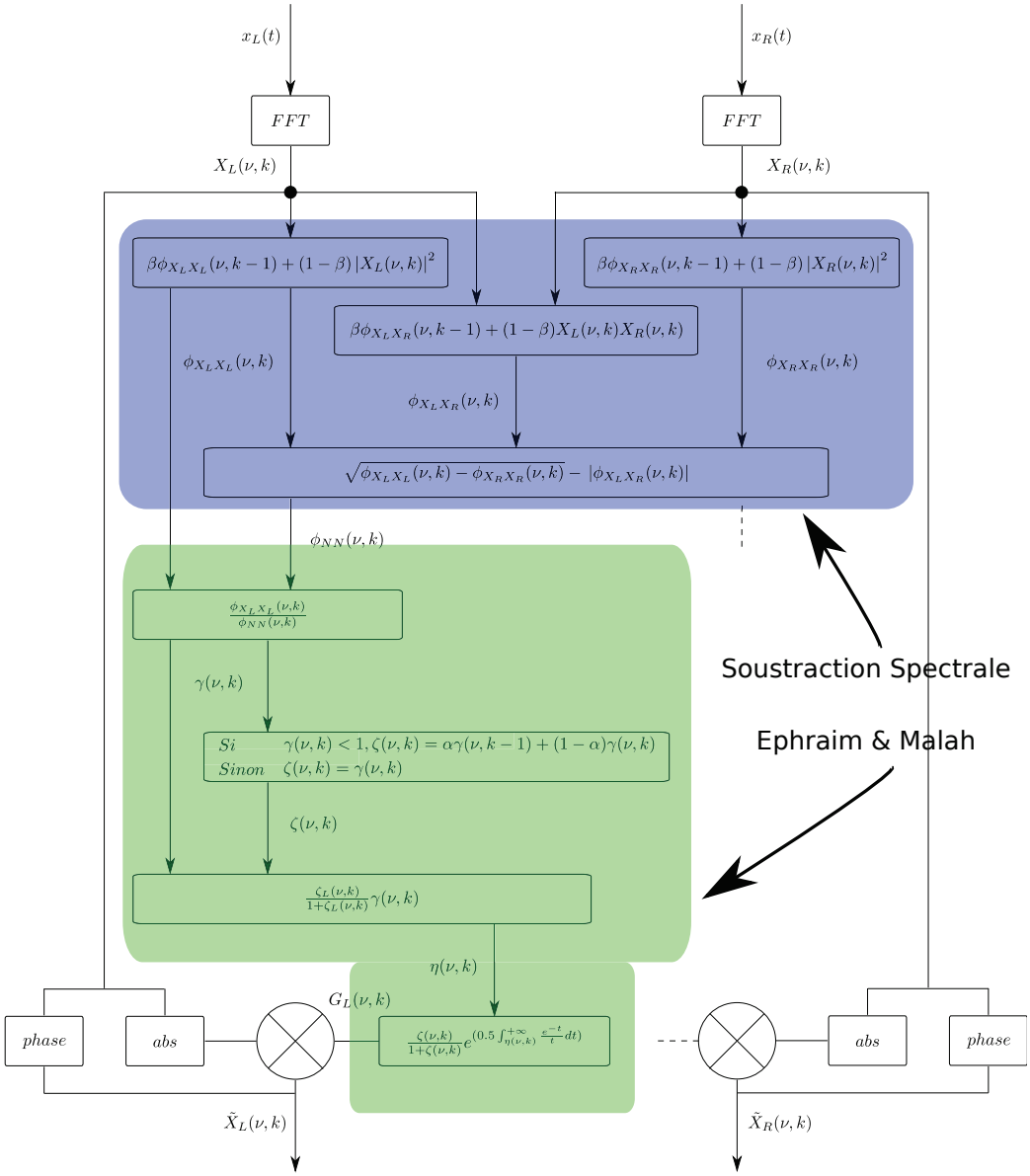


FIGURE 1.5 – Synoptique de Doerbecker combiné avec Ephraim & Malah

– $R_{post}(f)$ est le RSB calculé a *posteriori*
– M est une fonction qui est définie par l'équation 1.45

$$M(x) = e^{\frac{-x}{2}}\left[(1+x).I_0\frac{x}{2} + x.I_1\frac{x}{2}\right] \tag{1.45}$$

Où I_0 et I_1 représente les fonctions de Bessel modifié d'ordre 0 et 1.

Cette equation, 1.43, ne doit pas avoir de valeurs négatives. Pour compenser les valeurs négatives de R_{post}, la fonction M effectue les calculs seulement sur la partie positive ; la condition est que $h(x) = x$ si $x \geq 0$ et $h(x) = 0$ sinon. La propriété de lissage fréquentiel (paramètre α) du gain permet d'atténuer le bruit musical ainsi que les distorsions.

Cappé, [13] fait une analyse du gain en fonction du R_{prio}. et il montre que pour des valeurs très faibles de R_{prio}, une très forte atténuation est effectuée sur le gain. Cette propriété est très utile pour les segments de parole à faible énergie.

L'amélioration apportée en 1985, [30] en prenant le logarithme du module du signal se rapproche de l'échelle subjective de la perception humaine. L'expression du gain devient alors :

$$G_k^{LSA}(f) = \left(\frac{R_{prio}(f)}{1 + R_{prio}(f)} \right) . e^{\frac{1}{2} \int_{\nu_k(f)}^{+\infty} \frac{e^{-t}}{t} dt} \tag{1.46}$$

Où :
– $\nu_k(f)$ est donné par l'équation : $\nu_k(f) = (\frac{R_{prio}(f)}{1+R_{prio}(f)}) R_{post}(f)$
– l'intégrale est appelée intégrale exponentielle.

Loizou, [56] a étudié les différences entre les deux correcteurs, avec et sans logarithme a montré que $G_k^{LSA}(f)$ est souvent inférieur de 3 dB par rapport à $G_k(f)$ pour les mêmes valeurs de RSB. Cette dernière permet alors de corriger le bruit musical sans introduire plus de distorsion.

Correction du bruit par la méthode de Scalart

Scalart, [74] a proposé une autre approche pour la correction de l'estimation du bruit. Il a créé un filtre semblable au filtre de Wiener.

$$G_k(f) = \sqrt{\frac{R_{post}(f) - 1}{R_{post}(f)}} = \sqrt{1 - \frac{1}{R_{post}(f)}} \tag{1.47}$$

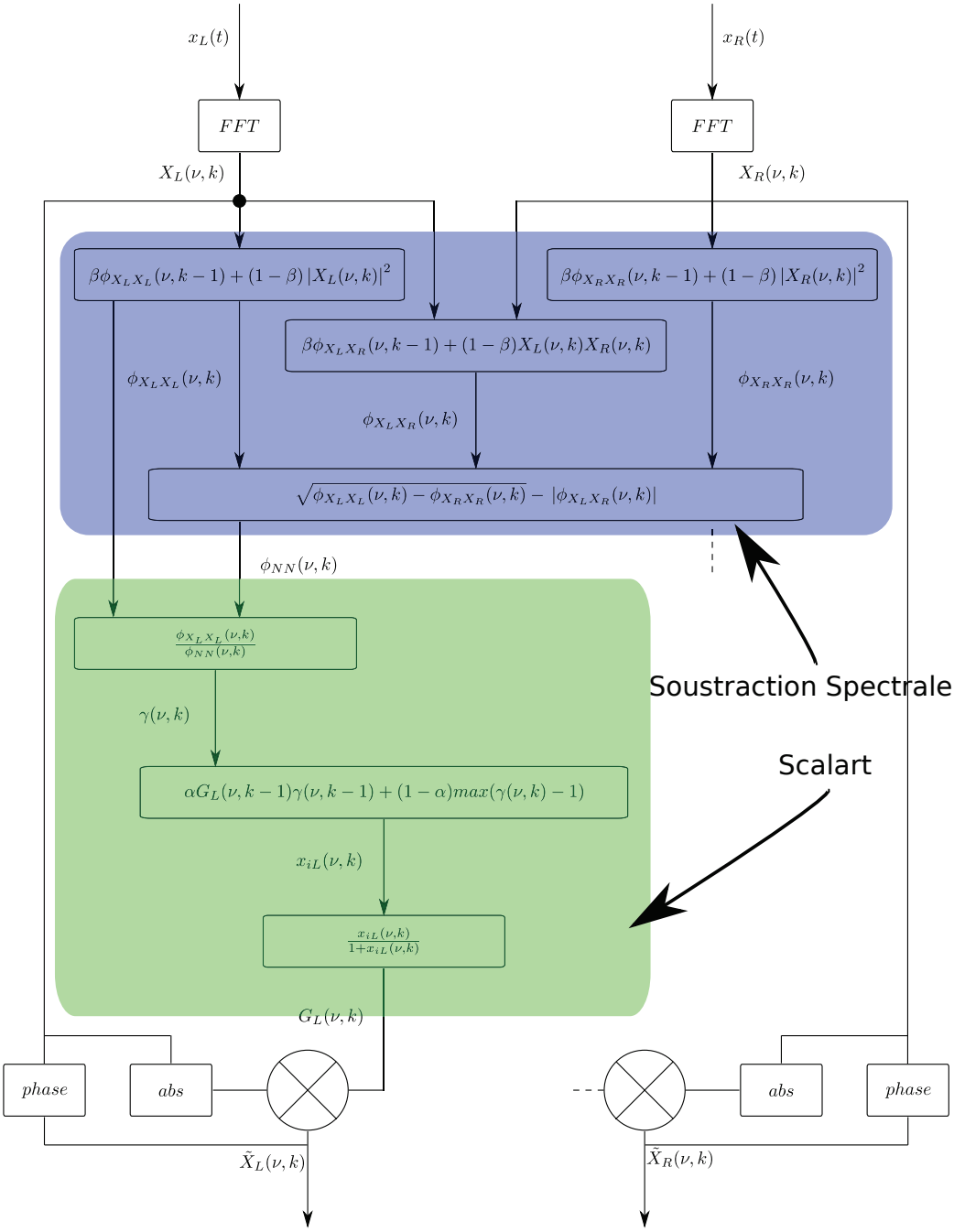


FIGURE 1.6 – Synoptique de Doerbecker combiné avec Scalart

Elle est basée sur la technique simple microphone. Le but de cette méthode, contrairement à la précédente, est qu'il faut pouvoir garder facilement le

temps réel pour les machines qui utilisent le système. Contrairement à Ephraim & Malah, Scalart base sa méthode sur le RSB et non sur les statistiques.

Correction du bruit résiduel par le filtrage de Wiener

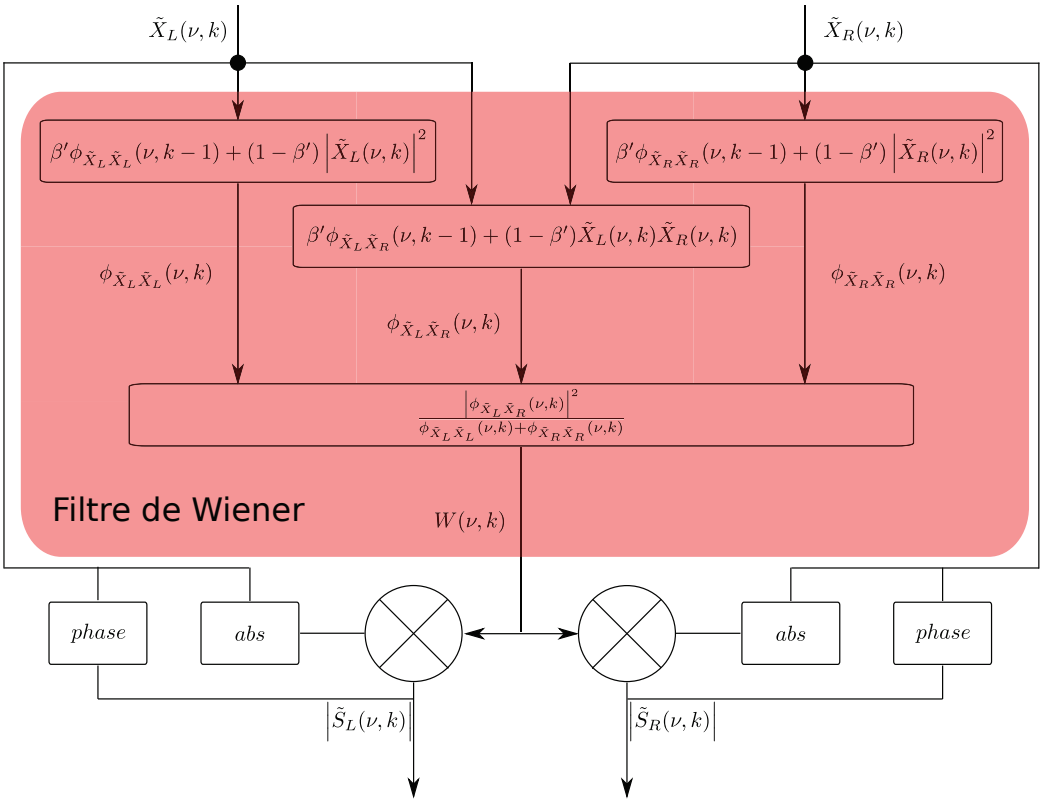


FIGURE 1.7 – Synoptique de Doerbecker, filtre de Wiener commun a Ephraim & Malah et Scalart

Une fois que l'une ou l'autre méthode est appliquée aux signaux bruités, il y a généralement une étape de filtrage de Wiener. Ceci, permet d'optimiser le RSB de la trace correspondante. La formule du gain est identique à la formule en mono-voie. L'estimation du bruit est le rapport des du produits croisés avec la piste correspondante.

1.3.5 Récapitulatif des formules de gain de réduction du bruit dans le cas du binaural

A l'instar de Scalart, [74], les différences de calcul de gain sont mieux visibles lorsque représentées dans un tableau, 1.1.

Méthode utilisée	Gain de correction
Estimation de la puissance [9, 74]	$G(f) = \sqrt{\frac{R_{post}(f)-1}{R_{post}(f)}}$
Filtre de Wiener [53]	$G(f) = \frac{R_{post}(f)-1}{R_{post}(f)}$
MMSE [13, 28, 29]	$G(f) = \frac{\sqrt{\pi}}{2}\sqrt{\frac{1}{1+R_{post}(f)}\frac{R_{prio}(f)}{1+R_{prio}}}(f)M\left[(1+R_{post}(f))(\frac{R_{prio}(f)}{1+R_{prio}}(f))\right]$
MMSE-LSA [13, 28, 30]	$G(f) = \left(\frac{R_{prio}(f)}{1+R_{prio}(f)}\right).e^{\frac{1}{2}\int_{\nu_k(f)}^{+\infty}\frac{e^{-t}}{t}dt}$

TABLE 1.1 – Récapitulatif des différentes méthodes de réduction de bruit

On peut remarquer les similarités entre le filtrage de Wiener et la proposition de Scalart. On pourrait étendre ce tableau à beaucoup d'autres fonctions de correction comme la technique de Mac Maulay, [59] ou bien avec sa modification plus tard avec Malpass, [59, 104]. Néanmoins, dans les futurs chapitres, les algorithmes développés seront basés sur le beamformer fixe sur la méthode de Doerbecker.

1.3.6 Coefficient de ré-injection

Le problème d'un algorithme binaural (qui peut être aussi un avantage) est qu'il localise une direction. Ce qui revient à dire que les algorithmes perdent ou atténuent les autres directions. Dans ce cas, l'avantage de la localisation risque d'être perdu si un algorithme trop « puissant » est appliqué aux signaux d'entrées.

Dans son article de 2008, van den Bogaert [91] propose une méthode simple pour essayer de diminuer la perte de la localisation. En effet, en ré-injectant

légèrement du signal bruité (qui contient les informations de localisation et les indices binauraux) dans le signal débruité (qui ne contient plus que ce qui provient de face), la localisation serait renforcée. La formule appliquée est la suivante 1.48 :

$$s(t) = (1 - \alpha)x'(t) + \alpha.x(t) \tag{1.48}$$

Où :
– s(t) est la sortie après la ré-injection
– x'(t) est le signal à la sortie des algorithmes
– x(t) est le signal d'entrée de l'algorithme

Chaque voie est indépendante et ce système peut être appliqué aussi en mono-voie. L'effet de ce facteur sera étudié dans le chapitre 4.

1.4 Et sans les mathématiques !

Ce premier chapitre montre beaucoup de mathématiques pour la réduction de bruit. Lorsqu'un signal arrive sur un microphone, il contient obligatoirement du signal utile et du bruit. Les algorithmes de débruitage ont pour but d'enlever le bruit en gardant soit une direction si on est dans le cas de plusieurs microphones, soit la parole, s'il n'y en a qu'un.

Il existe différents algorithmes qui permettent cette réduction du bruit comme la soustraction spectrale qui comme son nom l'indique permet de soustraire le bruit au signal. Ensuite, généralement après une soustraction spectrale, on place un filtre de Wiener qui optimise le RSB de la trace. D'autres algorithmes comme le beamformer, sélectionnent une cible sur un plan pour localiser une source.

Les problèmes des algorithmes de débruitage et qu'ils atténuent les signaux mais ajoutent des distorsions dans le signal qui parfois peuvent être plus gênantes que le bruit de départ. Des méthodes existent pour palier à ce problème bien que toutefois, il faut un compromis entre le débruitage et la distorsion.

Conclusion

Dans ce chapitre sur la réduction du bruit mono-voie et stéréophonique, le principe des algorithmes de débruitage a été expliqué. Chaque algorithme a un principe différent pour obtenir en sortie un signal contenant moins de bruit.

L'avantage de débruiter un signal permet de mieux comprendre la parole. Cependant, si l'on débruite « trop », l'information peut devenir distordue et introduire un biais dans les calculs. On appelle le bruit résultant, le bruit musical du fait de l'écoute (caractère) tonale. Les méthodes d'Ephraim & Malah ou encore de Scalart permettent d'atténuer ce signal. Les techniques de type Beamformer permettent quant à elles de localiser une source parmi un plan.

Dans le prochain chapitre, une présentation de l'implant cochléaire avec les différentes stratégies de codage seront expliquées puis dans le chapitre 3 et 4, l'utilisation des algorithmes décrient dans cette partie sera développée.

Chapitre 2

Implant Cochléaire, application à la binauralité et simulation.

Introduction

L'audition est un sens fragile. Elle nous permet de localiser, d'écouter, et de faire les activités quotidiennes. Elle subit cependant les attaques du temps qui de manière irréversible conduit souvent à une surdité. Ce handicap majeur a des effets sur la vie sociale des personnes.

Ces dernières années, de gros progrès ont été effectués pour réhabiliter ces personnes. L'arrivée du numérique, des nouvelles technologies, de la miniaturisation des composants permet pour les personnes atteintes d'une surdité, de « corriger » le problème. Les aides auditives permettent de réhabiliter les personnes sourdes et malentendantes alors que l'implant cochléaire permet de réhabilité les personnes sourdes profondes.

Il existe plusieurs types de surdité ; elles sont classées en trois classes.

– Les surdités de transmission, de type mécanique et qui sont dues à une atteinte du tympan, des osselets et des éléments anatomiques qui assurent la transmission dans l'oreille moyenne.
– Les surdités de perception, qui sont dues à une atteinte du fonctionnement

neuronal de l'audition. Dans ce cas, le problème est situé dans l'oreille interne.
– Les surdités mixtes, qui comme leur nom l'indique, sont mixtes entre la surdité de transmission et la surdité de perception.

2.1 Aides auditives

Les aides auditives couramment appelées prothèses auditives permettent une amplification et une correction du son et apportant une efficacité reconnue, mais variable suivant la surdité. Elles sont capables d'améliorer la communication, de débruiter le signal environnant, d'améliorer certaines fréquences ainsi que la compression en fonction de l'intensité du signal.

L'aide auditive reçoit un signal acoustique et renvoi un signal acoustique. Il n'y a pas de transformation physique du son.

Il existe deux types d'aide auditive :
– Les contours d'oreille ; ils sont portés sur la pavillon et reliés au conduit auditif par un tube dit « embout ». Cet embout assure l'étanchéité acoustique entre le tube et le tympan. Ceci a pour effet d'éviter l'effet Larsen.
– Les intra-auriculaires, qui sont positionnés directement dans le conduit auditif et quasiment invisibles. Ce dernier appareil ne peut être utilisé que pour des « petites surdités ». En effet, ils ne sont pas capables de fournir une amplification suffisante pour les sons élevés.

2.2 Implant cochléaire

2.2.1 Introduction

L'implant cochléaire permet de réhabiliter comme l'aide auditive les patients atteints de surdité profonde à totale et qui ne tirent pas d'effet d'une amplification classique. Il permet de compenser les impulsions électriques

délivrées par les cellules ciliées par l'intermédiaire d'électrodes placées dans la cochlée pour que les impulsions du nerf auditif soient interpréter par le cerveau.

2.2.2 Historique

Les premiers implants cochléaire ont vu le jour vers la fin des années 1960, Simmons, [78] est le premier à parler de l'implant cochléaire puis House, [39]. Au début, les implants ne possédaient qu'une seule électrode. Les techniques de chirurgie sont apparues ensuite pour affiner le positionnement de l'implant, [18]. Ensuite sont apparue les implants multi-électrodes, [19]. Vers la fin des années 1970, une nouvelle avancée avec l'implantation bilatérale a vu le jour, [41], malheureusement, les premiers résultats n'étant pas très bons en terme de performance par rapport au monaural, les conclusions ne donnaient pas un bon avenir au bilatéral, [3]. Avec l'arrivée du numérique, les performances des implants ont tout de suite été meilleurs, le chorimac possédait 15 électrodes, [17].

L'implant cochléaire est maintenant « une utilité lorsque les patients sont sourds totaux ». Il est généralisé et l'intervention est rapide.

2.2.3 Caractéristiques de l'implant cochléaire

Tous les types d'implant cochléaire possèdent deux parties : une partie externe qui permet de capter le son à l'aide d'un microphone, puis de traiter via le microprocesseur les signaux. Ensuite une partie interne va simuler le nerf auditif grâce aux électrodes. La figure 2.1 résume les différentes parties.

Description de la partie externe

Les accessoires principaux de la parie externe sont le microphone, le processeur et l'antenne. Le son est capté par le microphone, puis par une FFT, traité au niveau fréquentiel pour sélectionner les électrodes à simuler, la compression et le codage sont ensuite appliqués.

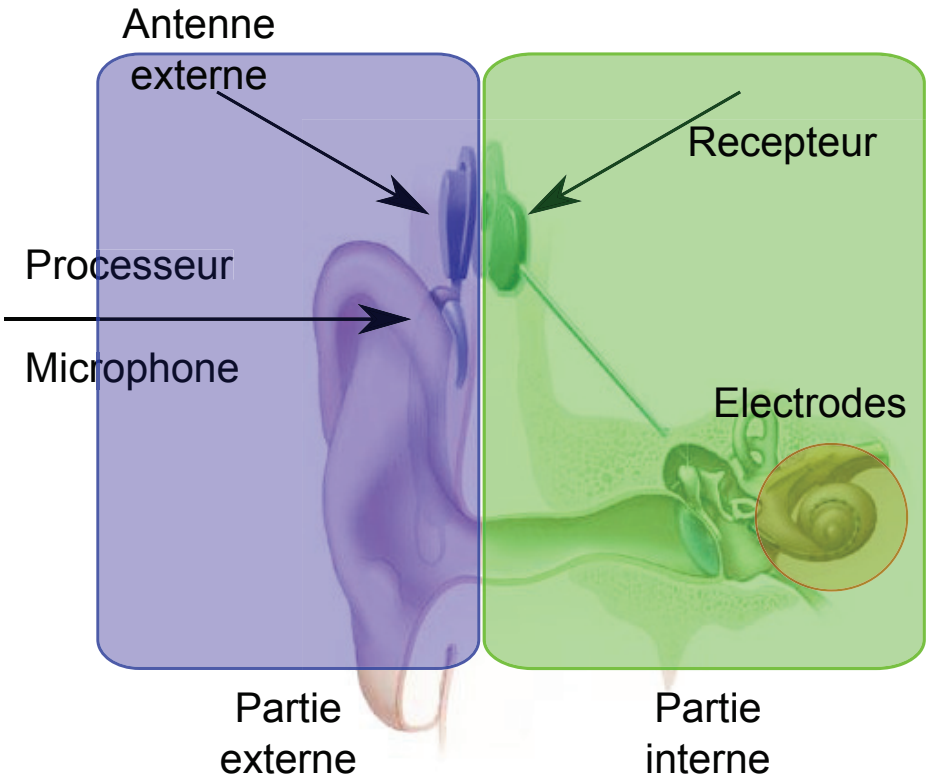


FIGURE 2.1 – Représentation des parties externe et interne de l'implant cochléaire

Description de la partie interne

La partie interne contient une bobine qui capte le signal émis pat l'antenne externe. Puis elle va simuler les différentes électrodes situées dans la cochlée.

2.2.4 Principe du fonctionnement de l'implant cochléaire

Le principe du fonctionnement de l'implant consiste à capter un signal sonore et à le transmettre sous forme d'impulsions au nerf auditif par l'intermédiaire des électrodes. Le but est de faire percevoir aux personnes porteuses de l'implant les différences acoustiques et ainsi de recréer un environnement audible. Il s'agit de faire percevoir la parole, suite à une stimulation électrique. La figure 2.2 montre les différentes étapes du

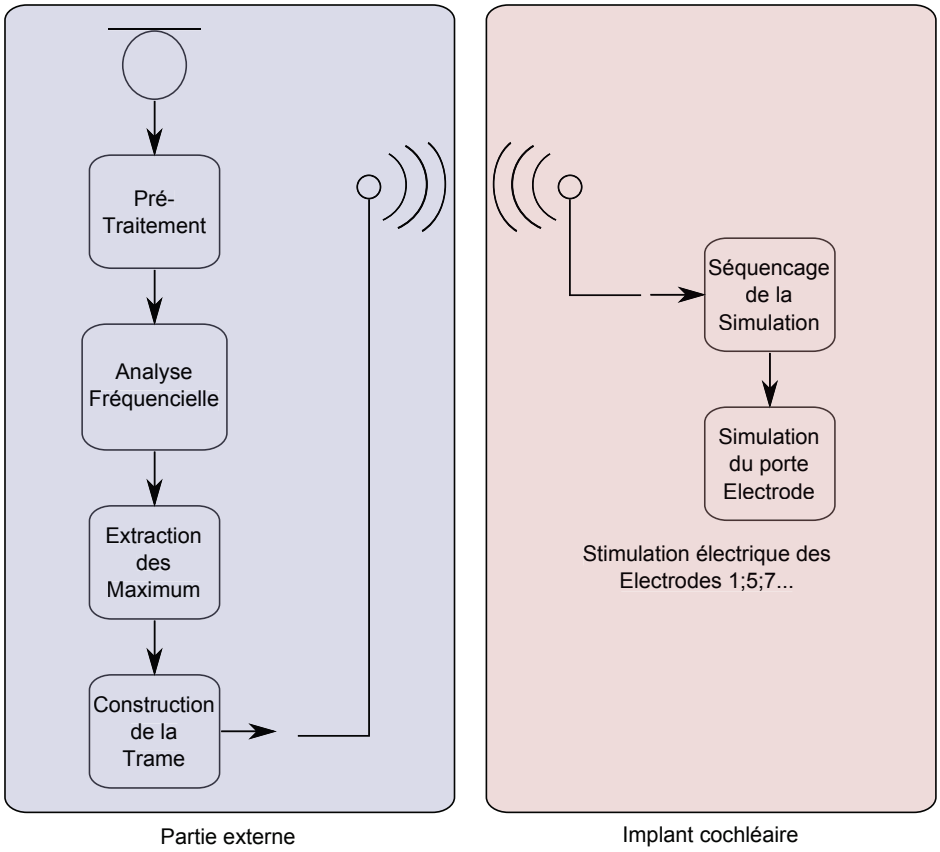


FIGURE 2.2 – Principe de fonctionnement de l'implant cochléaire

traitement.

Pré-traitement

La première phase de traitement consiste à enregistrer le son, l'amplifier et à apposer un filtre anti-repliement. Puis une pré-emphase (présentée dans le prochain chapitre) est appliquée pour que les hautes fréquences ne soient pas trop pénalisées.

Analyse fréquentielle

Deux possibilités sont possibles pour une analyse fréquentielle du signal. La première est de poser un banc de filtres qui contient N filtres passe bande suivant un découpage spécifique. La seconde solution est d'appliquer une FFT

sur le signal. Dans les deux cas, on connaît pour chaque largeur de bande les informations fréquentielles.

Ensuite, un regroupement par bandes de fréquences est effectué pour obtenir un nombre de filtres égal aux nombre d'électrodes. Chaque bande contient une F_{min} et une F_{max} et la fréquence centrale F_{centre} qui est la moyenne géométrique entre F_{min} et F_{max}. La non linéarité des fréquences (tonotopie) dans le système auditif impose un découpage linéaire/logarithmique.

Si l'on considère une zone linéaire de l'audition, la largeur de la bande est donnée par l'équation suivante :

$$\frac{F_{max} - F_{min}}{N} \tag{2.1}$$

Où N représente le nombre de bandes.

Il existe des découpes plus précises et plus ressemblantes à l'audition basées sur des formules lin/log. Par exemple la formule en barks décrite par Tjomov, [85] en 1971, qui considère les fréquences linéaires dans les basses fréquences et non linéaires pour les hautes fréquences pour arriver dans un nouvel espace (l'espace des barks) qui est linéaire. La formule est présentée si dessous :

$$B(f) = 6.7.argsinh(\frac{f-20}{600}) \tag{2.2}$$

Où :
– f représente les fréquences
– B(f) représente la transformée en bark de la fréquence.

La largeur du filtre a aussi été étudiée par Skinner, [79]. Trois types de découpage du filtre fréquentielle existent. L'échelle logarithmique, l'échelles de mels, [31], et les bandes critiques de Zwicker, [105]. Pour la reconnaissance de parole, le dernier type de filtre conduit aux meilleurs résultats ensuite l'échelle logarithmique puis l'échelle des mels. Le découpage suivant l'échelle

des mels présente beaucoup de bandes autour du formant F_1, il y en a environ 4. Le découpage suivant les bandes critiques en présente 6 au minimum.

Une fois le découpage en bandes de fréquence effectué, l'enveloppe du signal est calculée pour obtenir son énergie qui sera transmise aux électrodes. Deux façons différentes permettent de la calculer. La première méthode est de redresser le signal puis de filtrer passe bas. Elle est très rapide, robuste et le temps de calcul est faible. La fréquence de coupure du filtre passe bas contrôle la profondeur de la modulation des enveloppes. Aucune différence significative n'a été observée sur la fréquence de coupure, [76]. La seconde façon de détecter l'enveloppe est la transformée de Hilbert. Si l'on considère un signal de parole x(t) et sa transformée de Hilbert, y(t) est la réponse d'un filtre de quadrature dont la réponse impulsionnelle est la suivante, 2.3 :

$$h(t) = \frac{1}{\pi t} \tag{2.3}$$

Le signal de sortie y(t) s'écrit alors :

$$y(t) = h(t) * s(t) \tag{2.4}$$

Où, l'opérateur $*$ désigne le produit de convolution. Dans le domaine spectral l'équation devient alors :

$$Y(f) = H(F).S(F) \tag{2.5}$$

La Transformée de Fourier de h(t) est donnée dans l'équation 2.6 :

$$H(f) = -j.signe(f) \tag{2.6}$$

tel que $j^2 = -1$ et signe(f) est donné par l'équation 2.7 :

$$signe(f) = \begin{cases} 1 & \text{si} f > 0 \\ 0 & \text{si} f = 0 \\ -1 & \text{si} f < 0 \end{cases} \tag{2.7}$$

La détermination de l'enveloppe se fait en calculant le module du nombre complexe calculé dans l'équation 2.8.

$$z(t) = x(t) + jy(t) \tag{2.8}$$

Puis l'enveloppe :

$$e(t) = |z(t)| = \sqrt{x(t)^2 + y(t)^2} \tag{2.9}$$

Les deux techniques de détection de l'enveloppe fonctionnent correctement. La différence entre ces deux vient du fait qu'il n'y a pas de déphasage lors du filtre avec la transformée de Hilbert alors qu'il y en a un pour le filtrage passe bas.

Une fois l'extraction de l'enveloppe effectuée, le calcul de l'énergie de celle-ci par bande de fréquence est effectuée, de la même façon que lors du traitement du bruit. Puis vient l'extraction des maxima. Ils correspondent aux bandes de fréquence ayant le plus d'énergie. Dans le cas de l'implant monaural, si 18 maxima sont choisis, alors les 18 bandes de fréquences les plus fortes seront sélectionnées.

Enfin le train d'impulsions pour les électrodes est alors fabriqué. La stimulation des électrodes se fait ensuite suivant la stratégie de codage.

2.2.5 Descriptions des différentes stratégies de codage

Plusieurs stratégies de codage existent (suivant les constructeurs), elles peuvent être séparées en trois catégories qui vont différer par des cadences de simulation, extraction de pics...

Stratégies basées sur l'extraction des formants

Multi-Peak ou Mpeak ,
Le processeur de l'implant réalise l'extraction des formants F_0, F_1 et F_2 et utilise par la suite trois filtres passe-bande pour transmettre quelques

informations supplémentaire dans les hautes fréquences. Dans cette stratégie, la cadence de stimulation est liée au fondamental de la voix (sons voisés) ou aléatoire (sons non voisés).

Spectral Maxima Sound Processor ou SMSP ,
Cette stratégie n'est plus utilisée actuellement. Les six bandes d'amplitude maximum sont extraites et stimulent les six électrodes correspondantes.

Stratégies qui favorisent la représentation fréquentielle

Adaptative Simulation Rate ou ASR ,
La vitesse de stimulation des électrodes est adaptative. C'est une évolution de la stratégie SPEAK avec une détection des passages transitoires (comme les consonnes) et stationnaires (les voyelles).

Spectral Peak Strategy ou SPS ,
Le signal du microphone est découpé en vingt bandes de fréquences. La vitesse de stimulation est fixe. Le principe est de simuler les plus grandes énergies. Généralement, six maxima sont sélectionnés.

Stratégies qui favorisent la représentation temporelle

Compress Analog ou CA ,
Quatre électrodes simulent quatre zone de la cochlée. L'étage de filtrage est formé par un banc de filtres passe bande analogique. L'énergie du signal issu de chaque bande sera convertie en stimulus électrique.

Continuous Interleaved Sampling ou CIS ,
C'est la stratégie de traitement la plus utilisé actuellement. Les électrodes sont stimulées séquentiellement (les unes après les autres) ce qui diminue les interactions entre les électrodes. La structure temporelle du signal est bien conservée. Mais la rapidité de traitement fait que les informations fréquentielle sont moindres.

Simultaneous Analog Stimulation ou SAS ,
L'oreille traite le son simultanément dans l'ensemble de la cochlée. Pour imiter cette conception naturelle, l'implant peut fournir des impulsions électriques simultanément à divers endroits de la cochlée.

Stratégies dites hybrides

Spectral Peak Strategy ou SPEAK ,
Le signal capté par le microphone est découpé en vingt bandes de fréquences. Ensuite un certain nombre d'électrodes correspondant aux maximums d'énergie sont utilisées pour la stimulation. Les maxima sont déterminés par le niveau sonore du signal d'entrée et de composition spectrale. Une variante existe en basant la stimulation sur la fréquence fondamentale de la voix.

Advanced Combination Encoders ou ACE ,
La dernière stratégie de codage de l'implant est ACE. Elle prend le meilleur de CIS et de SPEAK. Dans CIS, la vitesse de stimulation importante est gardée et dans SPEAK on conserve le système de stimulation sélectionnant les sites d'activation de manière dynamique en fonction de l'énergie spectrale.

Le tableau 2.1 présente les quatre constructeurs d'implant cochléaire ainsi que leur stratégie de codage de l'implant.

Constructeurs	Processeur	Stratégie	Nombre d'électrodes
Advanced Bionics	HiRes 120	CIS	16
Cochlear	Nucleus Freedom	SPEAK	22
MedEL	COMBI 40+	CIS	12
Neurelec	Digisonic SP	ACE	20

TABLE 2.1 – Implants cochléaires commercialisés

Le principe de l'implant monaural, vient d'être revu dans cette section. L'arrivée d'un second implant ou bien l'implant binaural améliore non seulement la reconnaissance de la parole mais aussi la localisation dans l'espace des sources environnantes.

2.3 Implant cochléaire bilatéral

L'implant bilatéral est en fait un implant monaural sur chaque oreille. Son avantage est qu'il permet de localiser en plus de comprendre la parole dans le bruit, [89]. Un gain de $3dB$ dans le bruit est obtenu en reconnaissance de parole, [83]. En effet, il est logique que les résultats soient meilleurs qu'en monaural car l'Homme possédant deux oreilles, il est normal de mettre un implant sur chaque oreille si les deux sont défectueuses, [75, 97, 99]. De nombreuses études ont montré les bienfaits de la localisation, [32, 63, 88, 94], l'erreur est beaucoup plus faible lorsque les implants sont utilisés. Dunn, [27] compare les résultats entre bilatéral et monaural ; la conclusion est que la perception et la localisation des sons est supérieures dans le cas bilatéral.

Une étude proposée par Poeppel, [67] montre les différences de compréhension entre l'oreille gauche et l'oreille droite. Ainsi, il en déduit que l'hémisphère gauche et l'hémisphère droit du cerveau ont des informations différentes à traiter pour la compréhension du son. L'hémisphère droit, correspondant à l'oreille gauche, capterait et interpréterait préférentiellement les informations environnantes alors que l'hémisphère gauche, correspondant à l'oreille droite aurait une préférence sur la parole. Ces travaux ont ensuite été appliqués sur des personnes implantées cochléaires mais la variabilité entre les sujets n'a pas montré de résultats significatifs, [73].

2.4 Implant cochléaire binaural

2.4.1 Introduction à l'implant binaural

L'implant cochléaire binaural est un dérivé du monaural et du bilatéral. En effet, il s'agit d'un seul appareil, qui possède un seul processeur et qui va simuler les électrodes dans les deux oreilles. Il possède actuellement douze électrodes positionnées dans chaque oreille soit un total de vingt-quatre électrodes. La stratégie de codage utilisée est ACE. Il n'y pas beaucoup d'articles qui parlent de cet implant. Le premier article de Dillier, [20] en 1993

s'intéresse à la localisation et aux indices binauraux que peut apporter cette technologie. Les papiers suivants sont basée sur la technique chirurgicale, [86, 87] qui propose de passer par le sommet de la tête. Le dernier article en date est celui de Verhaert, [101] qui montre qu'il n'y a pas de différence de reconnaissance de la parole dans le bruit ainsi que de la localisation par rapport à l'implant bilatéral.



FIGURE 2.3 – Implant binaural

La figure 2.3 montre l'implant cochléaire binaural. Un microphone est placé sur chaque oreille. Un fil relie le microphone controlatéral au processeur puis il y a une seule antenne pour les deux porte-électrodes. Tout comme l'implant bilatéral, la localisation des sources et la perception des sons dans le bruit sont meilleurs que pour l'implant monaural.

Les avantages et les inconvénients de l'implant binaural et de l'implant bilatéral sont présentés dans le tableau 2.2.

Le principal avantage vient du fait de la synchronisation entre les deux voies. En effet sur l'implant bilatéral, les stimulations sont totalement indépendante et parfois avec des cadences différentes. Elles peuvent introduire

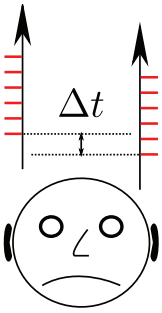
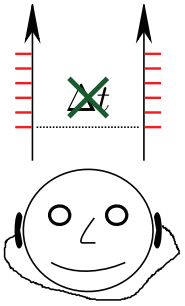
Implant Bilatéral	Implant Binaural
2 implants 2 traitements indépendants 2 stimulations asynchrones	1 implant 1 traitement 2 Stimulations synchronisées
	

TABLE 2.2 – Comparaison bilatéral VS binaural

des décalages « non perceptibles » mais qui peuvent induire une différence de temps interaural qui n'existe pas dans la réalité, [68]. Cependant la synchronisation pose des questions, Brown, [11], mesure l'effet de la synchronisation en mettant un jitter. Cependant, un biais est à prévoir car il effectue les mesures sur un implant bilatéral autrement dit, il n'y a pas de synchronisation entre les deux oreilles ; van Hoesel propose d'ajouter un jitter variable entre les deux oreilles, [95] qui améliorerait la localisation.

2.4.2 Traitement du signal effectué par l'implant bilatéral

Actuellement cet implant fonctionne comme deux implants monauraux, ou l'implant bilatéral. La figure 2.4 montre les voies du traitement du son.

Le son envoyé par la source est capté par les deux microphones, (flèches vertes) puis le son est envoyé au processeur, (flèche jaune). Le traitement du signal est effectué puis le son est envoyé à l'implant (flèche rouge) pour stimuler les deux cochlées. Ce qui est reçu sur le microphone de gauche stimule la cochlée à gauche et ce qui est reçue à droite stimule la cochlée à droite, (flèche bleue). Il n'y a pas de traitement croisé des informations.

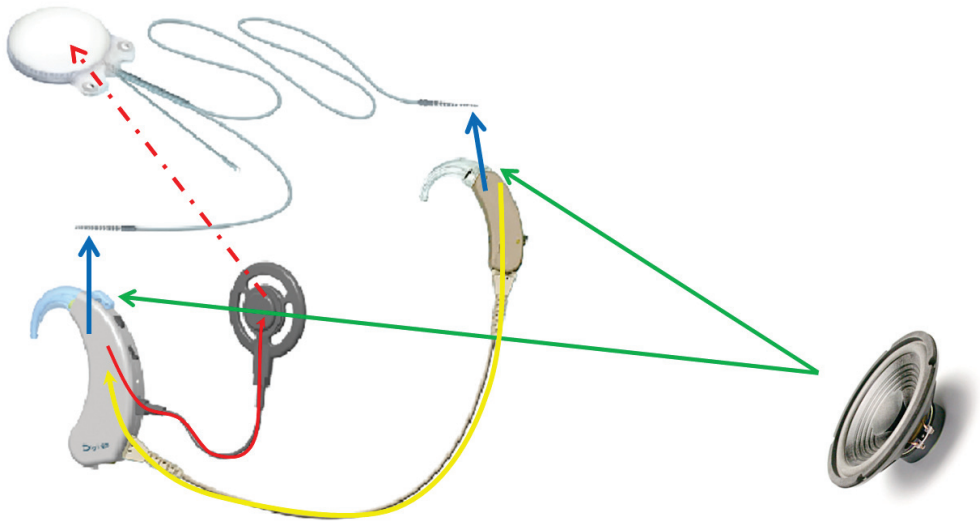


FIGURE 2.4 – Traitement du son dans l'implant binaural

2.5 Vocodeur de l'implant cochléaire

Du fait du trop petit nombre de personnes implantées en binaural, il est difficile d'établir un modèle pour les algorithmes de débruitage. De plus la variabilité inter-sujet étant très grande, il est impossible de définir un algorithme de réduction de bruit valable pour tous. Dans les études suivantes, la simulation du codage effectué par le vocodeur prenant les paramètres de base de l'implant binaural a été effectué. De nombreuse études ont recours à cette pratique pour obtenir un modèle avant de l'appliquer aux patients, [14, 15, 37, 36, 84]

Le synoptique 2.5 présente les caractéristiques de base d'un simulateur de l'implant cochléaire binaural. La fréquence d'échantillonnage est de $Fe = 16kHz$, la longueur de chaque trame d'acquisition est de $8ms$ soit 128 échantillons, avec un recouvrement (overlap) de $2ms$. Le spectre est exploré depuis la fréquence minimum est de $f_{min} = 325Hz$ à la fréquence maximum de $f_{max} = 6kHz$ suivant la répartition de l'échelle de bark, [85]. Il y a douze

électrodes par oreille et l'extraction de huit maxima.

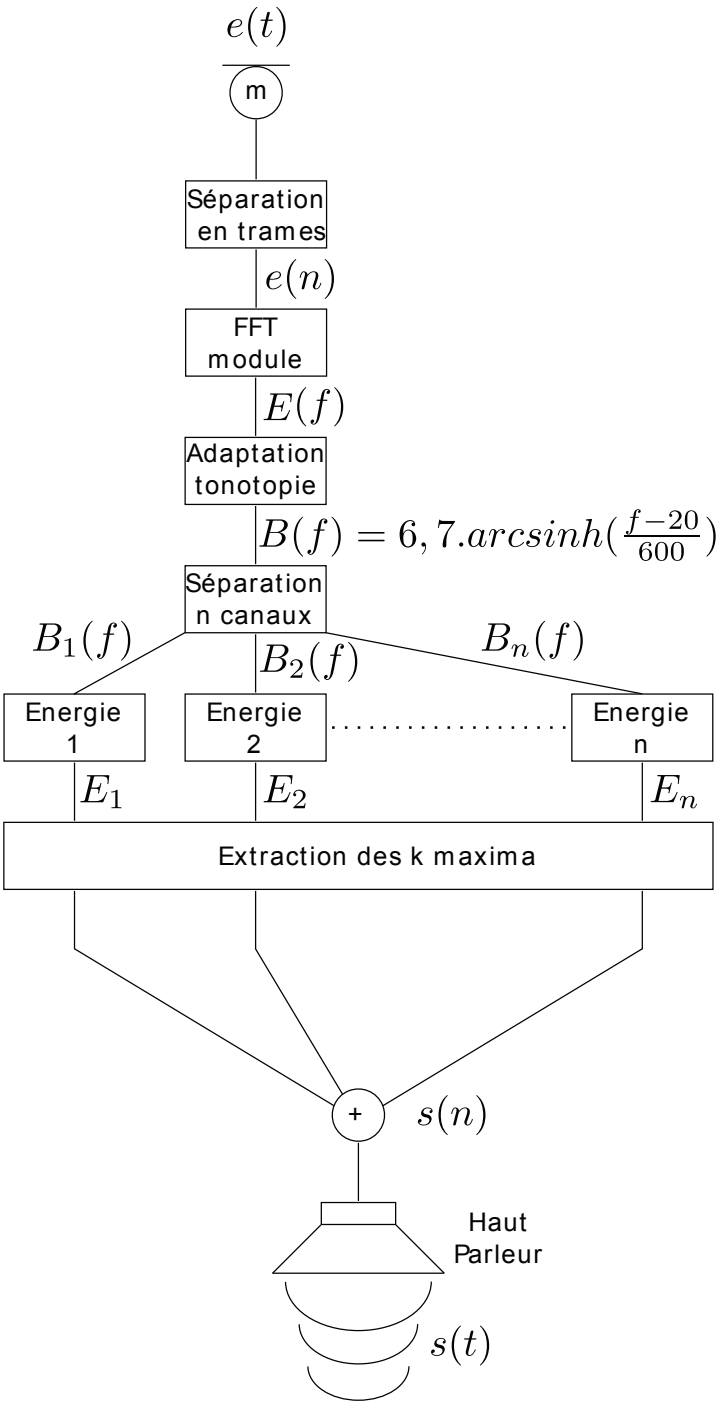


FIGURE 2.5 – Vocodeur simulant l'implant cochléaire

Le tableau 2.3 représente les douze différents bancs de filtres avec la fréquence centrale. Sur la figure 2.6, on remarque la répartition linéaire puis logarithmique de l'audition.

Electrodes	fréquence centrale		fréquence de coupure	
	bark	$f_{centrale}$ (Hz)	f_m(Hz)	f_M(Hz)
1	1,4	392	325	473
2	2,4	550	473	641
3	3,4	732	641	836
4	4,5	944	836	1067
5	5,5	1197	1067	1344
6	6,5	1502	1344	1679
7	7,5	1871	1679	2087
8	8,5	2322	2087	2585
9	9,5	2873	2585	3195
10	10,5	3550	3195	3945
11	11,5	4381	3945	4866
12	12,5	5402	4866	6000

TABLE 2.3 – Tableau des bancs de filtres pour la séparation des douze bandes de fréquences

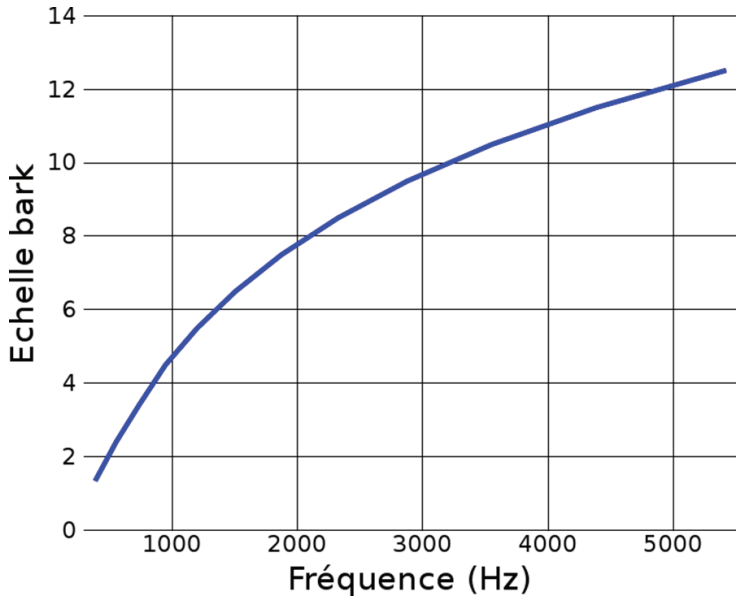


FIGURE 2.6 – Répartition fréquentielle suivant l'échelle de bark

Il existe plusieurs façons pour « remplir » les bandes de fréquences, [57]. La première est d'utiliser un bruit à bande étroite entre les fréquences min et max de la bande correspondante. La seconde est de remplacer les canaux par une sinusoïde à fréquence centrale avec un déphasage nul, c'est à dire que la sinusoide est la même pour toutes les trames. Les performances de la deuxième méthode ne sont pas différentes de la première méthode pour la compréhension de la parole.

La reconstruction du signal acoustique se fait en additionnant les différentes sinusoides.

Une fois le son reconstitué et modélisé par un vocodeur, on positionne un casque sur la tête de la personne et effectue la tâche que l'on souhaite.

Conclusion

L'implant cochléaire monaural, bilatéral ou bien binaural permet avant tout de réhabiliter les personnes atteintes de surdités profondes à totales. Plusieurs types de stimulation au niveau des électrodes existent, soit dans le domaine fréquentiel, soit dans le domaine temporel ou bien dans les deux domaines pour certaines stratégies. En mode monaural, la reconnaissance de la parole domine mais la localisation des sources sonores est un vrai challenge. L'implant bilatéral et l'implant binaural permettent de remettre l'aspect localisation. Cependant, l'implant binaural peut avoir un avantage sur l'implant bilatéral car il possède un seul processeur et peut permettre ainsi des traitements croisés entre les deux voies. Ce qui amène logiquement à ajouter un algorithme de traitement de réduction du bruit. Ces algorithmes sont exploités en reconnaissance de parole dans le chapitre 3 et en localisation dans le chapitre 4.

Chapitre 3

Reconnaissance de phonèmes après un débruitage binaural par l'intermédiaire d'un vocodeur.

Introduction

Depuis le début des années 2000, l'importance de « nettoyer » le signal arrivant aux oreilles est devenue primordiale. En effet, tous les domaines acoustiques (audibles) ont été modifiés ces dernières années. Par exemple, le téléphone portable doit isoler la voix qui parle et atténuer le reste. De même dans les voitures qui deviennent de plus en plus silencieuses. Ainsi, le traitement du signal de l'aide auditive et de l'implant cochléaire a largement été modifié pour essayer d'optimiser au maximum le signal arrivant aux oreilles. Naturellement, les méthodes de réduction de bruit ont aussi été portée par la généralisation et les capacités de calcul des processeurs numériques.

Le challenge pour les années à venir est d'arriver à récupérer et traiter le meilleur signal pour que l'oreille puisse correctement interpréter les

informations que l'on souhaite entendre. Depuis de nombreuses années, [3, 51], les implants cochléaires sont classiquement utilisés pour les personnes sourdes profondes. Cependant, dans les situations bruyantes du quotidien, les performances tombent très rapidement, [24, 99]. Comprendre la parole dans le bruit est devenue primordiale, [26].

Dans la littérature, van Hoesel, [99], a testé l'effet de la perception des mots dans le bruit en monaural et bilatéral sur cinq personnes implantées cochléaires. Lorsque le stimulus à retrouver était une phrase, les scores étaient proches de zéro lorsque la source utile provenait de face et le bruit provenant de plusieurs directions ($\pm 90^o$ et 0^o). Il démontre alors que la perception est meilleure dans le cas bilatéral que monaural notamment grâce aux différentes informations binaurales, et à l'ombre de la tête, [3, 26, 50]. Le port de deux implants permet alors une meilleure compréhension dans le bruit.

Les performances de reconnaissance de la parole et de localisation en se servant des informations binaurales, permettent de mieux comprendre l'environnement. On peut classer ses informations suivant trois types, [60, 61, 62, 70, 96].

– les « Différence de Temps Interaural », DTI, qui sont une mesure de la différence de phase entre les deux oreilles.
– les « Différence de Niveau Interaural », DNI, qui sont une mesure de la différence d'intensité entre les deux oreilles.
– l'ombre de la tête permet de différencier les sons ; elle est reliée au DNI

Ces indices permettent de mieux localiser les sons. En effet, les DTI, principalement situées dans les basses fréquences (inférieur à la demi-onde) de la tête permettent de situer les sons provenant d'une oreille à l'autre. Si un signal arrive à gauche, alors l'oreille gauche capte le son en premier puis l'oreille droite ensuite. En ce qui concerne, les DNI, ils agissent pour des fréquences beaucoup plus élevées, car elles arrivent en même temps aux oreilles. Si le son provient de gauche, alors l'ombre de la tête aura tendance a atténuer le signal sur l'oreille droite. Le son perçu à gauche sera plus fort

qu'à droite.

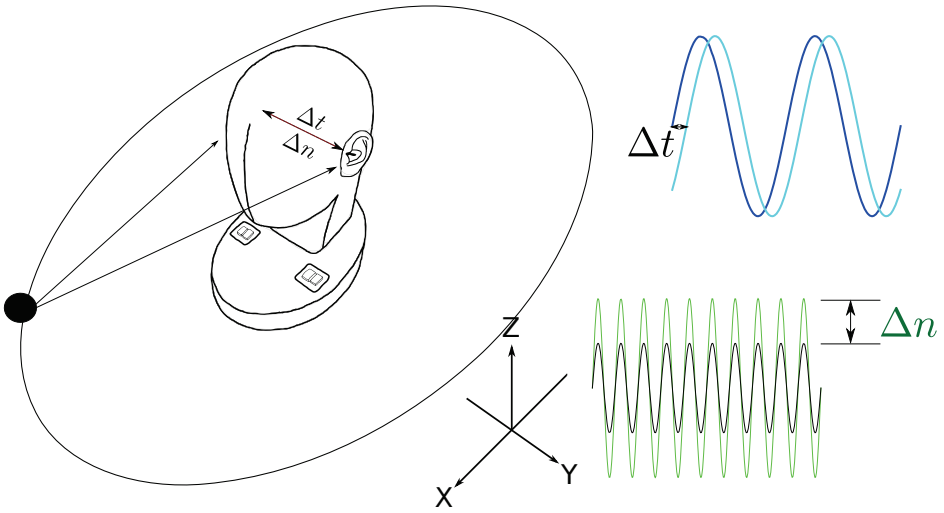


FIGURE 3.1 – Illustration des informations binaurales.

Ces informations sont particulièrement importantes dans les milieux bruités. L'absence de ces informations pour les personnes implantées cochléaires est un désavantages pour la compréhension des sons, [33, 50]. C'est pourquoi l'implantation bilatérale apporte de meilleurs résultats, notamment parce qu'il est plus facile de comprendre un signal si l'on connait sa provenance, [12, 99]. Il est bon de faire une localisation du son pour pouvoir mieux comprendre.

Pour permettre cette amélioration chez les sourds profonds, deux moyens peuvent être utilisés. Le premier est de mettre un implant cochléaire dans chaque oreille ; il y a alors deux implants, cette solution est dite bilatérale, [3, 12, 26, 99, 98]. La seconde solution, est d'avoir un seul implant cochléaire pour les deux oreilles, cette une condition est dite « binaurale », car il n'y a qu'un seul processeur pour les deux oreilles. L'avantage de ce dernier est qu'il permet de mettre un ou plusieurs algorithmes synchronisés dans le processeur et ainsi de créer et de traiter de façon binaurale le signal arrivant aux deux oreilles. Peu d'études parlent de cet implant, [50, 86, 87]. Cependant dans

la dernière étude sur l'implant binaural, les résultats ne montrent pas de différence avec l'implant bilatéral, [101].

Dans la mesure où l'implant binaural donne les mêmes résultats (performances) que l'implant bilatéral, l'ajout d'un algorithme de réduction de bruit mérite d'être envisagé. En utilisant un seul processeur, les techniques utilisées dans l'aide auditive vont pouvoir être exploitées. Les techniques conventionnelles comme la soustraction spectrale [9], et le filtrage de Wiener, [25, 93], ou encore la combinaison des deux, [23, 47] peuvent être appliquées aux traitements de l'implant. Les stratégies devront être légèrement différentes pour l'implant car ce n'est plus un traitement « acoustique-acoustique » mais un traitement « acoustique-électrique ».

Un algorithme binaural de réduction de bruit fonctionne différemment d'un algorithme classique mono-voie. En effet, celui-ci va localiser une direction, améliorer la parole et s'appuyer sur le décalage entre les microphones. Les performances de reconnaissance de parole dans le cas d'une simulation de l'implant binaural vont être testées dans ce chapitre. Dans le suivant, nous verrons les performances de localisation.

3.1 Matériel & Méthodes

Dans cette expérience, le signal utile sera toujours envoyé de face (0^o, ce qui la situation classique de la cible de face). En principe nous regardons, ce que nous écoutons, et généralement il provient de devant. Le bruit que nous testerons proviendra de cinq directions : -90^o à $+90^o$ par un pas de 45^o.

3.1.1 Matériel acoustique

Le matériel utilisé pour les enregistrements sont les listes monosyllabique de Lafon, [49]. Chaque liste contient dix-sept mots de trois syllabes. Il y a vingt listes en tout. En ce qui concerne le bruit, il est fabriqué de la façon suivante. Tous les mots de chaque liste sont mis bout à bout, puis le spectre

global est calculé. En mettant une phase aléatoire, le signal temporel est reconstruit par une iFFT. De cette façon, nous avons un speech noise qui contient le spectre de la parole des mots pour chaque liste.

3.1.2 Enregistrement du signal

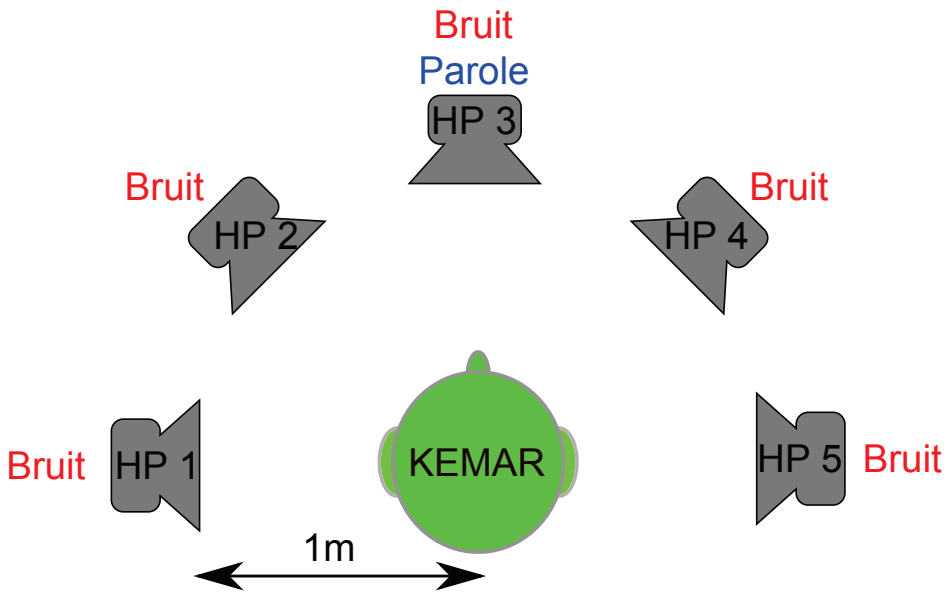


FIGURE 3.2 – Plan d'enregistrement des différentes pistes

L'enregistrement des différentes pistes se passe de la façon suivante, figure 3.2. La parole est toujours envoyée à 70 dB SPL et de face, l'intensité est mesurée par un sonomètre au niveau du HP 3 correspondant à l'angle de 0^o. Puis suivant la condition, le bruit est envoyé parmi les cinq haut-parleurs au rapport signal sur bruit désiré. Si par exemple le RSB est de $+3dB$, alors l'intensité du bruit mesuré au sonomètre sera de 67 dB SPL.

L'enregistrement se fait à l'aide d'un KEMAR, (figure 3.3, « **K**nowles **E**lectronic **M**anikin for **A**coustic **R**esearch »). Il permet de garder les informations binaurales. L'écart entre les deux oreilles est de $20cm$, figure 3.2. Les cinq haut-parleurs sont situés à $1m$ du centre de la tête. Un angle de

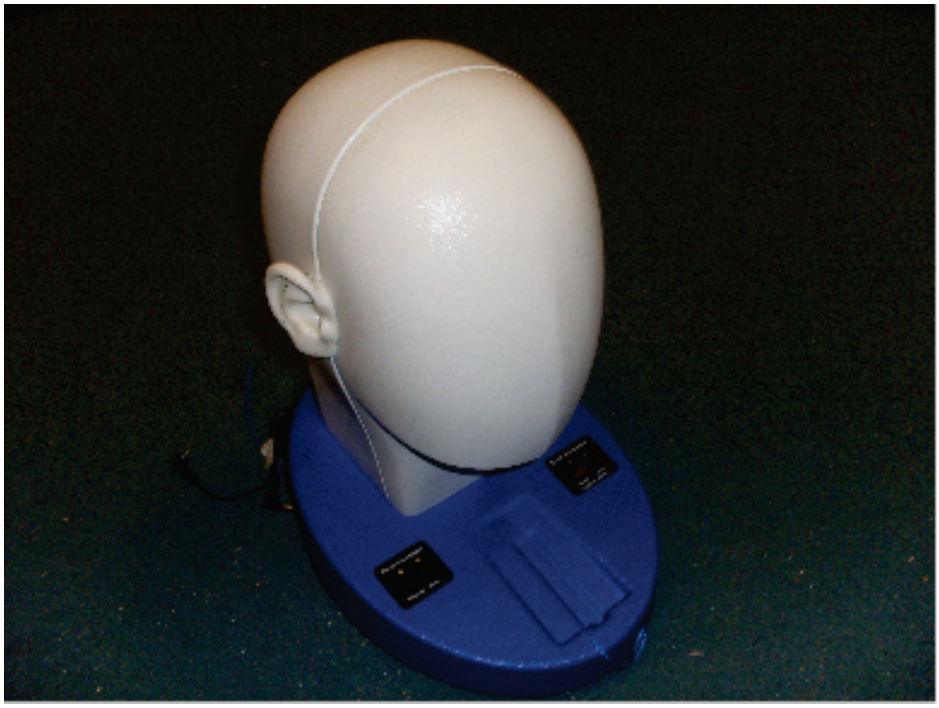


FIGURE 3.3 – Mr KEMAR

45^o les séparent. Sur la tête artificielle on receille un signal électrique stéréo correspondant au signal capté par les oreilles. Une carte son enregistre le signal avec une fréquence d'échantillonnage $Fe = 44100Hz$.

Au départ, le son capté est sous la forme tridimensionnelle (3D) puis traité en bidimensionnelle (2D). C'est à partir de ces enregistrements que les signaux vont être analysés. Ils sont tous composés par des informations binaurales avec un mixage entre le signal utile et le bruit.

3.1.3 Algorithmes de traitement du signal

Classiquement dans l'aide auditive, le signal capté est ensuite passé dans un filtre de pré-emphase, qui généralement compense le filtrage de la bouche et évite de toujours stimuler les basses fréquences. C'est un filtre de type

passe-haut, avec pour fonction de transfert 3.1, 3.4 :

$$H_z = 1 - 0.9z^{-1} \tag{3.1}$$

Représentant la combinaison : $y(n) = x(n) - 0.9x(n-1)$

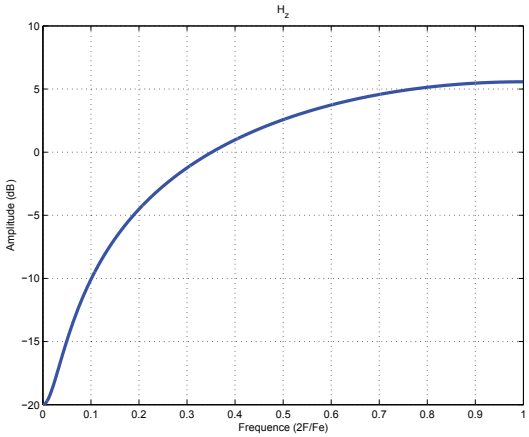


FIGURE 3.4 – Fonction de transfert $H(f)$ de la pré-emphase

Ainsi, lors du calcul pour les amplitudes des énergies des canaux du vocodeur, les électrodes stimulées ne seront pas toujours en basses fréquences.

Dans l'aide auditive, l'algorithme de débruitage est placé ensuite. Notre choix s'est porté sur deux algorithmes « classiques », le beamformer et le traitement proposé par Doerbecker.

Beamformer

Le beamformer, décrit dans le premier chapitre, permet de localiser une direction dans un plan. Dans ce cas, la direction par l'intermédiaire d'une cardioïde sera centrée sur le signal provenant de face, et les retards sur les deux cotés seront identiques et correspondant à la distance entre les deux oreilles. Les hautes fréquences seront rajoutées ensuite.

Algorithme de Doerbecker

Classiquement l'algoritme de Doerbecker peut se décomposer en une soustraction spectrale suivit d'un filtrage de Wiener. Dans la soustraction spectrale, plusieurs estimateurs de bruit ont été utilisés, celui d'Ephraim & Malah, [29], et celui de Scalart, [74]. Le premier à la particularité d'atténuer le bruit musical alors que le second améliore le RSB sur les différentes voies.

Ré-injection du signal d'entrée

Afin de minimiser les difficultés de localisation, van den Bogaert, [92], a proposé de ré-injecter une partie du signal d'entrée sur la sortie. En effet, les algorithmes de débruitage favorisant une source, il est alors difficile de localiser les autres sources environnantes. Cette ré-injection du signal d'entrée sur la sortie peut permettre de compenser. Ce paramètre est plus spécifiquement étudié dans le chapitre 4 sur la localisation. L'équation de la ré-injection est de la forme suivante :

$$s'(t) = (1 - \alpha)s(t) + \alpha.e(t) \tag{3.2}$$

Cette équation a déjà été vue dans le chapitre 1.

3.1.4 Vocodeur

Après le débruitage du signal, le vocodeur est utilisé pour « simuler » le codage introduit par l'implant cochléaire. L'implant binaural est composé de douze électrodes de chaque côté. Les trames d'acquisitions sont de $8ms$ avec un chevauchement de 75% soit $2ms$. Les huit maxima de chaque côté sont alors envoyés aux électrodes. La simulation acoustique sur normo-entendants reprend les mêmes caractéristiques.

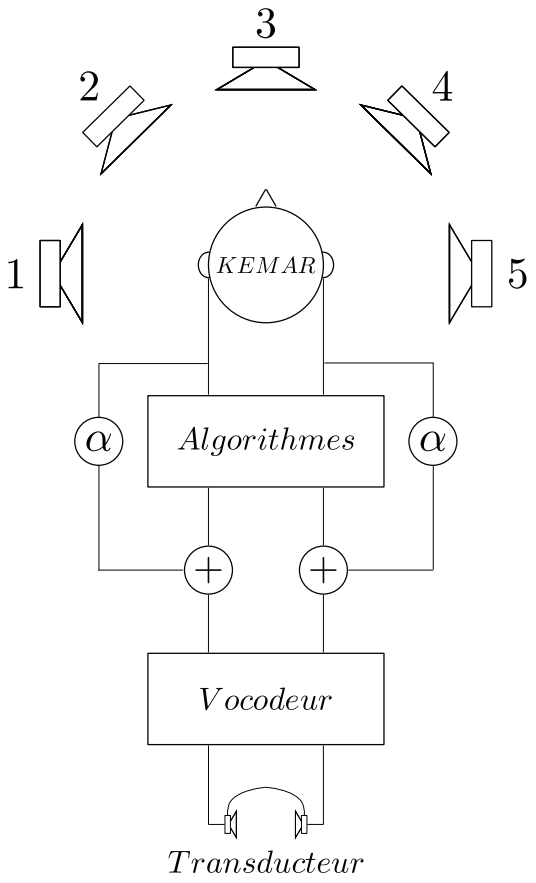


FIGURE 3.5 – Récapitulatif de l'expérience

3.1.5 Session

Sujets

Vingt-six personnes ont participé à cette étude, quatorze garçons et douze filles âgés entre vingt et vingt-six ans avec une moyenne d'âge de vingt-quatre ans. Tous les sujets avaient une audition normale, c'est à dire que tous les seuils sur l'audiométrie tonale étaient inférieurs à $20dBHL$.

Tous les sujets participants ont eu ensuite un entraînement au vocodeur puis, est réalisé la tâche proprement dite.

Entraînement

La tâche d'entraînement a consisté à répéter les mots composés de trois phonèmes, entendus dans le casque (transducteur, figure 3.5). Tant que le score mesuré en pourcent n'est pas supérieur à 80%, la tâche d'apprentissage est répétée. L'entraînement durait environ vingt minutes.

Reconnaissance de phonème

Les sujets devaient ensuite écouter un total de 150 listes, durant environ trois heures trente minutes, chaque liste correspondant à une condition expérimentale différente (détaillée ci-dessous). Les listes ont été choisies au hasard dans la banque de données (20 listes différentes), et chaque liste était également présentées dans les différentes conditions.
Les conditions sont :
- 3 algorithmes (Beamformer, Doerbecker + Ephraim & Malah, Doerbecker + Scalart)
- 3 pourcentages de réinjection ($\alpha = 0\%$, 20% et 40%)
- 5 angles pour le bruit ($\theta = -90^o$, -45^o, 0^o, $+45^o$ et $+90^o$) ; le bruit provenait d'un seul haut-parleur
- 3 SNR ($-6dB$, $0dB$ et $+6dB$)

Ceci a conduit à 3x3x5x3 = 135 possibilités.

Aussi, la perception de la parole a été testée avec le vocodeur seulement, sans les algorithmes, produisant alors quinze autres conditions expérimentales (3 SNR * 5 angles de bruit).

3.1.6 Analyses statistiques

Un modèle mixte linéaire, de la bibliothèque (lmer) du paquet R lme4, [4, 69], à l'aide de la famille quasi-binomiale, a été ajusté aux données, pour tenir compte des multiples facteurs répétés, c'est-à-dire le RSB, la ré-injection α, la direction du bruit (θ) et l'algorithme. Une ANOVA a été réalisées suivie par

Algorithmes	Estimation du bruit	Re-injection	Vocodeur	Conditions
.	.	.	Oui	VOC
Beamformer	.	00%	Oui	BEAM 00
Beamformer	.	20%	Oui	BEAM 20
Beamformer	.	40%	Oui	BEAM 40
Doerbecker	Ephraim & Malah	00%	Oui	DOEM 00
Doerbecker	Ephraim & Malah	20%	Oui	DOEM 20
Doerbecker	Ephraim & Malah	40%	Oui	DOEM 40
Doerbecker	Scalart	00%	Oui	DOS 00
Doerbecker	Scalart	20%	Oui	DOS 20
Doerbecker	Scalart	40%	Oui	DOS 40

TABLE 3.1 – Récapitulatif des différentes conditions

des comparaisons post-hoc afin d'évaluer toutes les situations. Les paramètres de la lmer étaient :
– Algorithme (VOC, BEAM, DOEM, DOS)
– α (00%, 20% et 40%)
– SNR ($-6dB$, $0dB$ et $+6dB$)
– Angle de bruit ($\theta = -90^o$, -45^o, 0^o, $+45^o$ et $+90^o$)

La variable dépendante est le pourcentage de reconnaissance de phonèmes (pourcentage de phonèmes correctement répétés : PC).

3.2 Résultats

Les résultats de reconnaissance des phonèmes sont indiqués sur les figures 3.6 à 3.9 et dans le tableau 3.2.

Lorsque toutes les conditions sont réunies, le score moyen et pourcentage de reconnaissance de la parole a été de 41% (sd = 19%) variant de 24% à 49% selon la condition (en fonction de la moyenne des résultats sur les 26 sujets).

L'étude statistique a montré une influence significative des facteurs suivants :

Algorithmes	RSB (dB)	Moyenne (%)	Erreur Standard (%)
VOC	-6	29,47	1,57
VOC	0	32,25	1,43
VOC	6	35,10	1,49
BEAM	-6	44,77	2,57
BEAM	0	46,61	2,60
BEAM	6	48,42	2,53
DOEM	-6	42,45	2,02
DOEM	0	45,01	2,17
DOEM	6	47,58	2,19
DOS	-6	36,99	1,95
DOS	0	40,35	1,78
DOS	6	43,59	1,76

TABLE 3.2 – Résultats en fonction du RSB par condition

- Le rapport signal sur bruit (SNR), ($t = 13,2$, $p < 0,001$)
- Les algorithmes, avec une différence significative entre le beamformer et les deux autres algorithmes. Ils sont significativement meilleurs pour la condition BEAM que pour DOEM, ($t = 3.18$, $p < 0,01$), mais aussi pour DOS ($t = 2,29$, $p < 0,01$).
- Une différence significative a aussi été observée avec l'angle de provenance du bruit (θ) ; la différence étant située entre les angles de -90^o et $+45^o$($t = 4,83$, $p < 0,05$).
- Aucune différence significative n'a été obtenue avec le facteur α. Cependant, il y a une interaction significative de α avec les algorithmes : une augmentation significative pour le DOS et est apparue lors de la réinjection entre $\alpha = 0\%$ ou $\alpha = 20\%$, ($p < 0,01$) et elle n'était plus significative ensuite.

Dans l'ensemble, DOEM et DOS n'ont pas été statistiquement meilleurs en reconnaissance de parole que le VOC seul.

Les comportements spécifiques observés selon le RSB sont indiqués ci-dessous. Les différences dans les PC obtenus avec le Vocodeur (pris comme la situation de référence) et les différents algorithmes ont été calculé. Ils

représentent l'influence introduite par les différents algorithmes (figure 3.9).

Les variations diffèrent en ce qui concerne le RSB ($t = 5,26$, $p < 0,05$).

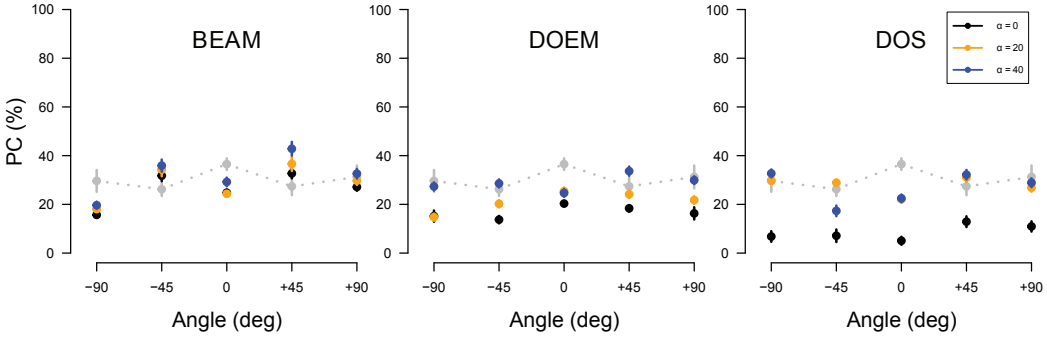


FIGURE 3.6 – RSB = $-6dB$

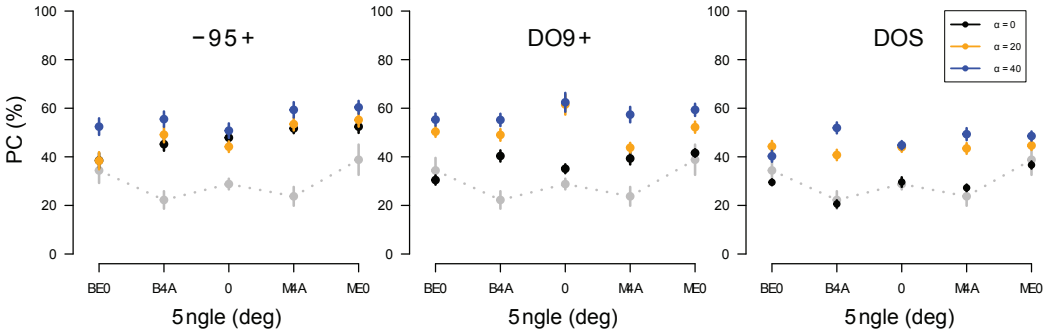


FIGURE 3.7 – RSB = $0dB$

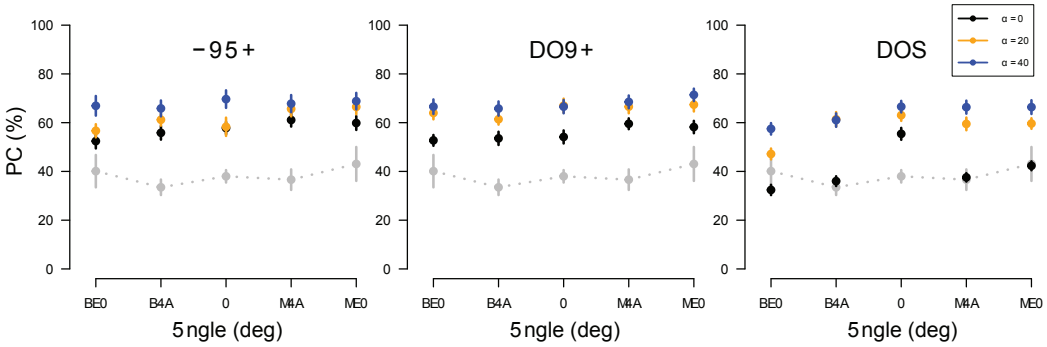


FIGURE 3.8 – RSB = $+6dB$

Pour $RSB = -6dB$, les algorithmes n'améliorent pas les résultats observés avec le vocodeur seul. Une seule amélioration significative, a été observée, avec la condition BEAM, pour les angles ($\pm 45^o$), ($t = 2.46$, $p < 0,05$).

Pour $RSB = 0dB$, une amélioration a été observée par tous les algorithmes. Elle est très significative : ($t = 24,1$, $p < 0,001$). La meilleure amélioration a été obtenue pour les angles $\pm 45^o$ ($t = 6,75$, $p < 0,01$). BEAM et DOEM conduit à des résultats meilleurs que DOS, mais la différence n'était pas significative.

Avec un $RSB = +6dB$, une amélioration significative a été obtenue pour tous les algorithmes ($t = 10,0$, $p < 0,001$). BEAM a des résultats nettement meilleurs que le DOEM ($t = 2,28$, $p < 0,05$) et DOS ($t = 2,28$, $p < 0,05$). Il n'y a pas eu de différence significative entre DOEM et DOS.

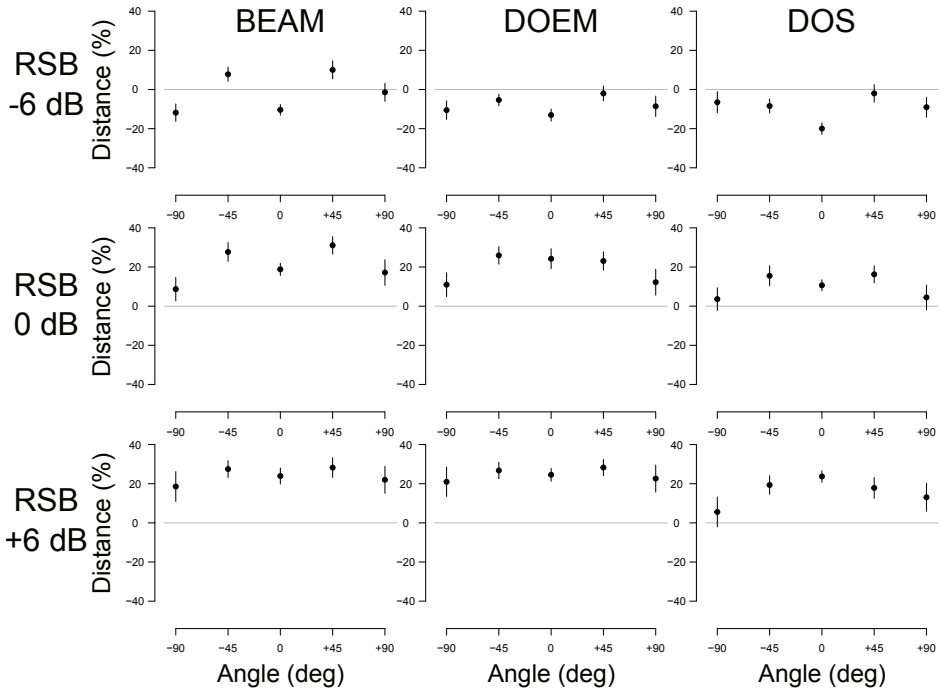


FIGURE 3.9 – Pourcentage obtenus par rapport à la condition VOC

3.3 Discussion

Par rapport à la condition VOC seul, les scores ont été significativement meilleurs avec les trois différents algorithmes dans le cas $RSB = 0dB$ ou $SNR = +6dB$.

Au contraire, aucune amélioration n'a été notée pour la valeur $SNR = -6dB$, surtout avec les algorithmes DOEM et DOS. La détérioration a été principalement observée avec un angle de bruit de 0^o, c'est à dire lorsque le bruit avait le même angle que le signal utile. Comme nous l'avons indiqué précédemment, les algorithmes favorisent l'orientation de face et le bruit et le signal sont difficiles à séparer s'ils proviennent tous les deux de face. Pour les valeur de $RSB = 0dB$ et $RSB = +6dB$, l'amélioration a été la meilleure pour les angles intermédiaires $\pm 45^o$. Elle a été plus faible pour les angles $\pm 90^o$. Il est possible que, pour ces angles extrêmes, le vocodeur détériore les indices binauraux pour la localisation sonore, surtout l'ITD, [34].

Une ré-injection du signal d'entrée par le paramètre α, [34, 91] améliore la performance des algorithmes et plus particulièrement zvec les algorithmes DOEM et DOS. L'algorithme DOS, a donné les moins bons résultats. La meilleure amélioration avec cet algorithme a été pour $\alpha > 0$ et pour les $RSB > 0dB$, figure 3.7, 3.8, 3.9. La performance de BEAM, (qui a été la meilleure), n'a pas été significativement modifiée par le facteur α.

Les résultats peuvent provenir de la soustraction spectrale utilisée dans les stratégies DOEM et DOS. Ces procédures utilisées pour la suppression de bruit, permettent de réhausser efficacement le signal. La ré-injection du signal d'entrée s'est avérée utile. Ce très bon résultat montre qu'on peut améliorer la localisation sans déteriorer le signal débruité, [33, 91].

La condition BEAM, qui repose principalement sur le délai entre les deux oreilles, n'a pas profité d'une ré-injection pour la reconnaissance de la parole.

Les scores obtenus pour un bruit venant de la gauche ont été légèrement inférieurs qu'avec un bruit venant de droite. Les préférences hémisphériques peuvent peut-être expliquer cette différence, [67, 73]. L'hémisphère gauche principalement alimentée par l'oreille droite, préfère classiquement les transitions rapides qui correspondent à la parole. L'hémisphère droit est traditionnellement plus efficace avec les transitions lentes liées à l'environnement.

Si plusieurs sources de bruit arrivent aux microphones, [26, 42, 99]. La compréhension de la parole risque encore diminuer.

Conclusion

L'efficacité de certains algorithmes de réduction de bruit (Doerbecker Vs beamformer) dans le cadre d'un codage de l'implant cochléaire binaural, a été testée dans la présente étude.

Les stratégies de type Beamforming et Doerbecker ont amélioré la perception des phonèmes dans des situations modérément bruyantes ($SNR = 0dB$ ou $+6dB$) et présentent les mêmes performances. L'amélioration moyenne a été d'environ 20 %. L'amélioration a été très faible (et parfois même négative) pour la valeur $RSB = -6dB$.

La condition BEAM a conduit aux meilleurs résultats (l'augmentation de la PC a été de 30 %), surtout pour un bruit provenant des angles $\pm 45^o$. Ici aussi, les résultats ont été médiocres pour $RSB = -6dB$. La technique de Doerbecker a amélioré également la perception de la parole. Avec la correction d'Ephraim & Malah, les résultats ont été meilleurs que pour la stratégie Scalart.

La ré-injection du signal (α) a réintroduit une partie de la localisation. Lorsque le pourcentage de ré-injection a été plus élevé (20 % et 40 %) la

reconnaissance a été peu améliorée. La condition BEAM s'est montré peu sensible à la ré-injection du signal.

Chapitre 4

Performance de localisation de source sonore après un débruitage binaural par l'intermédiaire d'un vocodeur.

Introduction

Pour comprendre et interpréter un son, les personnes normo-entendantes utilisent leurs deux oreilles. En effet, nous avons déjà vu qu'il existe des indices binauraux qui améliorent la perception des sons dans l'espace qui sont basés sur la différence de temps interaural (DTI) et sur la différence de niveau interaural (DNI), [46, 62, 70]. Le premier indice est principalement présent dans les basses fréquences ($f < 800Hz$) et le second pour les hautes fréquences ($f > 3000Hz$). Ces indices permettent de localiser les sons dans l'espace, et aussi de sélectionner une source sonore à partir des différents rapports signal sur bruit (RSB) qui arrivent aux oreilles. Le meilleur RSB sera a priori sélectionné et interprété.

Si on considère un sujet implanté cochléaire, comprendre la parole en milieu bruité devient difficile. Avec l'implant le traitement de l'information

se fait différemment. Le pavillon de l'oreille étant court-circuité, un facteur de détection des sources est perdu. Le traitement du signal peut restituer cette information. Il existe plusieurs façons d'améliorer la perception dans le bruit. Dans ces conditions, les traitements binauraux effectués dans l'aide auditive conventionnelle peuvent être adaptés à l'implant binaural [23].

Actuellement lorsque l'implant binaural n'effectue aucun traitement entre les deux voies. Les premiers résultats obtenus avec cet implant ne montrent pas de différence significative avec les personnes portant un implant bilatéral. Ces résultats peuvent être améliorés avec des méthodes de réduction de bruit (débruitage) déjà utilisé dans l'aide auditive, [9, 23, 30, 43, 74].

Nous reprendrons les traitements binauraux tels que la soustraction spectrale, [9, 43] ou bien le filtrage de Wiener, [103]qui favorisent le signal venant de face. Nous regardons, ce que nous écoutons. Les algorithmes de débruitage détectent et rehaussent ce qui provient de face, [23]. On a vu que la soustraction spectrale binaurale améliore le signal provenant de face, mais elle introduit un bruit musical dû à la mise à zéro de certaines fréquences. On sait que pour atténuer le bruit musical, il existe une méthode développée par Ephraim & Malah, [29, 30]. On peut aussi utiliser une autre méthode d'estimation qui est basée sur le modèle de Scalart, [74] et qui améliore le rapport signal sur bruit de la trace.

L'effet pervers de ces méthodes de débruitage est qu'elles favorisent ce qui provient de face, et perturbent la localisation, [61], des sources sonores. Van den Bogaert a proposé d'introduire un facteur (α) de ré-injection du signal d'entrée sur la sortie pour restituer une meilleure localisation, [91]. Nous nous proposons d'étudier cet aspect de localisation dans ce chapitre.

4.1 Matériel & Méthodes

La base du débruitage du signal d'entrée dans l'implant cochléaire binaural se situe dans le domaine spectral.

L'algorithme de Doerbecker [23], classiquement utilisé dans l'aide auditive, ainsi que le beamformer seront considérés ci-dessous. Le premier est composé de deux modules. Une soustraction spectrale [9] et le filtrage adaptatif de Wiener [103]. La soustraction spectrale est basée sur l'introduction d'un estimateur de bruit de type Ephraim & Malah, [29] ou de type Scalart, [74]. Le beamformer quand à lui est le beamformer fixe présenté dans le chapitre 1.

4.1.1 Soustraction spectrale

La soustraction spectrale binaurale est composée d'un module estimateur de bruit et ensuite d'une soustraction spectrale binaurale.

Une FFT est appliquée aux deux signaux d'entrée. La puissance sur chaque voie est ensuite calculée, ainsi que la puissance croisée des deux côtés. Ensuite on soustrait au signal d'entrée le bruit pour obtenir un signal débruité. On rajoutera la formule d'Ephraim & Malah et celle de Scalart. Une fois le module soustraction spectrale effectué, le filtrage de Wiener sera appliqué.

4.1.2 Filtrage de Wiener

Le filtrage de Wiener applique le meilleur RSB possible entre le bruit estimé et le signal utile. Le filtrage de Wiener est une multiplication du signal par un filtre $W(f)$.

Nous avons testé cette stratégie avec un groupe de sujets normo-entendants (20 personnes).

4.1.3 Ré-injection

La ré-introduction du signal d'entrée présenté par van den Bogaert, [91] est une méthode simple pour restituer des informations binaurales.

Les proportions α prisent sont de 00%, 20% et 40%.

4.1.4 Vocodeur

Le vocodeur est un traitement du signal qui rappel le codage de l'implant. Les caractéristiques fréquentielles et temporelles de l'implant sont une fréquence d'échantillonnage de 16 kHz, des trames d'acquisitions de $8ms$ (128 échantillons) et un recouvrement de $2ms$ (75%). L'implant considéré comporte 12 électrodes sur chaque porte électrodes. On sépare le signal suivant la tonotopie de l'oreille, linéaire jusqu'à $800Hz$ puis logarithmique au-delà. Selon la formule de bark, [85]. Le signal est ensuite séparé en 12 bandes de fréquences. Les huit maxima d'énergie sont conservés puis ils sont multipliés par la fréquence centrale de la bande de fréquence correspondante. Le signal est ensuite recomposé en additionnant les différentes fréquences.

Tout le traitement du signal (algorithme + vocodeur) est rappelé sur la figure 4.1.

4.1.5 Sujets

Vingt personnes normo-entendantes ont participé à cette expérience. La moyenne est de 25 ans, le minimum est de 22 ans et le maximum est de 29 ans. Le seuil de l'audition est inférieur à 20 dB HL sur les fréquences caractéristique de l'audiométrie tonale. Toutes les personnes sont rapportées droitières par le test d'Hedinburg, [65]. Cette étude a été approuvée par le comité d'éthique sous la référence.

4.1.6 Protocole

La première étape a consisté à enregistrer à l'aide d'une tête artificielle (KEMAR, figure 3.3) les informations binaurales. A la sortie des microphones, placés dans les oreilles du KEMAR, le signal est numérisé. Ensuite les deux algorithmes de débruitage sont appliqués avec la ré-injection puis intervient le vocodeur. Les auditeurs reçoivent le signal via un casque et non en champ libre. Le son est délivré par le casque à $55dBSPL$ ce qui est un niveau confortable. Le signal utilisé est un speech noise (bruit de parole) composé

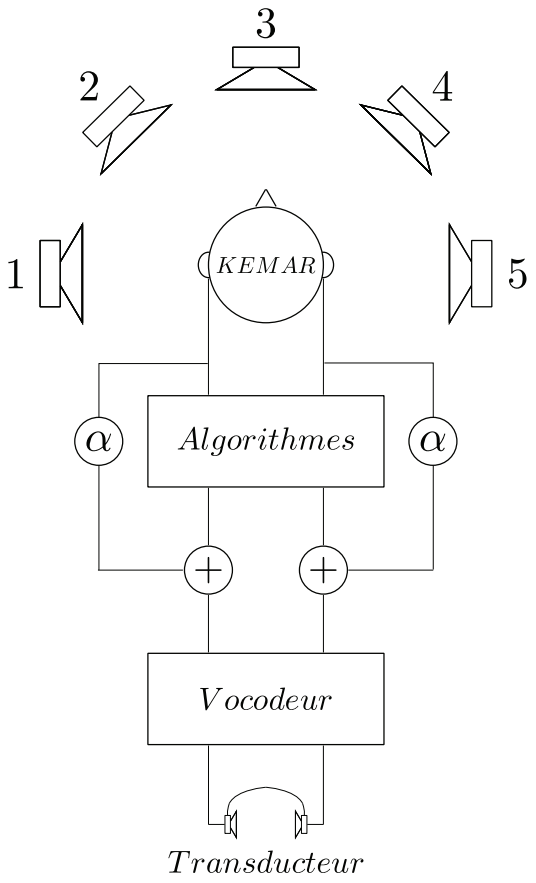


FIGURE 4.1 – Récapitulatif de l'expérience

des sommes des mots des listes de Lafon avec la phase randomisée. Il dure 4 secondes et contient tous les spectres des mots. Chaque situation est répétée trois fois. Le signal provient de 5 haut-parleurs placés sur le plan horizontal, figure 4.2. 150 sons ont été proposés par session.

Un logiciel de traitement a été développé pour faciliter le traitement des résultats ainsi que pour la randomisation des stimuli envoyées aux personnes. Les personnes devait cliquer sur « start » une fois qu'elles étaient prêtes puis cliquer sur l'un des cinq boutons qui sont assimilés chacun à un angle une fois le son entendu.

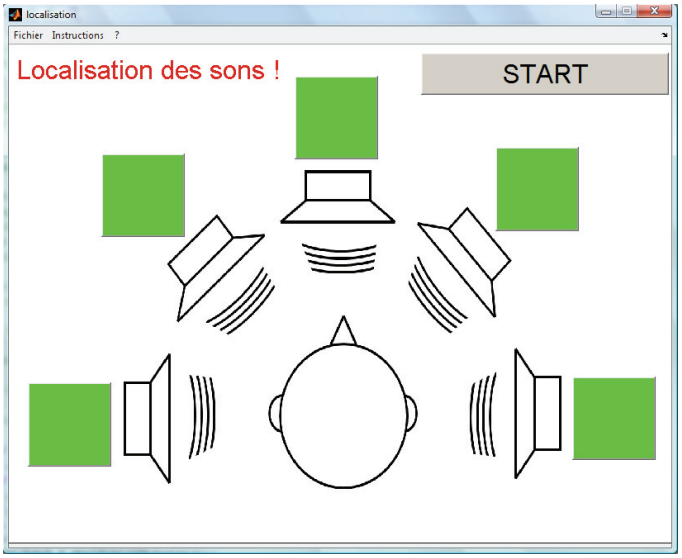


FIGURE 4.2 – Logiciel de localisation

Le plan d'experience a comporté, tableau 4.1 :

- 3 algorithmes
- 5 directions
- 3 répétitions
- 3 ré-injections
- 1 vocodeur
- 1 non modifié

Algorithmes	Estimation du bruit	Re-injection	Vocodeur	Conditions
.	.	.	Non	Normal
.	.	.	Oui	Voc
Beamformer	.	00%	Oui	Beam 00
Beamformer	.	20%	Oui	Beam 20
Beamformer	.	40%	Oui	Beam 40
Doerbecker	Ephraim & Malah	00%	Oui	Doem 00
Doerbecker	Ephraim & Malah	20%	Oui	Doem 20
Doerbecker	Ephraim & Malah	40%	Oui	Doem 40
Doerbecker	Scalart	00%	Oui	Dos 00
Doerbecker	Scalart	20%	Oui	Dos 20
Doerbecker	Scalart	40%	Oui	Dos 40

TABLE 4.1 – Récapitulatif des différentes conditions

La tâche de l'auditeur a été de trouver (localiser) parmi les 5 directions, la provenance du son. L'effet de la ré-injection du signal sur la localisation sera plus particulièrement regardé.

4.2 Résultats

Les résultats sont présentés sur les figures 4.3 à 4.8.

4.2.1 Courbes de localisation

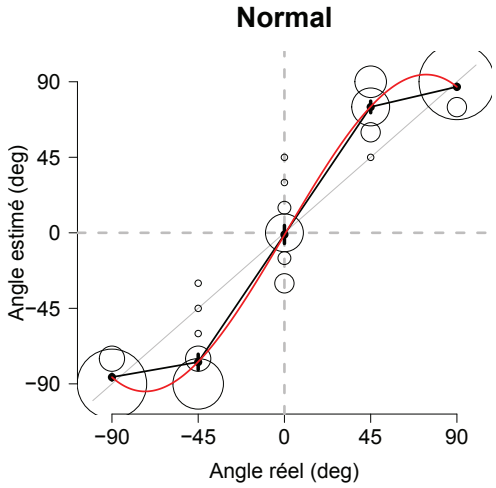


FIGURE 4.3 – Courbe de localisation pour la condition « Normale »

Les des différentes courbes sont organisées de la façon suivante. En abscisse, l'angle réel d'émission est affiché, en ordonnée, la réponse des différentes personnes. La courbe grise sur la diagonale, représente la détection parfaite. Les cinq angles d'émissions sont représentés. La courbe en noir représente les résultats de la moyenne des réponses avec pour chaque angle, l'erreur standard. La courbe en rouge représente la régression linéaire pour les statistiques.

A.)

Pour la figure 4.3, représente la condition « Normale », c'est à dire lorsque le son est non modifié par les algorithmes et le vocodeur ; « il est tel que nous nous l'entendons ». On peut remarquer que les angles $\pm 90^o$ et 0^o sont bien indiqués. Les cercles sont de grosse taille, ce qui montre qu'il y a beaucoup de réponses pour ces angles. En revanche, $\pm 45^o$ n'est pas souvent cité. On voit que la sélection des réponses est plus proche des angles extrêmes que du bon angle. Une explication peu provenir du fait que le son est plus fort d'un côté que de l'autre. ITD et ILD amènent les informations binaurales mais avec un avantage en faveur de ILD. Donc un son plus fort d'un côté que de l'autre entraine comme réponse des individus, les angles extrêmes.

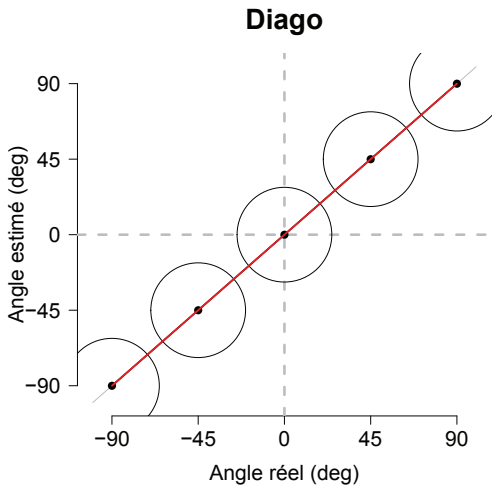


FIGURE 4.4 – Courbe de localisation pour la condition « Diago » de référence

B.)

La figure « Diago », (4.4), représente pour les vingt personnes, le score parfait, avec les dimensions des cercles maximum.

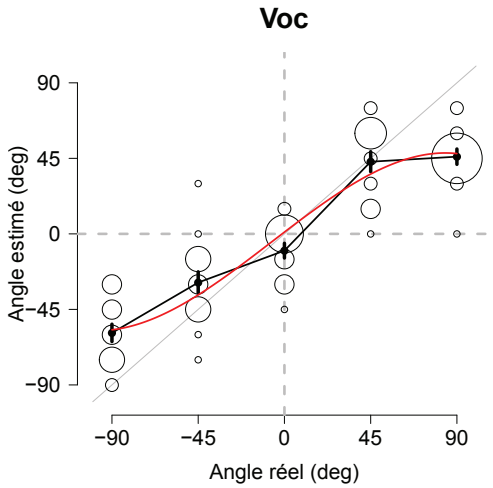


FIGURE 4.5 – Courbe de localisation pour la condition « Voc »

C.)

La condition « Voc », figure 4.5, représente le vocodeur seul. C'est ce qui est actuellement obtenu avec le traitement du signal de l'implant. La suite des résultats sera à comparer par rapport à cette figure de référence.

La direction de face 0^o est bien indiquée. Pour les angles extrêmes, les personnes ont tendance à ne pas répondre au-delà de $\pm 60^o$. On peut remarquer qu'il y a une grande variabilité pour les angles extrêmes contrairement à 0^o.

D.)

Les trois types de Beamformer, figures 4.6 conduisent à des résultats très différents. Lorsque $\alpha = 00\%$, les résultats obtenus montrent absence de bonne localisation. Les angles sont au niveau de 0^o. C'est sur cet angle que la cardioïde est centrée. La ré-injection d'un α restitue une localisation. Pour $\alpha = 40\%$, les résultats obtenus sont proches de la condition « Voc ».

Dans la figure 4.6, la condition Beam 20, est totalement différente de la condition Beam 00. La courbe est en ré-injectant légèrement ($\alpha = 20\%$) du

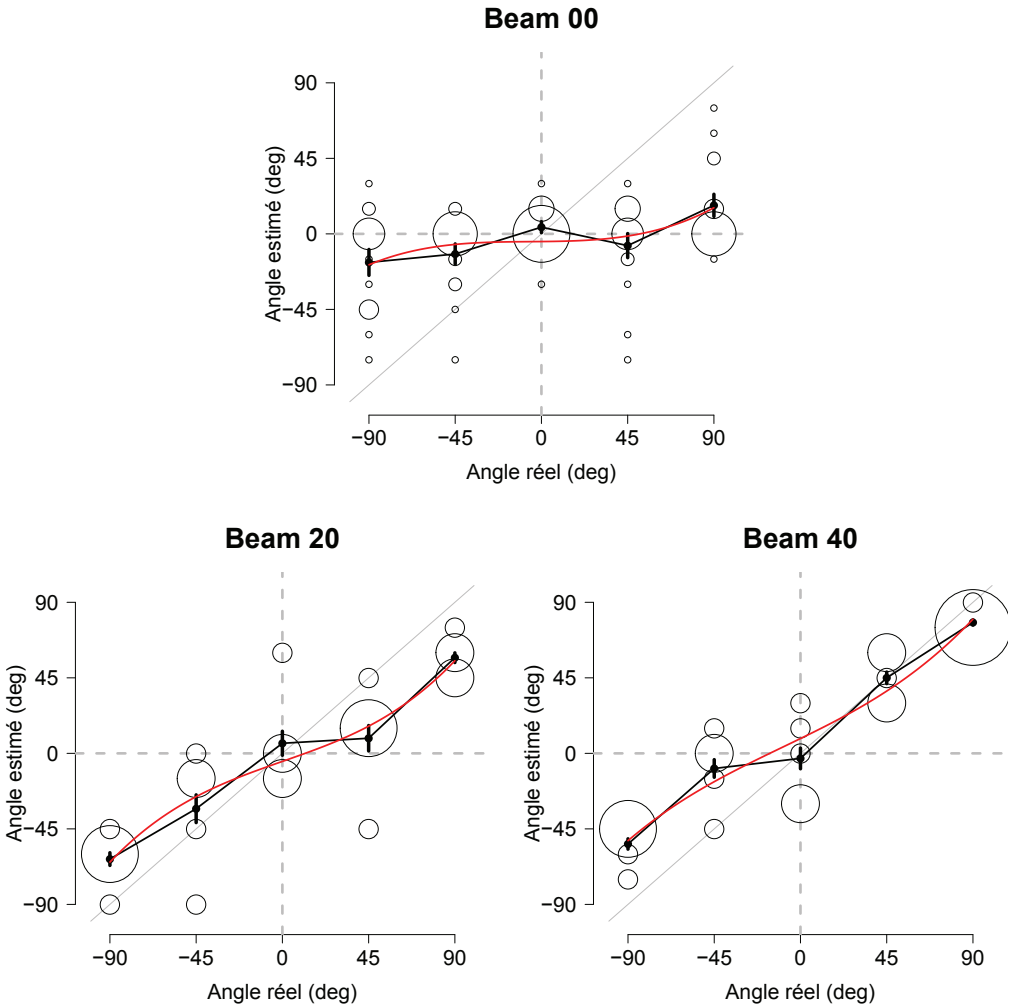


FIGURE 4.6 – Courbe de localisation pour la condition « Beam »

signal d'entrée, beaucoup plus proche de la diagonale que pour $\alpha = 0\%$. La plage des réponse est d'environ $\pm 60^o$.

Lorsque l'on passe à $\alpha = 40\%$, la bonne localisation augmente encore par rapport à 20%. La courbe moyenne est encore plus proche de la diagonale que pour la condition précédente.

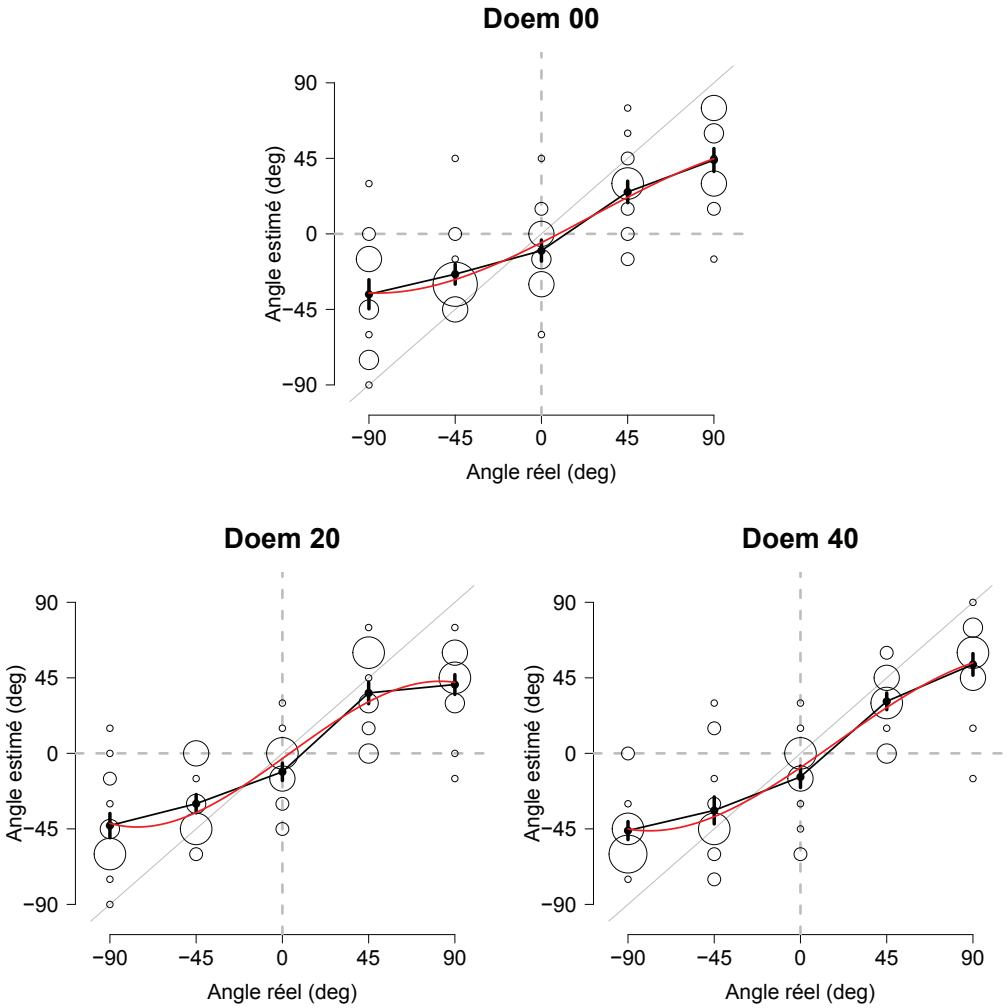


FIGURE 4.7 – Courbe de localisation pour la condition « Doem »

E.)

Les courbes de localisation avec l'algorithme « Doem », pour les trois α, donne des résultats différents de la condition *Beam*. Les trois courbes sont similaires. Seule la dispersion pour les angles extrêmes est plus faible lorsqu'un α est grand.

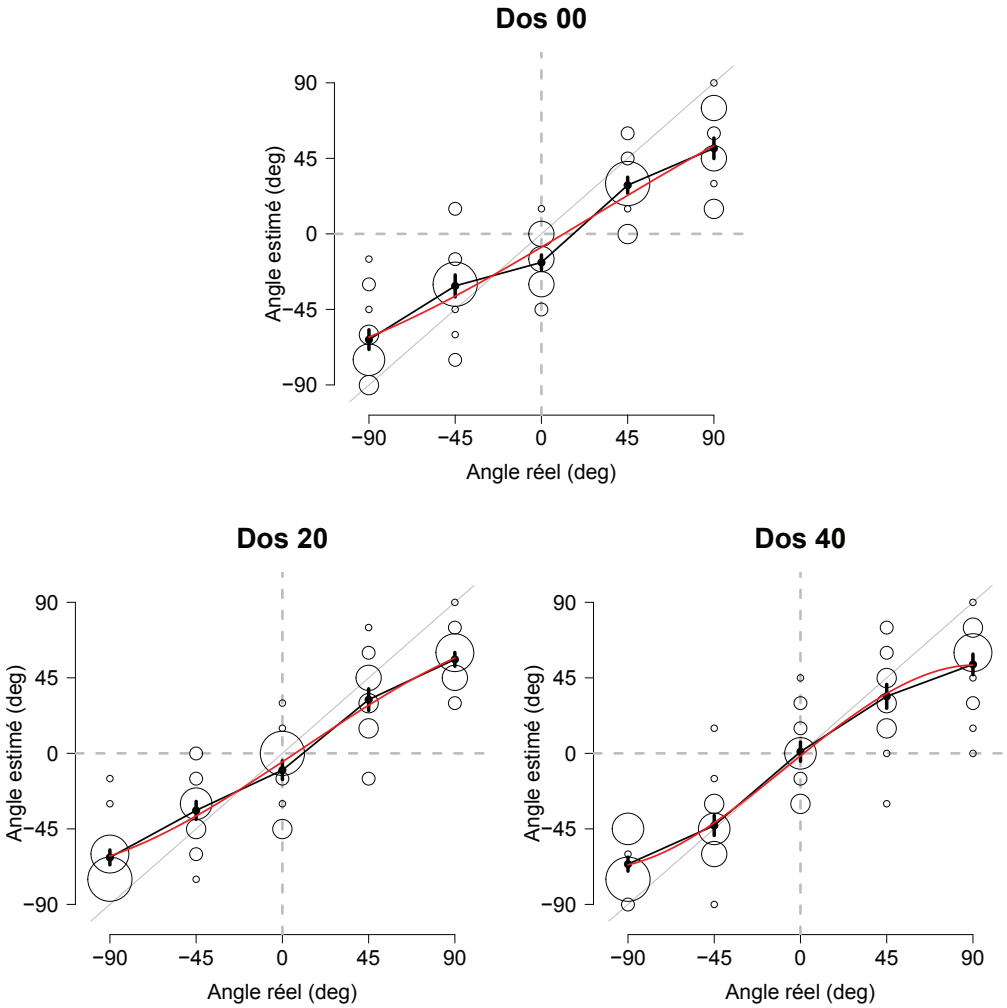


FIGURE 4.8 – Courbe de localisation pour la condition « Dos »

F.)

Les courbes obtenues pour la condition « Dos », figure 4.8, donnent des résultats proches de la diagonale.

La dernière condition pour « Dos 40 », est celle qui est le plus proche de la localisation de la condition « Voc ».

4.2.2 Statistiques

Décomposition de la courbe par une fonction élémentaire

Les courbes en rouge sur les différentes courbes de localisation, représentent un fitting des résultats ; dans le cas d'une approche analytique. Une régression a été envisagé pour représenter tous les algorithmes. On peut proposer cette régression par l'équation suivante (mode additif) :

$$y(x) = c_0.x^0 + c_1.x^1 + c_2.x^2 + c_3.x^3 \tag{4.1}$$

Soit,

$$y(x) = c_0 + c_1.x + c_2.x^2 + c_3.x^3 \tag{4.2}$$

Ou encore :

$$y(x) = constante + droite + parabole + cubique \tag{4.3}$$

Avec :
- x représente l'angles d'émission et y la réponse des sujets
- Le cas idéal est $y(x) = x$, soit $c_0 = 0$, $c_1 = 1$, $c_2 = 0$ et $c_3 = 0$

En pratique, on peut dire que :
- c_0 représente l'ordonnée. Il indique la réponse pour une émission de face.
- c_1 est le facteur d'ordre 1 un du facteur, c'est la pente à l'origine. Plus c_1 est proche de 1, plus on est proche de la diagonale.
- c_2 et c_3 sont les facteurs de distribution par rapport à la droite (réponse non linéaire)

Sur les figures représentant les coefficients de la régression, figure 4.9 à 4.12, on représente la valeur de chaque coefficient selon la stratégie (moyenne et intervalle de confiance à 5%). Ce n'est pas la variabilité qui est représentée mais les 95% des valeurs des coefficients.

Le coefficient c_0

Le coefficient c_0, sur la figure 4.9, représente le passage à l'ordonnée. On remarque que tous les algorithmes sauf « Beam 40 » sont proche de 0.

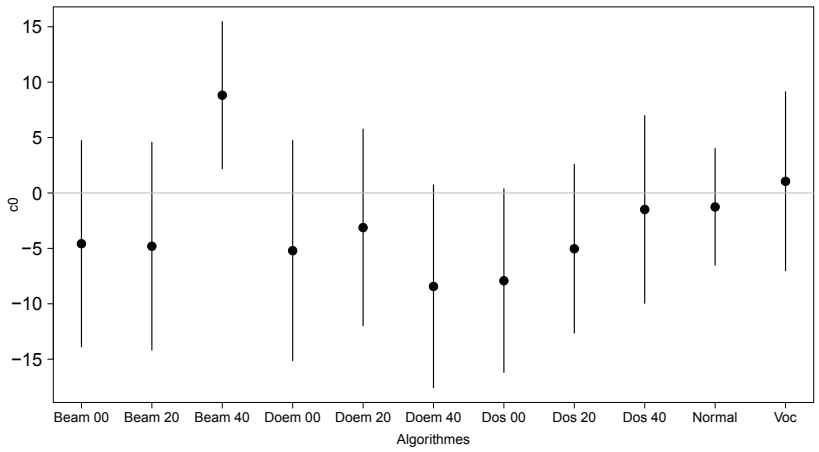


FIGURE 4.9 – Paramètre c_0

Le coefficient c_1

Pour la condition « Beam », on voit que α rectifie c_1 et qu'on commence à reconnaitre la direction. Deux algorithmes conduisent à des valeurs très différentes de 1 :
– « Beam 00 » qui ne permet pas de récupérer la source.
– « Normal » avec une surélévation de l'angle qui est trop latéralisé.

Le coefficient c_2

Le coefficient c_2 représente la composante quadratique. Une bonne reconnaissance implique qu'il est proche de zéro. c_2 a été comparé à 0 et l'hypothèse nulle à été acceptée dans tous les cas, ce qui indique une absence ou une faible influence de ma composante parabolique.

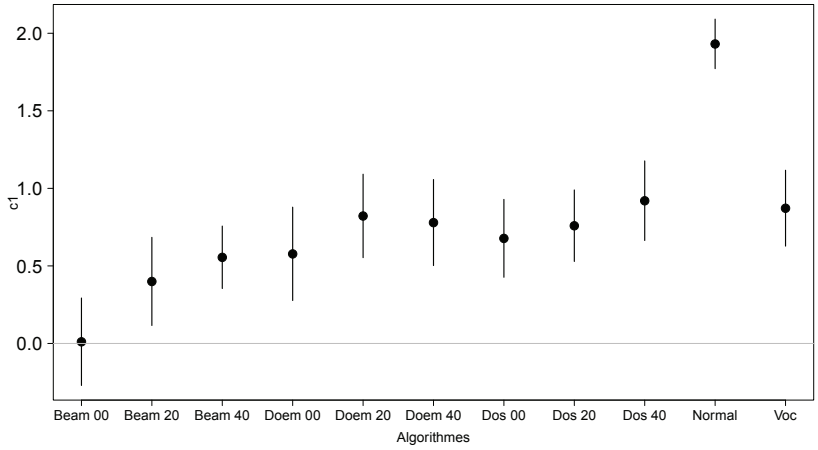


FIGURE 4.10 – Paramètre c_1

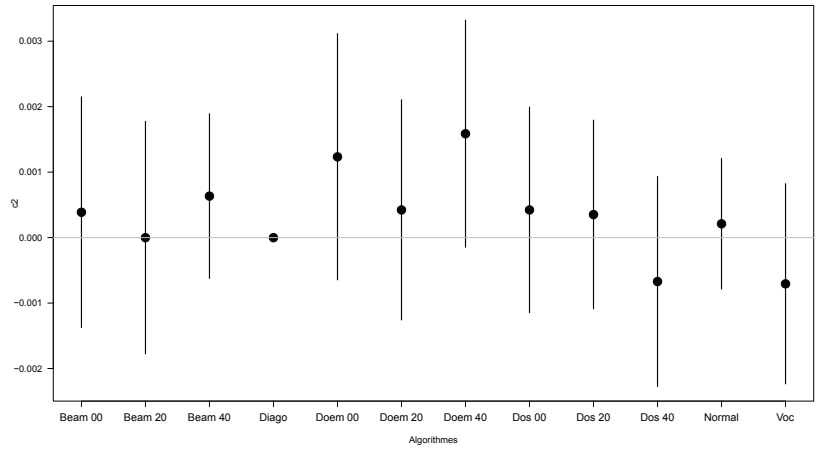


FIGURE 4.11 – Paramètre c_2

Le coefficient c_3

Le dernier coefficient c_3 représente la fonction cubique (figure 4.11). Il impose plus le résultat pour les angles intermédiaires $\pm 45^o$. S'il est positif, il aura tendance à éloigner de la réponse 0^o alors que s'il est positif, il indique un rapprochement de $\pm 90^o$. Les trois conditions sont à priori les plus délétaires mais on doit signaler que, dans notre travail, c_3 n'a pas été statistiquement

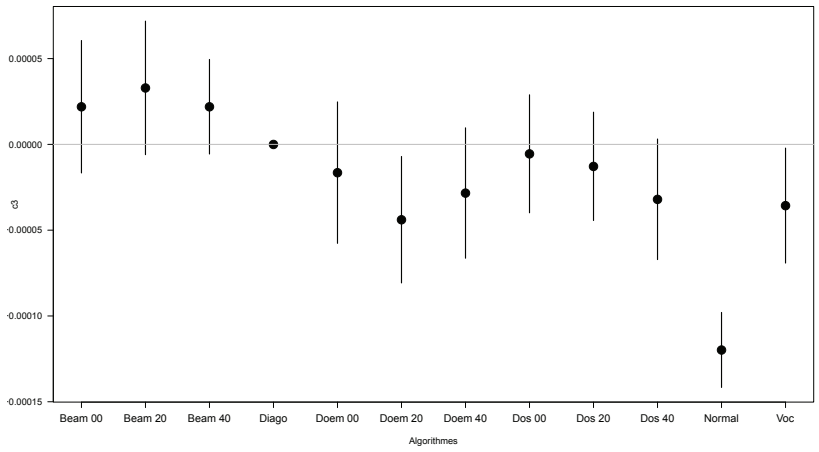


FIGURE 4.12 – Paramètre c_3

différent de 0 (à l'exception de la condition « Normal »).

Distance au cas idéal

Pour comparer les effets des différents paramètres, l'erreur de localisation E (distance à la diagonale) a été calculée. On a utilisé l'équation suivante :

$$E = \frac{1}{N} \sum_{i=1}^{N} \sqrt{(y_i - diago)^2} \tag{4.4}$$

Où :
– N : est le nombre total de personne, $N = 20$

Précisons l'équation 4.4 :

$$E = \frac{1}{N} \sum_{i=1}^{20} \sqrt{\sum_{j=1}^{3} \sum_{k=1}^{5} (y_{i,j,k} - diago_{i,j,k})^2} \tag{4.5}$$

Où :
– i : indique l'auditeur
– j : représente la répétition
– k : est l'indice de l'angle

Les résultats obtenus suivant cette formule sont montrés dans la table 4.2 et la figure 4.13 :

La figure 4.13 est décomposée de la façon suivante. Les différentes couleurs des rectangles correspondent aux différentes conditions. Puis la moyenne et l'erreur standard pour chaque condition sont affichés. Les différences significatives sont nombreuses et sont répertoriée dans le tableau 4.2.

Algorithmes	Moyenne (deg)	Déviation standard (deg)
Beam 00	49	6.4
Beam 20	27	6.7
Beam 40	22	6.3
Doem 00	33	11
Doem 20	30	5.6
Doem 40	28	7.1
Dos 00	28	4.4
Dos 20	24	4.1
Dos 40	22	5.5
Normal	16	3.1
Voc	24	4.1

TABLE 4.2 – Erreur par rapport à la diagonale

Les erreurs montrent l'influence d'α notamment pour l'algorithme du beamformer. En effet, la distance par rapport à la diagonale (donc l'erreur) diminue parfois fortement, lorsque ce dernier paramètre augmente. Les erreurs des algorithmes par rapport aux angles sont montrées dans la figure 4.14.

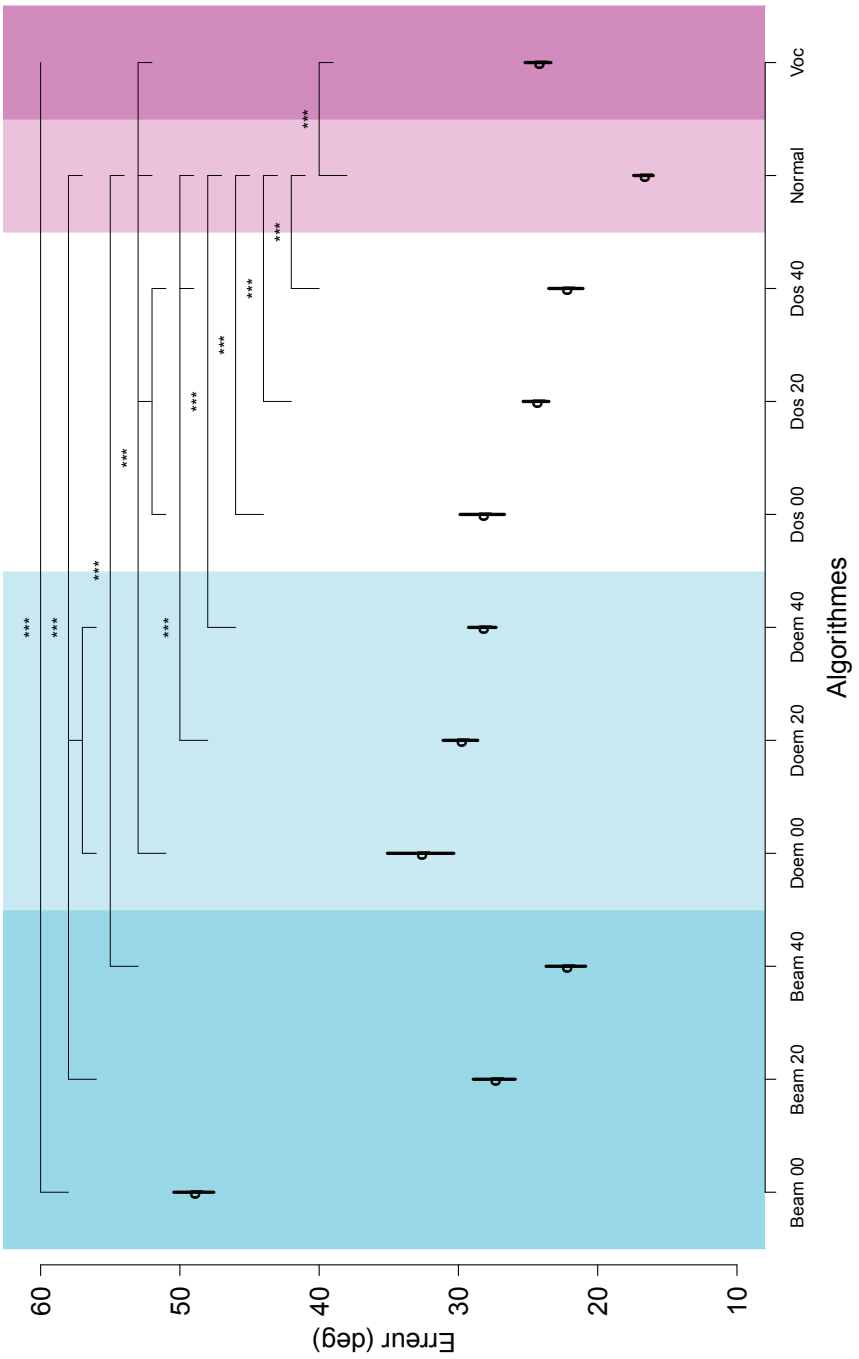


FIGURE 4.13 – Erreur de localisation

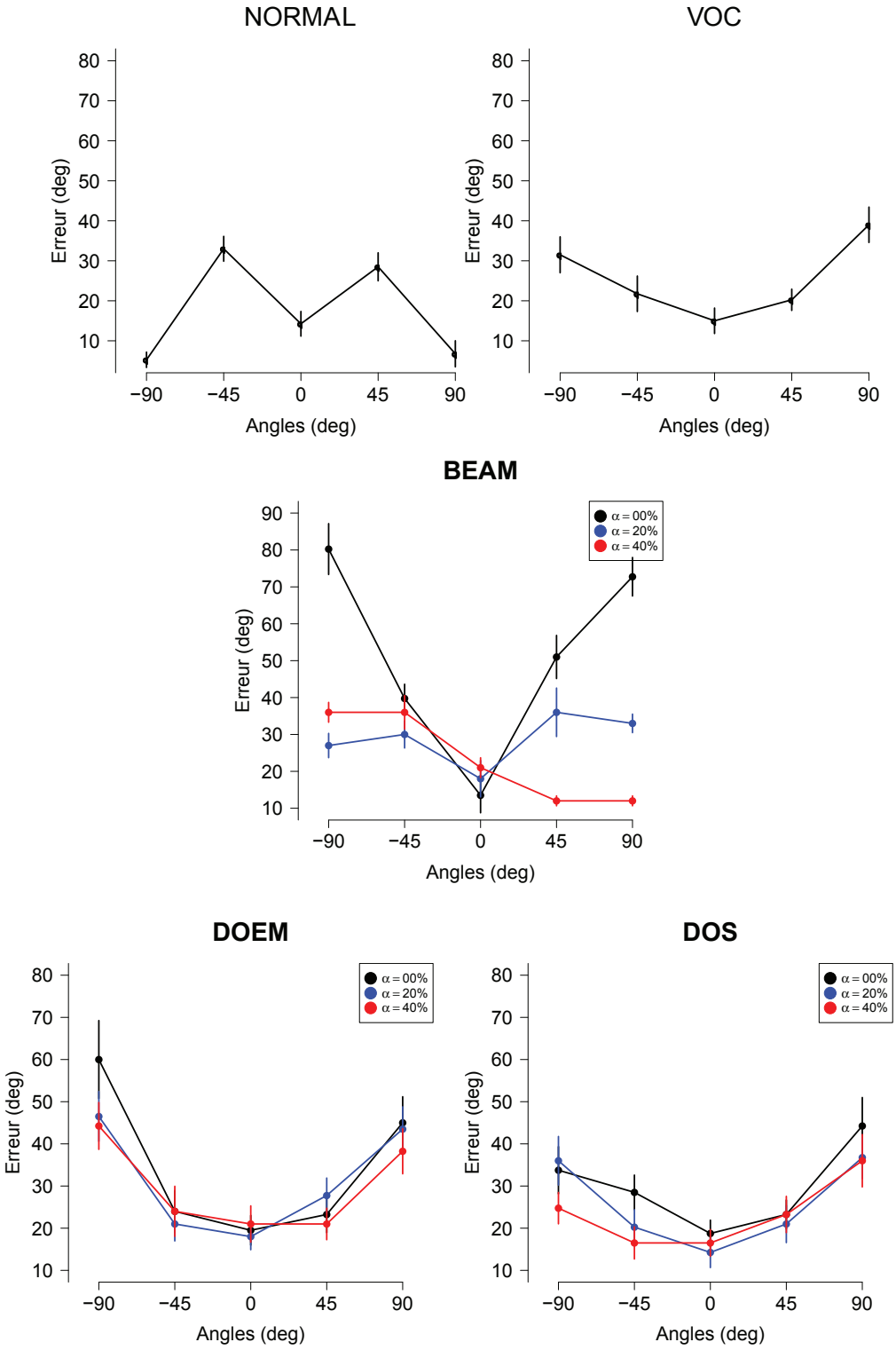


FIGURE 4.14 – Erreur pour les différentes conditions en fonction de l'angle

4.3 Discussion

Les résultats obtenus montrent une dégradation de la localisation lorsqu'on favorise une direction. L'amélioration de la reconnaissance de la parole s'effectue au détriment de la localisation. La condition « Normal » est celle qui est la plus proche de la diagonale ; elle correspond au signal non modifié, pris directement en sortie du KEMAR. La condition « Voc », correspond à l'implant cochléaire binaural seul. Voc détériore la localisation, et cette localisation est encore plus détériorée lorsque l'on introduit les algorithmes de débruitage.

Ensuite, l'erreur diminue lorsque l'on ré-injecte en sortie une fraction du signal d'entrée. Dans ce cas, l'angle $\pm 45^o$ est plus facilement indiqué. On remarque de plus que $\pm 90^o$ est rarement indiqué. L'algorithme de Scalart a moins déterioré l'orientation de la détection de la source que celui d'Ephraim & Malah.

Ces résultats sont en accord avec la littérature. Lorsque le signal est traité par le vocodeur, les ITD sont perdus car cette différence de temps est inférieure à la longueur des trames d'analyse. Il ne reste alors que les ILD, [61]. Les ILD conduisent à des localisations jusqu'à des valeurs de $\pm 60^o$; les valeurs au delà sont peu citées [33, 77]. La ré-injection α améliore la localisation. On peut assimiler α à un algorithme de rehaussement des informations binaurales jouant sur les amplitudes des signaux [91].

Conclusion

L'étude a porté sur la perte de la localisation due au débruitage binaural. Si les algorithmes classiques de rehaussement de la parole permettent d'améliorer dans le bruit, la localisation est détériorée. La proposition de ré-injecter le signal d'entrée par van den Bogaert permet, par l'intermédiaire du facteur α, de rétablir une partie de la localisation. En effet, avec $\alpha = 40\%$, on s'approche des résultats du vocodeur simple pour la condition Dos. Ceci

indique qu'avec une ré-injection du signal d'entrée, la localisation peut être restaurée jusqu'à $\pm 45^o$.

Conclusion

L'implant cochléaire binaural est une technologie qui permet de réhabiliter les personnes atteinte de surdité bilatérale profonde. Il permet de stimuler, par le biais d'électrodes placées le long de la cochlée, le nerf auditif. Suite à une rééducation et avec l'aide de l'orthophonie, les personnes porteuses de ce type d'appareil peuvent comprendre et localiser les sources sonores. Les travaux présentés dans ce manuscrit de thèse ont montré différents aspect.

Dans le premier chapitre, les principaux algorithmes de débruitage associés à une aide auditive ont été décrits. Que ce soit en monaural ou binaural, le but est de rehausser la parole ou un signal utile par rapport à l'environnement. Les principaux algorithmes étudiés ont été la soustraction spectrale, le filtrage de Wiener et la technique du beamforming. Ils ont tous des actions différentes mais permettent au final d'améliorer la qualité du signal acoustique. L'apparition d'un bruit musical lors de la correction introduit un biais dans la perception car le bruit étant atténué, il en apparait un autre. Des techniques existent pour limiter les distorsions et le bruit musical. Ensuite, la question s'est posée pour essayer d'adapter ces systèmes à l'implant cochléaire. Il a été présenté dans le second chapitre.

L'avantage de mettre un implant cochléaire sur chaque oreille n'est plus à démontrer. Les résultats sont significativement supérieurs à un implant monaural, que ce soit en reconnaissance de parole ou bien en localisation. De plus, les résultats de l'implant binaural montrent qu'il n'y a pas de différence entre bilatéral et binaural en localisation et perception des sons

dans le bruit lorsque les implants sont indépendants. Ce qui montre qu'avec moins d'électrodes (classiquement $\simeq$ 20 électrodes pour l'implant bilatéral et 12 électrodes par coté pour l'implant binaural), les performances ne sont pas détériorées. Un avantage existe avec l'implant binaural car il possède un seul processeur et permet le traitement des deux voies. Ainsi, la mise en place d'un vrai algorithme de réduction de bruit binaural peut améliorer la reconnaissance de parole ou la localisation.

Une façon simple a été choisie pour tester les algorithmes en simulant le codage de l'implant. Le signal acoustique a d'abord été modifié par les algorithmes puis un vocodeur binaural a été positionné en aval. Cette séquence que l'on peut voir dans le chapitre 3 et 4 a ainsi été traitée.

Les résultats obtenus dans le chapitre 3 ont retrouvés des résultats classiquement décrit dans la littérature. En effet, si l'on active un algorithme de débruitage sur un signal bruité, il est normal que le signal de sortie soit de meilleure qualité. Un facteur de ré-injection, α n'a pas dégradé significativement les résultats de perception, ce qui montre que l'on peut ré-injecter du signal bruité sans trop atténuer les résultats. Mais si la reconnaissance est augmentée, c'est au détriment de la localisation. Cette relation a été étudiée dans le dernier chapitre.

Dans le chapitre 4, on a vu que les performances de localisation diminuent si on focalise une direction ; c'est le but des algorithmes de débruitage. Les résultats obtenus avec les différents algorithmes ne sont pas les mêmes. En effet, le beamformer sans ré-injection donne les meilleurs résultats de reconnaissance de parole, mais il ne permet pas de localiser un son. Par contre la ré-injection, joue un grand rôle dans ce cas, car elle permet de restituer jusqu'à un angle de $\pm 45^o$ une localisation correcte. Pour les autres algorithmes, l'action de la ré-injection n'est pas aussi significative car le domaine de travail des algorithmes est fréquentiel et la phase du signal est un peu oubliée dans ce cas.

Donc avec l'implant cochléaire binaural, l'apport d'un algorithme de débruitage permet d'améliorer la qualité de la reconnaissance de la parole mais au détriment de la localisation. Si on désire conserver la localisation et la reconnaissance de parole, alors un compromis entre les deux paramètres est nécessaire et il doit être adapté aux besoins du patient.

Pour continuer...

Au cours de nos travaux de thèse, plusieurs questions se sont posées suite aux expériences, discussion en congrès et lecture de la bibliographie.

– On pourrait étudier le facteur de ré-injection plus profondément. Car en effet, il se pourrait qu'il y ait un maximum puis une décroissance de la localisation. Cette étude permettrait d'étudier les valeurs maximum et minimum du facteur α.

– Au lieu de distinguer deux blocs distincts, les algorithmes et le vocodeur, la mise en place directement dans le vocodeur de l'algorithme permettrait de gagner du temps de calcul mais aussi, de travailler avec beaucoup moins d'informations (regroupement fréquentiel). Ceci diffère de l'aide auditive. Le travail canal par canal peut être aussi envisagé.

– Dans le mesure où les algorithmes ne sont pas trop difficiles à mettre en oeuvre, tester en condition réelle avec des personnes implantées cochléaires permettraient encore plus précisément de connaitre le comportement des algorithmes.

– Pour améliorer la localisation chez les personnes implantées cochléaires, il faudrait mesurer les effets de DTI et de DNI. Pour cela, on pourrait répéter les expériences de Moore, [62] qui travaille sur l'ombre de la tête et le démasquage binaural.

Annexes

Les annexes présentent le CPP, les différentes publications et congrès auxquelles j'ai participé ainsi qu'une partie neurosciences portant sur les Speech ABR.

Tables des matières Annexes

-A.) Comité de Protection des Personnes
-B.) Publications

A.)Comité de Protection des Personnes (CPP)

La première chose à effectuer lorsque l'on veut faire des expériences sur des personnes humaines, est d'obtenir un CPP. Il est nécessaire de se couvrir sur le plan de l'éthique. Un grand merci à Michel et Evelyne qui m'ont aidé pour la finalisation du projet.

Les personnes doivent remplir le consentement puis la notice d'informations. Si les personnes renoncent à être payée alors, elles doivent signer le papier sur le renoncement de l'indemnisation.

Tous les résultats doivent être inscrits dans un cahier d'observation (CO).

Le numéro de référence du CPP est le 0100630314037.

B.) Publications

Revues

Revue publiées

[71]

Céline Richard, **Arnaud Jeanvoine**, Evelyne Veuillet, Annie Moulin, Hung Thai-Van. **Exploration électrophysiologique des voies auditives sous-corticales chez l'humain : du clic au son de parole**. *Neurophysiologie Clinique*. 2010 ; 40(5–6) : 267 - 279.

[2]

Idrick Akhoun, Annie Moulin,, **Arnaud Jeanvoine**, Mikael Menard, François Buret, Christian Vollaire, Riccardo Scorretti, Evelyne Veuillet, Christian Berger-Vachon, Hung Thai-Van. **Speech auditory brainstem response (Speech ABR), characteristics depending on recording conditions, and hearing status : an experimental parametric study**. *J. Neurosci. Methods*. 2008 ; Nov 15 ; 175(2) : 196 - 205.

Revue « in press »

Revue STH, Elsevier
Arnaud Jeanvoine, Dan Gnansia, Eric Truy, Christian Berger-Vachon. **Identification of the source location in the case of binaural cochlear implant coding ; contribution of denoising algorithms.**, *Science et Technique du handicap, STH*

Revue en examen

Revue Measurment, Elsevier
Arnaud Jeanvoine, Dan Gnansia, Rafael Laboissière, Hung Thai-Van, Christian Berger-Vachon.**Source detection : Influence of two**

algorithm's noise reduction ; Simulation in the case of binaural cochlear implant coding., *Measurment*

Revue en attente

AAATE
Arnaud Jeanvoine, Dan Gnansia, Eric Truy, Christian Berger-Vachon. **Efficiency of noise reduction algorithms in a binaural cochlear implant coding environment.**

Congrès

Durant ma thèse, j'ai eu l'occasion de participer à plusieurs congrès :

- Juin 2012, *Influence du débruitage spatial sur la localisation dans le cas du codage de l'implant cochléaire binaural. Simulation sur normo-entendants.*, Handicap 2012, Paris, **Proceeding + Présentation orale**

- Novembre 2011, *Validation d'un test de perception de la parole à choix forcé : Audimots*, SFA 2011, Montpellier, **Poster**

- Octobre 2011, *Analyse en ondelette des potentiels évoqués auditifs en réponse à des sons de parole (SABR) : une nouvelle analyse objective de l'encodage sous-cortical des sons de parole voisés et non voisés*, SFORL 2010, Paris (75), **Présentation orale**

- Avril 2011, *Improved speech perception in noise for cochlear implantees using binaural signal processing*, ESPCI 2011, Athènes (Grèce), **Poster**

- Novembre 2010, *Débruitage du signal de parole, évaluation de différents traitements binauraux appliqués à l'implant cochléaire binaural*, JJCAAS, Paris (75), **Présentation orale + Poster**

- Octobre 2010, *Etude d'algorithmes de débruitage binauraux en stimulation*

d'implant cochléaire binaural, SFORL 2010, Paris (75), **Présentation orale**

– Juillet 2010, *Binaural denoising algorithms associated with the Digisonic SP binaural cochlear implant. Simulations on normal-hearing subjects*, CI 2010, Stockholm (Suède), **Poster**

– Juin 2010, *Détection d'indices caractérisant un Speech ABR par la méthode de détection de forme*, Handicap 2010, Paris (75), **Proceeding + Présentation orale**

– Avril 2010, *Détection des indices pertinents caractérisant le Speech ABR*, CFA 2010, Lyon (69), **Proceeding + Présentation orale**

– Octobre 2009, *L'enveloppe d'un signal de parole peut-elle suffire à l'obtention d'un speech ABR ?*, SFORL 2009, Paris (75), **Présentation orale**

– Mai 2008, *Le speech ABR est-il une réponse neurale ou un artéfact électromagnétique ?*, JJCAAS, Toulouse (31), **Présentation orale + Poster**

Voici l'articles portant sur la thèse, les procedings et les posters :

Revue STH, Elsevier

Arnaud Jeanvoine, Dan Gnansia, Eric Truy, Christian Berger-Vachon. **Identification of the source location in the case of binaural cochlear implant coding ; contribution of denoising algorithms.**, *Science et Technique du handicap, STH.*

Identification of the acoustic source location in the case of binaural cochlear implant coding; contribution of denoising algorithms and signal re-injection.

Arnaud Jeanvoine[*,**,****] **—Dan Gnansia**[****]**—Eric Truy**[*,**,*****]**—Christian Berger-Vachon**[*,**,***]

* *INSERM U1028, Research Centre of Neuroscience in Lyon, team PACS (Speech, Audiology, Communication Health), Lyon, F-69000, France*

** *CNRS UMR5292, Research Centre of Neuroscience in Lyon, team PACS (Speech, Audiology, Communication Health), Lyon, F-69000, France*

*** *University Lyon 1, Lyon, F-69000, France*

**** *Neurelec, 2720 Chemin Saint-Bernard, Vallauris, F-06224, France*

***** *Audiology and Department ORL, Edouard Herriot Hospital, Lyon, F-69437, France*

arnaudjeanvoine@yahoo.fr

ABSTRACT. *Cochlear implants provide a good speech intelligibility in quiet, but speech understanding in noisy environment remains difficult. The lack of binaural information may explain this observation. Thus, spatial information may help to separate the sources arriving to the ears and improve signal intelligibility. Binaural signal processing techniques are commonly used in conventional hearing aids to reduce the noise. This is less known with binaural cochlear implant. A bilateral implant provides information to both ears but the processors are independant. In the case of binaural implant, a single processor is used and both informations can be connected allowing the application of noise reduction algorithms.*

Several signal processing methods exist to improve the signal to noise ratio and binaural methods are a good way to improve the directivity of the microphones. Algorithms tested in this study are based on this assumption. A study with normal-hearing subjects and the use of a vocoder simulating binaural implant processing has been conducted. Several parameters such as the re-injection, and the angle of the source have been studied.

The results show that the source localization without signal processing is good. The introduction of denoising algorithms affects the identification of the source position. Re-injection of the input signal improved the source localization without heavily deteriorating speech perception.

This study suggests that the introduction of a binaural processing in the implant may not heavily impair the source detection when noise reduction algorithms are used.

RESUME. *Les implants cochléaires fournissent une bonne intelligibilité vocale dans le calme, mais la compréhension de la parole dans un environnement bruyant est difficile. Le manque d'information binaural peut expliquer cette observation. Ainsi, l'information spatiale peut aider à séparer les sources qui arrivent aux oreilles. Les techniques binaurales de traitement du signal sont couramment utilisées dans les prothèses auditives pour réduire le bruit. Un implant bilatéral fournit des informations sur les deux oreilles, mais les processeurs sont indépendants. Dans le cas d'un implant binaural, un seul processeur est utilisé pour la stimulation des deux oreilles et permette l'application d'algorithmes de réduction du bruit.*

Plusieurs méthodes de traitement du signal existent pour améliorer le rapport signal sur bruit. Les méthodes de débruitage binaurale sont un bon moyen d'améliorer la reconnaissance de la parole et la localisation. Une étude avec des sujets normaux entendants et l'utilisation d'un vocodeur simulant le traitement binauralimplant a été réalisée. Plusieurs paramètres tels que la ré-injection, et l'angle de la source en fonction des différents algorithmes ont été étudiés.

Les résultats montrent que la localisation de la source sans traitement du signal est bonne. La mise en place d'algorithmes de débruitage affecte l'identification dela position de la source. La réinjection du signal d'entrée améliore la localisation de la source sans détériorater fortement perception de la parole.

Cette étude suggère que l'introduction d'un traitement binaural dans l'implant peut pas fortement nuire à la détection de source lorsque algorithmes de réduction de bruit.

KEYWORDS: *Binaural Cochlear Implant, Vocoder, Binaural Denoising, Source Localization.*

1. Introduction

To understand and interpret sounds, normal hearing people use their two ears and binaural indices allow the perception of sounds in space. These indices are mainly based on the interaural time difference (ITD) and the interaural level difference (ILD), (Kistler, 1992, Morre, 1995, Rayleig, 1907). ITD is mainly present in low frequencies (f < 800 Hz) and ILD in high frequencies (f > 3000 Hz). Then sounds may be located in space, and also favor a source relatively to another improving its signal-to-noise ratio (SNR). Automatically, best SNR are selected and interpreted as the useful signal.

For cochlear implant (CI) subjects, speech understanding in a noisy environment is difficult. First, speech processing is altered. The pinna effect in the ear is lost, and the detection of the sources is affected. There are several ways to deal with speech perception in noise. First two independant cochlear implants can be fitted one on each ear. This case, is called "bilateral" because no cross information processing between the two implants is done. Another way is provided by the "binaural" implant, (Truy, 2002); a single processor stimulates both ears. In this case, a joint treatment between the sounds arriving on the left and the right microphones can be done. This binaural processing is already used in conventional hearing aids and it can be adapted to binaural implant (Doerbecker, 1996).

In the case of a bilateral cochlear implant both ears are independent. When a binaural cochlear implant is used without using cross information between the two ears, no significant difference with people wearing a bilateral implant is seen (Verhaert, 2012). Also, speech recognition is improved when noise reducing methods based on binaural processing are used in hearing aid (Boll, 1979, Doerbecker, 1996 , Ephraim et al, 1994, Kamath, 2002, Scalart, 1996).

Several binaural treatments such as spectral subtraction, (Boll, 1979, Kamath, 2002) or Wiener filtering are used in hearing aids, when it is decided that the useful signal comes from the front; sight and audition are linked (Boll, 1979, Doerbecker, 1996). Thus noise reduction algorithms enhance what comes from the front (front target). Binaural spectral subtraction methods improve the signal coming from the front, but they bring in a “musical” noise due to the introduction of zeros or negative modules for some frequencies. To lower the musical noise, a method was developed by Ephraim & Malah, (Ephraim et al, 1994). Another method suggested by Scalart, (Scalart, 1996) improves the performance in noisy environments.

As mentioned previously, noise reduction methods promote the front direction, but they lower the localization, (Kistler, 1992). A re-injection factor (factor α) can be used to bring back some input at the end of the processing and to restore some localization, (Bogaert, 2008). This action is tested in this paper.

2. Material & Method

The basis of signal denoising in binaural cochlear implant is located in the spectral domain. Doerbecker algorithm (Doerbecker, 1996), typically used in the hearing aid, can favor speech front target. It is composed of two modules. The first is a spectral subtraction (Boll, 1979) and the second is an Adaptive Wiener filtering. In the spectral subtraction, an estimator of noise was introduced (figure 1). In the middle a reduction of the musical noise is introduced. Two types of musical noise reduction will be discussed, Ephraim & Malah, (Ephraim et al, 1994) or Scalart, (Scalart, 1996).

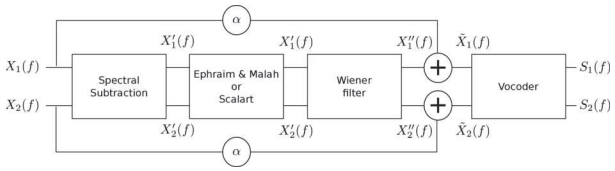


Figure 1. *Block diagram of the signal processing*

2.1. Spectral Subtraction

Binaural spectral subtraction is based on a noise estimator followed by a spectral subtraction (figure 2).

First, a FFT is applied to the two input signals. The power in the spectrum is then calculated by taking the energy of the input signal $P_{X1}(f) = X_1(f)^2$. The cross-power is obtained by the product of the two inputs, $P_{X1X2}(f) = P_{X1}(f).P_{X2}(f)$. What is common to the two inputs is kept and it represents the useful signal. The noise is defined by the difference between the input spectrum and the cross-power spectrum: $B_1(f)^2 = |P_{X1X2}(f) - P_{X1}(f)|$. Then, the clean signal is obtained by substrating the noise to the input signal. This approach is the basic technique. Ephraim & Malah formula or Scalart suggested further corrections. Fanally, the Wiener filtering is applied.

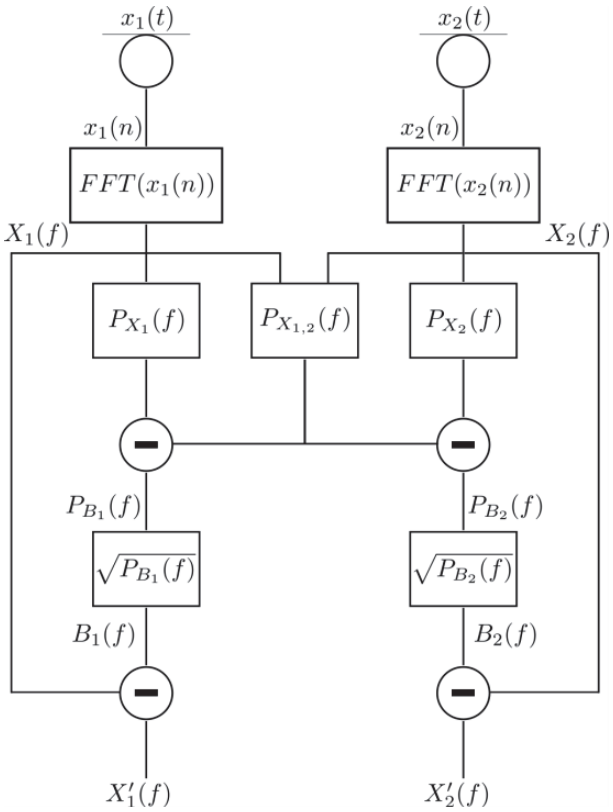


Figure 2. *Synoptic binaural spectral subtraction*

2.2. *Ephraim & Malah or Scalart*

The formula suggested by Ephraim & Malah (Ephraim et al, 1994) for the noise correction is a multiplication by a gain G_L and G_R (on each way):

$$G(f) = \frac{\sqrt{\pi}}{2}\sqrt{\left(\frac{1}{1+R_{post}}\right)\left(\frac{R_{prio}}{1+R_{prio}}\right)} * M\left[(1+R_{post})\left(\frac{R_{prio}}{1+R_{prio}}\right)\right] \quad [1]$$

Where:

- G is the gain that is applied to each short time spectrum
- R_{post} is the a posteriori SNR
- R_{prio} is the a priori SNR, it represents the information of the unknown spectrum (this value defines the attenuation of the musical noise and widely discussed in Cappe (Cappe, 1994))
- M is composed by the modified Bessel functions of zero and first order

The Scalart strategy, similar to the Wiener's approach, was developed to improve the signal to noise ratio in the frequency domain:

$$G(f) = \sqrt{\frac{SNR(f)-1}{SNR(f)}} = \sqrt{1-\frac{1}{SNR(f)}} \quad [2]$$

G(f) is low when the SNR(f) is high.

2.3 *Wiener filter*

The Wiener filter yields an optimal linear filtering. It estimates the SNRs for each frequency. Wiener filtering consists of a multiplication of the signal by a weighting factor W(f).

The Wiener coefficients are obtained by the ratio between the cross-power and the power for each frequency. They strengthen the spectral subtraction and are given by:

$$W_1(f) = P_{X'1.X'2}(f) / P_{X'1}(f) \quad [3]$$

$$W_2(f) = P_{X'1.X'2}(f) / P_{X'2}(f) \quad [4]$$

The higher $W_1(f)$, the more useful the frequency.
Then W(f) is the average of $W_1(f)$ and $W_2(f)$:

$$W(f) = (W_1(f)+W_2(f)) / 2 \quad [5]$$

W(f) is the Wiener filter coefficient for f. It is then applied to the input signal.

$$X''_1(f)^2 = P_{X'1}(f)*W(f) \quad [6]$$

$$X''_2(f)^2 = P_{X'2}(f)*W(f) \quad [7]$$

Thus frequencies less affected by the noise will be favored.
These two strategies have been tested with a population of normal hearing subjects (20 persons).

2.4. *Re-injection*

The re-introduction of the input signal was suggested by van den Bogaert, (Bogaert, 2008). This method returns some intial information. The re-injection of a percentage of unprocessed input signal may restore some of the initial cues and may be used by the ear to locate the source. The formula re-injection is indicated below:

$$\tilde{X}_1(t) = (1- \alpha).X''_1(t) + \alpha.X_1(t) \quad [8]$$

Where $X''_1(t)$ is the denoised signal and $X_1(t)$ is the original signal.

$\tilde{X}_1(t)$ is the corrected signal (figure 1).

The α values 00%, 20% and 40% have been tested. The general organisation of the experiment is indicated on figure 3.

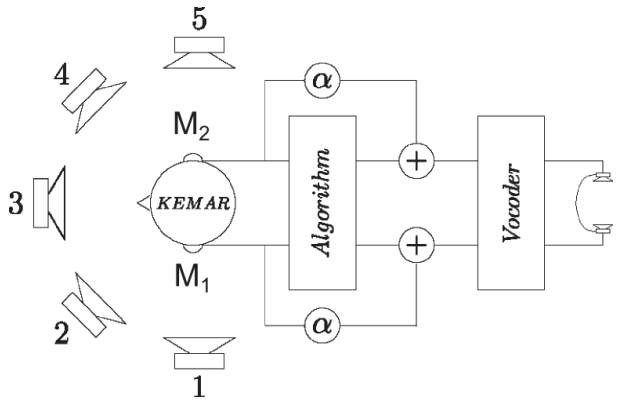


Figure 3. *Synoptic of the experiment*

2.5. *Vocoder*

A vocoder has been used to simulate the cochlear implant (CI) coding. The following characteristics were used in the current work:

- 16 kHz sampling frequency,
- Acquisitions frames: 8ms (128 samples) and an overlap: 2ms (75%).
- 12 electrodes on each ear (12 channels).
- Only the 8 channels above a given energy were kept (8 maxima).

First the frequency range (325 Hz -6000 Hz) was parted into 12 frequency bands of the ear. Cochlea tonotopy is linear up to 800Hz and then logarithmic above. The bark formula (Tjomov, 1971) reproduces this behavior.

$$B(f) = 6.7\ Arcsinh(\frac{f-20}{600}) \quad [9]$$

Where :

- B(f) is in bark

- f in Hz

The eight best energy channels were kept. Then these energies multiplied the center frequency of the corresponding frequency bands (Table 1) according to the sine vocoder strategy. The final signal was recomposed by adding these frequencies.

Channel	Center		Cut off	
	bark	Fréquency (Hz)	Min (Hz)	Max (Hz)
1	1.4	392	325	473
2	2.4	550	473	641
3	3.4	732	641	836
4	4.5	944	836	1067
5	5.5	1197	1067	1344
6	6.5	1502	1344	1679
7	7.5	1871	1679	2087
8	8.5	2322	2087	2585
9	9.5	2873	2585	3195
10	10.5	3550	3195	3945
11	11.5	4381	3945	4866
12	12.5	5402	4866	6000

Table 1. *Frequency distribution of the 12 electrodes; each channel was represented by its center frequency*

2.5. *Listeners*

Twenty normal hearing people participated to the experiment, and their age ranged from 22 years to 29 years (mean 25 years). Their hearing threshold was not higher than 20 dB HL and it was checked by tonal audiometry (half octave) in the ORL department of the Edouard-Herriot hospital. All people were right-handed according to the Hedinburg test. This study was approved by the Ethics Committee (CPP, 0100630314037).

2.6. *Listening protocol*

The listeners task was the localization of the sound source in the anechoic room of the Edouard-Herriot hospital (figure 3). First, the signal was recorded with an

artificial head (KEMAR), M_1 and M_2 were the input microphones. The signal was digitized. Secondly, the denoising algorithms were applied with the re-injection and the vocoder coding. The sound was delivered to the listeners by the headset at 55 dB SPL which was a comfortable level. The source signal to be located was built up by the sums of the words of the Lafon lists [8] with a randomized phase. It lasted 4 seconds. Each situation was repeated three times. The signal came from one of the 5 speakers placed on the horizontal plane. 120 sounds have been proposed during each session (table 2).

Parameters are indicated below:

- 2 algorithms (Ephraim & Malah and Scalart),
- 5 directions (- 90 °, - 45 °, 0 °, 45 °, + 90 °),
- 3 repetitions,
- 3 re-injections (α = 00%, 20% and 40%.)

Then two reference situations have been considered with 3 repetitions and 5 directions each:

- 1 vocoder,
- 1 non-modified

Algorithms	Estimation of noise	Re-injection	Vocoder	Conditions
.	.	.	.	NORMAL
.	.	.	Yes	VOC
Doerbecker	Ephraim & Malah	00 %	Yes	DOEM00
Doerbecker	Ephraim & Malah	20 %	Yes	DOEM20
Doerbecker	Ephraim & Malah	40 %	Yes	DOEM40
Doerbecker	Scalart	00 %	Yes	DOS00
Doerbecker	Scalart	20 %	Yes	DOS20
Doerbecker	Scalart	40 %	Yes	DOS40

Table 2. *Summary of the different conditions.*

The listener's task was to find ("to guess") among the 5 directions, where the source was. The, signal reinjection effect was also considered.

3. Results

The results (figure 4a, 4b) are presented in the following manner: the abscissa indicates the angle of emission of the signal and the y-axis the angle perceived (restituted) by the listener. If a point is on the bisector (line in gray), the emission angle is correctly determined. The calculation of the error E [10], (distance to the diagonal) is calculated and given in table 3.

Figure 4b shows that the error was rather (but not completely) symmetrical.

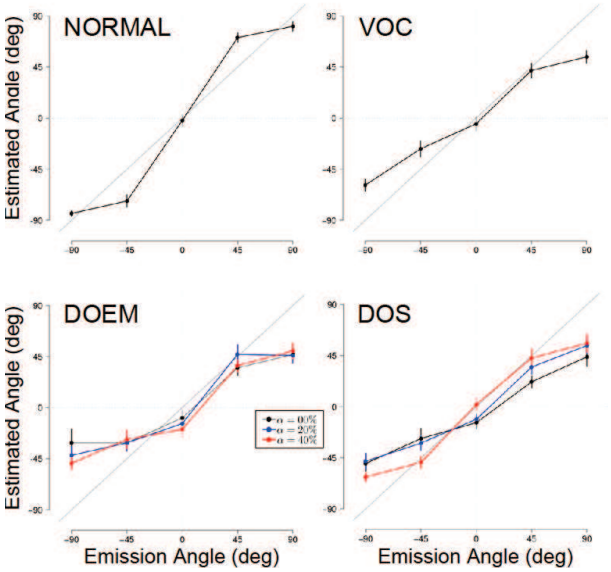


Figure 4a. *Results of localization for different conditions.*

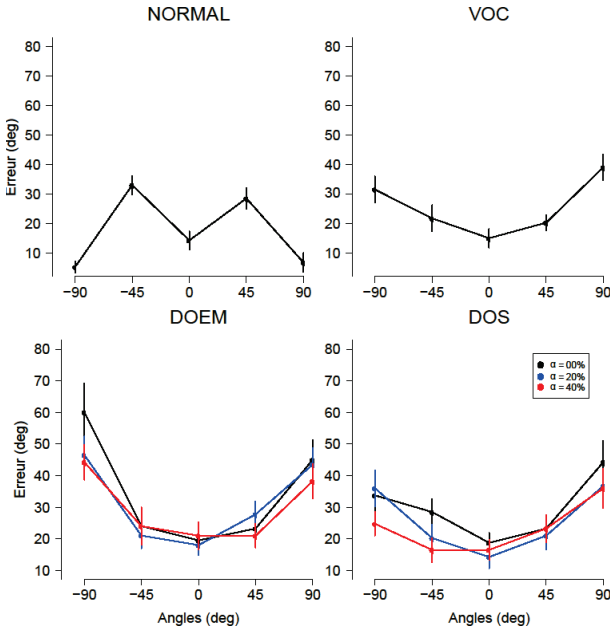


Figure 4b. *Error localization for real angle*

The error is given by:

$$E = \sqrt{\sum_{j=1}^{20} \sum_{k=1}^{5} \left(y_{jk} - diago_k\right)^2} \qquad [10]$$

Where, y_{jk} is the mean (on the three repetitions) response of the person j for a given angle k and diagonal k is the emission angle.

The error, represented in degrees, calculates the distance to the diagonal; the greatest the error, the, worse the strategy. Results are summed up on table 3.

The best strategy was when the signal was not modified (NORMAL).

The VOC condition indicates the vocoder only. The results obtained with the denoising algorithms should be compared to this curve as these algorithms are followed by the vocoder.

From table 3, it can be seen that re-injection brings the curves near to the diagonal. The error is still high for the extreme angles, figure 4b.

Conditions	Errors
NORMAL	212°
VOC	283°
DOEM00	397°
DOEM20	354°
DOEM40	329°
DOS00	355°
DOS20	307°
DOS40	271°

Table 3. *Error according to the condition*

An analytical analysis of the results can be done.

- NORMAL: The distance E obtained (E = 212°) was the smallest among all the conditions, meaning the localization was the best in this case, (without any signal processing).
- When the sound was coming from the front (Θ = 0°), responses were mainly zero; the distances indicated on figure 4, were not significantly different from 0 ($p > 0.05$).
- Directions ±90° were also frequently indicated, when the sound come from the first loudspeaker or the last. Also ±90° was often indicated when the emitting angle was ±45°. When the sound was emitted from the side (Θ = ±90° or Θ = ±45°), the listener tended to indicate full "right" or full "left", without caring about the angle. The extreme angle was the usual decision.
- VOC: The distance E obtained (E = 283°) was situated between the noise-reduction algorithms and the NORMAL condition. It can be seen that noise reduction algorithms deteriorated the localization.
- If the sound was coming from the front, the answer was mainly 0°. The extreme angle was not often decided and responses hardly exceeded ±60°.

3.1. *DOEM*

- DOEM00: The distance E obtained (E = 397°) was the worst among of all the conditions. The dispersion of the responses, for each angle, was very large.
- Responses for the front target were fairly correct.
- The maximum localization for the input angle ±90° was often less than ±45°. Responses for the emission angles ±90° and ±45° were rather similar (p-value > 0.05). These results can be explained by the large dispersion of the responses for the ±90° angles.
- DOEM20: The distance E obtained (E = 354°) is slightly lower than the one obtained in the previous condition (DOEM00) but it was still high. The localization of the sound seemed to be better than DOEM00.
- The error seen for ±90° was slightly inferior to the one observed with DOEM00. The answer +45° was favored. The error was lower for +45° than for -45° (emission angles). ±90° was almost never indicated.
- DOEM40: The error E obtained (E = 329°) was slightly lower than the other DOEM conditions.
- Again ±90° was not often indicated and a fairly high proportion of the subjects indicated a mean response ranging between ±90° and ±45°.

3.2. *DOS*

- DOS00: Responses obtained with the Scalart processing are rather similar to the results seen with the Ephraim & Malah algorithm. The localization with no re-injection was lower than with re-injection. In comparison with DOEM00, the error was smaller.
- For ±90° the error was much smaller than for DOEM00.
- Also, decision 0° was favored.
- DOS20: The distance E obtained (E = 307°) was lower than for DOEM20. The localization was close to the VOC.
- The front target was also indicated by many listeners.
- DOS40: This condition led to good results (EC = 271°), better than those obtained with VOC. This score may come from the ±45° and front target emission, which led to results close to the bisector. Also the angles ±90°were still often forgotten.

4. Discussion

Results indicated that the noise reduction algorithms deteriorated the localization of the sources: "Improvement of speech recognition is done at the expense of localization".

The NORMAL condition gave the most occurate results; it corresponded to the signal taken directly from the KEMAR. The VOC condition, corresponded to the vocoder alone. VOC deteriorated the location, and this was worse when denoising algorithms are used.

With the algorithms, the error decreased when some re-injected signal was introduced. This was mainly true with, the angles +45° or -45°. Then, +90° or -90° was not often indicated. Scalart algorithm performed better than the Ephraim & Malah's approach. When the signal is coded by the vocoder, the ITD is probably lost because the time difference between the two microphones (0.6ms) is small compared to the 8ms frame; the spectral differences are very small and likely to be not significant. About the ILD, (Mills, 1960), it is useful up to 60°, where the localization is good, (Francart, 2011, Shaw, 1974). The re-injection α improves this localization. α restitutes the relative amplitudes of the signal (between the two ears) and helps to restore some localization (Bogaert, 2008). It can be pointed out that the results obtained with DOS 40 were better than those attached to VOC only. Further studies, on the significance of this finding and on the influence of the re-injection percentage, in this situation (cochlear implant simulation), should be carried out both for speech recognition and for the source localization.

5. Conlcusion

This study assessed the loss of localization introduced by a binaural denoising. Conventional speech enhancement algorithms improve the recognition of the speech in noise, but they deteriorate the location of the sources. Some input re-injection after the algorithms allowed, some restoration of the location. The re-injection percentage α = 40%, led to results close to the performance obtained with the vocoder only, mostly when the Scalart algorithm was used. This improvement was mostly efficient for the emission angles +45° or -45°. Thus, re-injection may be a good way to restore source localization when noise reduction algorithms are used.

6. Bibliography

Boll, S. F. « Suppression of acoustic noise in speech using spectral subtraction ». *IEEE* ; 27(2), 113 – 120, 1979.

Cappe, O. « Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor ». *IEEE_J_SAP* 1994, 2(2), 345-349.

Doerbecker, M. and S. Ernst. « Combination of two-channel spectral subtraction and adaptive wiener post-filtering for noise reduction and deverberation ». *European Signal Processing Conference*, 1996.

Ephraim, Y. and D. Malah. « Speech enhancement using a minimum meansquare error log-spectral amplitude estimator ». *IEEE Transaction on Acoustics, Speech, and Signal Processing* ; 33(2), 443 – 445, 1985.

Francart T, Lenssen A, Wouters J. « Enhancement of interaural level differences improves sound localization in bimodal hearing ». *J Acoust Soc Am.*, Nov ; 130(5) : 2817 - 26, 2011

Kamath S, Loizou P. « A multi-band spectral subtraction method for enhancing speech corrupted by colored noise ». In: *Proceedings of ICASSP-2002*, Orlando; 2002.

Kistler DJ, Wightman FL. « A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction ». *J Acoust Soc Am.* ; Mar ; 91(3) : 1637 - 47, 1992

Lafon, J. C. (1958), *Le test phonétique*, Mason, ed. Paris.

Mills, A. W. « Lateralization of high-frequency tones ». *J Acoust Soc Am.* ; 32(1), 132 –134, 1960

Moore, B. C. *An introduction to the Psychology of Hearing*, Fifth Edition. *Emerald Group Pub Ltd*, (book) 1995.

Rayleig, L. « On our perception of sound direction ». *Physiolosophical Magazine Series* ; 6 13(74), 214 – 232, 1907.

Scalart, P. and J. Viera Filho. « Speech enhancement based on a priori signal to noise ratio estimation ». *IEEE* ; 629 – 632, 1996.

Shaw EA. « Transformation of sound pressure level from the free field to the eardrum in the horizontal plane ». *J Acoust Soc Am.* ; Dec ; 56(6) : 1848 - 61, 1974.

Tjomov, V. and G. Gershuni. « A model to describe the results of psychoacoustical experiments on steady-state stimuli ». In *Analiz Rechevykh Signalov Chelovekom, Nauka* ; 36 – 49, 1971.

Truy, E., Ionescu, E., Ceruse P., and Gallego, S.. « The binaural digisonic cochlear implant: surgical technique ». *Otol Neurotol* ; 23(5), 704 – 709, 2002.

van den Bogaert, T., S. Doclo, J. Wouters, and M. Moonen. « The effect of multimicrophone noise reduction systems on sound source localization by users of binaural hearing aids ». *J Acoust Soc Am* ; 124(1), 484 – 497, 2008.

Verhaert, N.; Lazard, D.; Gnansia, D.; Bébéar, J.; Romanet, P.; Meyer, B.; Péan, V.; Mollard, D. & Truy, E. (2012), « Speech performance and sound localization abilities in Neurelec Digisonic® SP binaural cochlear implant users »., *Audiol Neurootol.* 17(4), 256--266.

Juin 2012,

Influence du débruitage spatial sur la localisation dans le cas du codage de l'implant cochléaire binaural. Simulation sur normo-entendants.,

Handicap 2012, Paris,

Proceding + Présentation orale

Influence du débruitage spatial sur la localisation dans le cas du codage de l'implant cochléaire binaural. Simulation sur normo-entendants

Arnaud Jeanvoine[1,2,4], Dan Gnansia[4], Eric Truy[1,2,5], Rafael Laboissière[1,2], Christian Berger-Vachon[1,2,3]

1 : INSERM U1028, Centre de recherche de Neuroscience de Lyon, équipe PACS (Speech, Audiology, Communication Health), Lyon, F-69000, France
2 : CNRS UMR5292, Centre de recherche de Neuroscience de Lyon, équipe PACS (Speech, Audiology, Communication Health), Lyon, F-69000, France
3 : University Lyon 1, Lyon, F-69000, France
4 : Neurelec, 2720 Chemin Saint-Bernard, Vallauris, F-06224, France
5 : Audiologie et Département ENT, Edouard Herriot Hospital, Lyon, F-69437, France
arnaudjeanvoine@yahoo.fr

Résumé : Les implants cochléaires fournissent une bonne intelligibilité de la parole dans le calme, mais comprendre la parole dans un environnement bruyant est toujours difficile pour les utilisateurs. Le manque d'information binaurale peut être une des explications possible. En effet, les informations spatiales permettent de séparer les sources arrivant aux oreilles et donc d'extraire le bruit du signal que l'on voudrait entendre. Les techniques de traitement du signal par débruitage binaural sont couramment utilisées dans les aides auditives conventionnelles, mais pas encore dans l'implant cochléaire. En ce qui concerne l'implant bilatéral, il permet d'obtenir une information venant de chaque oreille mais les processeurs ne communiquent pas entre eux. L'implant binaural possède un unique processeur permettant un croisement des informations arrivant sur chaque oreille et ainsi, appliquer des débruitages binauraux comme dans l'aide auditive.

Actuellement plusieurs méthodes de traitement du signal existent pour améliorer le rapport signal sur bruit. Ces méthodes binaurales ont pour but d'améliorer la directivité des microphones. Un des principes utilisés par ces algorithmes est que les sources sont séparées spatialement les unes des autres. A priori, plus les signaux sont séparés entre eux, plus le débruitage sera efficace. En contre-partie, la directivité risque de perturber la localisation des sons. Afin de définir des paramètres les efficaces dans l'implant, une étude avec des sujets normo-entendants et l'utilisation d'un vocodeur reproduisant le traitement binaural de l'implant, plus un algorithme binaural, a été réalisée. Ainsi plusieurs paramètres tels que la ré-injection, l'angle de provenance du bruit et le rapport signal sur bruit ont été étudiés en localisation.

Les résultats obtenus montrent que la localisation sans algorithme binaural donne une localisation proche d'une condition réelle. L'ajout d'un algorithme influe sur la localisation des sources. La ré-injection du signal d'entrée permet d'améliorer la localisation des sons sans détériorer la perception de la parole.

Cette première approche suggère qu'un algorithme binaural dans l'implant, pourrait permettre de retrouver une localisation et une perception des sons de parole en milieu bruité.

Mots clés : Implant cochléaire binaural, vocodeur, débruitage binaural, localisation.

I. INTRODUCTION

Pour comprendre et interpréter un son, les personnes normo entendantes utilisent de leurs deux oreilles. En effet, il existe des indices binauraux qui améliorent la perception des sons dans l'espace. Ces indices sont basés sur la différence de temps interaural (DTI) et sur la différence de niveau interaural (DNI), [6,8,9]. Ce premier indice est principalement présent dans les basses fréquences (f < 800 Hz) et le second pour les hautes fréquences (f > 3000 Hz). Ces indices permettent de localiser les sons dans l'espace, et aussi de sélectionner une source sonore par rapport à une autre à partir des différents rapports signal sur bruit (RSB) qui arrivent aux oreilles. Le meilleur RSB sera à priori sélectionné et interprété.

Si l'on considère un sujet implanté cochléaire, comprendre la parole en milieu bruité devient difficile. Tout d'abord, le traitement de l'information se fait différemment. Le pavillon de l'oreille étant court-circuité, l'aspect détection des sources est alors perdu. Un traitement du signal peut retransmettre ces informations. Il existe plusieurs façons d'améliorer la perception dans le bruit. La première est de mettre un implant cochléaire sur chaque oreille. Dans ce cas, l'audition est dite bilatérale car il n'y pas de traitement de l'information entre les deux implants. Une autre façon est l'implant binaural, [13], qui a un seul processeur pour simuler les deux oreilles. Celui-ci permet d'effectuer un traitement conjoint entre les sons arrivant sur le microphone gauche et le microphone droit. Dans ces conditions, les traitements binauraux effectués dans l'aide auditive conventionnelle peuvent être adaptés à l'implant binaural [2].

Actuellement cet implant n'effectue aucun traitement entre les deux voies. Ce qui est reçu sur le microphone droit et envoyé sur les électrodes à droite et ce qui est reçu à gauche est envoyé sur les électrodes de gauche. Les premiers résultats obtenus avec cet implant ne montrent pas de différence significative avec les personnes portant un implant bilatéral. Ces résultats peuvent être améliorés avec des méthodes de réduction de bruit (débruitage) utilisés dans l'aide auditive, [1,2,3,5,10].

Il existe plusieurs traitements binauraux qui existent à ce jour tels que la soustraction spectrale, [1,5] ou bien le filtrage de Wiener. La principale hypothèse est que l'on considère que la source utile provient de face. Nous regardons, ce que nous écoutons. Les algorithmes de débruitage devront détecter et rehausser ce qui provient de face. Ces algorithmes ont déjà démontré leurs performances dans l'aide auditive, [1,2]. La soustraction spectrale binaurale améliore le signal provenant de face, mais elle introduit un bruit musical dû à la mise à zéro de certaines fréquences. Pour atténuer le bruit musical, il existe une méthode développée par Ephraim & Malah, [3]. On peut aussi utiliser une autre méthode d'estimation qui est basée sur le modèle de Scalart, [10] et qui améliore le rapport signal sur bruit de la trace.

Les méthodes de débruitage ont un effet pervers ; elles favorisent ce qui provient de face, et perturbe la localisation, [7], des sources sonores environnantes. Un facteur α de ré-injection peut être utilisé pour ré-introduire le signal d'entrée sur la sortie pour restituer une meilleure localisation, [14]. Ceci est testé dans cet article.

II. MATERIEL ET METHODE

La base du débruitage du signal d'entrée dans l'implant cochléaire binaural se situe dans le domaine spectral. L'algorithme de Doerbecker [2], classiquement utilisé dans l'aide auditive, permet d'améliorer la parole provenant de face. Il est composé de deux modules. Le premier est une soustraction spectrale [1] et le second est un filtrage adaptatif de Wiener. Dans la soustraction spectrale, l'introduction d'un estimateur de bruit de type Ephraim & Malah, [3] ou de type Scalart, [10] a été effectuée.

A. La soustraction spectrale

La soustraction spectrale binaurale est composée d'un module estimateur de bruit et ensuite d'une soustraction (Fig. 1).

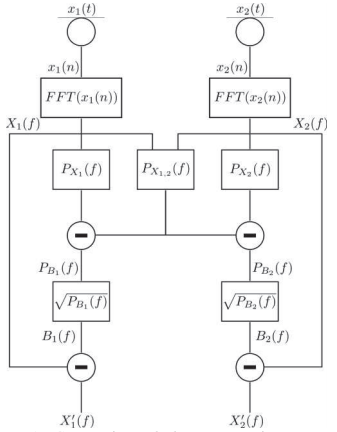


Figure 1 : Synoptique de la soustraction spectrale binaurale

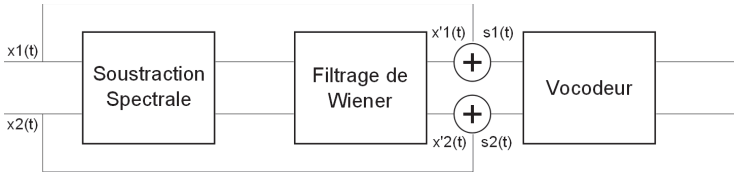


Figure 1b : Synoptique des différentes étapes du traitement du signal

Une FFT est appliquée aux deux signaux d'entrée. La puissance est ensuite calculée en prenant l'énergie du signal d'entrée $P_{X1}(f) = X_1(f)^2$, ainsi que la puissance croisée des deux côtés, $P_{X1X2}(f)$ qui et commun aux deux parties et représente donc le signal. La dernière étape consiste à soustraire au signal d'entrée au bruit pour obtenir un signal débruité. L'estimation présentée ici est la technique de base. On peut rajouter la formule d'Ephraim & Malah dans l'estimation du bruit ou celle de bien Scalart. Une fois que le module soustraction spectrale est effectué, le filtrage de Wiener est alors appliqué.

B. Le filrage de Wiener

Le filtrage de Wiener est le filtrage linéaire optimal. En fonction des caractéristiques d'entrée du signal il va appliquer le meilleur RSB possible entre le bruit estimé et le signal utile. Le filtrage de Wiener est contrairement à la soustraction spectrale une multiplication du signal par un filtre W(f).

Les coefficients du filtre en binaural sont obtenus en effectuant le rapport entre la puissance croisée et la puissance de chaque voie obtenue, de la même façon que pour la soustraction spectrale. Ainsi :

$$W_1(f) = P_{X'1,2}(f) / P_{X'1}(f) \quad (1)$$
$$W_2(f) = P_{X'1,2}(f) / P_{X'2}(f) \quad (2)$$

Où, $W_i(f)$ est le pourcentage de signal utile.

En prenant la W(f) comme étant la moyenne de $W_1(f)$ et $W_2(f)$, on obtient :

$$W(f) = (W_1(f) + W_2(f)) / 2 \quad (3)$$

On appelle W(f) les coefficients du filtre de Wiener. Ensuite ils sont appliqués au signal d'entrée.

$$S_1(f)^2 = \sqrt{P_{X'1}(f) * W(f)} \quad (4)$$
$$S_2(f)^2 = \sqrt{P_{X'2}(f) * W(f)} \quad (5)$$

On voit donc que les fréquences peu affectées par le bruit sont favorisées.

Nous avons testé cette stratégie avec une population de sujets normo-entendants pour disposer d'un échantillon de taille raisonnable (20 personnes).

C. La ré-injection

La ré-introduction du signal d'entrée a été présenté par Van den Bogaert, [14]. C'est une méthode simple qui permettrait de restituer des informations binaurales. En effet en ré-injectant un peu de signal bruité d'entrée, donc contenant des informations binaurales, le signal de sortie pourrait retrouver un peu de localisation. La formule utilisée en temporel est :

$$s_1(t) = x'_1(t) + \alpha.x_1(t) \quad (6)$$

Où, $x'_1(t)$ est le signal débruité et $x_1(t)$ est le signal bruité.

Les proportions α choisis sont de 00%, 20% et 40%. La méthode permet de tester sans ré-injection, de ré-injecter plus ou moins du signal d'entrée.

D. Le vocodeur

Le vocodeur est un traitement du signal qui permet de se rapprocher de l'implant. En effet, celui-ci reproduit les caractéristiques fréquentielles et temporelles de l'implant qui sont une fréquence d'échantillonnage de 16 kHz, des trames d'acquisitions de 8 ms (128 échantillons) et un recouvrement de 2 ms (75%). La technologie de l'implant utilise les maxima d'énergie. Il possède sur chaque porte électrode 12 électrodes.

La première étape consiste à séparer le signal suivant la tonotopie de l'oreille. Celle-ci étant linéaire jusqu'à 800 Hz puis logarithmique au-delà. La formule utilisée est la formule de bark, [12]. Le signal est ensuite séparé en 12 bandes de fréquences. Les huit maxima d'énergie sont conservés (méthode d'extraction des maxima). Puis ils sont multipliés par la fréquence centrale de la bande de fréquence correspondante (Tab. 1). Le signal est ensuite recomposé en additionnant les différentes fréquences.

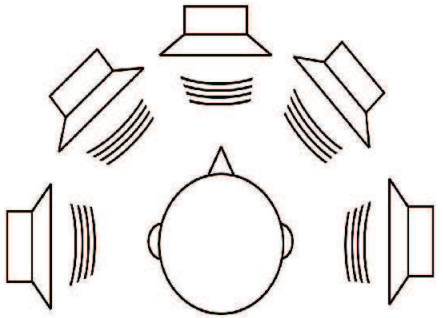
Canal	Centrale		Coupure	
	bark	Fréquence	Min	max
1	1.4	392	325	473
2	2.4	550	473	641
3	3.4	732	641	836
4	4.5	944	836	1067
5	5.5	1197	1067	1344
6	6.5	1502	1344	1679
7	7.5	1871	1679	2087
8	8.5	2322	2087	2585
9	9.5	2873	2585	3195
10	10.5	3550	3195	3945
11	11.5	4381	3945	4866
12	12.5	5402	4866	6000

Tableau 1 : Répartition fréquentielle des 12 électrodes

E. Auditeurs

Vingt personnes normo-entendantes ont participé à cette expérience. La moyenne est de 25 ans, le minimum est de 22 ans et le maximum est de 29 ans. Le seuil de l'audition n'est pas supérieur à 20 dB HL sur les fréquences caractéristique de l'audiométrie tonale (par demi octave). Toutes les personnes sont raportées droitières par le test d'Hedinburg. Cette étude a été approuvée par le comité d'étique.

F. Protocole d'écoute


Figure2 : Représentation du placement des hauts parleurs

L'étude effectuée dans cet article est un test de localisation des sources sonores. La première étape consiste à enregistrer à l'aide d'une tête artificielle (KEMAR) qui permet de récupérer les informations binaurales car elle possède les caractéristiques d'une tête humaine. En sortie des microphones, placés dans les oreilles du KEMAR, le signal est numérisé. Ensuite les deux algorithmes de débruitage sont appliqués avec la ré-injection puis le vocodeur. Le son passe alors d'un environnement trois dimensions pour avoir en sortie un environnement à deux dimensions. Les auditeurs ont entendu le signal via un casque et non en champ libre. Le son délivré par le casque est de 55 dB SPL qui est un niveau confortable. L'utilisation de parole n'étant pas neutre (stationnaire dans le temps) et pour réduire la durée de l'expérience, un speech noise (bruit de parole) composé des sommes des mots des listes de Lafon avec une phase randomisée a été effectué. Il dure 4 secondes et contient tous les spectres des mots. Chaque situation est répétée trois fois. Le signal provient de 5 haut-parleurs placés sur le plan horizontal. 150 sons ont été proposés par session (tableau 2).

- 2 algorithmes (Ephraim & Malah et Scalart)
- 5 directions (-90°, -45°, 0°, +45°, +90°)
- 3 répétitions
- 3 ré-injections (α = 00%, 20% et 40%.)
- 1 vocodeur
- 1 non modifié

Algorithmes	Estimation du bruit	Re-injection	Vocodeur	Conditions
.	.	.	.	Normal
.	.	.	Oui	Voc
Doerbecker	Ephraim & Malah	00 %	Oui	Doem00
Doerbecker	Ephraim & Malah	20 %	Oui	Doem20
Doerbecker	Ephraim & Malah	40 %	Oui	Doem40
Doerbecker	Scalart	00 %	Oui	Dos00
Doerbecker	Scalart	20 %	Oui	Dos20
Doerbecker	Scalart	40 %	Oui	Dos40

Tableau 2 : Récapitulatif des différentes conditions.

La tâche de l'auditeur est de trouver parmi les 5 directions, la provenance du son. L'effet de la ré-injection du signal sur la localisation sera plus particulièrement regardé.

III. RESULTATS

Les résultats (figure 3) sont présentés de la façon suivante : l'abscisse indique l'angle d'émission et l'ordonnée, l'angle perçu par l'auditeur. Si un point se trouve sur la diagonale, l'angle d'émission est correctement déterminé. Le calcul de l'erreur (7), par rapport à la diagonale est calculé et donné dans le tableau 3.

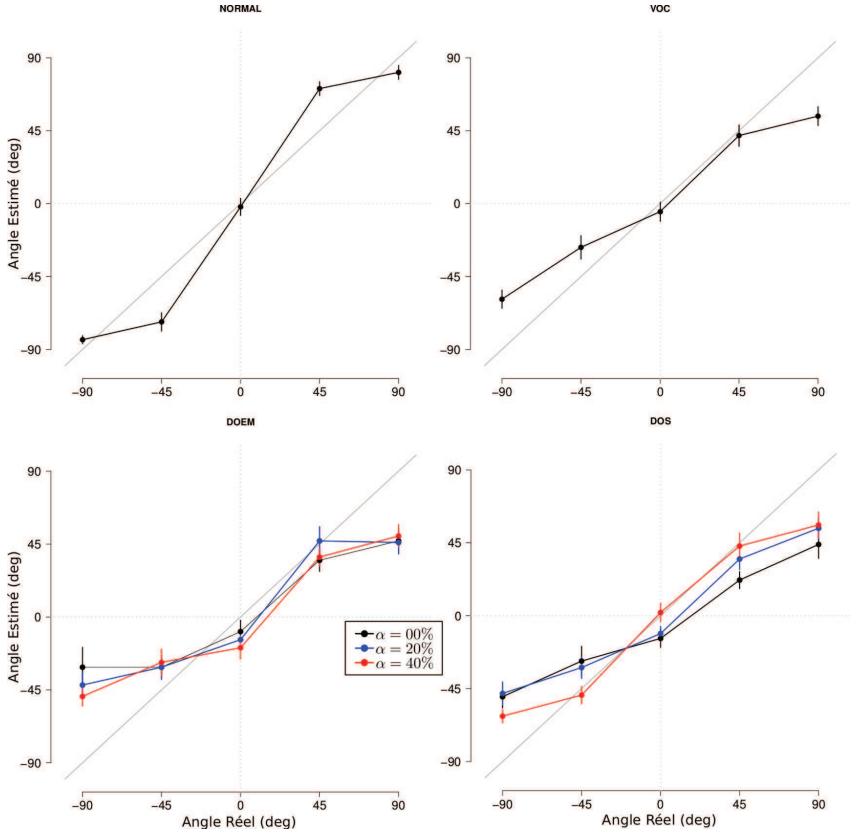


Figure 3 : Résultats de localisation pour les différentes conditions.

$$\sqrt{\sum_{J=1}^{20} (y_j - \text{diagonale})^2} \qquad (7)$$

Où, y_j est la réponse de la personne pour un angle donné et diagonale est l'angle de la source d'émission.

L'erreur représentée en degré indique la distance par rapport à la diagonale. Plus elle est grande, plus la stratégie conduit à des erreurs.

La condition la plus proche de la diagonale est celle qui est non modifiée (Normale). On indique plus facilement la direction de coté que les angles intermédiaires. Le coté joue peu : + ou – 45° n'a pas été privilégié. La condition Voc est celle de l'implant binaural seul. C'est donc vers cette courbe que les résultats des algorithmes de débruitage doivent converger.

Pour les algorithmes, on remarque que plus la ré-injection est grande, plus la courbe se rapproche de la diagonale. Seuls les angles extrêmes sont peu indiqués.

Conditions	Erreurs
Normal	212°
Voc	283°
Doem00	397°
Doem20	354°
Doem40	329°
Dos00	354°
Dos20	307°
Dos40	271°

Tableau 3 : Erreurs en fonction des Conditions

IV. DISCUSSION

Les résultats obtenus permettent de mettre en avant une dégradation de la localisation lorsque l'on favorise une direction. L'amélioration de la reconnaissance de la parole s'effectue au détriment de la localisation.

La condition « Normal » est celle qui est la plus proche de la diagonale ; elle correspond au signal non modifié, et il est pris directement en sortie du KEMAR. La condition « Voc », correspond à l'implant cochléaire binaural seul. Voc détériore la localisation, cette localisation est encore plus détériorée lorsque l'on couple avec des algorithmes des débruitage.

Pour les algorithmes, l'erreur diminue lorsque l'on ré-injectedu signal non modifié. Dans ce cas, l'angle + ou – 45° est plus facilement indiqué. On remarque de plus que + ou – 90° n'est jamais indiqué. L'algorithme de Scalart a moins déterioré l'orientation que celui d'Ephraim & Malah. Ces résultats sont en accord avec la littérature. Lorsque la trame du signal est décomposée pour le vocodeur, les ITD sont perdus car cette différence de temps est inférieure à la longueur de la trame et de son recouvrement. Il ne reste alors que les ILD, [7]. Les ILD permettent de localiser un son jusqu'à 60°, ensuite, la localisation n'est plus bonne, [4,11]. La ré-injection α améliore la localisation. On peut assimiler α à un algorithme de rehaussement des informations binaurales ne jouant que sur les amplitudes des signaux [14].

V. CONCLUSION

L'étude a porté sur la perte de la localisation due au débruitage binaural. Si les algorithmes classiques de rehaussement de la parole permettent d'améliorer celle-ci dans le bruit, la localisation des sources environnantes est alors détériorée. La proposition de ré-injecter le signal d'entrée et non modifié par van den Bogaert permet, par l'intermédiaire du facteur α, de rétablir une partie de la localisation. En effet, avec α = 40 %, on s'approche les résultats du vocodeur sans algorithme pour la condition dos. Ceci indique qu'avec une ré-injection du signal d'entrée, la localisation peut être restauré jusqu'à + ou – 45°.

VI. BIBLIOGRAPHIE

[1] Boll, S. F. Suppression of acoustic noise in speech using spectral subtraction. *IEEE* ; 27(2), 113 – 120, 1979.

[2] Doerbecker, M. and S. Ernst. Combination of two-channel spectral subtraction and adaptive wiener post-filtering for noise reduction and deverberation. *European Signal Processing Conference*, 1996.

[3] Ephraim, Y. and D. Malah. Speech enhancement using a minimum meansquare error log-spectral amplitude estimator. *IEEE Transaction on Acoustics, Speech, and Signal Processing* ; 33(2), 443 – 445, 1985.

[4] Francart T, Lenssen A, Wouters J. Enhancement of interaural level differences improves sound localization in bimodal hearing. *J Acoust Soc Am.*, Nov ; 130(5) : 2817 - 26, 2011

[5] Kamath S, Loizou P. A multi-band spectral subtraction method for enhancing speech corrupted by colored noise. In: *Proceedings of ICASSP-2002*, Orlando; 2002.

[6] Kistler DJ, Wightman FL. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. J Acoust Soc Am. ; Mar ; 91(3) : 1637 - 47, 1992

[7] Mills, A. W. Lateralization of high-frequency tones. *J Acoust Soc Am.* ; 32(1), 132 –134, 1960

[8] Moore, B. C. An introduction to the Psychology of Hearing, Fifth Edition. *Emerald Group Pub Ltd*, (book) 1995.

[9] Rayleig, L. On our perception of sound direction. *Phiolosophical Magazine Series* ; 6 13(74), 214 – 232, 1907.

[10] Scalart, P. and J. Viera Filho. Speech enhancement based on a priori signal to noise ratio estimation. *IEEE* ; 629 – 632, 1996.

[11] Shaw EA. Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. *J Acoust Soc Am.* ; Dec ; 56(6) : 1848 - 61, 1974.

[12] Tjomov, V. and G. Gershuni. A model to describe the results of psychoacoustical experiments on steady-state stimuli. In *Analiz Rechevykh Signalov Chelovekom, Nauka* ; 36 – 49, 1971.

[13] Truy, E., E. Ionescu, P. Ceruse, and S. Gallego. The binaural digisonic cochlear implant: surgical technique. *Otol Neurotol* ; 23(5), 704 – 709, 2002.

[14] van den Bogaert, T., S. Doclo, J. Wouters, and M. Moonen. The effect of multimicrophone noise reduction systems on sound source localization by users of binaural hearing aids. *J Acoust Soc Am* ; 124(1), 484 – 497, 2008.

Avril 2011,

Improved speech perception in noise for cochlear implantees using binaural signal processing,

ESPCI 2011, Athènes (Grèce),

Poster

IMPROVED SPEECH PERCEPTION IN NOISE FOR COCHLEAR IMPLANTEES USING BINAURAL SIGNAL PROCESSING

Arnaud JEANVOINE[1], Christian BERGER-VACHON[1], Dan GNANSIA[2], Eric TRUY[1,3], Hung THAI-VAN[1,3]

1. INSERM U1028, CNRS UMR5292, Lyon, France
2. MXM – Neurelec, Vallauris, France
3. Hospice Civils de Lyon, Hôpital Edouard Herriot, Département ORL, de chirurgie Cervico-Maxillo Faciale et d'Audiophonologie, Lyon, France

Introduction :

Cochlear implants provide good speech intelligibility in quiet, but speech understanding in noisy environments is still a challenge.
Because spatial information allows to separate different sources and thus improves the speech to noise ratio, providing binaural cues seems essential. Binaural denoising, which is the result of different algorithms interaction, is currently used in conventional hearing aids, but not yet in cochlear implants, as such techniques cannot be implemented on two independent processors (as used in bilaterally-implanted patients). However, the Neurelec® binaural CI has a single processor, allowing the implementation of a true binaural processing technique.
In order to identify the best algorithm for binaural processing, 15 normally hearing subjects underwent tests in vocoded condition, on two modalities: a localization task and a perception task (speech perception in noise).

Material & Method :

Recorded Material:

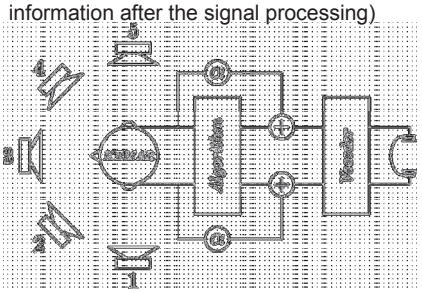
- We used monosyllabic word lists (Lafon, recorded using a Kemar head), in a speech condition (SNR: -6; 0 and 3 dB), with respectively signal at 0° and noise presented at -90°, -45 °, 0 °, 45 ° and 90 ° azimutal positions.
- speech noise is obtained by the randomized the phase of the words
- Stimuli is presented at 65 dB SPL

Task: Speech perception in noise

The percentage of intelligibility was obtained by evaluating each time the word produced by the headphone was correctly said by the listener.

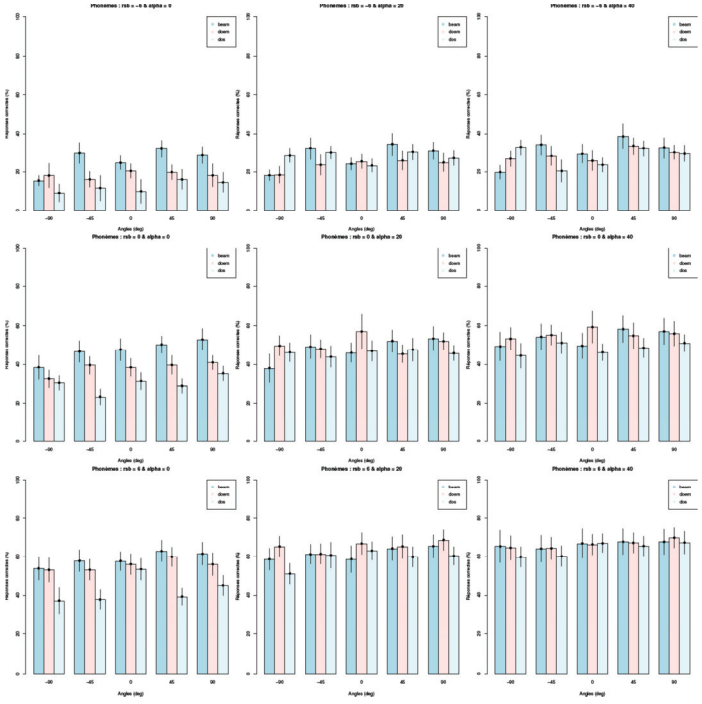
Conditions:

- Non-modified speech (unproc)
- Vocoded speech
- Vocoded speech with different algorithms:
- Beamformer
- Doerbecker (Estimation by Ephraim & Malah and Scalart with α = 0.0, 0.2, 0.4 (α is used to reintroduce the binaural information after the signal processing)

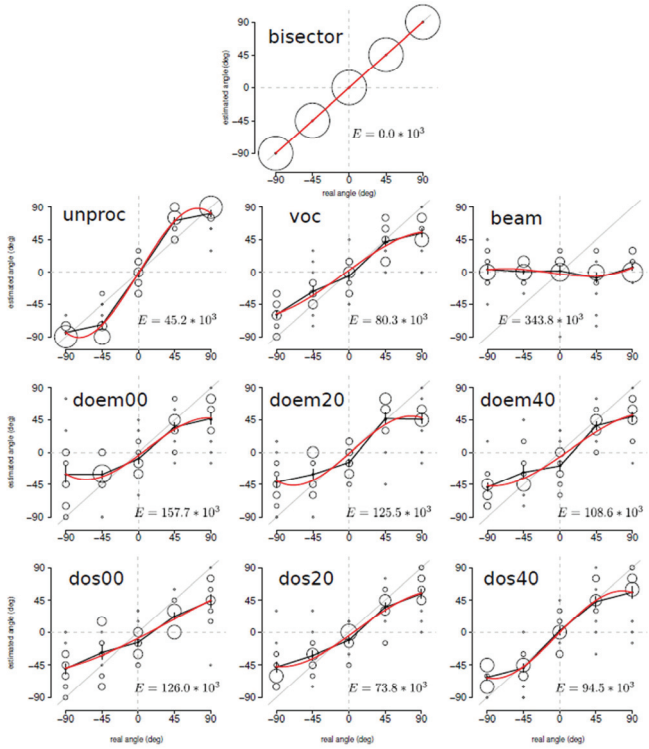


Task: Localization

The sound was presented through one out of the five headphones and the listener has to click on the button corresponding to the direction from which he believes the sound comes.



Results & Discussion :



Localization: The direction 45° was found with difficulties for the control condition, whereas the -90° and 90° directions seemed easily located. No difference was found between the different conditions concerning the 0° azimuth direction.
The sounds originating from the different directions are attenuated by the use of the Beamformer technique excepted for the 0° azimut direction. Concerning the two last conditions (DOS and DOEM algorithms), the more α increases, the better the localization was. Besides, localization tasks results were better with the Dos04 algorithm or the Vocoder than with the beamformer technique.
For the localization task, the best algorithm was the Doerbecker one (with a Scalart estimation and α = 0.4), permitting an azimuth identification closer to Vocoder 's one.

Speech perception in noise: The best results were obtained using DoEM algorithm with different α. The percentage of recognition was better as the signal to noise ratio increased ($p<0.001$). Beamformer technique at α = 0 gave best results, in particular for the -45° and +45° directions.

The best algorithm for the monosyllabic words recognition was the Dorbecker 's one (estimation by Ephraim & Malah with α = 0.2);

Conclusion :

Because of the limited differences observed between the different algorithms and the vocoder performances, the determination of the best algorithm remains challenging. However, it seems that the Doerbecker algorithm (with estimation Ephraim and Malah by an α equal to 0.2) would be the best compromise for sound recognition and localization.

References :

- M. Dorbecker and S. Ernst, "*Combination of Two Channel Spectral Subtraction and Adaptive Wiener Post Filtering for Noise Reduction and Dereverberation*", Proc EUSIPCO 1996, Triest 1996
- W.R. Drennan et al., "*Localization and Speech-Identification Ability of Hearing-Impaired Listeners Using Phase-Preserving Amplification*", Ear & Hearing 2005; 26 : 461–472
- V. d. Bogaert et al., "*Horizontal localization with bilateral hearing aids : Without is better than with*", J. Acoust. Soc. Am., 119, (1), January 2006 : 515–526
- V. d. Bogaert *et al.*, "*The effect of multimicrophone noise reduction systems on sound source localization by users of binaural hearing aids*", J. Acoust. Soc. Am., 124 (1), July 2008 : 484–497
- V. d. Bogaert et al., "*Speech enhancement with multichannel Wiener filters techniques in ultimicrophone binaural hearing aids*", J. Acoust. Soc. Am., 125, (1) January 2009 : 360–371

Juillet 2010,

Binaural denoising algorithms associated with the Digisonic SP binaural cochlear implant. Simulations on normal-hearing subjects,

CI 2010, Stockholm (Suède),

Poster

Binaural Denoising Algorithms associated with the Digisonic® SP Binaural Cochlear Implant

Simulations on normal-hearing subjects

Arnaud Jeanvoine[1], Christian Berger-Vachon[1], Dan Gnansia[2], Hung Thai-Van[1]

1: UMR CNRS 5020 – HEH, 5 place d'Arsonval, 69003 Lyon

2: MXM - Neurelec, 2720 Chemin St Bernard, 06224 Vallauris

INTRODUCTION

Cochlear implants provide good intelligibility in quiet, but speech understanding in noisy environment remains challenging.

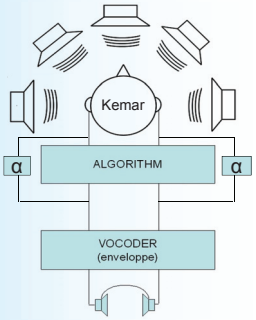
Providing binaural cues could be a possible solution, because spatial information allows to separate different sources and therefore improves the speech to noise ratio. Binaural denoising, performed using different algorithms, is used in conventional hearing aids, but not in cochlear implants, as such techniques cannot be implemented on two independent processors, used in bilaterally-implanted patients. However, the Neurelec binaural CI has a single processor, allowing the implementation of a true binaural processing technique, in order to improve perception in noise. However, in order to maintain some degree of localization cues, a proportion (α) of the input signal is added to the output of the algorythms.

In order to identify the best algorithm (best compromise denoising/localization) and the proportion α, we tested, using vocoded sounds, 15 normally hearing subjects in two tasks: sound localization and speech perception in noise, using 9 different conditions (7 combinations of algorithms and α, no binaural processing (vocoder), control).

MATERIAL & METHODS

Recorded Material:
Monosyllabic word lists (Lafon), in speech noise with a SNR of -6, 0 and 3 dB, recorded using a Kemar head, with signal at 0° and noise presented at -90°, -45°, 0°, 45° and 90° azimutal positions.

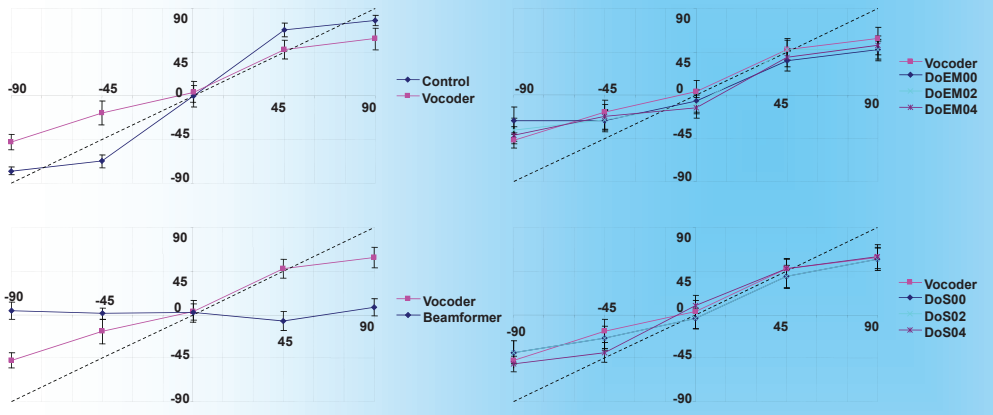
Conditions:
- Non-modified speech (control)
- Vocoded speech
- Vocoded speech with different algorithms:
- Beamformer
- Doerbecker (Estimation by Ephrahim & Malah (DoEM) and Scalart (DoS) with α = 0.0, 0.2, 0.4)



Sound localization was assessed using a 5 alternatives forced choice task, where the subject listened to each sound 3 times, in each condition, presented in a randomized order.

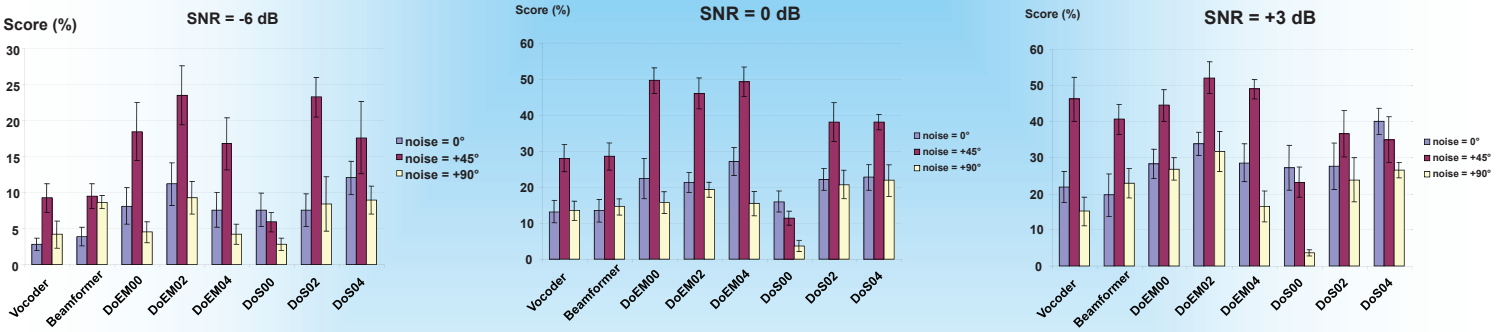
Speech perception in noise was assessed in 3 directions x 3 SNR x 8 conditions presented in random order. Subjects were asked to repeat the words they heard, and percentages of correctly repeated phonemes were calculated.

RESULTS & DISCUSSION



Localization: A general linear model (quasibinomial family), was fitted to the data. In the control condition, the 45° direction was significantly the most difficult to identify. For vocoded sounds, DoS (for all α), DoEM00 and DoEM04 algorithms, no significant differences were obtained according to the different azimutal positions. However, the Beamformer algorithm yielded very high performance at 0°, contrasting with significantly poorer performances in all other directions ($p<0.001$). For DoEM02, significant better performances were obtained at 0° versus 45 and 90° azimuthal positions ($p<0.05$).

Further analysis was conducted on the subset of DoS and DoEM algorithms: significant differences in scores were obtained depending on the azimut ($p<0.001$), and on the algo ($p<0.05$), without significant interaction azimut x algo. Comparisons between the different DoS and DoEM algorithms, for each azimuthal position, didn't show any significant differences.



Speech perception in noise: A general linear model (quasibinomial family), was fitted to the data. Significant differences in scores were obtained according to the algorithm, direction, and SNR chosen, with significant interactions between algorithm and direction ($p<0.0001$), but no interaction with the different SNRs. Except for the DoS04 algorithm, all scores were significantly greater for azimut 45 than for the other two angles.

For a 0° azimut, the Dos04 yielded a significantly greater score than the other algos. For a 45° azimut, the DoEM02 yielded a significantly greater score. For a 90° azimut, a significant difference was obtained for the DoS00 algo.

Further analysis was performed on the subset of DoEM and DoS algorithms: For 0° azimut, no significant differences were obtained between the different DoS and DoEM algorithms. For 45° and 90° azimut, the DoS00 algorithm gave significantly poorer performance than the DoEM algorithms ($p<0.01$ with DoEM00).

CONCLUSION

Those results do not allow to identify the best algorithm for all conditions. As expected, the Beamformer abolishes localization abilities and centers the auditory image.The DoEM02 showed greater localization scores for an azimut of 0°, and, in addition, it gave the greatest speech perception scores for the 3 angles tested (0, 45 and 90°). Therefore, the Doerbecker with estimation Ephrahim and Malah with α = 0.2 seems to be the best compromise between localization and speech perception in noise.

References

- M. Dorbecker and S. Ernst, "*Combination of Two Channel Spectral Subtraction and Adaptive Wiener Post Filtering for Noise Reduction and Dereverberation*", Proc EUSIPCO 1996, Triest 1996
- W.R. Drennan et al., "*Localization and Speech-Identification Ability of Hearing-Impaired Listeners Using Phase-Preserving Amplification*", Ear & Hearing 2005; 26 : 461–472
- V. d. Bogaert et al., "*Horizontal localization with bilateral hearing aids : Without is better than with*", J. Acoust. Soc. Am., 119, (1), January 2006 : 515–526
- V. d. Bogaert *et al.*, "*The effect of multimicrophone noise reduction systems on sound source localization by users of binaural hearing aids*", J. Acoust. Soc. Am., 124 (1), July 2008 : 484–497
- V. d. Bogaert et al., "*Speech enhancement with multichannel Wiener filters techniques in ultimicrophone binaural hearing aids*", J. Acoust. Soc. Am., 125, (1) January 2009 : 360–371

Acknowledgements : Dr Annie Moulin, Dr Céline Richard, Nicolas Bouillard, All 15 subjets.

Juin 2010,

Détection d'indices caractérisants un Speech ABR par la méthode de détection de forme,

Handicap 2010, Paris (75),

Proceding + Présentation orale

Détection automatique des indices caractéristiques d'un Potentiel Evoqués Auditif en réponse à un son de parole par une utilisation de la méthode de détection de forme

A. Jeanvoine[1, 2], C. Richard[1, 3], A. Moulin[1, 2], E. Perrin[2, 4], R. Laboissière[2, 5], Thai-Van[1, 2], C. Berger-vachon[1, 2]
1 : Laboratoire Neurosciences Sensorielles, Comportement et Cognition, UMR 5020, H.E.H., 5 place d'Arsonval, 69437 Lyon cedex 03
2 : Université Claude Bernard Lyon1, 43 Bd du 11 Novembre 1918, 69100 Villeurbanne
3 : Service d'ORL et de chirurgie cervico-faciale, CHU de Saint Etienne, Hôpital Nord, 42055 Saint Etienne cedex 02
4 : CREATIS-LRMN, CNRS INSA UCBL UMR 5220 INSERM U630, bâtiment 308, CPE Lyon, 43 Bd du 11 Novembre 1918, 69622 Villeurbanne cedex
5 : INSERM U864, Espace et Action, 16 Avenue Doyen Lépine, 69676 Bron cedex
arnaudjeanvoine@yahoo.fr

Résumé : - Un Potentiel Evoqué Auditif en réponse à un son de parole (PEASP) est formé de deux parties distinctes : la réponse impulsionnelle (ou Onset Response, OR) et la réponse soutenue en fréquence (ou Frequency Following Response, FFR). Dans l'état actuel des connaissances, nous ne pouvons pas caractériser précisément les paramètres pertinents du PEASP (amplitudes, durées et latences). L'analyse actuelle étant effectuée visuellement, et donc observateur-dépendante, l'objectif de ce travail est la mise au point d'une méthode de détection automatisée des différentes caractéristiques de la réponse PEASP, avec une application clinique future. - La mise au point d'une méthode automatisée nécessite tout d'abord la création d'un modèle mimant les caractéristiques des différentes parties du PEASP. Ainsi l'OR est caractérisée par une variation bipolaire rapide et est modélisée au moyen de trois fonctions linéaires différentes. La FFR est modélisée par un ensemble de sinusoïdes, correspondant aux fréquences fondamentale et harmoniques. - La comparaison de ce modèle type avec la réponse, sur une fenêtre glissant sur l'ensemble de la réponse, permet de déterminer les points de plus forte corrélation entre le modèle et la réponse, identifiant par là même, les indices au sein de la réponse, de manière objective.

I. INTRODUCTION

Les potentiels évoqués auditifs précoces, en réponse à des stimuli acoustiques de type clics, constituent une technique électrophysiologique d'exploration des voies auditives centrales, bien codifiée, et couramment employée en pratique clinique. L'utilisation récente de son de parole de type « consonne + voyelle » a l'avantage d'être plus écologiquement valide, et d'explorer, de manière plus fine, la perception de sons de parole le long des voies auditives. En revanche, les potentiels évoqués auditifs obtenus en réponse à ces sons de parole (PEASP), n'ont pas encore fait l'objet d'une validation et codifications objectives, préalables à toute utilisation clinique. Les réponses de type PEASP sont recueillies au niveau de 3 électrodes de surfaces placées sur le scalp du sujet, en réponse à la syllabe « /BA/ ». La forme typique du PEASP se compose de deux parties (fig. 1) : une réponse impulsionnelle, ou Onset Response (OR), qui est similaire à l'onde V des PEAP évoqués par clics et une réponse soutenue en fréquence ou Frequency Following Response (FFR) qui suit la périodicité de la voyelle du stimulus [1][2][3]. La latence, durée et amplitude de la FFR, reflètent certains aspects du traitement des sons de parole au niveau du tronc cérébral, et en particulier, le degré de synchronisation des voies auditives sur la périodicité du stimulus (pour une revue des applications cliniques [4][5][6]. Le but de ce travail est de caractériser, de manière objective, les paramètres des réponses de type PEASP, afin de déterminer les indices qui seront pertinents pour une utilisation clinique future.

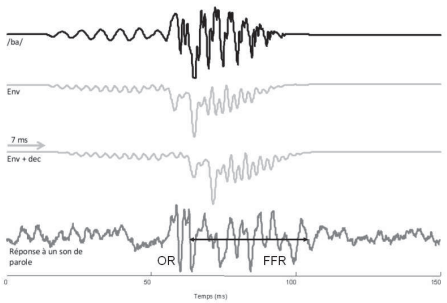


Figure 1 : Potentiel Evoqué Auditif en réponse à un son de parole.

II. MATERIEL ET METHODE

A. Méthode de recueil de PEAPSP

Le PEASP est une réponse électrophysiologique en réponse à un stimulus /BA/, présenté binauralement. Le matériel de recueil se compose de (fig. 2) [7][8]:

- Une chaîne de stimulation composée d'un ordinateur portable et d'une carte son externe, dont l'un des canaux est, après amplification, connecté aux écouteurs de type insert, insérés au niveau du conduit auditif externe du sujet, un autre canal étant connecté au trigger.
- Une chaine de recueil, composée de 3 électrodes de surface, positionnées sur le scalp du sujet, au niveau du vertex (électrode « + »), du front (masse), et de chaque mastoïde (électrode « - ») (fig. 2). La mesure du signal, de l'ordre du micro-volt, s'effectue par la différence de potentiel entre l'électrode « + » et l'électrode « - », par rapport à la masse commune de la tête. Le signal des électrodes est amplifié (boite tétière) avant d'être recueilli par un appareil couramment utilisé pour le recueil des PEAP : le Centor USB, qui assure le moyennage des réponses, synchronisées sur le stimulus.
- Du fait de la longueur du stimulus (>50 ms), de son occurrence simultanée au recueil, et de sa similitude avec la réponse de type PEAPSP, (fig. 1), le recueil nécessite une isolation dans une cage de Faraday stricte reliée à la masse, ainsi que l'utilisation d'écouteurs de type d'inserts plutôt que de type casque (Akhoun et coll.), permettant d'isoler le plus possible les électrodes de la transduction electro-acoustique des écouteurs. Un moyennage de plus de 2000 traces en polarité alternée est habituellement nécessaire, avec une intensité de stimulation variant de 50 à 80 dB HL selon les équipes (richard et al.). La fréquence d'échantillonnage est celle maximale admise par le Centor USB (13,3 kHz).
- Malgré ces précautions, il est parfois difficile d'obtenir des enregistrements peu bruités, surtout en milieu hospitalier, où les sources électromagnétiques interférentes peuvent être nombreuses, d'où l'intérêt de développer des méthodes permettant d'extraire et de quantifier les réponses de manière plus objective, que la méthode visuelle actuellement employée.

B. Caractéristiques du PEAPSP

Le PEAP se compose de l'OR, qui est une réponse impulsionnelle ressemblant fortement à l'onde V de PEAP classiques, et qui est suivie de la réponse fréquentielle soutenue (FFR). Cette réponse de type FFR est d'énergie quasi-constante, et présente la même enveloppe que celle du stimulus, avec un décalage. La figure 1 montre les traces temporelles du stimulus /BA/, de son enveloppe, de cette même enveloppe décalée de 7 ms, et de la réponse de type PEAPSP. Ce décalage, ou latence de la FFR, est obtenu en corrélant l'enveloppe inversée du stimulus, obtenue après un filtrage passe-bas au dessous de 500 Hz d'ordre un, avec la réponse du type FFR [9][10][11].

La transformée de Fourier du signal sur la partie FFR (fig. 2), montre des pics correspondant à la fréquence du fondamental (F0), ainsi que des premières harmoniques (2F0), de manière similaire au spectre de la voyelle. L'amplitude de ces pics donne une image de la qualité de synchronisation des voies auditives périphériques, sur la périodicité du stimulus et pourrait donc refléter la qualité du traitement des signaux de parole au niveau du tronc cérébral [12]. Identifier, de manière objective, la latence et la durée de la FFR, permettrait donc une meilleure objectivation de la qualité de cette synchronisation, calculée sur la durée de la FFR uniquement.

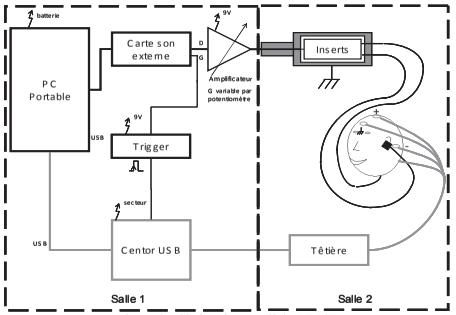


Figure 2 : Chaine de recueil du PEASP

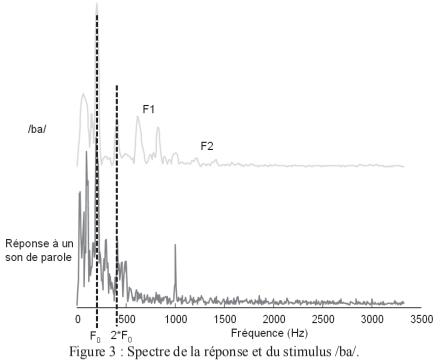


Figure 3 : Spectre de la réponse et du stimulus /ba/.

C. Modélisation d'un PEAPSP type.

Pour identifier les paramètres (OR et FFR) au sein de la réponse recueillie, nous avons créé un modèle imitant ces indices. La comparaison de ce modèle type avec la réponse, sur une fenêtre glissant sur l'ensemble de la réponse, permet de déterminer les points de plus forte corrélation entre le modèle et la réponse, identifiant par là même, les indices au sein de la réponse, de manière objective.

L'OR a été modélisée par la concaténation, dans le temps, de 3 fonctions linéaires correspondants aux temps de montée, descente et montée respectivement, sur une durée très courte (1 ms). La fig. 4 présente le modèle de l'OR, superposé sur une OR réelle moyennée sur 15 sujets. (La durée de l'OR étant très réduite par rapport à la fréquence d'échantillonnage, avec un rapport signal sur bruit de l'ordre de 6 dB, la méthode de la dérivée n'est pas pertinente pour isoler cette OR.)

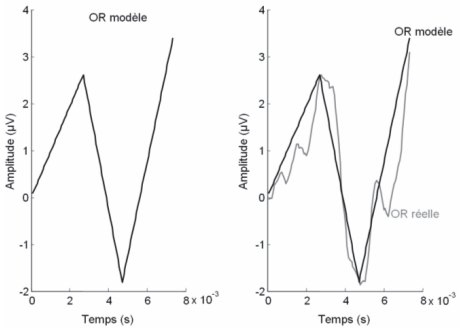


Figure 4 : Modèle de l'OR (gauche) et modèle + réelle (droite)

La FFR est une onde stationnaire qui suit la périodicité de la voyelle. Nous l'avons donc modélisée par un ensemble d'une ou plusieurs sinusoïdes, dont l'une correspond au pic fréquentiel de plus haute amplitude retrouvé sur la réponse de type FFR, par une transformée de Fourier. La fréquence de ce pic correspond à la fréquence fondamentale de la voyelle (F0, fig. 2). Il est possible d'ajouter des harmoniques d'ordre n, avec une amplitude $A/2^n$, A étant l'amplitude de F0, avec n généralement inférieur à 3, étant donné que la réponse est filtrée passe-bas en dessous de 500 Hz. Dans l'exemple présenté fig. 2, F0=206 Hz, 2*F0=412 Hz. La FFR modélisée correspondante est présentée fig. 5. Le nombre de périodes est un compromis entre une durée longue (7-8 cycles), avec une sensibilité moindre au bruit, au détriment de la précision temporelle, et une durée courte (2 cycles), permettant une plus grande précision temporelle, mais au détriment d'une plus grande sensibilité au bruit. Le modèle sinusoïdal de la FFR est pondéré par une fenêtre de hanning, pour éviter, lors de la corrélation, les effets de bord.

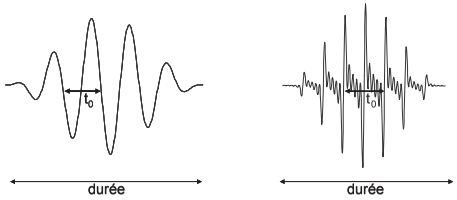


Figure 5 : Modèle sinusoïdal de FFR, à gauche avec une seule harmonique et à droite avec quatre harmoniques, correspondant à la FFR présentée fig. 1 et 3.

D. Méthode de détection

Le principe de la détection de forme est la même que celui utilisé dans la méthode des ondelettes, c'est-à-dire qu'une corrélation est calculée entre le modèle créé, et le signal recueilli, sur une fenêtre glissante. Le signal est, au préalable, filtré par un banc de filtres passe-bandes de type Butterworth d'ordre un et de largeur de bande de trois fois la plus petite fréquence que l'on peut mesurer sur le recueil du signal (fréquence d'échantillonnage/nombre de points du recueil). Ceci nous permet d'obtenir une force de corrélation entre le modèle et le signal, de manière indépendante pour chaque bande fréquentielle du signal, et de le visualiser sous forme d'un diagramme temps/fréquence.

III. RESULTATS ET DISCUSSIONS

La modélisation présentée a été effectuée sur la réponse PEASP moyenne obtenue chez 15 sujets normo-entendants, en réponse à un stimulus /BA/ de 45 dB HL d'intensité, et de 100 ms de long, avec un moyennage de 2400 réponses et une fréquence d'échantillonnage de 13,3 kHz, sur une fenêtre de recueil de 150 ms. Le PEASP recueilli est présenté fig. 1 et 3. La précision fréquentielle des recueils est de 6 Hz, et les calculs de corrélations ont été effectués sur 18 Hz, du fait du nombre limité de points.

La fig. 6 montre le résultat sur un modèle de l'OR. L'OR est difficile à distinguer, dans la réponse, du fait de sa brièveté (et donc un nombre de points très faible). En revanche, le fait qu'il s'agisse de la première onde à localiser, et qu'elle soit de grande énergie, permet de la repérer sous forme d'une ligne de corrélation présente sur l'ensemble des fréquences du signal.

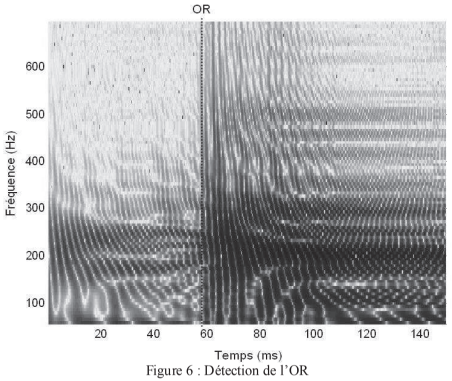


Figure 6 : Détection de l'OR

La figure 7 montre le résultat, en utilisant un modèle sinusoïdal de FFR, composé de F0 et de deux harmoniques, sur une durée de 7 cycles (fig. 7 haut), et sur une durée de 2 cycles (fig. 7 bas). Les 2 modèles permettent d'identifier le début de la FFR, par une augmentation du nombre de fréquences montrant de fortes corrélations entre modèle et signal. En revanche, la durée de la FFR est surtout visible sur le modèle à 7 cycles, montrant des corrélations constantes sur une même bande de fréquence, avant de se réduire à la fréquence du fondamental.

Il est donc possible, à partir d'un modèle simple, de déterminer objectivement les latences de l'OR et de la FFR d'un PEASP, ainsi que la durée de la FFR, et par conséquent son amplitude. Ce travail, effectué sur des données moyennées, et donc, avec un bon rapport signal/bruit, demande à être élargi à des données plus artéfactées. Cependant, en utilisant différents niveaux de seuil de corrélation, nous pensons qu'il est possible de s'adapter à différents rapports signal/bruit lors du recueil de PEASP, et donc, d'obtenir des paramètres objectifs sur des données recueillies individuellement pour chaque sujet.

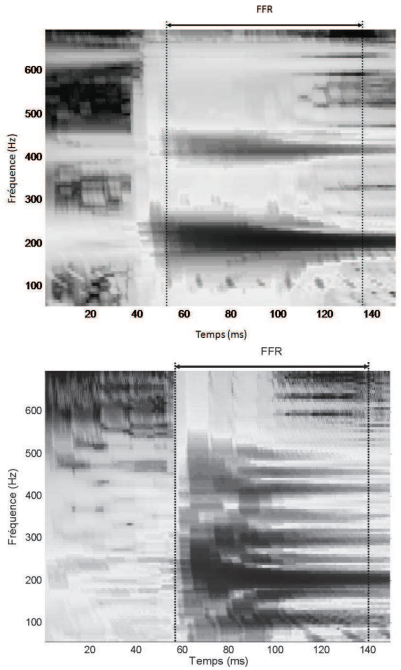


Figure 7 : Détection de la FFR avec 7 oscillations pour le graphique haut et 2 oscillations pour le graphique du bas

IV. CONCLUSION

Les PEASP représentent une technique d'avenir d'exploration du codage des signaux de parole par les voies auditives du tronc cérébral. La modélisation mathématique permet d'objectiver les paramètres des réponses, pertinents pour le neurophysiologiste. A plus long terme, ceci permettra au clinicien, de comparer des données individuelles à des valeurs normatives, de manière observateur-indépendante.

V. BIBLIOGRAPHIE

[1] C.L. Faingold, D.M. Caspary. «Frequency-following responses in primary auditory and reticular formation structures». *Electroencephalogr Clinical Neurophysiol*, 1979; 47: 12_20

[2] J.W. Hall. «Auditory brainstem frequency following responses to waveform envelope periodicity». *Science*, 1979; 205: 1297_1299

[3] N. Russo, T.G. Nicol, G. Musacchia, N. Kraus. «Brainstem responses to speech syllables». *Clinical Neurophysiol*, 2004; 115: 2021_2030

[4] C. Richard, A. Moulin, A. Jeanvoine, I. Akhoun, E. Veuillet, L. Collet, H. Thai-Van. «Exploration électrophysiologique des voies auditives sous-corticales chez l'Humain : du clic au son de parole». (communication personelle). *Neurophysiologie Clinique*

[5] K.L. Johnson, T.G. Nicol, N. Kraus. «Brain stem response to speech: a biological marker of auditory processing». *Ear Hear*, 2005; 26: 424_434.

[6] E. Skoe, N. Kraus. «Auditory brain stem response to complex sound : a tutorial». Ear Hear, 2010; 31: 1_23

[7] I. Akhoun, A. Moulin, A. Jeanvoine, M. Ménard, F. Buret, C. Vollaire, R. Scorretti, E. Veuillet, C. Berger-Vachon, L. Collet, H. Thai-Van. «Speech auditory brainstem response (speech ABR) characteristics depending on recording conditions, and hearing status an experimental parametric study», *Journal of Neuroscience Methods*, 2008; 175: 196_205.

[8] J.T. Marsh, F.G. Worden, J.C. Smith. «Auditory frequency-following response: neural or artefact?». *Science*, 1970; 169: 1222_1223

[9] I. Akhoun , S. Gallégo, A. Moulin, M. Ménard, E. Veuillet, C. Berger-Vachon, L. Collet, H. Thai-Van. «The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults». *Clinical Neurophysiol*, 2008; 119: 922_933.

[10] S.J. Aiken, T.W. Picton. «Envelope following responses to natural vowels». *Audiol Neurotol*, 2006; 11: 213_232

[11] D. A. Abrams, T. Nicol, S.G. Zecker, N. Kraus. «Auditory Brainstem Timing Predicts Cerebral Asymmetry for Speech». *Journal of Neuroscience*, 2006; 26(43): 11131_11137

[12] J. Hornikel, E. Skoe, N. Kraus. «Subcortical laterality of speech», *Audiol Neurotol*, 2009; 14: 198_207.

Avril 2010, *Détection des indices pertinents caractérisant le Speech ABR*,
CFA 2010, Lyon (69),
Proceding + Présentation orale

10ème Congrès Français d'Acoustique
Lyon, 12-16 Avril 2010

Détection Automatique d'Indices Pertinents dans les Réponses Electrophysiologiques de Type Potentiels Evoqués Auditifs recueillis en réponse à un son de parole (Speech ABR).

A. Jeanvoine[a], C. Richard[a,b], A. Moulin[a], C. Berger-Vachon[a], et H. Thai-van[a]

a : Laboratoire Neurosciences Sensorielles, Comportement et Cognition, CNRS UMR 5020, HEH, 5 place d'Arsonval, 69437, Lyon cedex 03
b : Service d'ORL et de chirurgie servico-faciale, CHU de Saint Etienne, Hopital Nord, 42055 Saint Etienne cedex 02
arnaudjeanvoine@yahoo.fr

Le Speech ABR est la réponse du tronc cérébral à un son de parole /ba/. Elle est composée de deux indices pertinents qui sont : l'Onset Response (OR) et la Frequency Following Response (FFR). L'OR est caractérisée par une impulsion de haute énergie similaire à une « Onde V » dans un Potentiel Evoqué Auditif Précoce évoqué en réponse à un clic. La FFR est une onde stationnaire qui suit la périodicité de la voyelle /a/ jusqu'à 500 Hz. Nous avons utilisé plusieurs méthodes mathématiques qui permettent de détecter ces indices pertinents de manière objective, afin de pouvoir caractériser la variabilité inter-individuelle de l'aspect des Speech ABR.

1 Introduction

Les potentiels évoqués auditifs précoces, en réponse à des stimuli acoustiques de type clics, constituent une technique électrophysiologique d'exploration des voies auditives centrales, bien codifiée, et couramment employée en pratique clinique. L'utilisation récente de son de parole de type « consonne + voyelle » a l'avantage d'être plus écologiquement valide, et d'explorer, de manière plus fine, la perception de sons de parole le long des voies auditives. En revanche, les potentiels évoqués auditifs obtenus en réponse à ces sons de parole (PEASP ou Speech ABR), n'ont pas encore fait l'objet d'une validation et codifications objectives, préalables à toute utilisation clinique.

Les réponses de type Speech ABR sont recueillies au niveau de 3 électrodes de surfaces placées sur le scalp du sujet, en réponse à la syllabe « /BA/ ». La forme typique du Speech ABR se compose de deux parties (figure 1a) : une réponse impulsionnelle, ou Onset Response (OR), qui est similaire à l'onde V des PEAP évoqués par clics et une réponse soutenue en fréquence ou Frequency Following Response (FFR) qui suit la périodicité de la voyelle du stimulus [1]. Les latences, durée et amplitude de la FFR, reflètent certains aspects du traitement des sons de parole au niveau du tronc cérébral, et en particulier, le degré de synchronisation des voies auditives sur la périodicité du stimulus [2].

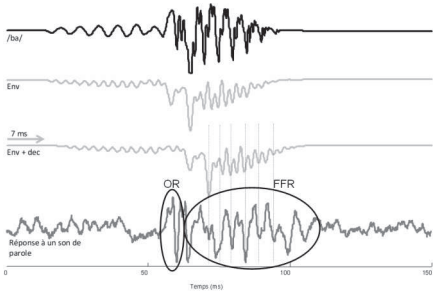


Figure 1a : Potentiel Evoqué Auditif en réponse à un son de parole (domaine temporel).

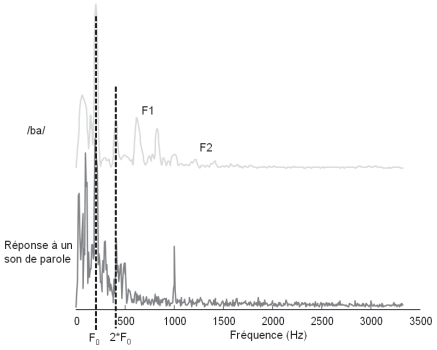


Figure 1b : Potentiel Evoqué Auditif en réponse à un son de parole (domaine fréquentiel).

La réponse qui est recueillie sur les électrodes cutanées est une réponse qui est similaire à l'enveloppe inversée du stimulus filtrée passe bas avec une fréquence de coupure de 500 Hz et décalée d'une dizaine de millisecondes (figure 1a). De plus la réponse suit exactement le contenu fréquentiel de la voyelle /a/ jusqu'à 500 Hz [1].

Actuellement la méthode employée pour la localisation temporelle des deux indices qui sont l'OR et la FFR est une méthode visuelle. Il peut avoir dans ce cas un biais observateur dépendant, en particulier dans les cas où le signal est bruité.

Le but de ce travail est de caractériser, de manière objective, les paramètres des réponses de type Speech ABR, afin de déterminer les indices qui seront pertinents pour une utilisation clinique future.

2 Matériel et méthodes

2.1 Méthode de recueil du Speech ABR

Le Speech ABR est une réponse électrophysiologique en réponse à un stimulus de parole /ba/, présenté binauralement. Le matériel de recueil se compose de (figure 2) :

- Une chaîne de stimulation composée d'un ordinateur portable et d'une carte son externe, dont l'un des canaux est, après amplification, connecté aux écouteurs de type insert, insérés au niveau du conduit auditif externe du sujet, un autre canal étant connecté au trigger.
- Une chaîne de recueil, composée de 3 électrodes de surface, positionnées sur le scalp du sujet, au niveau du vertex (électrode « + »), du front (masse), et de chaque mastoïde (électrode « - »). La mesure du signal, de l'ordre du micro-volt, s'effectue par la différence de potentiel entre l'électrode « + » et l'électrode « - », par rapport à la référence « masse » de la tête. Le signal des électrodes est amplifié (boîtier Têtière) puis numérisé avant d'être recueilli par un appareil couramment utilisé en clinique pour le recueil des PEAP : le Centor USB° (Racia Alvar°), qui assure le moyennage des réponses, synchronisées sur le stimulus à l'aide du trigger.
- Du fait de la longueur du stimulus (>50 ms), de son occurrence simultanée au recueil, et de sa similitude avec la réponse Speech ABR, (figure 1a), le recueil nécessite une isolation de la transduction électrique/acoustique dans une cage de Faraday reliée à la masse du système, ainsi que l'utilisation d'écouteurs de type d'inserts plutôt que de type casque classique, permettant d'isoler le plus possible les électrodes des écouteurs [3,4]. Un moyennage de plus de 2000 traces en polarité alternée est habituellement nécessaire, avec une intensité de stimulation variant de 50 à 80 dB HL selon les équipes [5,6] pour obtenir un recueil avec un rapport signal sur bruit suffisant. La fréquence d'échantillonnage du recueil est celle maximale admise par le Centor USB (13,3 kHz).

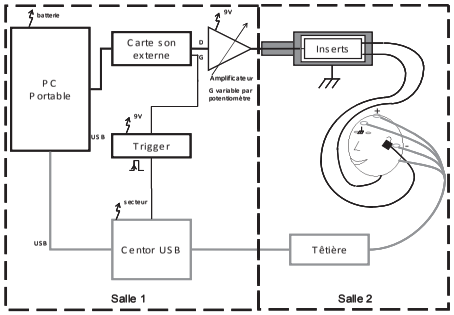


Figure 2 : Chaîne de stimulation et de recueil d'un Speech ABR

Malgré ces précautions, il est parfois difficile d'obtenir des enregistrements peu bruités, surtout en milieu hospitalier, où les sources électromagnétiques interférentes peuvent être nombreuses, d'où l'intérêt de développer des méthodes permettant d'extraire et de quantifier les réponses de manière plus objective, que la méthode visuelle actuellement employée.

2.2 Caractéristiques du Speech ABR

Le premier indice pertinent du Speech ABR est une réponse impulsionnelle ressemblant fortement à « l'onde V » des Potentiels Evoqués Auditifs Précoces classiques (ou OR). Elle est facilement décelable car l'amplitude est largement supérieure à celle du bruit moyen de la trace. Elle est suivie de la réponse soutenue en fréquence (FFR) [5]. La figure 1a montre les traces temporelles du stimulus /ba/, de son enveloppe [7], de cette même enveloppe décalée de 7 ms, et de la réponse électrophysiologique du stimulus [1,5]. Ce décalage d'une dizaine de milliseconde (dépendant du sujet) donne la latence de la FFR, elle est obtenue en corrélant l'enveloppe inversée du stimulus [1], obtenue après un filtrage passe-bas avec une fréquence de coupure de 500 Hz d'ordre un, avec la réponse de type FFR [8,9,10]. Les caractéristiques des indices sont répertoriées dans le tableau 1.

La transformée de Fourier du signal sur la partie FFR (figure 1b), montre des pics correspondant à la fréquence du fondamental (F_0), ainsi que la première harmonique ($2xF_0$), de manière similaire au spectre de la voyelle. Il n'y a plus d'informations au delà de 500 Hz.

	Amplitude	Impulsion	Régime		Fréquences		Énergie	Périodicité	Latence		Durée
			Transitoire	Stationnaire	Basse	Haute			Brève	Longue	
OR	+	++	+	-	-	+	++	-	+	-	-
FFR	+	-	-	+	+	-	+	++	-	+	+

Tableau 1 : Récapitulatif des caractéristiques des indices pertinents.

2.3 Utilisation de méthodes mathématiques pour la détection des indices pertinents

Le premier travail est de séparer l'OR de la FFR dans le domaine fréquentiel, par un filtrage. L'OR est plutôt caractérisée par des hautes fréquences et la FFR par des basses fréquences (tableau 1). La séparation fréquentielle se fait par un filtre passe bas de type Butterworth ordre un, avec une fréquence de coupure de 300 Hz pour la FFR et l'utilisation d'un filtre passe bande de type Butterworth ordre un avec une bande passante de 300 Hz à 550 Hz pour l'OR. Tout le signal est préalablement filtré à 50 Hz pour s'affranchir des composantes très basses fréquences ainsi que celle du réseau électrique.

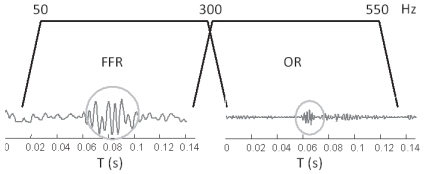


Figure 3 : Séparation de l'OR de la FFR fréquentiellement, la trace temporelle en bas montre les différences temporelles pour l'OR et la FFR.

2.3.1 Méthode de détection de l'OR

Le calcul du bruit du signal se fait sur les dix premières millisecondes car le sitmulus n'est pas entendu par la personne pendant cette période de la trace par la méthode du RMS. Si le rapport signal sur bruit de la trace est de 0 dB, alors la réponse mesurée est discutable.

Pour l'OR la trace utilisée sera celle des hautes fréquences (figure 3). La méthode utilisée pour la détection de l'OR permet de rester dans le domaine temporel.

La dérivée permet, à chaque instant de la trace, de connaître le coefficient directeur de l'information temporelle haute fréquence, un coefficient directeur grand montre une impulsion et donc une montée d'énergie très courte. Il suffit de prendre l'instant où le coefficient directeur est le plus grand négativement car la pente recherchée est décroissante. La partie correspondante est alors isolée par un encadrement de 10 ms centré sur le coefficient directeur trouvé.

L'énergie est ensuite calculée autour de ce point, et qui correspond à l'intégrale du signal. L'isolation de cette partie permet de déterminer la position précise de l'indice de long du signal recueilli.

2.3.2 Méthode de détection de la FFR

La trace utilisée sera celle des basses fréquences (figure 3). De la même façon que pour la détection de l'OR, le traitement se fait dans le domaine temporel. Il faut trouver cette fois-ci une énergie constante pendant plusieurs périodes d'oscillation. Le calcul de l'énergie se fait en prenant l'intégrale de la trace basse fréquence. Le signal est ensuite isolé lorsque l'énergie est constante et située au dessus du bruit de fond de la trace. De plus le calcul de la dérivée donne une meilleure précision sur les oscillations car l'amplitude de celle-ci est plus constante et la fréquence fondamentale est multipliée par deux, ce qui donne deux fois plus d'informations pour la même durée.

3 Résultats

3.1 Détection de l'OR

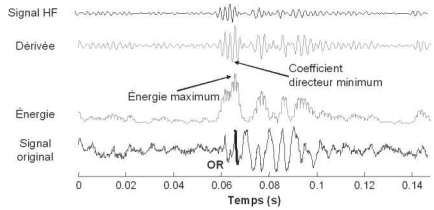


Figure 4 : Détection de l'OR (en gras sur la courbe du bas)

Sur la trace originale, l'OR est le premier indice qui intervient dans la réponse, elle est toujours située avant la FFR. C'est pourquoi dans le calcul de l'énergie il peut y avoir ambiguïté car on remarque plusieurs impulsions (figure 4), surtout si le signal est bruité (rapport signal sur bruit proche de 0).

Une fois l'OR trouvée, les informations suivantes sont alors connues : la latence et l'amplitude, ainsi que sa durée. La latence varie selon l'intensité et le type des sons de parole utilisés (Kraus revue).

3.2 Détection de la FFR

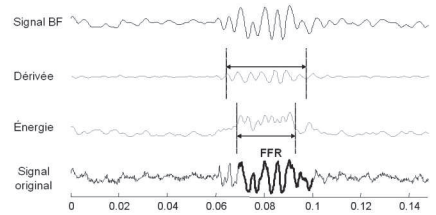


Figure 5 : Détection de la FFR (en gras sur la courbe du bas)

En ce qui concerne la FFR, il est beaucoup plus facile de l'isoler même en présence de bruit car l'oscillation et l'énergie sont plus importantes pendant un certain temps qui correspond à la durée de la FFR. On remarque que, sur la courbé dérivée, les amplitudes des oscillations sont constantes et qu'elles oscillent au double de la trace basse fréquence, par conséquent, dès que l'oscillation est supérieure au bruit moyen de la trace, il est possible d'effectuer un encadrement de la partie correspondant à la FFR.

Dès que la FFR est trouvée, comme pour l'OR, l'amplitude, la latence et la durée de celle-ci peuvent être obtenues facilement, et sont des préalables à l'analyse de la FFR par FFT.

4 Discussion

Il est très facile de déceler les indices dans une réponse qui n'est pas artéfactée. Cependant, lorsque la trace est bruitée, il est très difficile de déceler l'impulsion de l'OR car les changements brusques de la courbe induisent d'énormes coefficients directeurs qui peuvent être assimilés à la pente de l'OR lors des différents calculs. De plus le rayonnement du réseau électrique ainsi que ses harmoniques peuvent contaminer de façon importante le recueil du signal, en particulier si l'isolation des électrodes n'est pas correcte. La réponse électrophysiologique peut être alors fortement perturbée et induire sur la réponse une périodicité qui n'est pas la réponse à la voyelle /a/. En revanche si le signal recueilli est de bonne qualité avec un rapport signal sur bruit de l'ordre de 6 dB, alors les indices sont facilement décelable dans le Speech ABR par une méthode automatique.

5 Conclusion

Les résultats montrent que l'on peut trouver dans une réponse de type Speech ABR, les deux indices qui la caractérisent, autrement que par une méthode visuelle qui était beaucoup moins précise. La précision temporelle maintenant atteinte est de l'ordre de la milliseconde, du fait de la faible fréquence d'échantillonnage.

Le Speech ABR représente une technique d'avenir d'exploration du codage neuronal des signaux de parole dans le tronc cérébral. La technique mathématique présentée ici permet d'objectiver les deux paramètres de la réponse électrophysiologique. A plus long terme, ceci permettra au clinicien, de comparer des données individuelles à des valeurs normatives, de manière observateur-indépendante.

Références

[1] I. Akhoun , S. Gallégo, A. Moulin, M. Ménard, E. Veuillet, C. Berger-Vachon, L. Collet, H. Thai-Van. "The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults", *Clinical Neurophysiol*, 2008; 119: 922_933.

[2] K.L. Johnson, T.G. Nicol, N. Kraus. "Brain stem response to speech: a biological marker of auditory processing". *Ear Hear*, 2005; 26: 424_434.

[3] I. Akhoun, A. Moulin, A. Jeanvoine, M. Ménard, F. Buret, C. Vollaire, R. Scorretti, E. Veuillet, C. Berger-Vachon, L. Collet, H. Thai-Van. "Speech auditory brainstem response (speech ABR) characteristics depending on recording conditions, and hearing status an experimental parametric study", *Journal of Neuroscience Methods*, 2008; 175: 196_205.

[4] J.T. Marsh, F.G. Worden, J.C. Smith. "Auditory frequency-following response: neural or artefact?", *Science*, 1970; 169: 1222_1223

[5] E. Skoe, N. Kraus. "Auditory brain stem response to complex sound : a tutorial". Ear Hear, 2010; 31: 1-23

[6] C. Richard, A. Moulin, A. Jeanvoine, I. Akhoun, E. Veuillet, L. Collet, H. Thai-Van. "Exploration électrophysiologique des voies auditives sous-corticales chez l'Humain : du clic au son de parole". (communication personnelle). *Neurophysiologie Clinique*

[7] J.W. Hall. "Auditory brainstem frequency following responses to waveform envelope periodicity". *Science*, 1979; 205: 1297_1299

[8] N. Russo, T.G. Nicol, G. Musacchia, N. Kraus. "Brainstem responses to speech syllables". *Clinical Neurophysiol*, 2004; 115: 2021_2030

[9] S.J. Aiken, T.W. Picton. "Envelope following responses to natural vowels". *Audiol Neurotol*, 2006; 11: 213_232

[10] Aiken SJ, Picton TW. "Envelope and spectral frequency-following responses to vowel sounds". Hear Res 2008;245:35-47.

Bibliographie

[1] S. M. Abel and V. H. Hay. Sound localization. the interaction of aging, hearing loss and hearing protection. *Scand Audiol*, 25(1) :3–12, 1996.

[2] Idrick Akhoun, Annie Moulin, Arnaud Jeanvoine, Mikael Ménard, François Buret, Christian Vollaire, Riccardo Scorretti, Evelyne Veuillet, Christian Berger-Vachon, Lionel Collet, and Hung Thai-Van. Speech auditory brainstem response (speech abr) characteristics depending on recording conditions, and hearing status : an experimental parametric study. *J Neurosci Methods*, 175(2) :196–205, Nov 2008.

[3] T. Balkany, W. Boggess, and B. Dinner. Binaural cochlear implantation : comparison of 3m/house and nucleus 22 devices with evidence of sensory integration. *Laryngoscope*, 98(10) :1040–1043, Oct 1988.

[4] Douglas Bates, Martin Maechler, and Ben Bolker. *lme4 : Linear mixed-effects models using S4 classes*, 2012. R package version 0.999999-0.

[5] C. Beaugeant, V. Turbin, P. Scalart, and A. Gilloire. New optimal filtering approaches for ands-free telecommunication terminals. *Signal Processing*, 64(15) :33–48, Jan 1998.

[6] M. Berouti, Bolt Beranek, Newman, R. Schwartz, and J. Makhoul. Enhancement of speech corrupted by acoustic noise. *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '79.*, 4 :208–211, Apr 1979.

[7] M. Bhatnagar. *A modified spectral subtraction method combined with perceptual weighting for speech enhancement*. PhD thesis, UT-Dallas, 2002.

[8] R. Boite, H. Boulard, T. Dutoit, J. Hancq, and H. Leich. *Traitement de la parole*. P U Polytec Rom, 2000.

[9] Steven F. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE*, 27(2) :113 – 120, Appril 1979.

[10] Christian Bouchard, David I. Havelock, and Martin Bouchard. Beamforming with microphone arrays for directional sources. *J Acoust Soc Am*, 125(4) :2098–2104, Apr 2009.

[11] Andrew D Brown and G. Christopher Stecker. Temporal weighting functions for interaural time and level differences. ii. the effect of binaurally synchronous temporal jitter. *J Acoust Soc Am*, 129(1) :293–300, Jan 2011.

[12] Kevin D. Brown and Thomas J. Balkany. Benefits of bilateral cochlear implantation : a review. *Curr Opin Otolaryngol Head Neck Surg*, 15(5) :315–318, Oct 2007.

[13] O. Cappe. Elimination of the musical noise phenomenon with the ephraim and malah noise suppressor. 2(2) :345–349, 1994.

[14] F. Chen. The relative importance of temporal envelope information for intelligibility prediction : a study on cochlear-implant vocoded speech. *Med Eng Phys.*, 33(8) :1033–8, Oct 2011.

[15] Fei Chen and Philipos C. Loizou. Predicting the intelligibility of vocoded speech. *Ear Hear*, 32(3) :331–338, May-Jun 2011.

[16] Teresa Y. C. Ching, Robyn Massie, Emma van Wanrooy, Emma Rushbrooke, and Colleen Psarros. Bimodal fitting or bilateral implantation. *Cochlear Implants Int*, 10 Suppl 1 :23–27, 2009.

[17] CH. Chouard and JL. Weber. Minimac : a totally numeric 15 channel implanted stimulator. *Pacing Clin Electrophysiol*, 12(4 Pt 2) :743–748, Apr 1989. Centre Hospitalo-Universitaye Paris-Saint-Antoine, France.

[18] GM. Clark. A surgical approach for a cochlear implant : an anatomical study. *J Laryngol Otol.*, 89(1) :9–15, Jan 1975.

[19] GM. Clark and RJ. Hallworth. A multiple-electrode array for a cochlear implant. *J Laryngol Otol.*, 90(7) :623–627, Jul 1976.

[20] N. Dillier, H. Bögli, and T. Spillmann. Speech encoding strategies for multielectrode cochlear implants : a digital signal processor approach. *Prog Brain Res*, 97 :301–311, 1993.

[21] H. Dillon. Advanced signal processing schemes in hearing aids. In *Thieme*, volume chapitre 7, pages 187–208, 2001.

[22] H. Dillon. Compression systems in hearing aids. *Hearing aids*, 6 :159–186, 2001.

[23] Matthias Doerbecker and Stefan Ernst. Combination of two-channel spectral subtraction and adaptive wiener post-filtering for noise reduction and deverberation. *European Signal Processing Conference*, 1996.

[24] MF Dorman, PC Loizou, J Fitzke, and Tu Z. The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels. *J Acoust Soc Am*, 104(6) :3583–5., Dec 1998.

[25] DJ. Doyle. Some comments on the use of wiener filtering for the estimation of evoked potentials. *Electroencephalogr Clin Neurophysiol*, 38(5) :533–4, May 1975.

[26] Camille C. Dunn, William Noble, Richard S. Tyler, Monika Kordus, Bruce J. Gantz, and Haihong Ji. Bilateral and unilateral cochlear implant users compared on speech perception in noise. *Ear Hear*, 31(2) :296–298, Apr 2010.

[27] Camille C Dunn, Richard S Tyler, Sarah Oakley, Bruce J Gantz, and William Noble. Comparison of speech recognition and localization performance in bilateral and unilateral cochlear implant users matched on duration of deafness and age at implantation. *Ear Hear*, 29(3) :352–359, Jun 2008.

[28] Y. Ephraim and D. Malah. Speech enhancement using optimal non-linear spectral amplitude estimation. *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '83.*, 8 :1118–1121, Apr 1983.

[29] Yariv Ephraim and David Malah. Speech enhancement using a minimum mean-square error short -time spectral amplitude estimator. *IEEE Transaction on Acoustics, Speech, and Signal Processing*, 32(6) :1109–1121, December 1984.

[30] Yariv Ephraim and David Malah. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Transaction on Acoustics, Speech, and Signal Processing*, 33(2) :443–445, April 1985.

[31] G. Fant. *Speech Sounds and Features*, (Cambridge, MIT Press), 1973.

[32] Tom Francart, Jan Brokx, and Jan Wouters. Sensitivity to interaural level difference and loudness growth with bilateral bimodal stimulation. *Audiol Neurootol*, 13(5) :309–319, Apr 2008.

[33] Tom Francart, Anneke Lenssen, and Jan Wouters. Enhancement of interaural level differences improves sound localization in bimodal hearing. *J Acoust Soc Am*, 130 (5) :2817–26, Nov 2011.

[34] Tom Francart, Tim Van den Bogaert, Marc Moonen, and Jan Wouters. Amplification of interaural level differences improves sound localization in acoustic simulations of bimodal hearing. *J Acoust Soc Am*, 126(6) :3209–3213, Dec 2009.

[35] R. Frosch. *Introduction to Cochlear Waves*. VDF, Hochschulverlag AG an der ETH Zürich, 2010.

[36] Soha N. Garadat, Ruth Y. Litovsky, Gongqiang Yu, and Fan-Gang Zeng. Role of binaural hearing in speech intelligibility and spatial release from masking using vocoded speech. *J Acoust Soc Am*, 126(5) :2522–2535, Nov 2009.

[37] Matthew J. Goupell, Piotr Majdak, and Bernhard Laback. Median-plane sound localization as a function of the number of spectral channels using a channel vocoder. *J Acoust Soc Am*, 127(2) :990–1001, Feb 2010.

[38] M. Hansen. Effects of multi-channel compression time constants on subjectively perceived sound quality and speech intelligibility. *Ear Hear.*, 23(4) :369–80, Aug 2002.

[39] WF House. Goals of the cochlear implant. *Laryngoscope*, 84(11) :1883–1887, Nov 1974.

[40] Yi Hu and Philipos C. Loizou. Subjective comparison and evaluation of speech enhancement algorithms. *Speech Commun*, 49(7) :588–601, Jul 2007.

[41] LG Johnsson, WF House, and FH. Linthicum. Bilateral cochlear implants : histological findings in a pair of temporal bones. *Laryngoscope*, 89(5 Pt 1) :759–762, May 1979.

[42] Fathi Kallel, Mondher Frikha, Mohamed Ghorbel, Ahmed Ben Hamida, and Christian Berger-Vachon. Dual-channel spectral subtraction algorithms based speech enhancement dedicated to a bilateral cochlear implant. *Applied Acoustics*, 73 :12–20, 2012.

[43] S. Kamath and P. Loizou, editors. *A multi-band spectral subtraction method for enhancing speech corrupted by colored noise.*, Orlando, 2002. Proceedings of ICASSP-2002.

[44] Gibak Kim, Yang Lu, Yi Hu, and Philipos C. Loizou. An algorithm that improves speech intelligibility in noise for normal-hearing listeners. *J Acoust Soc Am*, 126(3) :1486–1494, Sep 2009.

[45] B. P. Kimberley, R. Dymond, and A. Gamer. Bilateral digital hearing aids for binaural hearing. *Ear Nose Throat J*, 73(3) :176–179, Mar 1994.

[46] D. J. Kistler and F. L. Wightman. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J Acoust Soc Am*, 91(3) :1637–1647, Mar 1992.

[47] Kostas Kokkinakis and Philipos C Loizou. Multi-microphone adaptive noise reduction strategies for coordinated stimulation in bilateral cochlear implant devices. *J Acoust Soc Am*, 127(5) :3136–3144, May 2010.

[48] M. Kompis and N. Dillier. Performance of an adaptive beamforming noise reduction scheme for hearing aid applications. i. prediction of the signal-to-noise-ratio improvement. *J Acoust Soc Am*, 109(3) :1123–1133, Mar 2001.

[49] J. C. Lafon. *Le test phonétique*. Paris, 1958.

[50] D. T. Lawson, B. S. Wilson, M. Zerbi, C. van den Honert, C. C. Finley, J. C. Farmer, J. T. McElveen, and P. A. Roush. Bilateral cochlear implants controlled by a single speech processor. *Am J Otol*, 19(6) :758–761, Nov 1998.

[51] T. Lfukube. Discrimination of synthetic vowels by using tactile vocoder and a comparison to that of an eight-channel cochlear implant. *IEEE Trans Biomed Eng.*, 36(11) :1085–91, Nov 1989.

[52] Ning Li and Philipos C. Loizou. Effect of spectral resolution on the intelligibility of ideal binary masked speech. *J Acoust Soc Am*, 123(4) :EL59–EL64, Apr 2008.

[53] J. Lim and A. Oppeenheim. Enhancement and bandwidth compression of noisy speech. *IEEE*, 67 :1586–1604, 1979.

[54] L. Lin, W. H. Holmes, and E. Ambikairajah. Speech denoising using perceptual midification of winer filtering. *IEEE Electronic Letters*, 38 :1486–1487, Nov 2002.

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

[56] P.C. Loizou. *Speech enhancement : theory and practice*. Signal processing and communications. CRC Press, 2007.

[57] Philipos C. Loizou. Speech processing in vocoder-centric cochlear implants. *Adv Otorhinolaryngol*, 64 :109–143, 2006.

[58] R. Martin. Spectral subtraction based on minimum statistics. pages 1182–1185, 1994.

[59] R. McAulay and M. Malpass. Speech enhancement using soft-decision noise suppresion filter. *IEEE Trans. on ASSP*, 28(2) :137–145, Apr 1980.

[60] A. W. Mills. On the minimum audible angle. *J Acoust Soc Am*, 30(4) :237–246, April 1958.

[61] A. W. Mills. Lateralization of high-frequency tones. *J Acoust Soc Am*, 32(1) :132–134, January 1960.

[62] Brian CJ. Moore. *An introduction to the Psychology of Hearing, Fifth Edition*. Emerald Group Pub Ltd, 1995.

[63] Arlene C. Neuman, Anita Haravon, Nicole Sislian, and Susan B. Waltzman. Sound-direction identification with bilateral cochlear implants. *Ear Hear*, 28(1) :73–82, Feb 2007.

[64] R. Nouvian, D. Malinvaud, T. van den Abbeele, J-L. Puel, P. Bonfils, and P. Avan. Physiologie de l'audition. *Oto-rhino-laryngologie*, 10 :20–30, 2006.

[65] R. C. Oldfield. The assessment and analysis of handedness : the edinburgh inventory. *Neuropsychologia*, 9(1) :97–113, Mar 1971.

[66] Elizabeth S. Olson, Hendrikus Duifhuis, and Charles R. Steele. Von békésy and cochlear mechanics. *Hearing Research*, (0) :–, 2012.

[67] David Poeppel. The analysis of speech in different temporal integration windows : cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*, 41 :245–255, 2003.

[68] Becky B Poon, Donald K Eddington, Victor Noel, and H. Steven Colburn. Sensitivity to interaural time difference with bilateral cochlear implants : Development over time and effect of interaural electrode spacing. *J Acoust Soc Am*, 126(2) :806–815, Aug 2009.

[69] R Development Core Team. R Foundation for Statistical Computing, Vienna, Austria, 2010. ISBN 3-900051-07-0.

[70] Lord Rayleig. On our perception of sound direction. *Physiolosophical Magazine Series 6*, 13(74) :214 – 232, 1907.

[71] C Richard, A Jeanvoine, E Veuillet, A Moulin, and H Thai-Van. Exploration of the auditory system in humans : From click to speech auditory brainstem responses. *Neurophysiol Clin*, 40(5–6) :267–79, 2010.

[72] T. Ricketts and P. Henry. Evaluation of an adaptive, directional microphone hearing aid. *Int J Audiol*, 41 :100–112, 2002.

[73] Houda Saoud, Goulven Josse, Eric Bertasi, Eric Truy, Maria Chait, , and Anne-Lise Giraud. Brain–speech alignment enhances auditory cortical responses and speech perception. *J Neurosci*, 32(1) :275–281, Jan 2012.

[74] P. Scalart and J. V. Filho. Speech enhancement based on a priori signal to noise estimation. In *Proc. Conf. IEEE Int Acoustics, Speech, and Signal Processing ICASSP-96*, volume 2, pages 629–632, 1996.

[75] Bernhard U. Seeber, Uwe Baumann, and Hugo Fastl. Localization ability with bimodal hearing aids and bilateral cochlear implants. *J Acoust Soc Am*, 116(3) :1698–1709, Sep 2004.

[76] R. Shannon, F.-G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid. Speech recognition with primarily temporal cues. *Science*, 270 :303–311, 1995.

[77] EA Shaw. Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. *J Acoust Soc Am.*, 56(6) :1848–61, Dec 1974.

[78] FB. Simmons. Cochlear implant. *Arch Otolaryngol.*, 89(1) :66–69, jan 1969.

[79] M. Skinner, L. Holden, and T. Holden. Effect of frequency boundary assignment on speech recognition with the speack speech-coding strategy. *Annals of Otology, Rinology and Laringology*, 104(Suppl. 166) :307–311, 1995.

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

[81] A. Spriet. *Adaptive filtering techniques for noise reduction and acoustic feedback cancellation in hearing aids.* PhD thesis, K. U. Leuven, Belgique, 2004.

[82] Ann Spriet, Lieselot van Deun, Kyriaky Eftaxiadis, Johan Laneau, Marc Moonen, Bas van Dijk, Astrid van Wieringen, and Jan Wouters. Speech understanding in background noise with the two-microphone adaptive beamformer beam in the nucleus freedom cochlear implant system. *Ear Hear*, 28(1) :62–72, Feb 2007.

[83] R. Stapells, D., T. W. Picton, and A. D. Smith. Normal hearing thresholds for clicks. *J. Acoust. Soc. Am.*, 72(1) :74–79, 1982.

[84] Michael A. Stone, Christian Füllgrabe, and Brian C. J. Moore. High-rate envelope information in many channels provides resistance to reduction of speech intelligibility produced by multi-channel fast-acting compression. *J Acoust Soc Am*, 126(5) :2155–2158, Nov 2009.

[85] V. Tjomov and G. Gershuni. A model to describe the results of psychoacoustical experiments on steady-state stimuli. *In Analiz Rechevykh Signalov Chelovekom, Nauka*, pages 36–49, 1971.

[86] E. Truy, H. Thai-Van, L. Collet, S. Gallego, and E. Ionescu. Binaural cochlear implantation : general concept, surgical technique and electrophysiology. *Ann Otolaryngol Chir Cervicofac*, 119(3) :150–158, 2002.

[87] Eric Truy, Eugène Ionescu, Philippe Ceruse, and Stéphane Gallego. The binaural digisonic cochlear implant : surgical technique. *Otol Neurotol*, 23(5) :704–709, Sep 2002.

[88] Richard S. Tyler, Camille C. Dunn, Shelley A. Witt, and William G. Noble. Speech perception and localization with adults with bilateral sequential cochlear implants. *Ear Hear*, 28(2 Suppl) :86S–90S, Apr 2007.

[89] Richard S. Tyler, Camille C. Dunn, Shelley A. Witt, and John P. Preece. Update on bilateral cochlear implantation. *Curr Opin Otolaryngol Head Neck Surg*, 11(5) :388–393, Oct 2003.

[90] R. Udrea, N. Vizireanu, and S. Ciochina. An improved spectral subtraction method for speech enhancement using a perceptual weighting filter. *IEEE*, 18 :581–586, 2008.

[91] Tim Van den Bogaert, Simon Doclo, Jan Wouters, and Marc Moonen. The effect of multimicrophone noise reduction systems on sound source

localization by users of binaural hearing aids. *J Acoust Soc Am*, 124(1) :484–497, Jul 2008.

[92] Tim van den Bogaert, Simon Doclo, Jan Wouters, and Marc Moonen. The effect of multimicrophone noise reduction systems on sound source localization by users of binaural hearing aids. *J Acoust Soc Am*, 124(1) :484–497, Jul 2008.

[93] Tim van den Bogaert, Simon Doclo, Jan Wouters, and Marc Moonen. Speech enhancement with multichannel wiener filter techniques in multimicrophone binaural hearing aids. *J Acoust Soc Am*, 125(1) :360–371, Jan 2009.

[94] Tim Van den Bogaert, Thomas J. Klasen, Marc Moonen, Lieselot Van Deun, and Jan Wouters. Horizontal localization with bilateral hearing aids : without is better than with. *J Acoust Soc Am*, 119(1) :515–526, Jan 2006.

[95] Richard van Hoesel, Melanie Böhm, Jörg Pesch, Andrew Vandali, Rolf D. Battmer, and Thomas Lenarz. Binaural speech unmasking and localization in noise with bilateral cochlear implants using envelope and fine-timing based strategies. *J Acoust Soc Am*, 123(4) :2249–2263, Apr 2008.

[96] Richard van Hoesel, Richard Ramsden, and Martin Odriscoll. Sound-direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user. *Ear Hear*, 23(2) :137–149, Apr 2002.

[97] Richard J. M. van Hoesel. Exploring the benefits of bilateral cochlear implants. *Audiol Neurootol*, 9(4) :234–246, 2004.

[98] Richard J. M. van Hoesel, Gary L. Jones, and Ruth Y. Litovsky. Interaural time-delay sensitivity in bilateral cochlear implant users : effects of pulse rate, modulation rate, and place of stimulation. *J Assoc Res Otolaryngol*, 10(4) :557–567, Dec 2009.

[99] Richard J. M. van Hoesel and Richard S. Tyler. Speech perception, localization, and lateralization with bilateral cochlear implants. *J Acoust Soc Am*, 113(3) :1617–1630, Mar 2003.

[100] B. D. Veen and K. M. Buckley. Beamforming : A versatile apporach to spatial filtering. *IEEE ASSP Magazine*, 5 :4–24, 1988.

[101] N. Verhaert, D.S. Lazard, D. Gnansia, J.P. Bébéar, P. Romanet, B. Meyer, V. Péan, D. Mollard, and E. Truy. Speech performance and sound localization abilities in neurelec digisonic® sp binaural cochlear implant users. *Audiol Neurootol.*, 17(4) :256–266, May 2012.

[102] N. Virag. Single channel speech enhancement based on masking properties of the human auditory system. *IEEE*, 7 :126–137, 1999.

[103] DO. Walter. A posteriori "wiener filtering" of average evoked responses. *Electroencephalogr Clin Neurophysiol*, Suppl 27 :61+, 1968.

[104] J. Yang. Frequency domain noise suppression approaches in mobile telephone systems. *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, 2 :363–366, 1993.

[105] E. Zwicker and H. Fastl. *Psychacoustics : Facts and Models*, volume 22. Springer Series in Information Sciences, 2nd edition edition, 1999.

Résumé

Les implants cochléaires sont des appareilles destinés à la réhabilitation des surdités profondes et totales. Ils assurent la stimulation du nerf auditif en plaçant des électrodes dans la cochlée. Différentes études ont été établis afin d'améliorer l'intelligibilité de la parole dans le bruit chez le patient porteur de cet appareil. Les techniques bilatérales et binaurales permettent de reproduire une audition binaurale, car les deux oreilles sont simulées (comme pour les personnes normo-entendantes). Ainsi la localisation et la perception des sons environnants sont améliorées par rapport à une implantation monaurale. Toutefois, les capacités de reconnaissances des mots sont très vite limitées en présence de bruits. Nous avons développés des techniques de réduction de bruit afin d'augmenter les performances de reconnaissance. Des améliorations de 10% à 15% suivant les conditions ont été observées. Néanmoins, si la perception est améliorée par les algorithmes, ils focalisent sur une direction, et ainsi, la localisation est alors réduite à l'angle de localisation. Une seconde étude a alors été effectuée pour mesurer l'effet des algorithmes sur la localisation. Ainsi, le beamformer donne les meilleurs résultats de compréhension mais la moins bonne localisation. La ré-injection d'un pourcentage du signal d'entrée sur la sortie a permis de compenser la perte de la localisation sans détériorer l'intelligibilité.

Le résultat de ces deux expériences montre qu'il faut un compromis entre la perception et la localisation des sons pour obtenir les meilleures performances.

Mots clefs

Implant cochléaire binaural, Soustraction Spectrale, Filtre de Wiener, Beamformer, Ré-injection, localisation, intelligibilité

Abstract

Cochlear implants are to sail for the rehabilitation of deep and total deafness. They provide stimulation of the auditory nerve by placing electrodes in the cochlea. Various studies have been established to improve the speech intelligibility in noise in the patient of this device. Bilateral and binaural techniques allow reproducing a binaural hearing, since both ears are simulated (as for normal hearing people). Thus localization and the perception of the surrounding sounds are improved from a monaural implantation. However, the recognition of the words capabilities are limited very quickly in the presence of noise. We developed noise reduction techniques to increase the performance of recognition. Improvements of 10% to 15% depending on the conditions were observed. Nevertheless, if the perception is enhanced by algorithms, they focus on a direction, and thus the location is then reduced at the corner of localization. Then, a second study was made to measure the effect of localization algorithms. Thus, the beamformer gives the best results of understanding but the less good location. The re-injection of a percentage of the input to the output signal helped offset the loss of the location without damaging the intelligibility.

The result of these two experiments shows that it takes a compromise between perception and sound localization for best performance.

Key-words

Binaural cochlear implant, Spectral Subtraction, Winer filter, Beamforming, Re-injection, localization, speech intelligibility